Transportation Association of Canada

Changing Practices in Data Collection on the Movement of People

October 2014

DISCLAIMER

The material presented in this text was carefully researched and presented. However, no warranty expressed or implied is made on the accuracy of the contents or their extraction from reference to publications; nor shall the fact of distribution constitute responsibility by TAC or any researchers or contributors for omissions, errors or possible misrepresentations that may result from use of interpretation of the material contained herein.

Transportation Association of Canada
2323 St. Laurent Blvd.
Ottawa, ON K1G 4J8
Tel. (613) 736-1350 ~ Fax (613) 736-1395
www.tac-atc.ca

ISBN 978-1-55187-544-6

TAC REPORT DOCUMENTATION FORM

Title and Subtitle

Changing Practices in Data Collection on the Movement of People

Report Date

October 2014

Coordinating Agency and Address

Transportation Association of Canada
2323 St. Laurent Boulevard
Ottawa, ON K1G 4J8

ITRD No.

Author(s)

Eric Miller, Martin Lee-Gosselin, Khander Nurul Habib, Catherine Morency, Matthew J. Roorda, Amer S. Shalaby

Corporate Affiliation(s) and Address(es)

University of Toronto
35 St. George Street, Room 305
Toronto, ON M5S 1A4

Lee-Gosselin Associates Limited
275-A Chemin du Bout de l'Ile
Sainte-Pétronille
Ile d'Orléans, QC G0A 4C0

École Polytechnique de Montréal
C.P. 6079, succursale Centre-ville
Montréal, QC H3C 3A7

Abstract

The objective of the *Changing Practices in Data Collection on the Movement of People* project is to prepare a practical, comprehensive framework for the coordination, collection, processing and management of data on the movement of people by all modes in Canadian urban areas that is implementable and addresses data needs across the range of Canadian transportation agencies. This report presents the framework developed within this project.

It is essential that transport agencies have a clear and comprehensive understanding of the technical options available to them, of the strengths and weaknesses, benefits and costs of these options, and clear guidance in terms of cost-effective data collection programs that will best fit their local needs and budgets.

Keywords

Economics and Administration
Traffic and Transport Planning

- Data acquisition
- Data processing
- Data Transmission (Telecom)
- Global Positioning System
- Household
- In situ
- Interview
- Network (Transport)
- Planning
- State of the art report
- Telephone
- Traffic count
- Urban area
- Vehicle tracking (location)
- Web site

Supplementary Information

ACKNOWLEDGMENTS

The Changing Practices in Data Collection on the Movement of People report was developed with funding provided by several agencies. TAC gratefully acknowledges the following funding partners for their contributions to the project:

Agence Métropolitaine de Transport
Ontario Ministry of Transportation
City of Edmonton
City of Ottawa
Halifax Regional Municipality
Ministère des Transports du Québec
Metrolinx
TransLink
City of Mississauga
Transport Canada
Ville de Montréal

PROJECT STEERING COMMITTEE

This report was developed under the supervision of a project steering committee of volunteer members. The participation of these committee members throughout the project is gratefully acknowledged.

Timothy Spurr	Agence Métropolitaine de Transport (Co-Chair)
Lisa Salsberg	Metrolinx (Co-Chair)
Stéphane Brice	Ville de Montréal
William Hui and Lorie Srivastava	TransLink
Ahmad Subhani	City of Ottawa
Arthur Tai	Ontario Ministry of Transportation
Rhonda Toohey	City of Edmonton
Pierre Tremblay	Ministère des Transports du Québec
Bob Sasaki, Norbert Orzel, Joe Perrotta	City of Mississauga
David MacIsaac	Halifax Regional Municipality
Eric Sevigny and Arif Husain	Transport Canada
Katarina Cvetkovic	TAC (project manager)

CONSULTANT TEAM

Eric Miller	University of Toronto
Martin Lee-Gosselin	Lee-Gosselin Associates Limited
Khandker Nurul Habib	University of Toronto
Catherine Morency	Polytechnique Montreal
Matthew J. Roorda	University of Toronto
Amer S. Shalaby	University of Toronto

The consulting team gratefully acknowledges the assistance of Joshua Hoy in the preparation of the annotated bibliography and the analysis of survey results. The contributions of Taha Rashidi to the literature review are also acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

LIST OF TABLES

LIST OF FIGURES

APPENDICES

LIST OF FIGURES

LIST OF TABLES

LIST OF FIGURES

LIST OF TABLES

PART I – REPORT INTRODUCTION

1. PROJECT OVERVIEW

High quality, comprehensive data on travel behaviour, transport network performance and associated land use characteristics are absolutely essential to the planning and design of urban transport systems. Such information is derived from a variety of sources, but most comes from travel surveys and other data collection methods that provide the base data for both analysis and modelling of trip-making and transport system performance. Without adequate data we can neither understand our transport needs and issues nor can we design and evaluate service and policy alternatives. Given this fundamental need for data, transport agencies at all levels of government invest considerable resources in on-going data collection, using many methods. Changes in Information Technology, however, are rapidly introducing new data collection options (GPS, cell phones, internet, etc.), while at the same time reducing the effectiveness of some methods (e.g., the decline in effectiveness of land-line-based telephone interviews). Similarly, advances in statistical methods for fusing and mining diverse data sets offer the potential to make much better use of a variety of data sources. In addition, the emergence of a wide variety of new data providers, from commercial data retailers to open-source, web-based datasets introduces a potentially rich array of new, useful sources of transportation data. On the other hand, the demise of the mandatory Census Long Form represents a significant loss in high quality, detailed data about our urban regions and poses new challenges for the validation and weighting of travel surveys.

In parallel to this changing technical landscape, the needs and capabilities of our transport planning agencies are changing; they also vary considerably by city size, among other factors. Continuing urban growth and congestion, greenhouse gas emissions (climate change) and air quality, ecosystem health, economic productivity, safety and security, and capital and operating financial burdens are challenges that all urban areas face in varying ways. Increasingly difficult fiscal environments for agencies require them to be ever more cost-effective within typically shrinking budgets. Data collection is often viewed as a “costly frill”, despite the critical role that it plays in the planning and design of multi-billion dollar transport systems. Limited budgets also too often translate into limited staff resources for data collection and management.[1] And yet it is exactly in times of fiscal restraint such as we currently face that the need is greatest for cost-effective decision-making, and this can only be achieved if the right information (including model-based forecasts, with these models being derived from high quality base data) is available to inform and support these decisions.

For all of these reasons it is essential that transport agencies have a clear and comprehensive understanding of the technical options available to them, of the strengths and weaknesses, benefits and costs of these options, and clear guidance in terms of cost-effective data collection programs that will best fit their local needs and budgets.

---

[1] Budget limitations, of course, are not restricted to data collection and management tasks, but extend to entire programs for transportation investments and operations.

The objective of the *Changing Practices in Data Collection on the Movement of People* project is to prepare a practical, comprehensive framework for the coordination, collection, processing and management of data on the movement of people by all modes in Canadian urban areas that is implementable and addresses data needs across the range of Canadian transportation agencies. This report presents the framework developed within this project.

This framework is based upon several "building blocks" developed during the project:

- An extensive review of the survey and data collection literature that systematically identified current and emergency data collection issues, needs and methods. This review is documented in this report in *Part II: Literature Review*. In this review, data collection methods are divided into the following three major categories, with associated sub-topics:
    - **Population-based Surveys**
        - Definitions and Basic Concepts in Survey Design
        - Household Travel Surveys
        - Issues and Challenges of Telephone Interview Surveys
        - Methodological Advances
    - **Choice-based Sample Surveys**
        - Roadside Intercept Surveys
        - Transit User Intercept Surveys
    - **Technology-based Data Collection Methods**
        - Web-Based Surveys
        - GPS and Other Positioning Systems
        - Emerging Applications of Portable Technologies
        - Passive Data Streams (smart cards, etc.)
        - Media-enabled Social Networks (or Social Media Networks)

  Each type of data collection method is discussed in terms of:

    - Methods & applications.
    - Strengths, weaknesses, issues & opportunities.
    - Canadian applications.
- A review of data integration/fusion/synthesis methods that can be used to combine data from diverse sources together, thereby expanding the usefulness of available data. This review is documented in *Part III; Data Integration/Fusion Methods*. Although somewhat technically challenging, data fusion techniques are increasingly being used in a number of practical applications, including:

  - "Filling in" missing variables in survey datasets (e.g., imputing missing income information in travel surveys).
  - Synthesizing disaggregate agents as inputs into microsimulation models.
  - Generating large, combined datasets for largely descriptive analyses.
  - Combining revealed and stated preference survey data.
  - (Less commonly) converting repeated cross-section survey data into a pseudo-panel dataset to support dynamic, longitudinal analysis and modelling.
- An inventory and discussion of a variety of sources of data pertaining to urban passenger travel (Census, etc.) which Canadian transportation agencies might utilize, thereby reducing the need for original data collection efforts. This inventory is presented in *Part IV: Survey of Data Sources for Urban Transportation Applications*. As documented in this part of the report, Canadian transportation agencies use a wide variety of data sources, including Census and other Statistics Canada datasets, along with other federal, provincial, municipal, private sector and open-source (often web-based) data.
- Results from an original survey of Canadian transportation agencies undertaken by the project team to identify current Canadian data collection practice, issues and needs. The survey and its results are documented in *Part V: Design, Conduct and Results of a Survey of Canadian Transportation Data Collection Practice*. This survey was web-based. Transportation agencies across the country were contacted using mailing listed provided by TAC and the Canadian Urban Transit Association (CUTA). A total of 124 responses were received, of which 94 were complete and fully usable. Responses were received from every province and territory except PEI and Nunavut. Good distributions in responses across the size of the responding urban regions and the types of transportation agencies were also obtained. It is felt that the survey provides a good overview of Canadian data collection practice, and it provides an excellent complement to the literature review in terms of the identification of the current state of practice and the strengths, weaknesses and challenges associated with this current practice.
- Based on these "building blocks", the recommended framework for Canadian urban passenger transportation data collection and management is presented in *Part VI: A Framework for Urban Passenger Transportation Data Collection & Management*. This framework is summarized in the following section.

Parts II through VI of this report were originally submitted to the TAC Project Steering Committee (PSC) as five individual reports. This overall final project report consists of the assembly of these five individual reports into a single omnibus volume. Section and table/figure numbering for each of the original independent reports has been retained so as to maintain ease of readability. A detailed table of contents for each part is also provided at the beginning of each part.

2. FRAMEWORK SUMMARY

This recommended framework for Canadian urban passenger transportation data collection presented in detail in *Part VI* of this report is developed in seven parts:

- The "business case" for urban passenger data collection.
- A taxonomy of urban passenger data needs. Starting with very basic concepts in transportation planning and analysis, a comprehensive view of data needs is constructed. This culminates in the development of an "object-oriented model" for representing transportation systems and the data characterizing these systems. This comprehensive data model has many uses, in that it provides a:
    - Systematic, comprehensive representation of important data elements and their relationships, thereby providing a sound basis for determination of data needs for various applications.
    - Starting point for organizing data management systems.
    - Tool for model development.
- An overview of key data collection methods, which summarizes the findings of both the literature review (*Part II* of the report) and the survey of Canadian agencies (*Part V* of the report).
- A summary of trends, issues and options in meeting current and emerging data needs, building on earlier discussions in the report. Key issues and observations include:
    - Household surveys will remain in the future a primary data collection tool.
    - Current household survey methods, however, face major challenges and must adapt. Important problems with current methods that must be addressed include:
        - Problems with land-line-based sampling frames.
        - Problems with contacting and recruiting respondents.
        - Problems with retrospective/proxy reporting.
    - New technology offers numerous opportunities. Promising technologies include:
        - Multi-instrument surveys.
        - Web-based surveys.
        - Positioning (GPS / other), increasingly found on portable devices.
    - Data collection techniques must continuously adapt and evolve to address many trends in travel behaviour, technology and planning needs and policy issues.
- Issues in data management (including privacy issues).
- A new paradigm for coordinated, cost-effective survey design (the "core-satellite" paradigm) is developed and presented. The recommended core-satellite paradigm is a multi-instrument

approach to dealing with the complex, multi-faceted data needs of transportation agencies through a coordinated combination of data collection methods that attempts to match the best method to each data collection problem in cost effective ways. It is a very flexible/generalizable approach to data collection that is applicable to a wide variety of contexts. It is defined by the content that is required, not by commitment to a particular data collection method. (i.e., the methods will vary depending on the content required and the context within which this content is required). It also permits controlled experimentation with new methods and evolution of data collection programs over time so as to be able to flexibly adjust to changing needs and technological options.

- And finally, building upon the above, a recommended framework for urban transportation data collection and management is presented. This framework consists of six elements:
    - An institutional/political commitment to data collection.
    - Commitment to high quality data collection and management methods.
    - Adoption of a standard (object-oriented) model of urban transportation data.
    - Adoption of a core-satellite design paradigm for data collection.
    - Commitment to experimentation and evolutionary change in data collection methods.
    - (Possibly) growing urban surveys (into province-wide surveys?).

This framework is based upon the fundamental proposition that the *status quo* in Canadian urban passenger transportation data collection and management is not sustainable: practice must evolve to meet changing needs, opportunities and challenges. We are in a highly dynamic era in which technology is changing rapidly and picking “winners/losers” is difficult to do. In addition, appropriate methods vary by context and this context is also changing as urban areas grow and evolve. Thus, it is essential that transportation agencies be willing to experiment with new methods; they must learn how to learn and evolve. As a result of these needs, the framework presented in this report is not a static structure. Rather it is a blueprint for a process for adaptation and evolution in a changing world.

PART II – LITERATURE REVIEW

1. INTRODUCTION

This part of the report presents a survey of the current and emerging state of practice in urban transportation data collection methods in Canada and elsewhere. This literature review provides a starting point for the development of an urban data collection framework for Canadian transportation agencies, which is the primary product of the project.

Table 1.1: Data Type/Method Taxonomy

Data Type	**Household-Based Surveys**				**Choice-Based Sample Surveys**		**Standard Technology-Based Methods**			
	Face to face	Telephone	Mail-back	Web- based	Road-side	On-board	GPS	Roadside detectors	Smart-card	Mobile phone
Household activity / trip-making behaviour	X	X	X	X	X	X	X		X	X
Count data					X	X		X	X	
Transportation system characteristics							X	X	X	X
Transportation cost and service levels	X	X	X	X			X	X	X	X
Land use characteristics										
Population socio-economic information	X	X	X	X	X	X				
Attitudes / opinions / stated choices	X	X	X	X	X	X				
System impacts (e.g. emissions)								X		

	Emerging Technology Based Methods							
Data Type	Remote sensing (satellite / aerial)	Web "apps"	Social network software	Smart phone	Acceler-ometers	Personal health sensors	Environ-mental sensors	Blue-tooth
Household activity / trip-making behaviour		X	X	X	X			
Transportation system characteristics	X							X
Transportation cost and service levels	X			X				X
Land use characteristics	X	X					X	
Population socio-economic information		X	X	X		X		
Attitudes / opinions / stated choices		X	X	X				
System impacts (e.g. emissions)					X	X	X	

As illustrated in Table 1.1, many types of data are used in urban transportation planning, including:

- Household activity / trip-making behaviour.
- Count data (traffic, riders, etc.).
- Transportation system characteristics (speeds, lane widths, etc.[2]).
- Transportation cost and service levels.
- Land use characteristics (population, employment, etc.)
- Population socio-economic information (income, auto ownership, etc.).
- Attitudes / opinions / stated choices.
- System impacts (e.g., emissions).

As is also illustrated in Table 1.1, numerous data collection methods exist that are suitable for collecting various types of data, depending on the data type and the eventual usage of the data. These data collection methods, however, can be usefully grouped into three broad categories:

- Population-based surveys.
- Choice-based sample surveys.
- Non-survey data collection techniques (both "standard" and "emerging").[3]

Sections 2, 3 and 4 discuss each of these three types of data collection in detail, with particular attention being paid to Canadian data collection practice and issues. This overview discussion of the data collection methods is supplemented by an annotated bibliography of the urban transportation data collection literature presented in an appendix at the end of this part of the report.

In addition to actively collecting data through surveys and other methods, urban transportation agencies make use of a wide variety of independent data sources, which include (but are not limited to) census data, spatially indexed data for geocoding purposes, inventories of transit schedules and other road and transit network attributes, etc. Incorporating these various data with survey data involves a variety of *data fusion* techniques for melding diverse datasets together in statistical valid ways. Parts III (*Data Integration/Fusion Methods*) and IV (*Survey of Data Sources for Urban Transportation Applications*) address these issues in detail. Part V of this report (*Design, Conduct and Results of a Survey of Canadian Transportation Data Collection Practice*) further extends this discussion by providing a snapshot of current Canadian data collection practices and issues as reported in a national survey undertaken within this project.

---

[2] Both static (e.g., posted speed) and dynamic (e.g., observed speeds by time of day).

[3] The discussion of data collection technology in Section 4 also includes use of these technologies in survey applications.

2. POPULATION-BASED SURVEYS

2.1 INTRODUCTION

A *survey* involves requesting specific information from a *sample* of respondents selected from a well-defined *population* of agents of interest. A sample of the population, rather than the full population, is surveyed in most instances, since surveying an entire population is usually prohibitively expensive, normally practically impossible, and generally not necessary since valid statistical inferences concerning the population can be made from properly constructed samples.

Surveys, of course, are undertaken throughout our society for a wide variety of reasons (political polling, market research for new products, public opinion surveys, etc.). Well-established survey methodology exists and is well documented in a variety of texts and manuals (see, among many others, Cochran, 1977; Dillman, 1978, 2009), which will not be reviewed in detail here. In transportation planning, surveys are a primary (and for many purposes the only feasible) means of gathering information on trip-makers' travel behaviour, and so they constitute an essential data collection tool within the profession. Standard texts and manuals for transportation survey design and application also exist (Cambridge Systematics, 1996; Dillman, 1978, 2009; NCHRP, 2007, 2008; Transport Canada, 2009; TRB, 2011), and, again, while these manuals will be drawn on throughout this report, they are not reviewed in detail herein.

In transportation (among other fields), two fundamental ways of defining the group of agents to be surveyed exists:

1. *Population-based* surveys, in which a target group of agents (households, firms, etc.) are identified based on their *attributes* (all physically disabled persons, all registered voters, all households living in a study region, etc.).
2. *Choice-based sample survey*s, in which the target group to be surveyed is determined by their *behaviour* or choices (transit riders, airline passengers, roadway users, shopping mall patrons, etc.).

Both population-based and choice-based sample surveys are used extensively in transportation applications. This section discusses population-based methods, applications and issues, while Section 3 discusses choice-based sample surveys. In both cases a strong emphasis is placed on Canadian practice and needs. To begin this discussion, Subsection 2.2 defines key terms and issues in survey design that are essential to all subsequent discussion of survey methods and applications.

Both population-based and choice-based surveys can also be categorized in terms of whether they collect *revealed preference/behaviour* (RP) or *stated preference/response* (SP) data, where these terms are defined as:

- ***Revealed preference/behaviour*** surveys collect information on <u>actual, executed</u> choices made by respondents within actual, real-world conditions, for example, information on the actual trips made on a given survey day, including the actual start times, purposes, destinations and modes chosen for these trips.
- ***Stated preference/response*** surveys place respondents within a <u>hypothetical</u>/experimental context in which they are asked to "state" what their response would be if they found

themselves in the given hypothetical context as presented to them within the controlled experiment. A typical example of an SP survey is to ask respondents to select which travel mode they would chose for a given trip, given the hypothetical travel times, costs, etc. for competing modes for this trip.

While "revealed" versus "stated" is a widely understood distinction, the term "preference" is seen by some authors as too restrictive. What people do, or say what they would do, is not necessarily what they prefer. "Revealed behaviour" (RB) and "Stated response" (SR) have been suggested as general terms (e.g., for SR, Bradley and Hensher, 1992). In the case of SR, a number of other terms are found in the literature. "Stated Choice" is used rather interchangeably with "Stated Preference". Lee-Gosselin (1996) suggested that four sub-classes of SR (Stated Preference, Stated Adaptation, Stated Tolerance and Stated Prospect) be defined according to whether *constraints* and/or *behavioural outcomes* are mostly given or mostly elicited in the survey procedure. This taxonomy is also found in Cambridge Systematics (1996).

Stated preference/response methods are increasingly being used in a variety of operational transportation planning contexts, the two most common of which are situations in which:

- The choice of interest cannot, for one reason or another, be observed in the real world. The most common example of this is when one wants to model choice of a "new mode" which currently does not exist within the study area (BRT and/or LRT in a city in which these services currently do not exist; toll road usage in the absence of existing toll roads; etc.).
- One wants to very precisely explore the trade-offs that people make between level of service (or other) attributes within a given choice context. For example, an SP survey can present trip-makers with time-cost trade-off combinations that do not exist in the real world that provides enhanced insights into their values of time.

Mixed RP/SP surveys are often undertaken, with the SP component used to explore aspects of travel behaviour that cannot be directly observed via RP methods, while the RP component "grounds" the SP results within observed real-world behaviour. A recent example is found in LeVine et al (2012). As is discussed in greater detail in Part III (*Data Integration/Fusion Methods*), standard methods exist for combining RP and SP data within travel model estimations.

The design of either a stand-alone SP or a mixed RP/SP survey involves all the issues and considerations that are discussed throughout this report. In addition, the design of the SP experiments (time-cost trade-offs, etc.) is an art/science onto itself, the detailed discussion of which is beyond the scope of this report. For more information concerning SP survey methods and design issues see, for example, (Kroes and Sheldon, 1988; Hensher, 1994; Louvière et al, 2002, Rose and Bliemer, 2012). The remainder of this report focusses on RP-type surveys while, again, recognizing that the methods and issues discussed also apply to SP applications.

Many types of population-based surveys are used in transportation planning. These include:

- Household surveys, in which a sample of households is selected and, usually, all household members above a certain minimum age are surveyed about their travel behaviour.
- Place-of-work surveys, in which workers are randomly selected at their place of work and surveyed concerning their work trip commuting behaviour.

- Employment surveys, in which firms are surveyed concerning their number of employees, floorspace occupied, etc.
- Other special-purpose populations (the elderly, school children, the physically disabled, etc.).

Of these, by far the most important is household-based travel surveys. Given their importance, Subsections 2.3-2.5 discuss in detail household surveys. The methods and issues discussed with respect to household surveys generally apply to other types of population-based surveys and so these are not explicitly discussed in this report.

2.2 DEFINITIONS AND BASIC CONCEPTS IN SURVEY DESIGN

In order to make statistically valid inferences about a population's attributes or behaviours from a sample of respondents surveys must be very carefully constructed so as to ensure that a *representative sample* is drawn that adequately characterizes the distribution of population attributes and behaviours. In other words, great care must be taken to avoid the introduction of *bias* in the survey results. That is, the purpose of any survey is to estimate one or more *parameters* of the population of interest, where these parameters might be population attributes (average income, labour force participation rates, auto ownership rates, etc.) or behaviours (travel mode splits, home-based trip rates, etc.). The true values of such parameters can never be known with certainty (without surveying the entire population). Given randomness in sample selection, measurement errors, etc. sample estimates of population parameters are *random variables*, that will have some distribution of possible outcomes around the true population values (i.e., if a given survey was run many times it would generate a distribution of outcomes around the true population values). An *unbiased* sample estimate is one whose *expected value* is the true population parameter value (i.e., if the survey was run many times, the average of the estimates obtained would equal the true population parameter value). An *efficient* sample estimate is one whose distribution (variance) around its expected value is as small as possible, so that any one estimate (which is all one generally has in practice) can be expected to lie "close" to the true value. Thus, both efficiency (precision) and unbiasedness (accuracy) are important in survey design: one absolutely wants to avoid introducing systematic error (bias) into the survey results, but, given that one has only a single shot at obtaining these results, one wants them to be as precise as possible as well.

Regardless of the application, all survey designs involve the following 11 steps (Cochran, 1977):

1. Establish a clear statement of the survey objectives.
2. Define the population to be sampled and the target groups to be focused on.
3. Identify the specific data that are relevant for the purpose of the survey.
4. Specify the degree of precision required from the survey results (i.e., how much error can be tolerated in the results).
5. Determine the methods to be used in obtaining the survey results.
6. Divide the population into sampling units and list the units from which the sample will be drawn.
7. Select the sampling procedure and the sample size.
8. Pretest the survey and field methods to ensure that the procedures are workable and the survey is understandable.

9. Establish a good supervisory structure for managing the survey.
10. Determine the procedures for analyzing and summarizing the data.
11. Store the data and analysis results for future reference.

Each of these steps is critical to reducing bias (increasing accuracy), increasing efficiency/precision and maximizing the overall effectiveness and usefulness of the survey results. Surveys are inevitably reasonably costly to undertake: it is critical that their utility within the planning and decision-making process is maximized in order to ensure their cost-effectiveness. In particular, it is essential that the survey be designed and executed in a manner that ensures that it does, indeed, provide the data that are needed for the task(s) at hand and these data are as accurate and precise as possible. This sounds like a very self-evident statement, but the field is rife with examples of surveys that have failed in a variety of ways due to failure to follow Cochran's 11 steps as rigorously and as effectively as possible.

Survey design elements that need special discussion within this review include the following:

- Population, sampling frame and sampling unit definitions.
- Sampling procedure.
- Sample size determination.
- Interview type.
- Sample recruitment, contact methods and incentives.
- Temporal considerations.
- Questionnaire design.

Each of these issues is defined and discussed in the following sub-sections. Throughout this discussion a key emphasis is on maximizing survey *response rate* (i.e., the percentage of persons selected within the survey sample who actually complete the survey), since the larger the response rate, the less likely it is that the survey data will contain major biases. Keys to high response rates include:

- Minimizing *respondent burden*, i.e., the amount of time and level of effort required for the respondent to complete the survey.
- Maximizing the respondent's motivation to complete the survey.

2.2.1 Population, Sampling Frame and Sampling Unit Definitions

In survey design a *population* is the set of agents for which inferences are to be drawn from the survey data. Typical populations of interest within urban transportation planning include:

- All individuals living within the study area.
- All households living within the study area.
- All firms located within the study area.
- All employees working within the study area.

- All members of a group of special interest (physically disabled persons, the young or the elderly, low-income households, new immigrants within the region, etc.).

The appropriate definition of the population to be surveyed obviously depends on the questions to be addressed by the survey and the information required to answer these questions.

A *sampling frame* is the operational method of (at least implicitly) enumerating/listing the population from which the survey sample is to be drawn. For example, if the population is the set of all households in an urban region, the sampling frame in principle should be a physical list of all these households from which individual households can be selected to interview. In practice, sampling frames are often not literally complete but are assumed to be sufficiently complete and representative (i.e., bias free) to be useful. In a typical Canadian telephone-based home interview survey, for example, the sampling frame is usually the list of all telephone land lines listed in the telephone directory. This list clearly excludes households with unlisted numbers and who do not possess a land-lane telephone[4], but is used because it is a cost-effective, simple means of constructing a sampling frame that is assumed to be relatively representative.

Having cost-effective access to a complete (accurate) sampling frame is an essential component in every survey design. Without a well-defined sampling frame, sample weightings for the construction of population inferences cannot be computed and the representativeness of the sample cannot be guaranteed. In addition to enumerating the population, the sampling frame generally must provide the contact information (telephone number, mailing address, e-mail address, etc.) that is required to contact the selected survey respondents. The inability to construct an appropriate sampling frame is often one of the limiting factors in determining transportation survey feasibility.

The *sampling unit* is literally the elementary unit to be surveyed within the sampling frame. If the population consists of households, then the sampling unit is a single household within the sampling frame, if the population consists of employees, the sampling unit is a single employee, and so on. A *weight* (or *expansion factor*) must be applied to each unit sampled that represents the contribution of this unit to the calculation of population attribute values (population totals and averages, etc.). The calculation of these weights depends on the sampling procedure used, discussed below.

By far the most common population of survey interest in urban transportation is the set of households residing within the urban region. Most transportation agencies in Canada and elsewhere have traditionally relied on household-based surveys for the collection of origin-destination travel demand data. Household survey travel behaviour data are used for direct planning analysis purposes, the construction of travel demand forecasting models, and transit service planning applications, among many other uses. These surveys have long been the *sine qua non* for effective urban transportation planning across Canada. Household travel surveys are discussed in detail in Subsections 2.3-2.5.

2.2.2 Sampling Procedures

For any given sampling frame a number of procedures exist for actually drawing a sample of units (respondents) from this frame. These include (Meyer and Miller, 2001):

---

[4] As is discussed further in Section 2.4.1, this exclusion of households without landlines (typically because they exclusively use cell phones) is a growing problem with this method.

- Simple random sampling.
- Sequential sampling.
- Stratified random sampling.
- Cluster sampling.

In all cases, the objective is to achieve an unbiased sample in a cost-effective manner. The “standard” or “base” method is *simple random sampling*, in which the sampling units are literally randomly selected from the sampling frame. Simple random sampling ensures that no bias is introduced into the sampled set through the sampling process. Any of the other sampling procedures listed above are only used when they provide a more cost-effective (efficient) method for achieving a similarly unbiased result. In particular, they may provide a means to reduce sample size (while maintaining precision levels) or to select sampling units in cases in which an explicit sampling frame is difficult to enumerate.

Sequential sampling is a simple variant of simple random sampling in which the sampling frame can be safely assumed to be randomly ordered. In such cases the list can be sequentially, rather than randomly, sampled, thereby generally increasing the efficiency (and simplicity) of the sampling process. Examples of sequential sampling include selecting every n^{th} person arriving at a transit stop to interview (under the assumptions that the order of arrival at the stop is a random process that introduces no bias into the sample selection process) or selecting every n^{th} listing in a telephone directory (under the assumption that the alphabetical listing of names in the directory does not introduce any bias in the selection process). In all such cases, “n” is the inverse of the sampling rate. E.g., if a 10% sample is required, then every 10th person would be selected.

Stratified sampling permits the controlled and cost-effective over-sampling of important target sub-populations that may be difficult to observe in statistically useful numbers through a simple random sample without requiring excessive sample sizes. In this approach the population is stratified into a set of mutually exclusive and collectively exhaustive categories or strata. Each category is assigned its own sampling rate so as to optimize the number of observations per category required to achieve accuracy and precision targets within each category. Examples of stratified sampling include oversampling transit users in heavily auto-dominated travel markets and the oversampling of small sub-populations of particular policy interest (e.g., physically disabled persons, senior citizens, etc.). In all cases, care must be taken to ensure that sub-population sizes and sampling rates are known for each category so that proper weighting across the categories can be maintained so that correct population totals can be constructed.

Cluster sampling provides a means for cost-effectively surveying sampling units for which a sampling frame is difficult to directly construct. For example, place of work surveys generally use cluster sampling, in which a sampling frame of business establishments (BEs) is constructed and a sample of BEs is then randomly drawn. These BEs are clusters of workers (the actual sampling unit of interest), which can be enumerated and sampled (interviewed) within each selected BE. It is also important to note that households are clusters of individuals (each and every individual belongs to exactly one household), and so household-based surveys are implicitly cluster samples of individual persons.

2.2.3 Sample Size Determination

In general, the precision of population estimates derived from survey samples improves as sample sizes increase. Diminishing returns, however, exist, and there is always a point at which increased sample size is not cost-effective. Standard methods for computing sample sizes exist, but in practice sample size determination is a complicated process, for several reasons, including:

- Surveys generally consist of many questions, each one of which typically will have its “optimal” sample size. Sample size selection is therefore almost always a balance between the competing requirements of different components of the survey.
- Sample sizes are often driven by available budgets. In such cases it must be recognized that the survey precision is being driven by the budget rather than by desired precision targets.

A fundamental three-way trade-off exists in every survey among sample size (which determines survey precision), survey complexity (number of questions, etc., which determines the survey applicability) and survey cost. For a fixed budget, size-complexity trade-offs must be made; for a fixed level of survey complexity, size-cost trade-offs need to be examined; etc. Given that cost is inevitably a concern, these trade-offs play a critical role in every survey design.

Travel surveys are even more complex in terms of sample size determination, since two more “dimensions” exist in these trade-off calculations: spatial and temporal precision. A few hundred observations may be sufficient to determine region-wide mode shares. But in order to understand mode shares at the fine-grained spatial scale of traffic zones (as is typically required for transportation analysis and modelling), then much larger sample sizes are required. Similarly, accurate representation of trip rates by time period requires larger sample sizes than if 24-hour totals are only required.

As a very practical example of these trade-offs and of how different decisions concerning these trade-offs can be made, consider the difference between typical Canadian and US household survey practice. As discussed in more detail in Subsection 2.3, in Canada, very large (up to 5%) samples are collected so that (a) reliable origin-destination (O-D) trip matrices can be constructed at a traffic zone level and (b) sufficient numbers of transit trips are observed so that detailed transit mode and route choice models can be constructed. In order to keep survey costs within manageable limits, the survey questionnaire is kept as simple as possible (with, in particular, limited socio-economic data being collected) and only a single respondent is interviewed for the entire household (inevitably leading to under-reporting of some trip-making by non-respondents). In the US on the other hand, typically very detailed questionnaires are completed by all household members, but sample sizes are very small (typically a few thousand households for multi-million-person urban regions). While providing very detailed trip generation information, such surveys are not sufficiently large to reliably estimate detailed O-D matrices (thereby imposing constraints on the nature and quality of the trip distribution models that can be developed from such data), nor do they often provide sufficient numbers of transit trip observations to permit multi-modal mode choice models to be statistically estimated without augmentation by other data (such as onboard ridership survey data, see Section 3.3). And so, decisions concerning sample size, survey complexity and survey budget fundamentally affect the strengths, weaknesses and appropriate uses of the survey data collected.

2.2.4 Interview Type

Interviews of respondents can be undertaken in a variety of ways, which can be categorized broadly into four classes:

- *Face-to-face (personal) interview*. The interviewer meets in person with the respondent, directly puts the questions to the respondent and records the respondent's answers. When the interviewer uses a laptop to ask questions and input data, this is commonly referred to as Computer Assisted Personal Interviewing, or CAPI for short (Market Research World, 2011). This method allows for real-time verification of data, which further enhances the data quality and minimizes data gaps.
- *Mail-back survey*. The respondent is given (or sent via the mail) the survey questionnaire along with instructions on how to fill it out. The respondent completes the questionnaire and mails it back to the survey team using a stamped, addressed envelope that had been supplied to the respondent along with the questionnaire.
- *Telephone interview*. The interviewer conducts the interview of the respondent over the telephone. This type of interview is typically conducted *via* CATI (Computer Assisted Telephone Interview) systems, which enable the interviewer to enter data directly into the computer and verify the validity of the input information.
- *Web survey*. The respondent is given access to a web site that presents the survey questionnaire in an interactive format to the respondent. The respondent completes the survey on-line in response to questions and other prompts provided to the respondent from the computer program.

Detailed information on the design, testing, organization and implementation of the above survey methods is well documented in the literature (TRB Travel Survey Methods Committee, 2012; Transport Canada, 2009; NCHRP 2007, 2008). Face-to-face, mail-back and telephone interview methods are discussed further in Section 2.3, while web-based surveys are discussed further in Section 4.

2.2.5 Sample Recruitment, Contact Methods and Incentives

Once a sample has been drawn from the sampling frame, contacting and recruiting each respondent is the next critical step in the process. In most surveys the majority of non-respondents arise from failure to recruit them to the survey. Failure to recruit selected respondents results in extra survey costs if these respondents are to be replaced, and, more worrying, can introduce significant non-respondent bias into the survey if non-respondents are atypical in their travel behaviour relative to respondents. A classic example of this sort of bias is the case in which successful contacts are never achieved with very frequent trip-makers while very low frequency trip-makers are much more likely to be contacted and willing to complete the survey.

Contact methods often depend on the type of survey being conducted but generally include letters mailed to the respondents and telephone or e-mail contacts. Contact and interview media can be mixed. It is common for example, for telephone surveys to use a letter mailed to the respondent a week prior to the interview as the initial contact. This forewarns the respondent that they have been selected for the interview, helps differentiate the travel survey interview from other, routine "market research" calls, and provides an opportunity to impress upon the respondent the importance of the survey and the value of their participation in it. On the other hand, it is being found in recent surveys that it is increasingly difficult to get full addresses for telephone sampling frames, particularly for apartment dwellers, for whom apartment numbers are often missing. In the most recent Quebec City survey, it was decided not to send advance letters to households due to the difficulty in obtaining complete addresses, with a compensating increase in ongoing publicity efforts to increase public awareness and interest in completing the survey.

Incentives (lottery tickets, cash, small gifts, etc.) are often used as additional inducement for respondents to participate in the survey. These are generally not used for large-scale travel surveys but are often used in smaller-scale surveys, especially those involving considerable respondent burden. Mixed results are reported in the literature with respect to the cost-effectiveness of incentives on response rates (Tooley, 1996; Zimowski et al., 1997; Kurth, et al., 2001; Singer, 2002), with results clearly varying with the specifics of the survey and the population being surveyed.

For mail-back and web surveys, which depend upon the respondent self-completing the questionnaire, *follow-up* contacts with persons who have not yet responded to the survey can often significantly increase response rates by reminding the respondents about the survey and reinforcing the importance of their participation. These follow-ups can, again, be by mail, telephone or e-mail, depending upon the survey, the original contact medium, etc.

Dillman (1978, 2009) emphasizes the need for a “total design” approach to recruitment, incentives, follow-up, questionnaire design and all other aspects of Cochran’s survey design process in order to maximize response rates. Researchers who closely follow Dillman’s methods routinely report much higher response rates than in cases where much less attention is paid to design details (e.g., among many others, Miller and Crowley, 1989 and Hoddinott and Bass, 1986).

2.2.6 Temporal Considerations

Numerous temporal considerations exist in survey design. These include:

- Survey observation period.
- Day(s) of the week to be surveyed.
- Seasonality effects.
- Observation method.
- Cross-sectional versus time-series surveys.

Each of these issues is discussed briefly below.

Survey observation period: The most common travel survey observation period is one (24-hour) day, but both shorter time periods (morning peak period, etc.) and longer time periods (two or more days, one week, multi-week) are also used, depending on the survey purpose. Longer observation periods obviously provide additional information concerning a given respondent’s travel behaviour but with associated increases in survey cost and respondent burden (Pendyala and Pas, 2000; Chalansani and Axhausen, 2004).

***Day(s) of the week*:** Most travel surveys focus on weekday travel, given the emphasis in transportation planning on weekday peak-period travel as the dominant concern in facility and service design. Weekend travel, however, has very different travel patterns than weekday travel and therefore stresses the transportation system in different and important ways. Furthermore, there is considerable day-to-day variation in travel between weekdays and between weekend days, as well as important interactions between weekday and weekend travel decisions. In particular, some aspects of seven-day patterns, notably shopping trips, have evolved considerably in the past two decades. Understanding travel behaviour and needs over the entire week is also of importance in designing transit services that provide "real" alternatives to the private, car, particularly in suburban neighbourhoods. Activity-based travel models also are drawing increasing attention to the interplay between weekday and weekend travel decision-making.

As noted above, the most common type of household survey collects travel behaviour information for a single 24-hour day, usually a weekday. In large surveys, respondents are interviewed over a period of several weeks or even months, with each respondent being interviewed on a randomly selected day. The result is observations of travel distributed approximately uniformly across different days of the week (e.g., Monday through Friday) and, typically, across multiple weeks. These data are then generally averaged to yield a description of travel behaviour for a "typical" or "average" day.[5] While universally adopted, this approach does represent a form of temporal aggregation. Each day of the week has its own characteristic travel behaviour (trip rates, O-D patterns, etc.) and the "typical" day we construct from our survey data in a literal sense never truly occurs. Similarly, variations in weather and other factors will introduce week-to-week variations in travel which, it is assumed, "average out" when the survey data are combined to compute "typical" travel behaviour.

Seasonality effects: Over and above day-to-day variations, travel behaviour obviously varies by season. Summer travel patterns are different from the rest of the year due to schools generally not being in session and due to prevalence of summer vacations. Winter destination and modal choices may be different than in the spring and fall due to weather effects. If seasonality is explicitly of interest for one reason or another, then some form of time-series[6] survey is required to observe the season-to-season variations of interest. Most travel behaviour surveys, however, are cross-sectional[5] in nature, in which case the season in which the survey is to be conducted must be explicitly considered. By far the most common practice in Canada is to conduct surveys during the fall season, between Labour Day and early December. A number of factors underlie this decision, including:

- Fall and spring are usually assumed to represent "typical" travel conditions, in that they avoid both "atypical" summer conditions and winter weather effects.
- Weather conditions are hopefully moderate and have little overall impact on the observed travel behaviour.

---

[5] Or, equivalently, when building travel models, data from all days are used to construct models of travel behaviour for a "typical" day.

[6] See below for definition and discussion of time-series versus cross-sectional surveys.

- Fall is often preferred to spring since it has fewer holidays (only the Thanksgiving long weekend) and schools are in session throughout the season.

Regardless of season chosen, if travel surveys are to be repeated over time (e.g., every 5 years, as is the practice in urban regions such as Toronto and Montreal, see Subsection 2.3.3 below), then the same season should be used in each repetition of the survey so as to maintain comparability of the data gathered over time.

Also, given the Canadian climate, weather effects should always be considered in the design of any data collection effort, as well as in the analysis of the data collected. This may include ensuring that any instrumentation used can withstand the range of weather conditions under which it is expected to operate, possibly suspending surveying (or other data collection efforts) on extreme weather days, and, at a minimum, always recording the actual day(s) on which the survey/data collection occurs so that weather information can be attached to the data records for possible use in analyzing the data. In particular, increasing interest in weather effects on travel behaviour exists (e.g., Saneinejad et al., 2011), while weather effects are very important to account for when modelling transportation air pollution dispersion and exposure (Hatzopoulou and Miller, 2010).

Observation method: Respondents can be asked to record their travel behaviour in two ways:

- *Retrospectively*, in which they are asked to recall their travel behaviour over some prior time period. For travel behaviour, the prior time period is typically the previous day ("what did you do yesterday?"), although longer historical time periods may be useful for some purposes ("in your previous job, what mode did you use to travel to work?). When questioning retrospectively, two modes of questioning are possible: *"typical behaviour",* in which the respondent is asked to self-report their typical or "most likely" behaviour ("what mode to work do you typically use?"), and "*actual behaviour*" or "*revealed preference*" in which respondents are asked what they actually did in a given situation. The Canadian Census work trip mode question is an example of the "typical behaviour" approach. Most household travel surveys employ the "actual behaviour" approach ("what trips did you make yesterday"). A strong argument exists for requesting actual behaviour whenever it is feasible to do so, since (a) it avoids people "integrating" over their past behaviour to come up with what they think is their typical behaviour, which may generate reporting biases of various types, and (b) it allows for the observation of "random events" (someone who usually drives to work taking transit because the car is in the repair shop, etc.) that, with sufficient sample size, will yield a statistically more valid representation of a "typical day's" trip-making (i.e., a typical day includes a certain number of "untypical" events).
- Using *diaries* to record travel behaviour as it occurs, i.e., a trip (or activity) diary is filled out by the respondent as the behaviour occurs (or very shortly afterwards). Diaries are usually more expensive to process and may impose additional respondent burden but they generally produce more detailed and accurate data (Golob et al., 1997).

***Cross-sectional versus time-series surveys*:** A cross-sectional survey is one that is executed at effectively one point in time.[7] A time-series survey is one that it is repeated over time, typically at regular intervals. Four basic types of time-series surveys exist:

- *Repeated cross-sections*. This involves repeating the same survey at multiple points in time with independent samples of respondents being drawn each time. This approach allows behaviour within the system to be tracked over time, but the behaviour (and changes in this behaviour) or individual persons, households, etc. cannot be tracked over time. The Transportation Tomorrow Survey (TTS) program in Toronto (DMG, 2007) and the series of Montreal surveys dating back to 1970 (http://www.transport.polymtl.ca/eodmtl/) are both examples of on-going repeated cross-section surveys in Canada.
- *Panel surveys*. In a panel survey, a set of respondents is recruited and then repeatedly interviewed over time. This allows changes in behaviour among these respondents to be tracked over time, which is extremely useful for building dynamic models of travel behaviour. Panels, however, are expensive to construct and maintain, respondent burden tends to be high (with corresponding attrition in panel membership often occurring), and agencies may be unwilling to wait for the results of a long-term panel to materialize. As a result, panels are relatively rare in transportation planning (Golob, et al., 1997), although a few notable examples exist such as the Dutch Mobility Panel (Wissen and Meurs, 1989), the Puget Sound Transportation Panel[8] and the Swedish Panel Study[9] Examples of Canadian travel panels include the Toronto Travel Activity Panel Survey (, Doherty et al., 2004; Roorda and Miller, 2004) and the Quebec City Travel and Activity Panel Survey (Lee-Gosselin, 2005; Roorda et al., 2005). Multi-day/week surveys can be thought of as short panels, among the most notable of these is the German six-week Mobidrive survey (Chalasani and Axhausen. 2004).
- *Retrospective surveys.* Retrospective surveys ask respondents to recall past activities. Such surveys have been shown to produce useful data for “longer-term” decisions such as auto ownership (Roorda et al., 2000), residential location processes (Hollingworth and Miller, 1996; Haroun and Miller, 2004) and employment careers, providing a cost-effective, attractive alternative to panel surveys for such long-term processes. They are not generally useful for short-term, day-to-day travel decisions, except, as already discussed, for the “what did you do yesterday” type of questions typical of many travel surveys.
- *Continuous surveys.* An alternative to repeated cross-sectional surveys is a continuous survey approach in which smaller samples of respondents are surveyed on an on-going basis. Potential advantages of the continuous approach include avoidance of massive swings in budget and staffing and the constant availability of “current” data. The statistical challenges of “integrating”

---

[7] A large travel survey for a large urban region will typically take several weeks or months to complete. Such surveys are generally still considered as being cross-sectional in nature, since the observations are typically pooled to describe the average behaviour that occurred during the survey interval.

[8] http://psrc.org/assets/1484/PSTP_summary.pdf.

[9] The Swedish Panel Study gathered data on a variety of household market and nonmarket activities, including two detailed time-use surveys. http://www.nek.uu.se/faculty/klevmark/hus.htm.

over this continuous smaller-sample data stream to construct "current snapshots" of system behaviour, however, are non-trivial. Numerous examples of continuous surveys exist in both Europe and Australia (Zmud et al., 2011).

2.2.7 Questionnaire Design

Questionnaire design is critical to the quality of data that are obtained through the survey and to minimize respondent burden and understanding of the questions being asked. A detailed discussion of good questionnaire design is well beyond the scope of this review. Excellent manuals on questionnaire design include (Dillman, 1978, 2009).

2.3 HOUSEHOLD TRAVEL SURVEYS

2.3.1 Introduction

Undoubtedly, the "workhorse" of population-based travel surveys is household surveys. These are commonly used world-wide to gather primary data concerning urban person travel behaviour. Subsection 2.3.2 provides an overview of household travel surveys, including definition of key terms and discussion of typical methods. Subsection 2.3.3 then provides a brief history of household travel survey practice in Canada.

2.3.2 Overview of Household Travel Survey Methods

Statistics Canada defines a *household* as "a person or group of persons who occupy the same dwelling and do not have a usual place of residence elsewhere in Canada or abroad. The dwelling may be either a collective dwelling or a private dwelling. The household may consist of a family group such as a census family, of two or more families sharing a dwelling, of a group of unrelated persons or of a person living alone."[10]

The household is an extremely useful sampling unit for general purpose travel surveys for a number of reasons, including:

- Although travel is executed by individuals, travel decisions are made to a large degree within a household context, both in terms of the resources and constraints available to individuals for their trip-making (income, car availability, opportunities to ride-share, etc.), and in terms of household-based activities (joint shopping and recreation trips, serving dependents, household chores, etc.) (Miller, 2005).
- Individual-based sampling frames are not generally used in transport surveys,[11] whereas household-based sampling frames are commonly constructed. As noted in Section 2.2, the household represents a "cluster" structure for individuals, permitting cluster-bases sampling of individuals to be achieved through the household survey. That is, by enumerating and randomly

---

[10] http://www.statcan.gc.ca/concepts/definitions/household-logement-eng.htm.

[11] Although individual sampling frames are used for special-purpose surveys, such as surveys of driver licence holders.

sampling households we can implicitly enumerate individuals and construct a statistically valid sample of individual trip-makers. Traditionally, household sampling frames have included:

  - Assessment rolls.
  - Utility billing lists.
  - Telephone directories.

Thus, in a household travel survey a representative sample of households in the survey target area is selected from an appropriate sampling frame. The sampling procedure used is usually a simple random sample or else a spatially stratified sample, in which a target sampling rate is applied within each of a set of districts within the study area to ensure a geographically balanced sample (see, for example, DMG, 2007). Census data are often used to provide population totals that are used to establish sample weights for population level estimates. Detailed data are gathered on each household member aged above a pre-defined threshold. The data collected vary from one survey to another but they can broadly be classified into three types:

- Household data, including location, type, number of persons, number of vehicles, etc.
- Personal data, including age, gender, employment status, occupation, possession of valid driver's license, etc.
- Trip data, including origin and destination of trip, trip purpose, trip mode(s), start time of the trip, etc. In some surveys, further details are collected on transit trips (e.g., access stop, route, number of transfers, etc.).

Depending somewhat on the interview type (discussed further below) and other design considerations, the trip data can be collected either retrospectively ("what trips did you make yesterday?") or by means of a trip diary that is filled out by the respondent for the survey day(s). Increasingly, particularly in the US and Europe, trip diaries are being replaced by *activity* diaries, in which emphasis is placed on recording the attributes of out-of-home activities and the trips associated with participation in these activities rather than just on the trips per se. This activity-based approach facilitates the development to activity-based travel models, but also, it is argued, results in higher reported trip rates than those obtained in traditional trip-based surveys, in which non-work/school and/or non-home-based trips tend to be more readily forgotten (DMG, 1991, 1993; Ashley et al., 2009).

Activity-based surveys also facilitate the collection of *in-home* activities in addition to the *out-of-home* activities (trips) which have been the traditional focus of travel surveys (Doherty and Miller, 2000; City of Calgary, 2002). In-home – out-of-home activity trade-offs are becoming increasingly important for understanding:

- Trip generation rates for entertainment, shopping and social activities (among others).
- The impact of ICT (information communications technology) on work (and school) tele-commuting.
- The increasing propensity for people to work at home rather than an out-of-home location.

The collection of in-home activities is still not common, even in activity-based surveys, and the practical modelling of in-home activities is in its infancy, but the need to consider in-home activity effects in our analyses and modelling is growing. At a minimum, information concerning “work-at-home” and tele-commuting is required and is generally collected even in conventional travel surveys.

Household surveys have been conducted in a variety of ways which, as already discussed in Section 2.2, can be categorized broadly into four classes:

- Face-to-face (FTF) interviews.
- Mail-back surveys.
- Telephone interviews.
- Web surveys.

In addition to the interview type, another critical design issue is the determination of the respondents within each selected household who undertake the interview. Two options exist:

- A single respondent completes the interview on behalf of all household members. This single respondent is typically arbitrarily self-selected on the basis of who answers the door or the telephone (FTF or telephone interviews) or volunteers to complete the questionnaire (mail-back and web surveys), although targeting a particular household member is also an option.
- Every household member is interviewed.

Single-respondent surveys are much easier to administer, generally cheaper to execute and typically will have higher response rates than all-member surveys, which are difficult and expensive to coordinate and impose more burden collectively on the household. They do, however, have serious potential to generate significant non-respondent bias in that the respondent often will not have complete information concerning the non-respondents’ travel behaviour, unless special care is taken to minimize this risk.[12] Options for reducing non-respondent biases include:

- Having the respondent prepare in advance for the interview by collecting the required travel data from the other household members. It is generally difficult to monitor the effectiveness of this strategy.
- Adopting a mixed strategy, in which all household members fill out a trip diary, but only one household member provides the diary information to the survey staff.
- In telephone interviews have each household member come to the phone in turn to complete his/her portion of the interview. This increases response accuracy but obviously can pose interview scheduling issues (a randomly placed call to the household has a good chance of not finding all household members available). Multiple household member responses do occur informally in current telephone surveys, and at one time were systematically encouraged by MTQ in their surveys, but this practice was discontinued for unknown reasons.[13]

---

[12] This potential bias may also increase with household size, which increases the burden on the respondent to report the actions of more household members.

[13] Private communication from Pierre Tremblay, MTQ.

The first great round of household travel surveys occurred in the 1950's and 1960's throughout North America as part of the general emergence of comprehensive long-range urban transportation planning and the associated development of the first generation of computerized four-step travel demand forecasting model systems that occurred during this time period (Meyer and Miller, 2001). These surveys universally were face-to-face, reflecting both the resources available for surveys in this era as well as the relative lack of alternative technologies. These surveys were generally retrospective (previous day trips collected) from a single respondent, although FTF interviewing certainly does not preclude either the use of trip diaries or the interviewing of multiple household members. While the response rate and data quality of household FTF interviews are the highest relative to other methods of household surveys, it is the most expensive on a per-completed survey basis (Sharp and Murakami, 2005). As a result, FTF is rarely used today for surveys, except for highly specialized, small-sample surveys in which the complexity of the survey and the need for high-quality responses justifies the cost and effort involved.

Mail-back surveys are rarely used in North America[14] for large-scale household travel surveys, but they are commonly used in Australia, with self-completion questionnaires typically being personally delivered to respondent households (FTF travel diary interviews are also used in Australia) (Inbakaran and Kroen, 2011). The classic mail-out/mail-back survey involves mailing a questionnaire to each household in the sample, with the household then completing the survey and mailing it back. Variations, however, exist, including hand delivering the questionnaire (perhaps as part of an up-front, preliminary FTF interview) with the respondent mailing the completed questionnaire back to the survey team, or mailing the questionnaire to the household but then retrieving the completed survey via a telephone interview, among, undoubtedly other permutations. The survey itself can be either retrospective or a diary.

Mail-back surveys are generally considered to yield very low response rates with lower quality data but with the compensation of being cheap to execute. The costs of printing, envelopes and postage, however are in fact non-trivial (especially on a per-response basis if low response rates are, indeed, experienced and/or if multiple follow-up mailings are employed in an effort to boost the response rate), as are the data processing costs to get the completed questionnaires into machine-readable form. Also, as Dillman (1978, 2009) and others have consistently demonstrated, very adequate response rates are, indeed, achievable if proper care and effort is taken in survey design and execution. [15]

The third method, telephone interviews, generally strikes a balance between the two previous methods in terms of data quality, response rate and cost. The random sample is typically drawn from a comprehensive list of household landline numbers or via random digit dialing (RDD). Each household in the sample is contacted by an interviewer who collects the survey data over the phone. The respondent is asked questions concerning household, personal and travel characteristics, the latter typically on a retrospective basis, although retrieval of a previously completed trip diary over the telephone is also

---

[14] Although Canadian examples exist, such as the 1999 Vancouver survey, see http://vancouver.ca/ctyclerk/cclerk/020430/a7.pdf.

[15] Road intercept O-D surveys (see Section 3) were conducted in Troi-Rivières and Quebec in 2011 using mail-out, mail-back surveys based on a sampling frame of observed licence plates. Response rates were respectively 60% and 50%. (Private communication, Pierre Tremblay, MTQ).

possible. A CATI system is generally used, which greatly enhances the efficiency and accuracy of the data recording process. Household recruitment methods can vary, but Canadian practice is typically to mail a letter to the respondent household approximately a week before the planned telephone interview explaining the purpose of the survey and indicating the time they will be contacted for the interview.

2.3.3 Canadian Household Travel Survey Practice: A Brief History

Similar to practice in the US and elsewhere, large-scale household travel surveys in Canada were first undertaken in the 1960's as part of the emergence of comprehensive urban planning and modelling. As typified by the 1964 Metropolitan Toronto and Region Transportation Study (MTARTS)[16] survey, these early surveys were FTF surveys, used utility billing lists as their sampling frame, and were coded to a traffic zone level of spatial detail. Also, as was typical in the US, few Canadian cities undertook large-scale household travel surveys in the 1970s, largely reflecting a planning culture at that time that de-emphasized long-term planning and associated comprehensive data collection and modelling activities (Meyer and Miller, 2001).

A critically important exception to this generalization, however, is Montreal, which pioneered in 1970 an on-going travel survey program which has provided a model for Canadian household surveys to this day. This seminal1970 Montreal home interview survey was a telephone-interview survey using telephone white-pages listings as its sampling frame. This survey was conducted by the CTCUM (Commission de transport de la communauté urbaine de Montréal), with the primary purpose of measuring travel demand to optimise the transit network and to plan subway extensions. Given this purpose, a distinguishing feature of the survey was the collection of detailed route choice information for all transit trips taken to support transit service planning and modelling. Regional surveys were then held cyclically in the region, approximately every five years (1970, 1974, 1978, 1982, 1987, 1993, 1998, 2003, 2008), with increasing survey areas but comparable phone-based methodologies and a sample rate of approximately 5%. Starting in 1993, the Quebec Ministry of Transportation (MTQ) became a partner in the conduct of these surveys, which had previously been solely under the jurisdiction of the Montreal transit authority. It is also in 1993 that computers were introduced at the data collection phase to assist both data collection by the interviewers and the post-processing of the data. Since 1998, a consortium of partners has been responsible for the conduct and financing of the regional surveys. A new CATI tool is used for data gathering, sample management and data quality control since 1998, and it is upgraded for each new survey. The next survey is currently being planned for the fall of 2013.

---

[16] http://sda.chass.utoronto.ca/sdaweb/html/travel.htm; http://library.mcmaster.ca/catalogue/Record/582693; http://ao.minisisinc.com/scripts/mwimain.dll/287/DESCRIPTION_WEB_NOSRCH/REFD/RG~2014-35/$/DESCRIPTION_DET_REP_NOSRCH?JUMP

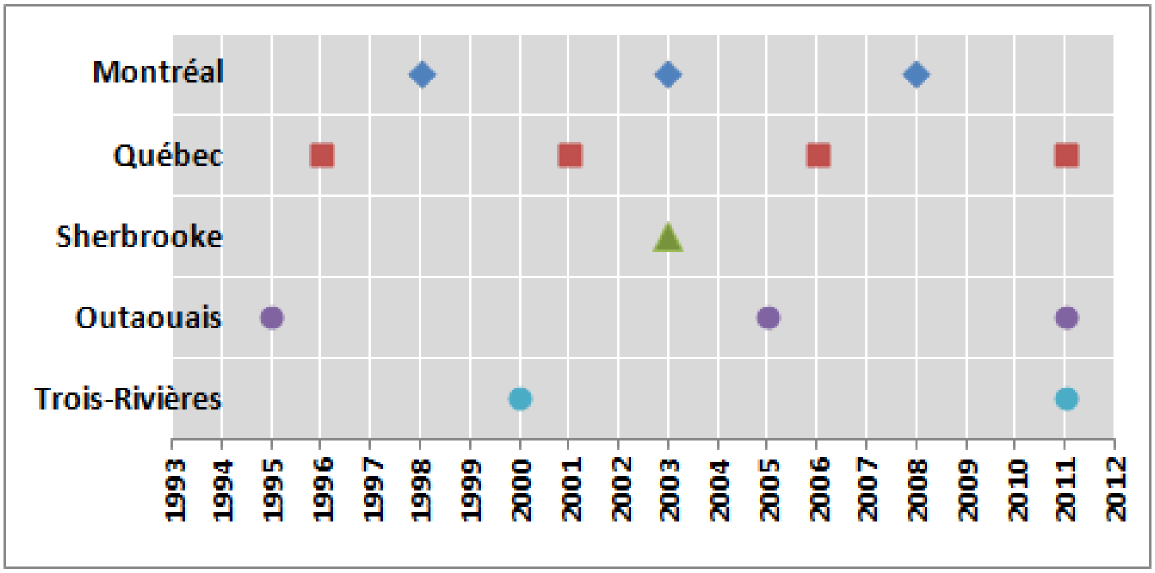


Figure 2.1: Timing of Quebec Household Travel Surveys

Household travel surveys are also routinely conducted in other regions of Quebec. In the Quebec City region, the first survey was conducted in 1972, covering a part of the region. Subsequently, regional surveys were conducted quite regularly, in 1977, 1981, 1986, 1991, 1996, 2001 and 2006. The latest survey was conducted in the fall of 2011 and was accompanied by a research project to test the potential of a web-based survey tool to gather data from people who refuse to answer to the phone survey and other hard-to-reach segments. Household surveys have also been undertaken in smaller regions of Quebec, namely in Sherbrook (2003) and Trois-Rivières (2000 and 2011). Figure 2.1 summarises the timeline of household travel surveys that have been conducted in the last 15 years in Quebec.

In the National Capital Region (Ottawa-Hull), household regional surveys have been conducted since 1986 by a consortium of partners called TRANS using a similar methodology. A second one was conducted in 1996, while the most recent one has just been conducted in the fall of 2011.

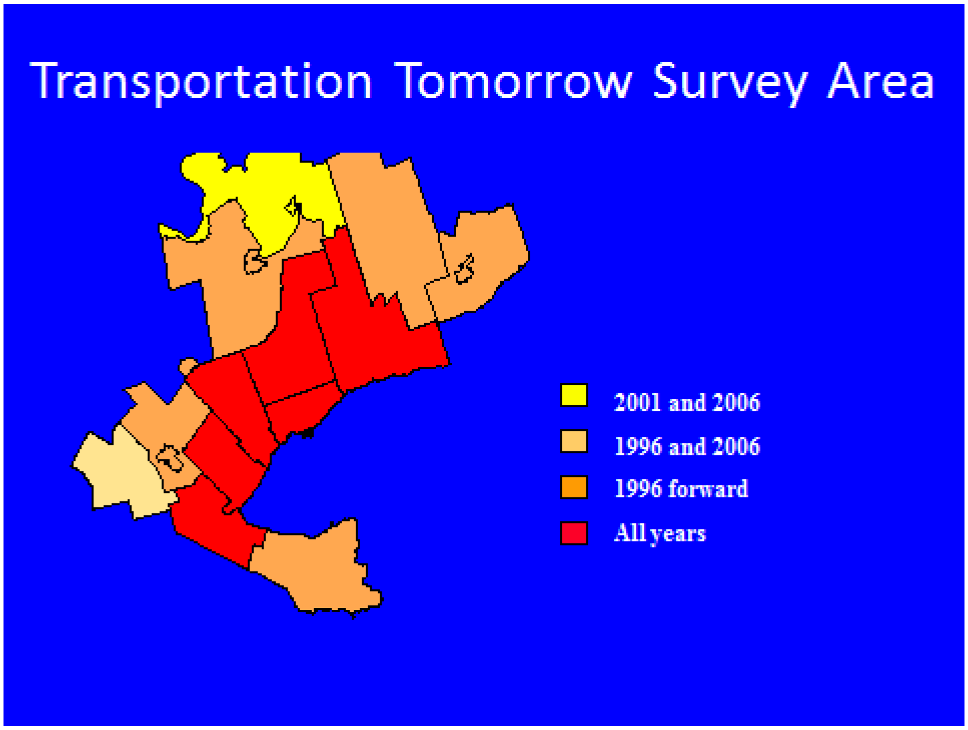


Figure 2.2: TTS Study Area

Inspired by the Montreal surveys, the Toronto region undertook in 1986 its first comprehensive household travel survey since the 1964 MTARTS survey. As in Montreal, the survey was telephone-based and emphasized the collection of transit route choice information, along with “normal” trip attributes (start time, mode, purpose, etc.). Called the Transportation Tomorrow Survey (TTS), the survey has been repeated every five years since 1986 (1991, 1996, 2001, 2006), with the most recent survey currently underway, split between the years 2011 and 2012. With the exception of 1991 (in which a 1% spatially stratified sample was used), the sample sizes have been approximately 5% of all households in the study region. As illustrated in Figure 2.2, the original 1986 survey area was the Greater Toronto-Hamilton Area (GTHA), but has expanded over the years to encompass essentially the entire Greater Golden Horseshoe (GGH) covering most of south-central Ontario. The combination of very large study area and large (5%) sample size makes TTS one of the largest (and quite possibly the largest) household survey program in the world.

A CATI system was first introduced in TTS in 1991, with upgrades to the computerized system occurring during each survey cycle. Residential, work and school locations, along with all trip origins and destinations have been geocoded since 1986 (first manually in 1986, with automated procedures being first introduced in 1991). Similarly Montreal has geocoded locations and trip ends, first at the postal code level in 1987, and then at the x-y coordinate since 1993 with various reference database available for the identification of spatial locations (trip generators, address, intersection). As in Quebec, the 2011-12 TTS is experimenting with a web-based version of the survey as an alternative to the traditional telephone interview.

Most other medium to large Canadian urban regions undertake household travel surveys on a semi-regular basis, including Vancouver (1985, 1994, 1999, 2004), Calgary, (2001, 2012), Edmonton (1994, 2005) and Winnipeg (2007), among others.[17]

While variations exist, typical Canadian household travel survey over the past two decades or more have been telephone-interview, CATI-based surveys. This trend continues, with the 2011 surveys in Toronto, Victoria and Ottawa all continuing to use CATI as the primary survey method (although, as noted above, web-based instruments are increasingly being experimented with), and Calgary will follow suit in its travel survey of 2012.

In addition to the use of CATI as the primary survey instrument, Canadian surveys are characterized by gathering one-day weekday travel data from very large sample sizes (up to 5% of household in the region) and relatively simplified questionnaires that are designed to minimize the amount of time that the respondent needs to spend on the phone (typically 15 minutes or less) while gathering the household, personal and travel data that are deemed to be of critical importance for planning and modelling purposes. While nominally all-day, all-mode surveys, given their typical design, these surveys generally do a better job in collecting peak-period travel information, especially for work and school trip purposes, as well as in collecting information for motorized (auto and transit trip-making) as opposed to non-motorized (walk and bicycle) travel modes. Also, as noted above, detailed transit route choice information is often gathered.

This approach can be contrasted with that which is taken in the US and Europe, among others, in which far more detailed surveys are undertaken with much smaller sample sizes. Trip/activity diaries (sometimes for 2 or more days) are completed, usually by all household members above a minimum age threshold. Households are often recruited via telephone (often using random-digit dialing as a contact method). Diaries are often mailed out but then collected over the telephone. The result is a relatively high quality, quite detailed dataset, but one that costs considerably more per household to collect and that does not necessarily provide the sample size required for some modelling work, among other planning applications. Also, detailed transit route choice information is typically not collected.

Given the prominence of the CATI household survey in the Canadian context, Subsection 2.4 discusses current issues and challenges facing this method, while Subsection 2.5 addresses methodological advances in this field that may help address these challenges.

2.4 ISSUES AND CHALLENGES OF TELEPHONE INTERVIEW SURVEYS

Despite their widespread use and popularity, telephone interview travel surveys have lately been faced with serious challenges that threaten their validity and outcomes. A recent study has investigated and summarized these challenges in the context of the Transportation Tomorrow Survey (TTS) conducted in the Greater Toronto and Hamilton Area (Roorda and Shalaby, 2008). Other phone travel surveys in Canada and elsewhere are faced with the same set of issues and challenges, which we discuss next.

---

[17] Additional information concerning Canadian household travel surveys will be presented in Report IV, *A Survey of Canadian Urban Transportation Data Collection Methods and Usage*.

2.4.1 Sample Selection Challenges

As noted above, sample selection is typically done by drawing a random sample from a comprehensive sampling frame which should ideally include a list of all population units being surveyed. In the case of the household surveys using telephone interviews, the most common sampling frame used has been the directory of residential telephone land-lines in the survey area. Although this sampling frame provided in the past an adequate base to draw a representative sample, it is increasingly becoming an incomprehensive list of all households in the survey area, potentially affecting the representativeness of the sample. Several emerging developments have contributed to this problem.

A major contributing factor has been the growing number of households with no land-lines, where members rely solely on their cell-phones. According to the 2010 Residential Telephone Service Survey of Statistics Canada, "…in 2010, 13% of households reported they used a cell phone exclusively, up from 8% in 2008". This phenomenon is even more evident for households with relatively young residents (18 to 34 years old) where 50% of such households reported their exclusive use of mobile telephones (up from 34% in 2008). Cellphone-only households are not listed in residential phone registries, and are therefore underrepresented in telephone survey samples. The growing numbers of cellphone-only households and the resulting sampling issues are noted in the literature on survey design of many developed regions, such as the Chicago Regional Household Travel Inventory (Bricka, 2007) and the National Survey of America's Families (Abi-Habib et al., 2003), to name two. Research has shown that the socioeconomic characteristics and trip patterns of individuals in cellphone-only households are different from those with land-lines, which makes this a significant sampling issue due to the potential sample bias that might be introduced if not treated carefully. Keeter and Kennedy (2006), through a comparison of land-line and cellphone-only samples, showed that cellphone-only Americans tend to be younger, less affluent, and less likely to be married or own their home. Another recent study has found that university students in Toronto with a land line are more likely to live in houses, with parents, and to live in suburban areas than students without a land line (Dumont, Shalaby and Roorda, 2012). They also make fewer trips in total, fewer discretionary trips, more transit and auto trips, and fewer active trips than students without a land line.

Another issue with phone land-line sampling is the growing subscription to phone services through Voice over Internet Protocol (VoIP). VoIP is a technology that allows users to make voice calls using a broadband internet connection instead of a regular (or analog) phone line service. VoIP providers allow users to keep their phone numbers, including the original area code, when moving to a different city or country, since the telephone service is provided through the internet. This causes issues for surveys using a phone land-line directory sampling frame, since households outside the survey area may be sampled and contacted while households with external phone numbers but residing in the survey area may not be captured.

The National Do Not Call List (DNCL) is also considered an emerging issue in telephone sampling. While no significant effect has been observed by the Do Not Call registry in the US, Stopher and Greaves (2007) suggests that the implementation of a similar registry in other countries may raise issues for public service oriented surveys.

The extent to which the issues discussed above will grow in the future is difficult to ascertain. Currently, cellphone-only households are probably the largest group of households without listed land-line phones.

Since such households are more likely to live in apartments and to consist of young residents, survey samples drawn exclusively from land telephone lists will under-represent these household types and their members, introducing some bias into the socioeconomic and travel characteristics of the sample. This warrants an explicit treatment of sampling cellphone-only households to avoid sample bias. At this point, there are no research results to show whether VoIP only households and potential DNCL subscribers have distinct characteristics that might introduce sample bias without explicit treatment. An additional practical consideration in sample selection is the availability of registers of active cellphone and VoIP subscribers for the construction of sampling frames used in RDD methods. This is discussed further below.

We expect the challenges associated with the selection of a representative sample from land telephone lists to grow in significance over the years if the above technologies and services take a bigger share of the market.

2.4.2 Respondent Contact and Recruitment Challenges

The proliferation of telemarketers and growing use of call screening services poses a serious challenge to recruit sample households directly by phone without prior notification. In an effort to address this challenge, some household surveys using telephone interviews have started a practice of sending an invitation letter by regular mail to each household in the sample prior to the interview in order to explain the objective and significance of the survey and to specify the targeted day for the interview. The invitation letter has proven effective in improving the response rate, specifically of households living in single family housing units. Nevertheless, it has been hard to reach out to apartments by regular mail because apartment numbers are not listed in the commonly used sampling frame (i.e., telephone land-line directories).[18]

If the first attempt to contact a household fails, the survey interviewer usually calls repeatedly at later dates until a successful contact is made or a maximum number of attempts is reached. While this approach works effectively with most household types, it has been a challenge to make successful contacts with apartment households, partly because of the invitation letter problem discussed above, and because young highly-mobile apartment dwellers are difficult to find at home.[19]

The above discussion points to the fact that representative survey samples are becoming increasingly harder to select, and sample subjects are proving more challenging to successfully recruit and interview using telephone as the main survey instrument. This is a particularly acute problem for households living in apartments and where mainly young people reside.

---

[18] In the fall, 2011 Quebec survey, for a sample of 1000 household to which mail invitations were sent over 25% were returned as undeliverable (incomplete address, missing apartment number, moved with no forwarding address, etc.) (personal communication, Pierre Tremblay, MTQ).

[19] In the 2011 Ottawa-Gatineau and Quebec surveys they could not make a connection with 25-30% of listed numbers despite repeated attempts (personal communication, Pierre Tremblay, MTQ).

2.4.3 Survey Instrument Challenges

Another limitation of the household telephone interview method is its collection of retrospective data of the respondent's travel on the previous day. In addition, the respondent is asked to report (by proxy) on the trips undertaken by each other household member on the previous day. The retrospective and proxy reporting commonly employed by household telephone survey methods has long been known to under-represent travel to destinations other than work and school, namely home-based discretionary trips and non-home-based trips. This travel market has grown over the years in significance (measured by size and impact), and is expected to continue to grow in the future. The growth in home-based discretionary and non-home-based trips reflects the increasingly complex trip chains and travel activity patterns in which people engage for a variety of reasons, including: increasing numbers of multi-worker households, increasing numbers of part-time workers and multi-job workers, increased standards of living (which encourage more travel), increased car-based mobility (which encourages trip-chaining), etc. (Miller and Shalaby, 2003).

2.5 METHODOLOGICAL ADVANCES

Many jurisdictions continue to adopt the telephone interview survey method in order to allow for trend analysis and to support legacy modelling systems. However, some efforts have been made recently to address some of the challenges listed above. For example, the dual-frame sampling technique has been implemented successfully in major jurisdictions (specifically Washington and Chicago) as an effective solution to some of the contemporary issues faced in surveys relying on landline-based samples and contact methods. Paskota (2004) suggests that the best approach to dealing with sample selection problems is to combine more than one sampling frame and target different types of people using various methods. There are generally two approaches to capturing cellphone-only households in dual-frame sampling. The first approach involves matching the names and addresses of all cell-phone users (assuming a comprehensive list is available) in the survey area to those in the telephone land-line directory in the same area so as to identify cellphone-only households. Subsequently, a sample of these households is selected and contacted (via cell-phone) to complete the survey using the cell-phone or another survey instrument option (e.g., internet) if one is made available. This sample augments the main sample of households with land-lines which is selected and surveyed using telephone interview or optionally another method (such as the internet). The challenge with this approach is the requirement of a comprehensive cell-phone list, which may not be possible to compile and obtain because of the potential reluctance of the numerous service providers to disclose the lists of their customers. Additionally, it may not be feasible to distinguish between residence-based and business-based numbers, which would pose a challenge to household survey samples.

The second approach involves identifying households without listed land-line services, which is achieved through address matching of a sample drawn from an address-based sampling frame (e.g., Census list of all residential households, compiled list of residential properties from municipal taxation and assessment agencies, or list of residential addresses from Canada Post) against the land-line phone list in the same survey area. Through the matching procedure, it is possible to identify households in the sample without listed land-line phones, which include not only households relying exclusively on cell-phones but also those having VoIP phone services, those subscribing to the DNC registry, and combinations of the three types. The sampled households without listed land-line phones would then be contacted through regular mail. This is known as "passive recruitment" because it does not follow the recruitment letter with another contact by phone (as no phone contact information is available at that point) but relies on the sampled household to respond to the letter. In order to improve the response rate, sampled households would be sent numerous reminders and possibly offered an incentive to provide their contact information. In addition, such households could be offered alternative methods to complete the survey (e.g., cell-phone, internet), which helps improve the response rate.

The Washington Council of Governments Household Travel Survey in 2007 is one example of a dual-frame sample survey, using the address-based method, and has been shown to provide considerable savings in costs compared to a single frame sample with a similar level of precision. In this survey, an address-based sample was obtained, and addresses and names were matched with the list of all land-lines. Following the address matching, sampled households were assigned to one of two groups, those with and those without land-lines. Households in both sample groups were contacted initially through mail, while the latter group was offered a $50 incentive for agreeing to participate, asking households to send further contact information (Zmud, 2007). Bricka et al. (2007) provides a description and assessment of the Chicago Regional Household Travel Inventory, which also had a similar dual-frame sample.

Other efforts to address some challenges faced in conducting telephone interview surveys include the hybrid use of this method with web surveys and GPS-based surveys. These types of surveys are described in Section 4, below.

SECTION 2 – REFERENCES

Abi-Habib, N., A. Safir and T. Triplett (2003). 2002 National Survey of America's Families: Methods and Data Reliability, Washington, DC: The Urban Institute.

Ashley, D., A. Richardson and D. Young (2009). Recent Information in the Under-Reporting of Trips in Household Travel Surveys, Australasian Transport Research Forum. http://www.tuti.com.au/atrf09-ashley.pdf

Bradley, M. and Hensher, D (1992). Workshop summary: Stated Preference Surveys. In Ampt, E.S., Richardson, A.J. and Meyburg, A.H. (Eds.): *"Selected Readings in Transport Survey Methodology".* Eucalyptus Press, Melbourne

Bricka, S., S. Sen and C. Arce (2007). *Chicago Regional Household Travel Inventory – Sampling Plan*, Unpublished manuscript. http://www.nustats.com/chicago/, accessed June 2, 2008.

Cambridge Systematics (1996). *Travel Survey Manual*, Washington, DC: US Department of Transportation, Federal Transit Administration, Federal Highway Administration, the Office of the Secretary and the U.S. Environmental Protection Agency.

Chalasani, V.S. and K.W. Axhausen (2004). *Mobidrive: A Six Week Travel Diary*, Zurich: Institute for Transport Planning and Systems, Swiss federal Institute of Technology Zurich. http://www.ivt.ethz.ch/vpl/publications/tsms/tsms2.pdf

http://www.fhwa.dot.gov/ohim/trb/mobidrive.pdf

City of Calgary (2002) 2001 Household Activity Survey, Calgary: Planning and Transportation Policy Forecasting Division. http://www.calgary.ca/Transportation/TP/Documents/forecasting/household_activity_survey_report.pdf

Cochran, W.G. (1977). Sampling Techniques, 3rd Edition, New York: John Wiley & Sons.

DMG (1991). Analysis of TTS Data Bias due to Use of Informants, Toronto: Data Management Group, University of Toronto.

DMG (1993*). Under-reporting of Trips in Telephone Interview Travel Surveys*, Toronto: Data Management Group, University of Toronto.

DMG (2007) *2006 Transportation Tomorrow Survey: Design & Conduct of the Survey*, Toronto: Data Management Group, University of Toronto.

Dillman, D.A. (1978). *Mail and Telephone Surveys: The Total Design Method*, New York: John Wiley & Sons.

Dillman, D,A. (2009). *Mail, Internet and Mixed-Mode Surveys: The Tailored Design Method*, New York: John Wiley & Sons.

Doherty, S.T., E. Nemeth, M.J. Roorda and E.J. Miller (2004). Design and Assessment of the Toronto Area Computerized Household Activity Scheduling Survey, *Transportation Research Records, Journal of the Transportation Research Board*, No. 1894, 140-149.

Dumont, J., A. Shalaby and M. Roorda (2012). GPS-Aided Survey for Assessing Trip Reporting Accuracy and Travel of Students without Land Lines, in press, *Journal of Transportation Planning and Technology*.

Golob, T., R. Kitamura and L. Long, (eds.) (1997). *Panels for Transportation Planning: Methods and Applications*, Dordrecht, the Netherlands: Kluwer Academic Publishers.

Haroun, A. and E.J. Miller (2004) Retrospective Surveys in Support of Dynamic Model-Building", presented at the 7th International Conference on Travel Survey Methods, Costa Rica, August.

Hatzopoulou, M. and E.J. Miller (2010). Linking an Activity-Based Travel Demand Model with Traffic Emission and Dispersion Models: Transport's Contribution to Air Pollution in Toronto, *Transportation Research D*, 15:6, 315-325.

Hensher, D. (1994). Stated Preference Analysis of Travel Choices: The State of Practice, *Transportation*, 21, 107-133.

Hollingworth, B. and E.J. Miller (1996) Retrospective Interviewing and Its Application in the Study of Residential Mobility *Transportation Research Record 1551*, pp. 74-81.

Hoddinott, S.N. and M.J. Bass (1986). The Dillman Total Design Survey Method: A Sure-Fire Way to Get High Survey Return Rates, *Canadian Family Physician*, 32, Nov. 2366-2368.

Inbakaran, C. and A. Kroen (2011). Travel Surveys – Review of international survey methods, Australasian Transport Research Forum 2011 Proceedings. http://www.atrf11.unisa.edu.au/Assets/Papers/ATRF11_0106_final.pdf

Keeter, S. and C. Kennedy (2006). *The Cell Phone Challenge to Survey Research*, Unpublished manuscript, http://people-press.org/reports/pdf/276.pdf, accessed December 26, 2011.

Kroes, E.P. and R.J. Sheldon (1988). Stated Preference Methods: An Introduction, *Journal of Transport Economics and Policy*, 22:1, 11-25.

Kurth, D.L., J.L. Coil and M.J. Brown (2001). Assessment of Quick-Refusal and No-Contact Nonresponse in Household Travel Surveys, *Transportation Research Record: Journal of the Transportation Research Board*, 1768, 114-124.

Lee-Gosselin, M.E.H. (1996): "Scope and potential of Interactive Stated Response data collection methods", Conference on Household Travel Surveys: New Concepts and Research Needs, Transportation Research Board, Conference Proceedings 10, pp. 115-133, National Academy Press, Washington DC.

Lee-Gosselin, M.E.H.: "A Data Collection Strategy for Perceived and Observed Flexibility in the Spatio-Temporal Organisation of Household Activities and Associated Travel", in Timmermans, H.J.P. *Progress in Activity- Based Analysis*. Elsevier, 2005, pp. 355-371

Le Vine, S., Lee-Gosselin, M.E.H., Sivakumar, A., Polak, J. (2012) "Design of a Strategic-Tactical Stated-Choice Survey Methodology Using a Constructed Avatar", forthcoming in *Transportation Research Record*.

Louviere, J.L., Hensher, D.A. and Swait J.D., *Stated choice methods: analysis and application*, Cambridge University Press

Market Research World (2011). http://www.marketresearchworld.net, accessed December 25, 2011.

Meyer, M.D. and E.J. Miller (2001). *Urban Transportation Planning: A Decision-Oriented Approach, 2nd Edition*, New York: McGraw-Hill.

Miller, E.J. (2005). Propositions for Modelling Household Decision-Making, *Integrated Land-use and Transportation Models: Behavioural Foundations*, M. Lee-Gosselin and S.T. Doherty (eds), Oxford: Elsevier, 2005, pp. 21-60.

Miller, E.J. and D.F. Crowley (1989) "Panel Survey Approach to Measuring Transit Route Service Elasticity of Demand", Transportation Research Record 1209, 1989, pp. 26-31.

Miller, E.J. and A.S. Shalaby (2003). Evolution of Personal Travel in Toronto Area and Policy Implications, *Journal of Urban Planning and Development*, 129:1, pp. 1-26.

NCHRP (2007) *Technical Appendix to NCHRP Report 571: Standardized Procedures for Personal Travel Surveys*, Contractor's Final Report for NCHRP Project 8-37, Web-Only Document 93, Washington, DC: National Cooperative Highway Research Program, revised December, 2007. http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_w93.pdf

NCHRP (2008) *Standardized Procedures for Personal Travel Surveys, NCHRP Report 571*, Washington, DC: National Cooperative Highway Research Program.

http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_571.pdf

Paskota, M. (2004). Sample Design and Survey Error, 7th International Conference on Travel Survey Methods, Costa Rica.

Pendyala, R.M. and E.I. Pas (2000) Multi-Day and Multi-Period Data for Travel Demand Analysis and Modeling, Resource Paper, Workshop in Multi-Day and Multi-Period Data, *TRB Transportation Research Circular E-C008: Transportation Surveys: Raising the Standard, Proceedings of an International Conference on Transport Survey Quality and Innovation, May 24-30, 1997,* Washington, DC: Transportation Research Board.

Roorda, M. J., M. Lee-Gosselin, S. T. Doherty, E.J. Miller and P. Rondier (2005). Travel/Activity Panel Surveys in the Toronto and Quebec City Regions: Comparison of Methods and Preliminary Results. CD Proceedings of the *PROCESSUS Second International Colloquium on the Behavioural Foundations of Integrated Land-use and Transportation Models: Frameworks, Models and Applications*. Toronto, June.

Roorda, M.J., E.J. Miller and A. Mohammadian. (2000). Toronto Area Car Ownership Study: A Retrospective Interview and its Applications. Transportation Research Record 1719. pp. 69-76.

Roorda, M.J., and E.J. Miller (2004). Toronto Activity Panel Survey: Demonstrating the Benefits of a Multiple Instrument Panel Survey. CD Proceedings of the *Seventh International Conference on Travel Survey Methods*. Costa Rica, August.

Roorda, M. and A. Shalaby (2008). *Transportation Data Collection in the Greater Golden Horseshoe: A Framework and Priorities for Improvement*, final report for the Data Management Group, UTRAC, University of Toronto.

Rose JM and Bliemer M (2012) 'Stated preference experimental desing strategies' in *Transport Economics: Critical Concepts in Economics (Volume 1)*, ed. David A. Hensher, Routledge, Oxford, United Kingdom, pp. 304-32

Saneinejad, S., M. J. Roorda and C. Kennedy (2011). Modelling the Impact of Weather Conditions on Active Transportation Travel Behaviour. *Transportation Research Part D: Transport and Environment*, 17(2) 129-137.

Sharp J. and E. Murakami (2005). Travel Surveys: Methodological and Technology-Related

Considerations, *Journal of Transportation and Statistics*, 8(3) 97-113.

Singer, E. *The Use of Incentives to Reduce Nonresponse in Household Surveys* (2002). Report no. 051, Ann Arbor: Survey Methodology Program, The University of Michigan Institute for Social Research Survey Research Centre. http://www.isr.umich.edu/src/smp/Electronic%20Copies/51-Draft106.pdf

Statistics Canada (2010). 2010 Residential Telephone Service Survey, http://www.statcan.gc.ca/daily-quotidien/110405/dq110405a-eng.htm, page consulted December 11h, 2011.

Stopher, P., and S. Greaves (2007). Household travel surveys: Where are we going? *Transportation Research Part A: Policy and Practice*, 41(5) 367-381.

Tooley, M.S. (1996). Incentives and Rates of Return for Travel Surveys, *Transportation Research Record: Journal of the Transportation Research Board*, 67-73.

Transport Canada (2009). *Canadian Guidelines for the Measurement of Transportation Demand Management Initiatives - User's Guide*.

TRB Travel Survey Methods Committee -ABJ40 (2011). *The On-Line Travel Survey Manual*, http://www.travelsurveymanual.org/, accessed December 26, 2011.

Wissen, L.J.G and H.J. Meurs (1989). The Dutch Mobility Panel: Experiences and Evaluation, *Transportation* 16:2, 99-119.

Zimowski, M., R. Tourangeau, R. Ghadialy and S. Pedlow (1997). *Nonresponse in Household Travel Surveys*, Washington, DC: US Federal Highway Administration.

http://tmip.fhwa.dot.gov/resources/clearinghouse/docs/surveys/nonresponse/non.pdf

Zmud, J. (2007). Washington Full Survey Design Documentation, Unpublished manuscript, Austin, Texas: NuStats.

Zmud, J., B. Chlond, T. Kuhnimhof and A. Richardson (2011). *Feasibility Study of a Continuous Household Activity Survey Program (CHASP),* prepared for PTV NuStats and City of Calgary.

3. CHOICE-BASED SAMPLE SURVEYS

3.1 INTRODUCTION

A choice-based sample survey is one in which the sampling frame consists of a set of people who have all made the same choice (mode, route, etc.). The most typical choice-based sample surveys used in urban transportation are roadside surveys (in which the sampling frame consists of trip-makers who have chosen to travel by car along a given route) and transit onboard surveys (in which the sampling frame consists of people who have chosen to take transit using a given transit line). These are discussed in some detail in Sections 3.2 and 3.3, respectively. Many other examples of choice-based sample surveys, however, exist, including air traveller surveys at airports, parking lot surveys, shopping mall surveys (or other “special generator” locations), etc.

Choice-based sample surveys are a very efficient data collection approach in cases in which information is only required for the targeted population (car drivers, transit riders, etc.). This is particularly the case for small or specialized populations that might be difficult and/or expensive to sample effectively through general-population-based sampling frames (such as are used for household surveys). The increasing cost of conventional household travel surveys, as well as the increasing need for a variety of reasons to study specific segments of trip-makers in detail, makes choice-based sample surveys an attractive option (Pendyala et al., 1993).

On the other hand, choice-based sample surveys obviously do not provide sufficient data for the analysis or modelling of travel choices that lie outside the survey context. For example, a roadside interview survey can provide information about the origin-destination pattern of trip-makers using the road at the survey point, but it cannot provide the basis for modelling the decision to drive relative to taking another mode of travel for the observed trips. Choice-based sample survey data, however, can be combined with other, more general data sets (e.g., a household travel survey) to provide needed details concerning the given process that may be missing or inadequately observed in the more general survey. A common example of this occurs in many US mode choice modelling exercises in which the household travel survey does not contain sufficient transit trips to adequately estimate the mode choice model parameters.[20] In such cases, it is common to use onboard transit ridership survey data to augment the household survey data (Cambridge Systematics 1996). In such applications, care must be taken to correctly combine the choice-based data within the model parameter estimation process, but standard procedures for doing so exist (Manski and Lerman, 1977).

Choice-based sample survey design generally involves the same steps and concerns as population-based surveys. The biggest differences between the two types of surveys include:

- Sampling frame definition is theoretically straightforward: it is the population of users of the facility being surveyed. Care, however, must be taken to accurately count this population, and cases exist in practice in which this is not a trivial task.

---

[20] This situation arises through the combination of the low transit mode splits (5% or less) experienced in many US cities combined with the small sample sizes typically employed in US household travel surveys.

- Sample selection and contact is accomplished through some form of *intercept* method, in which trip-makers are identified while in the process of making their trips. Different intercept/contact methods are used, depending on the specific type, constraints and needs of a given survey.
- The sampling methods most commonly used are generally either simple random sampling or sequential sampling (e.g., stop every 10th car), although stratification of trip-makers is also conceivable (e.g., over-sample families in an airport intercept survey).
- The most common interview methods include:
    - *Face-to-face* (FTF) is very commonly used, typically involving a quick interview at the point of interception. Paper-and-pencil or a computer-aided laptop system can both be used for data recording. In situations which permit the interviewer and respondent to sit down together and spend a few more minutes together (such as in an airport lounge) more detailed computer-aided questionnaires (including stated-preference/response type experiments) can be undertaken.
    - *Mail-back* surveys can be handed out to the respondent at point of contact, leaving the respondent to subsequently fill out and return the completed survey form. These questionnaires are typically very short and succinct in nature.
    - Having identified respondents at the intercept point, the respondents are later contacted and recruited for the survey. This subsequent contact is done generally by either mail or telephone[21], with the survey being either a mail-back or telephone interview. A typical example of this approach is the gathering of car license plates at various intercept points followed by contacting the car owners corresponding to the set of observed license plates.

3.2 ROADSIDE INTERCEPT SURVEYS

3.2.1 Methods and Design Issues

Roadside intercept surveys are one of the oldest and best established ways of collecting road-based travel data by intercepting and questioning people during the course of their travel (Lestina et al., 1999, Beirness and Beasley 2010, Cambridge Systematics 1996). Traditionally, such surveys have been used to collect auto origin-destination (O-D) trip patterns, and so they are frequently referred to by practitioners as “O-D surveys”.[22]

In cases in which transit and non-motorized travel is not of interest, roadside O-D surveys possibly can be used in place of household travel surveys to develop models of road-based trip generation, distribution and assignment. In all cases, such survey data can be used to supplement more general travel surveys, especially in situations in which data concerning types of trips is sparse within the general

---

[21] E-mail addresses typically are not known but could, of course, be used if they are known.

[22] Roadside intercepts can also be used to count and/or interview persons travelling by transit (which is one form of transit user intercept surveys, discussed in the next section) or even pedestrians and cyclists. The focus in this section, however, is on roadside intercepts of auto users, by far the most common application of this survey type.

survey (e.g., O-D patterns in small sample general surveys), as well as to help in travel model calibration/validation.

Cambridge Systematics (1996) presents a complete manual for roadside intercept surveys. It classifies intercept surveys into four main categories:

- *License plate surveys*, in which license plates are recorded (either manually or, more typically, by being automatically photographed) as vehicles pass by intercept points. The recorded license plate numbers are then matched with vehicle registry information to identify the home locations of the drivers, and either a telephone or mail-back survey is conducted to obtain detailed information concerning the observed trip.[23]
- *Roadside handout surveys*, in which vehicles are stopped at intercept points and a mail-back questionnaire is given to the driver to complete and return by mail at some subsequent point.
- *Roadside interview surveys*, in which vehicle drivers are interviewed “on the spot” once they have been stopped at an intercept point.
- *Combined roadside handout and interview surveys*, in which a quick roadside interview is combined with a more detailed mail-back handout.

In addition to the usual issues of sampling rate, questionnaire design, etc., important sample design issues for roadside surveys include:

- Choice of survey method (license plate survey, roadside handout, etc.).
- Selecting what road sections/links to sample in the survey.[24] Budget and other resource issues often limit the number of intercept points that can be feasibly used. These must, therefore, be chosen with care so as to maximize the information obtained about overall system performance.
- In the case of stopping vehicles, feasible locations for stopping the vehicles must be found that introduce minimal disruption to the vehicle stream and that are safe for drivers and interviewers alike. Police involvement is generally required to ensure safe and orderly implementation of the survey sites and compliance of drivers to stop for the interviews.
- Selecting the type(s) of road users to survey. If certain user types are to be over-sampled or excluded, then care must be taken to develop an appropriate stratified sampling procedure.

Typical roadside intercept survey data elements include:

- Travel data: trip purpose, arrival and departure time, travel time, vehicle type, origin and destination addresses, number of persons in the vehicle, travel routes and frequency of trips

---

[23] Note that if licence plate numbers are observed both entering and exiting a cordon around a study area then computerized number matching algorithms can be used to link together vehicle entry and exit points, thereby generating estimates of origin-destination flows passing through the cordon. The successfulness of this process depends on the having a large number of observation points and high licence recording rates, so that a large number of vehicles are successfully observed at both entry and exit.

[24] Roadside survey intercept points are often located at selected locations along a cordon or screenline so as to maximize the observation of flows into and out of the cordoned area.

- Demographic data: residential location, household size, occupation, household income, age and gender.
- Attitudinal data, such as perception about congestion, potential use of alternative routes, alternative means of travel etc.

3.2.2 Strengths & Weaknesses

Roadside interviews have the advantages that they provide immediate collection of the required data and they generate very high response rates – it is very difficult for a driver, once stopped, to refuse to do a quick interview. Disadvantages of the approach include:

- The method is clearly intrusive and disruptive.
- Setting up intercept stations is costly and requires extensive coordination with police, etc.
- The method generally is not practical for high-volume roadways.
- Relatively few interview stations usually are feasible to maintain.
- The interview generally needs to be very brief and so very limited information is typically collected.
- The legal authority to stop people to question them is required.

Roadside handouts share the disadvantages of roadside interviews in terms of needing to stop vehicles on the roadway. The mail-back handout questionnaire potentially can be somewhat more complex than the roadside interview survey since it is not completed on the spot. Response rates relative to the roadside interview, however, inevitably will be lower, especially since little generally can be done with respect to incentives and/or follow-up contacts to encourage completion of the questionnaire. The handout approach is less disruptive to traffic flow than roadside interviews, since the cars generally are stopped for shorter periods of time (just long enough to explain the survey and hand out the mail-back questionnaire) and may be less costly to run since road side survey staff are able to process a higher number of vehicles per unit time. This may also lead to higher sampling rates, which may at least partially compensate for lower response rates.

The combined roadside interview/handout method in some circumstances may combine the strengths of both individual approaches in that it ensures a good base of critical information collected during the short interview, complemented by more detailed information gathered through the mail-back handout. Motivation for completing the mail-back questionnaire may be higher, given that the respondent has already completed the short roadside interview. The challenges of stopping vehicles along the road, of course, remain with this method.

License plate surveys possess several attractive features, including:

- They do not require stopping vehicles and can be used in even very high-volume situations.
- They are relatively cheap and straightforward to implement.
- The telephone or mail-back survey can be reasonably complex in nature.

The license plate approach, however, is not without its own difficulties. These include:

- Privacy concerns. Many people object to their movements being monitored without their consent.
- Cooperation of the motor vehicle registry office to release vehicle owner addresses and/or telephone numbers is required.
- The driver of the vehicle during the recorded trip may not be the owner of the vehicle.
- The follow-up contact (by either mail or telephone) must occur as shortly as possible after the trip is recorded so that the respondent's recall of the trip is as fresh and accurate as possible.
- Response rates may be lower.

Regardless of the survey method employed, common limitations of roadside intercept surveys include:

- Only data concerning motor vehicle trips are obtained. Travel behaviour by other modes is usually ignored.[25]
- Given the survey methods employed, generally fairly limited information concerning the respondent's trip-making behaviour, personal attributes, etc. can be obtained, with this often being limited to just the observed trip.
- Data collection is limited to the (typically few) sites at which intercepts occur. Thus, information concerning O-D patterns (and road-based travel behaviour in general) within the urban area is inevitably limited to those trips which happened to choice a route passing by one of the intercept points. The approach thus is most effective when studying a specific corridor or when travel patterns are otherwise constrained in a way that they can be usefully observed using a relative handful of key observation points.

Cordon/screenline-based O-D surveys (for both road and transit, see the next section) can be combined with household travel survey data to improve the estimation of O-D matrices. For maximum effectiveness this obviously requires coordinating the timing of the roadside and household interview surveys. In addition to the use of household survey data, roadside data can also be combined with traffic count data and traffic route assignment models to generate improved O-D matrix estimates (Guy and Fricker 2005). The availability of WiFi technology and mobile phone network positioning capabilities also open possibilities for improved monitoring and modelling of road-based travel (see Section 4).

3.3 TRANSIT USER INTERCEPT SURVEYS

3.3.1 Methods and Design Issues

Transit users can be intercepted at the transit stations or stops while they are waiting for their bus or train or onboard while they are travelling on the vehicle. Such surveys can provide origin-destination information for transit riders together with personal and household characteristics. Such data are

---

[25] In principle pedestrian and bicycle trips can be surveyed using the same methods that are described for herein for motor vehicles, although this is rarely done in practice.

typically used for estimating transit ridership, understanding behavioural patterns of transit users and characteristics of transit riders. Onboard survey data are used for scheduling and operations planning, long-range planning and design, performance analysis, preparation of statistics and reports, and market evaluations (Cambridge Systematics 1996, TAC 2008). Surveying passengers onboard vehicles can be easier to implement compared to intercepting respondents at stations or stops since the respondent is “captive” while onboard the vehicle, although interviewing passengers onboard very crowded vehicles can be challenging. Regardless of intercept location, transit user intercept surveys are typically generically referred to as “onboard” surveys or “transit rider” surveys.

Onboard surveys are standard tools for virtually all transit agencies in Canada and elsewhere (Seskin and Stopher 1998). Hartgen (1992) explains the evolution of the concept of choice-based sample survey conducted by the transit agencies as opposed to relying upon large scale travel surveys. Stopher (1992) presents an example approach for developing a transit ridership forecasting method at the route level by using transit onboard survey data. Onboard surveys are also used for monitoring route-level ridership trips, transit rider travel patterns and attitudes, and before-and-after assessment of route service changes on transit ridership.

Typical approaches to conducting onboard surveys:

- The driver of the transit vehicle hands out the survey questionnaire to the passengers and passengers either return the completed form while leaving or mail it back to the designated address. This technique is suitable for buses or other vehicles without all-door boarding.[26]
- Surveyors aboard the transit vehicle distribute the survey and collect the completed responses. They may interview the passengers onboard as well as count the boarding passenger. This technique is suitable for trains, large buses, BRT or LRT.[27]
- Surveyors aboard the transit vehicle distribute the survey and passengers mail the completed responses later to the designated address.

Cambridge Systematics (1996) presents a complete manual for transit onboard survey. Typical data elements that can be collected through transit onboard surveys include:

- Travel data: boarding and alighting location/stop/station, trip purpose, arrival and departure time, travel time, origin and destination addresses, access and egress mode, transit routes, fare payment type, auto ownership and auto availability for the trip
- Demographic data: household size, occupation, household income, age and gender
- Attitudinal data, such as perception about transit service, customer satisfaction, etc.[28]

---

[26] Also providing union agreements permit such activities, there are not safety issues involved in the method and boarding volumes are low enough for this to be practical.

[27] This assumes that trips are long enough to complete the survey while the passenger is onboard.

[28] For example, Habib et al. (2009) used onboard survey data of Calgary transit to investigate riders’ attitudes towards transit services.

Baltes (2002, 2003) presents a best-practice manual for transit onboard surveys. It describes the necessary steps for conducting a successful onboard survey of public transit customers, and provides a clear understanding of the total customer surveying process and its importance in planning and transit service design. Specific items discussed include various methods of onboard data collection, questionnaire design, sample size determination, data entry procedures, data reporting procedures and data archiving procedures. Blash et al. (2002) argue that although standard manuals are available for designing onboard surveys, significant challenges typically face agencies in their design and execution of practical surveys. These are mostly sampling related, relating to the selection of the specific routes and time periods to survey.

TCRP (2005) presents a comprehensive discussion on transit survey techniques. It provides a summary of transit agencies' experiences in planning and implementing onboard and intercept surveys. It reports a survey of 52 agencies, 96 percent of which had experience in conducting onboard surveys. Survey results reveal that large agencies typically conduct five or more onboard/intercept surveys per year and small agencies typically conduct surveys every 1 to 3 years. Onboard surveys conducted by large agencies are primarily focused on specific routes or geographic areas, but surveys conducted by small agencies often involve the entire transit system. This report points out that proper guidance for the questionnaire design and incentives is often missing in existing practice.

Chu (2006) examines the feasibility of using local transit onboard surveys for the measurement of state-level transit performance levels. Their definition of performance level is defined by the proportion of choice and captive transit users among all transit riders. The report suggests that if such estimates of the local transit agencies are available and can be synchronized in time from all individual agencies, then the state-level measurement can be developed by using a stratified sampling technique by using the local agency-level weights. However, it seems that many transit agencies overlook the issue of defining captive and choice riders in their onboard surveys. Moreover, in many cases, onboard surveys do not cover the whole transit system of the corresponding agencies.

As one example of Canadian practice, transit authorities in the Montreal regularly conduct on-board surveys. AMT provides detailed information regarding the conduct and key facts from their annual surveys on-board their suburban rail lines (http://www.cimtu.amt.qc.ca/enquetes/trains/Index.asp). Typically, questionnaires are distributed at train stations along with counts of boarding passengers for further sample expansion. Information on the traveller (age, gender, home location, car ownership) and the current trip (boarding and alighting stations, time of departure, access mode, destination point, and fare type) are gathered. The sample rate is usually around 70%.

3.3.2 Strengths & Weaknesses

Transit user intercept surveys are less disruptive than roadside intercept surveys as the respondents are targeted while they are waiting at the station/stop or in-vehicle. Surveys that are filled out while the passenger is onboard (or while waiting at a stop/station) do not impose any extra time on the passengers (i.e., the response burden is very minimal) and so these surveys tend to have higher response rates than mail-back approaches.

As with road intercept surveys, the major limitations of onboard surveys are (a) they do not provide information on non-transit users and (b) they are limited to the transit routes/stations that are selected to be surveyed. For this reason some transit agencies conduct household telephone interviews to supplement their transit user intercept surveys (TCRP Synthesis 63, 2005).[29,30]

As is discussed further in Section 4, GPS methods and automatic passenger count systems on transit vehicles can be used to augment traditional onboard survey methods in a variety of ways.

---

[29] Indeed, a major motivation for initiation of the Toronto Tomorrow Survey (TTS) in Toronto in 1986 was the desire for the TTC to have detailed travel behaviour data for transit planning purposes.

[30] TransLink maintains an online panel called "TransLink Listens" to obtain customer feedback. This generates a contact list that can be used to target customers making specific choices (e.g., pass holders) for various surveys.

SECTION 3 REFERENCES:

Baltes M. A. (2002). *Costumer Surveying for Public Transit: A Design Manual for Customer On-Board Surveys*, final report for National Center for Transit Research NCTR, NCTR - 416 – 083

Baltes M. A. (2003). Conducting a Successful On-Board Survey of Public Transit Customers, Proceedings of the 2003 Mid-Continent Transportation Research Symposium, Ames, Iowa, August 2003.

Beirness D.J. and E.E. Beasley (2010). A Roadside Survey of Alcohol and Drug Use among Drivers in British Columbia. Traffic Injury Prevention 11: 215–221

Blash L., J.D. Rogers and R. LeGates (2002). Urban Public Transit Riders: Surveying a Population on the Move, Technical report, American Association of Public Opinion Research.

Cambridge Systematics (1996). Travel Survey Manual, Washington, DC: US Department of Transportation, Federal Transit Administration, Federal Highway Administration, the Office of the Secretary and the U.S. Environmental Protection Agency. http://www.travelsurveymanual.org/

Chu, X. (2006) Using Local Transit On-Board Surveys for State-Level Measurements, Report NCTR-576-07, BD549-16, National Center for Transit Research (NCTR) , University of South Florida

Guy, B.P. and J.D. Fricker (2005). Guidelines for Data Collection Techniques and Methods for Roadside Station Origin-Destination Studies, Publication FHWA/IN/JTRP-2005/27. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 2005. doi: 10.5703/1288284313368

Habib, K.M.N., L. Kattan and M.T. Islam (2009). “Model of Personal Attitude towards Transit Service Quality”, forthcoming in Journal of Advanced Transportation: 10.1002/atr.106

Hartgen D. T. (1992). Coming in the 1990s: The Agency-Friendly Travel Survey. Transportation 19: 79-95

Lestina D., M. Greene, R.B. Voas and J. Wells (1999). Sampling Procedure and Survey Methodologies for the 1996 Survey with Comparisons to Earlier National Roadside Survey, Evaluation Review 23: 28-46

Manski, C.F. and S.R. Lerman (1977). The Estimation of Choice Probabilities from Choice Based Samples. Econometrica 45 (8): 1977-1988.

Pendyala, R.M., K.G. Goulias, R. Kitamura and E. Murakami (1993). Development of Weights for a Choice-Based Panel Survey Sample with Attrition, Transportation Research Part A 27A: 477-492

Seskin I.M. and P.R. Stopher (1998). Spatial Variations in Attitude Towards Expanded Public Transit Service, Transportation 15: 211-232

Stopher P. (1992) Development of Route Level Patronage Forecasting Method. Transportation 19: 201-220

TCRP Synthesis 63 (2005). Conducting On-Board and Intercept Transit Survey Techniques, final project report, Washington, DC: US Federal Transit Administration, Project J-7, Topic SH-05.

Transportation Association of Canada (TAC) (2008). Best Practices for the Technical Delivery of Long-Term Planning Studies in Canada Final Report. Project Report ISBN 978-1-55187-261-7

4. TECHNOLOGY-BASED DATA COLLECTION METHODS

4.1 INTRODUCTION

For many years now, data collection methods have been deployed that rely on various types of technology. Usable data for planning, either strategic or operational, are being gathered using technologies as part of travel surveys or simply from observation and administrative systems.[31] Hence, technologies are helpful in various ways: enhancing classical travel survey methods, enriching sets of already available datasets, or generating new types of information that were not previously available. In addition, there is an increasing body of research suggesting that combining multiple survey instruments allows the survey process to reach a more complete and representative sample of individuals and households (Bonnel et al., 2009).

This section focuses on the various data collection methods that benefit from technology. The methods, application areas, strengths and weaknesses of the following technologies are discussed in this section for each of the following technologies:

- **Web–based surveys**: Web-based surveys are now quite extensively used to complement or replace traditional population-based and choice-based travel surveys. Multiple recent applications reflect an increasing proportion of people with access to an internet connection. Strengths and weaknesses are presented based on previous experimentation and the expected future role of web-based surveys as part of the larger set of methods to gather data on the movement of people is discussed.
- **GPS and other portable devices**: For more than two decades, GPS (Global Positioning System) technology has been implemented to allow monitoring of vehicles, such as buses or probe cars, aiming to assess travel times on specific routes. Portable GPS devices are also increasing used to understand spatial-temporal circumstances of person movements. Portable GPS devices are often used in combination with a survey (travel diary) or with other devices, such as accelerometers or health sensors, to simultaneously collect data on additional aspects of travel such as energy expenditure for health studies. The use of mobile telephones has also recently exploded. Many mobile telephones have built-in GPS devices that record movements of the telephone. The variety of portable devices that can be used to collect person movement data, including GPS, smartphones, accelerometer and personal health sensors are discussed.
- **Passive data streams:** Various types of passively collected, continuous datasets describing the use of transportation systems are examined. Examples of such data are transaction records from smart cards or onboard Automatic Passenger Counters (transit systems), carsharing and bikesharing systems, information from roadside detectors such as inductance loops, video and Bluetooth™ detectors, and automatic transit passenger counters using infrared beams or treadle mats. The collected data are not always well-suited for operational or strategic planning. But

---

[31] In this latter case, the technology typically is installed to address non-planning purposes. Usage of the data collected for transportation planning purposes represents an additional benefit derived from this technology, but may also mean that the technology and resulting database is not “optimized” for planning purposes.

with appropriate processing they can provide rich information on the use of transportation system and unit behaviours. The applications and potential outputs are discussed.

- **Social networks**: Social networks such as Facebook™ or Twitter™ are changing the way people interact and how information is shared. The role of social network software as a recruitment and survey tool are presented. At the same time, social networks define how people interact with one another and how, where and when they travel for social activities. Therefore, we also present data collection methods for capturing information about social networks as they relate to travel.

This review of data collection technology and methods is necessarily not exhaustive. Many standard methods for collecting transportation system performance data (floating car travel time studies, methods for observing pedestrian flows along sidewalks, among many others) are not included within this review. In general, emphasis within this review has been placed on new or emerging technologies, as well as on travel behaviour (as opposed to transportation system performance) applications.

4.2 WEB-BASED SURVEYS

4.2.1 Methods & Applications

Typically, travel surveys have been conducted using face-to-face interviews, telephone interviews or self-completion surveys (see Section 2). Recently, with the increasing number of mobile telephones, the mistrust towards telephone surveys, and increased use of call display on telephones, it has become harder to reach respondents using this medium and to convince them to participate in surveys. At the same time, access to the internet has become widespread in many industrialized countries. According to the 2009 Canadian Internet Use Survey, 80% of Canadians aged 16 and older, or 21.7 million people, used the internet for personal reasons, up from 73% in 2007 when the survey was last conducted (Statistics Canada, 2010). This same survey also reveals that "a*mong Canadians living in communities with a population of 10,000 or more, 83% used the internet compared with 73% of those from communities with fewer people*".

Various implementations of web-based surveys have gathered the sorts of trip or activity data from people and/or households that are collected when conducting travel surveys using other methods (phone, face-to-face, paper and pen). Sharp and Murakami (2005) compare data collection methodologies and define elements of internet surveys:

- Respondents complete the survey on the web (self-reporting survey).
- Only households with access to internet are covered.
- Response rates are usually lower than other methods.[32]
- Quality of data varies and depends on the validation functions, interface design and question formulation.

---

[32] Although, as with most survey instruments, response rates may possibly be increased through appropriate use of incentives.

- Data are rapidly available.
- Overall cost is low, but startup costs can be high compared with data-collection cost.

Collaborative tools such as Skype™ or Facetime™, which allowing teleconferencing, can allow for direct, real-time interaction between a respondent and an interviewer but this is currently not the state of practice. Usually, respondents fill the survey on their own, to their best knowledge. The clarity and user-friendliness is, in these circumstances, a key element. As for other technology-assisted tools, the web allows for real-time validation and can automatically provide feedback on the validity of the answer. A certain level of quality control at the time of the survey participation is then possible.

Web-based surveys have similar applications as other personal or household travel surveys, including the use of all transportation networks or only the use of a specific network (transit or carsharing for instance). They can assist in collecting information on households, people, trips and preferences.

4.2.2 Strengths & Weakness, Issues & Opportunities

Table 4.1, borrowed from Bourbonnais and Morency (2011), lists advantages and disadvantages related to web-based travel surveys. It also identifies probable trends that will enhance or worsen strengths and weaknesses.

In brief, limitations of web-based surveys that will probably remain over time are shared by other survey modes:

- **Survey design**. Designing a web interface that is clear for every respondent is not a simple task. It cannot be a simple transposition of a paper and pencil survey or an interviewer interface (from a CATI survey for instance). The interface has to be immediately understood and attractive. Hence, the initial investment is one of the important costs of web-based survey. Also, technology is changing fast in the web industry and it is difficult to keep a design operational for a long time (Bourbonnais and Morency, 2011). Using a web-based survey means that resources must be applied on a regular basis to update the application and redesign the interface.
- **Low response rates**. This is a general issue with travel surveys, notwithstanding the survey mode. The community will have to examine how to make sure that the survey tools are attractive and able to attract and retain participants. Efficient tools to recruit participants are required. For web-based surveys, social networks combined with interfaces for computers, tablets and smart phone will facilitate wider distribution of the call for respondents.
- **Sample representativeness**. This issue depends on various elements:
    - The quality of the sampling frame (i.e., the quality and representativeness of the list of people or households from which a random sample is drawn). In the case of web-based surveys, sampling can be performed by sending invitation letters or emails to selected households or people or by sending invitations at large. In the former case, there is still a certain control of the sample composition but not in the latter case. Also, gaining access to a representative list of emails is not trivial and emails are usually linked to people and not households.

  - Only people with access to internet can fill the questionnaire. The scale of this issue will decrease with time but there will probably always be people without access and they probably will not have similar attributes and behaviour to those who have access. Also, skills in the use of the internet are not uniformly distributed over the population. However, the general level of familiarity will increase over time.

- **Data quality**. The quality of data gathered using self-reported surveys highly depends on the quality of the interface. The web allows using various types of validation and graphical representation and it can help in ensuring good quality data. Post-processing procedures to further validate and impute, when necessary, are still required. As with many survey instruments, "call-backs" (by phone or email) can be used to verify responses if need be.

Table 4.1: Advantages and Disadvantages of Web-Based Travel Surveys

ADVANTAGES		**DISADVANTAGES**	
Description	**Trend**	**Description**	**Trend**
↑ / ↑↑: should become even better with time *→: should stay the same*		*↑ / ↑↑: should mitigate with time* *↓: could become even worse with time*	
Low marginal cost (Armoogum et al., 2009)	→	Uneven internet access and connection speed	↑↑
Less restrictive (the respondent chooses when to respond) (Armoogum et al., 2009)	→	Uneven usability of respondents (Alsnih, 2007)	↑
Rapid collection of data	→	Response rates are low and declining (high solicitation for a wide variety of surveys) (Armoogum et al., 2009)	↓
Real time validation of data (Armoogum et al., 2009, Timmermans and Hato, 2009)	↑	The concentration of respondents is decreasing (because of multitasking)	↓
Adaptability and flexibility (the questionnaire can be adapted to the respondent by analyzing previous answers)	↑↑	The questionnaire must be adapted to a web interface, which makes it more difficult to compare with other survey modes (Braunsberger et al., 2007)	→
Allows the study of respondent behaviour during the interview	↑↑	Email directories are of poor quality and/or difficult to obtain	?

ADVANTAGES		DISADVANTAGES	
Description	**Trend**	**Description**	**Trend**
Ability to change quickly while conducting the survey (to reduce ambiguity, add new questions or promote / reduce certain technical behaviour)	↑↑	Recruitment must often be performed using another mode	?
Makes administration of survey less arduous	↑	Technical differences from one respondent to another (browser, operating system, equipment, etc.). (Armoogum et al., 2009)	↑
Allows the dissemination of real-time results and presentation of preliminary statistical data to respondents at the end of their interview	↑↑	Several biases can be induced (coverage bias, sampling error, measurement error, non-response bias) (Alsnih, 2007) “Younger people have a strong preference of joining the CAWI survey while elderly or retired people clearly favour the telephone interview” (Kagerbauer et al., 2011)	↑
Allows a large number of simultaneous interviews (provided the equipment and connection are able to cope with higher traffic)	↑	Different understanding of some questions by respondents can cause poor quality or bias among some groups with different education, age or social profiles	→
Opportunity to ask random questions	→		
Increases the level of interactivity and visualization (Bonnel et al., 2009)(Armoogum et al., 2009)	↑↑		
Capability to survey hard-to-reach groups (Riandey and Quaglia, 2009)	↑↑		

Still, web-based surveys have potential and can enrich typical survey processes. TCRP 69 (2006) synthesizes how transit agencies use web-based surveys and provides a list of elements that agencies have found to be successful:

- "*Start simply with web-based surveys to learn the differences between web-based surveys and other survey methods*". It is necessary to understand differences with respect to response behaviour as well as travel behaviour and it is being examined by various authors: Bayart and Bonnel (2008) propose a comparison between face-to face and web for household survey; Potoglou and Kanaroglou (2011) compare telephone and web-based survey for collecting household information.
- "*Attempt to collect databases of e-mails from customers and potential customers to use as a sampling source for research.*" Recruitment is a critical task in the survey method process. There is currently no comprehensive email directory. For a specific transportation network such as a transit network, it is probably easier to access a representative sample at ticket selling points, at information kiosks, through information tools (internet pages, text messages).
- *"Apply web-based survey methods in a multi-method survey environment to improve response rates by providing response alternatives and to enable the transit researcher to gain the benefits of web-based survey data and techniques. This finding is made knowing that measurement error is an issue with multi-method surveys; therefore, this must be balanced against the benefits*." The importance of moving from single method to multi-method surveys and the resulting challenges has been discussed for quite a few years in the field. A review of the main questions arising in multi-method / multi-instrument settings was proposed by Goulias (2000) ten years ago. Challenges still remain.
- "*Research the issue of coverage error and try to minimize sampling bias".* This is directly related to the unequal distribution of internet access among the users as well as the difficulty to get access to good quality lists of emails or contacts for users, as discussed previously.
- "*Remain cautious but optimistic about including web-based surveys in research programs as the survey methods and the Internet mature*." The speed at which access to technology is increasing and changing the potential of the web to gather data from larger and more representative sets of people. Access and ease of use will improve but there will continue to be some population segments with limited ability to use the internet. This will increase the importance to provide other types of media to gather information and maybe settle for supervised filling of surveys. Still, there will be improvement and web will become one of the classical tools to conduct travel surveys.

Additional possible advantages of web-based surveys are: (1) questions are guaranteed to be posed to the interviewee in a consistent manner (tone, etc.); and (2) interviewees have the flexibility to complete the survey at a time of their own choosing.

Finally, it should be noted that web-based surveys can exploit direct access to "Google Map" capabilities for direct entry of trip O-D and other spatial data, thereby eliminating the need for post-survey geocoding (and the pre-survey assembly of geographic files required for geocoding).

4.2.3 Applications in Canada[33]

As experience with web-based surveys grows, they are becoming more common in Canadian urban travel survey applications. The 2011/2012 implementation of the Transportation Tomorrow Survey in the Greater Toronto and Hamilton Area, which is currently underway, includes a web version of the survey. This web survey has been completed by approximately 10,000 responding households although the quality of the data provided has yet to be determined (Steuart, 2011). These web survey completions are in addition to the standard telephone interviews that have been conducted in approximately the same manner since 1986.

In Quebec, a research project was started in 2010 to evaluate the potential of web-based surveys to complement regional surveys. A first test was conducted in the spring of 2011 during the Trois-Rivières regional survey where people who refused to participate in the phone survey were asked if they would agree to participate in a shorter web version (single person questionnaire). Also, a sample of mobile phone numbers was also processed to recruit respondents. A second test was performed in the fall of 2011 as part of the Quebec City regional survey. Again, people who refused to participate in the phone survey were asked if they would be willing to fill a web-survey but in this second case a full household-level survey was offered through the web. Bourbonnais and Morency (2011) discuss this tool, which was also used to conduct three trip generator travel surveys (two at Polytechnique Montreal and one at University of Montreal). Other implementations are under discussions.

A web-based travel survey was also developed in 2008 to gather data from Communauto members, in parallel with the conduct of the Montreal Regional survey. The data were used to compare behaviours of typical households and those with one person member of the carsharing system (Sioui et al., 2009). An additional web-based survey was developed in 2009 for Communauto members in Quebec City as a follow-up to the system's member satisfaction survey. It featured a number of innovations to give respondents a choice between street addresses, business names, or clicking on a map (with an aerial photograph in Google Maps) to code activity locations over a seven day observation period. A comparison with paper and pencil diary methods was made (Thériault et al, 2011). In addition, the 2011 Translink Trip Diary survey has used a web-based survey as its primary instrument.

4.3 GPS AND OTHER PORTABLE SYSTEMS

4.3.1 Methods & Applications

GPS units can be installed in vehicles or transported by individuals (either as a dedicated GPS device or within GPS equipped devices such as mobile telephones). GPS applications mainly differ with respect to the precision of the spatial location and the rhythm of data collection. GPS data can be collected continuously, every second, every minute or when a specific event occurs. Typical devices will provide: longitude/latitude of position, UTC (coordinated universal time) time/data, speed of travel, direction of travel (heading), altitude/elevation, and indicators of the quality of the estimated position (e.g., number of satellites). Devices have become smaller and more efficient with time, a trend that is likely to continue. A recent review of the evolving application of GPS and other movement-aware technologies

---

[33] The information reported in this and subsequent "Applications in Canada" subsections may well be supplemented with additional findings once the study's survey of Canadian transportation agencies is completed.

in travel surveys will be found in Lee-Gosselin, Shalaby and Doherty (2010). They point to parallel streams of development: *active* applications that interact with respondents to complete and/or validate automatically-logged data, and *passive* applications that give priority to extended observation periods using intelligent post-processing software, possibly calibrated during an “active” start-up phase.

GPS can be used in real-time applications (e.g., for operational control of highway facilities) or in planning applications (e.g., to analyze, *a posteri*, spatial-temporal traces that reflect the travel behaviour of people). They can also be used on floating-car studies to gather real-time travel time data for selected roads and paths within the transportation network.[34] Planning applications are the focus of discussion in this section. This discussion is organized into the following categories:

- In-Vehicle GPS Systems.
- Portable GPS Systems.
- Mobile Telephones with GPS .
- GPS Summary .
- Emerging Applications of Portable Technologies.

4.3.1.1 In-Vehicle GPS Systems

Depending on their temporal resolution (i.e., the temporal frequency of recorded GPS data points), data from in-vehicle GPS systems can provide travel times and/or speed data on segments of transportation networks. Data can be collected using predetermined routes (probe cars) or using random routes (equipped fleet of vehicles such as taxis, minibus, shared cars for instance). The following are example applications of in-vehicle GPS systems.

In Montreal, GPS traces from the Communauto shared cars are used to study travel conditions on the highway network (Saunier and Morency, 2011). In the Toronto Area, GPS devices are used on probe vehicles to determine average travel times on pre-determined routes (MTO, 2009). Greaves and Ellison (2011) use GPS to study driving behaviour of survey participants (speed, route and driving patterns) for a total of ten weeks. Sharman and Roorda (2011) have developed automated in-vehicle GPS processing methods for stop location identification and prediction of truck stop durations (Sharman et al., in press). de Fabritiis et al. (2008) reported on a GPS-based travel time study that made use of devices (installed for insurance purposes) on 600,000 privately owned cars. This unique use of large scale GPS data allowed for real time estimates of travel speed throughout the Italian road network.

4.3.1.2 Portable GPS systems

Portable GPS units are also available and can be used to collect continuous spatial-temporal data on the movements of people. These units are often used in a survey setting and many examples can be found in the recent literature. For instance, Abt SRBI (2011) conducted a prompted recall survey in which 1750 households provided GPS data (1250 with 3-day portable GPS data and 500 with 7-day vehicle GPS). Authors of this study claim that the device and data processing are well advanced and allow estimating spatial and temporal attributes of the conducted activities. Similarly, Bohte and Maat (2009), Greaves et

---

[34] Or travel times for other modes of travel (transit, walking, etc.).

al. (2010), Wolf et al. (2011) Frignani et al (2010) and Chung and Shalaby (2005) have used a combination of GPS logs, Geographic Information Systems (GIS) and interactive web based prompted recall to determine travel behaviour. Yang et al. (2010) demonstrate how accelerometer and GPS features may be used to recognize physical activity, in particular to classify a stationary, walking, running, cycling or in-vehicle mode. Kesten's research team (Kestens, 2011) developed a multi-sensor wearable device, which includes an accelerometer, GPS, GPSR (for data transmission) and algorithm to process the data and identify activity location and transportation modes. They currently use their device for health studies. Geostats has recently conducted a GPS-only travel survey of 6000 households in Jerusalem who used a portable device for one day and were then asked details regarding their activities (prompted recall) (Geostats, 2011). Stopher et al. (2011) also discuss a GPS-only household travel survey in which second-by-second data are initially collected using a GPS device and then processed to impute trip ends, travel mode and trip purpose. Respondents are also asked to fill a three-day internet survey. 2060 households provided fully complete GPS data which were then used to validate the processing and imputation model. It would not be surprising to see this method becoming state of practice.

4.3.1.3 Mobile Telephones with GPS

According to the 2010 Residential Telephone Service Survey of Statistics Canada, "w*ireless cell phones continue to grow in popularity in Canada. More than three-quarters (78%) of Canadian households indicated they had a cell phone in 2010, up from 74% in 2008*." This survey provides other relevant results, as follows:

- The proportion of households with mobile telephones is highest in the three western provinces with 87%, 83% and 82% for respectively Alberta, Saskatchewan and British Columbia.
- The proportion is also high in Ontario where 81% of households have mobile telephones.
- The lowest proportion, at 69%, is in Quebec.
- More households are abandoning their traditional landline telephones (13% in 2010 compared with 8% in 2008). This phenomenon is even more important for young households (18 to 34 years old) where 50% of households are using only mobile telephones (up from 34% in 2008).

The increasing presence of mobile telephones in our society is affecting the possibility to reach potential respondent using land-line telephones. But mobile telephones could provide interesting data on the travel habits of telephone holders. The data from mobile telephones typically contain the following information:

- **Unique identifier of the unit**. Data are anonymized but it is possible to follow a specific unit over space and time.
- **Sequence of points with location and time stamp**. Importantly, location information from mobile phones may be provided by a variety of assisting positioning technologies (involving GSM tower locations, WiFi networks, etc.) and not just GPS.

Some applications can be downloaded by users to allow access to their spatial-temporal information by another person but these are not usually developed by public agencies. Google Latitude™ is one example.

Gonzalez et al. (2008) studied the trajectory of 100,000 anonymized mobile telephone users, whose positions are tracked for a six-month period, to understand individual travel behaviour. Available data allowed them to measure whether individuals had regular patterns. They found that "*human trajectories show a high degree of temporal and spatial regularity, each individual being characterized by a time independent characteristic travel distance and a significant probability to return to a few highly frequented locations*".

In other applications, Charlton et al. (2011) used a smart phone application to collect bicycle routes traversed by city cyclists. Users of the application could record a bicycle trip and upload the resulting tracked route. They managed to gather usable data describing 5000 bicycle trips from hundreds of users. Licoppe et al. (2009) used mobile telephone data to link location information with communication patterns in an effort to reconstruct mobility and communication patterns.

4.3.1.4 GPS Summary

GPS is not the only positioning technology used to aid surveys. But it has "blazed a trail" of new data collection possibilities, notably a radical improvement in tracing route choices, and the extension of the period of observation on a large scale. Summarizing across the various applications:

GPS data from portable devices can be used to:

- Study travel behaviour of people and measure variability in space and time.
- Validate travel survey declarations (assess non-response for instance).
- Gather spatial-temporal circumstances of trips as part of a travel survey in which respondents are asked to provide additional information (reduce respondent burden by automatically gather partial information).

On the other hand, in-vehicle GPS data can be used to:

- Estimate travel times on various routes (probe cars).
- Measure variability of travel conditions on the road network (equipped fleet).
- Study schedule adherence of transit services (GPS-based Automatic Vehicle Location (AVL) systems on buses or trains).

4.3.1.5 Emerging Applications of Portable Technologies

Of growing significance is the miniaturisation of a number of technologies that, singly or in combination, promise to improve the travel survey toolkit. Some examples:

Kelly et al. (2011) investigated the potential efficacy of a new electronic measurement device, a wearable digital camera called SenseCam, to contribute to the study of travel behaviour. They concluded that there is some considerable potential but there are still some demonstrations and development work to do.

Using a GPS device with a built-in accelerometer, Schüssler et al. (2011) collected data for a pilot of 15 individuals over four weeks. Participants carried the device during the day and charged the device overnight, when the GPS data was transmitted to a central SQL database. Then it was post-processed to find the mode and stop points. They used a fuzzy logic method to determine the mode of transportation. As data were processed, participants could see the results on the web to confirm or correct them.

Noureldin et al (2009) have pioneered applications of micro-electro-mechanical systems (MEMS), notably Inertial Navigation Systems (INS), in the GEOIDE NCE (Geomatics for Informed Decisions – Canadian Network of Centres of Excellence) project "Multi-Sensors Systems for Tracking and Mobility Applications". Although currently applied to vehicle navigation and tracking, there is considerable potential for combining sensors in person-based devices to overcome the limitations of GPS.

Exner et al. (2011) conducted experimental tests that linked information from personal health sensors (including devices that detect perspiration and heart rate) to identify stress experienced when travelling, in an effort to identify desirable urban form. Their sensors collected information with the aid of mobile web technology.

Duncan et al. (2009) integrated the use of GPS units with heart rate monitors to assess energy expenditure associated with children's movements.

In the past decade, other survey researchers have experimented with merging data from triangulation on broadcast FM stations (especially for long-distance travel), ambient sensors (noise, temperature, etc) and the vertical (Z) parameter of the GPS data stream.

4.3.2 Strengths & Weaknesses, Issues & Opportunities

The main challenges related to the use of data from GPS and other portable devices are the following:

- **Processing of data**. Raw data need to be processed to output relevant information on travel behaviour. First, the data must be cleaned, as false readings can occur from a number of causes, such as reflected signals. Algorithms are typically used to detect and delete anomalies, such as impossible position sequences, speeds or changes in direction of travel. Processes are then required to segment trajectories into trips, identify activity locations, estimate duration and type of activity (inferred trip purpose), and determine modes of transportation. Teeuw et al. (2011) argue that "*it is not possible, however, to distinguish between different types of motorized transport (bus, car, train)*" but others have developed methods allowing this imputation using exogenous data (Tsui and Shalaby, 2006). Reddy et al. (2010) propose a classification system that uses mobile telephone with built-in GPS receiver and accelerometer to determine transportation mode. Various map-matching algorithms have been developed (e.g., Chung and Shalaby, 2005; Garcia et al., 2011; Dalumpines and Scott, 2011; Schüssler and Axhausen, 2008) and their success depends on the quality of the GPS positioning, the quality of the road network layer (and complexity of the road network in the region) and the temporal frequency of data points. Processing algorithms can also be enhanced through the use of data simultaneously collected by complementary technologies installed on the GPS devices, such as accelerometers or INS. It must also be noted that the sheer volume of data to be processed can also pose challenges in terms of transmission bandwidth, data storage and associated processing capabilities required. Careful design is required to deal with these technical challenges and to ensure that the data being collected is sufficient for the purpose at hand while not swamping data transmission, storage and processing capabilities.

- **Measurement errors**. GPS devices need a few seconds or minutes before they can provide a position (satellite detection, triangulation). If a respondent activates the device when he/she starts moving then part of the trip will not be recorded. This is what is called the time-to-fix problem. Also, depending on the device used, there may be some points missing for the imputation of route for instance (low-frequency GPS or missing data due to signal loss in buildings and tunnels/underground networks). Finally, as noted above, the use of GPS devices in areas with large buildings or obstacles can result in inaccuracies due to signal reflection. This is known as the multi-path effect. Attempts to solve some of the above issues have been made and reported in Tsui and Shalaby (2006) and Li and Shalaby (2008).
- **Confidentiality**. With the increasing awareness of users, availability of such data may decline. When respondents agree to provide data to a public agency to contribute to better management of transportation networks, they deserve assurances of confidentiality. Because of the high spatio-temporal precision of GPS data, it is difficult to anonymize a data record, and thus hide the identity of an individual respondent. Thus, clear consent from respondents and high standards of confidentiality and data security are essential.
- **Sample bias**. Using technological tools to gather data may lead to lower levels of response for some types of respondent (e.g., the elderly) and this could impact the representativeness of the sample.
- **Equipment costs and logistics**. If the GPS equipment need to be provided to respondents (or installed in their vehicles, etc.) then significant costs can be involved to purchase, distribute and collect these devices.

Also, if data are not specifically gathered through a survey setting, other issues include:

- **Access to data**. Currently, data are not freely available and most are held by various companies (e.g., mobile telephone companies), often without clear understanding by the user that data are being gathered. There are significant legal and ethical considerations that must be taken into account if such datasets are purchased and utilized for purposes other than the original intent for which the user gave (or perhaps did not give) their consent.
- **Linking GPS data to trip-maker socio-economics and activity location data.** GPS provides information about where people travel but not directly about the characteristics of the persons who are travelling, modes of transportation, or about the attributes of their trip destinations (i.e., trip purposes). This generally requires surveys of the persons being tracked (Schüssler et al. 2011; Chung and Shalaby, 2005; Tsui and Shalaby, 2006), although imputation of destination activities/trip purposes is increasingly being attempted using geocoded databases of land use, etc. (Schüssler and Axhausen, 2008).

However, data from such devices have multiple potential advantages:

- **Decrease respondent burden and improve quality of data**. Stopher et al. (2009) show that the average number of trips per person per day is significantly higher from GPS respondents than from diary respondents, confirming the known phenomena of underreporting of trips in surveys. Kohla et al. (2011) also confirm the under-reporting of trips in paper-pencil surveys using GPS data collection.

- **Low cost**. Data from portable devices such as smart phones are very low cost since the respondent usually has access to the device and downloads an application. Collected data are already in a numerical format and the interface can include multiple real-time validation checks, thus reducing the cost of coding and error checking.

4.3.3 Applications in Canada

Loustau et al. (2010) have used GPS traces from the Montreal carsharing fleet to estimate various indicators of travel conditions and compared results with travel times estimates from probe car data. Saunier and Morency (2011) compared travel time /speed estimation obtained using four technologies, namely outputs from the same GPS traces.

Harvey (2009) describes a GPS assisted household time-use survey, in which a sample of approximately 2000 households in Halifax were first asked to carry a custom GPS device for 48 hours, and then were called to complete a questionnaire. Thus, it is one of few surveys that combine time diaries with GPS tracking of travel, providing a more complete picture of space-time activity.

Clark and Doherty (2010) used a combination of GPS and web-based technology in an attempt to obtain detailed information about the activity and travel planning behaviour of individuals. They used GPS tracking, followed by a web-based prompted recall survey to obtain travel behaviour information. This was followed with in-depth interviews to further explore why and how planning decisions were made.

Trevor and Hildebrand (2011) used GPS travel diaries to assess the travel habits of rural older drivers in New Brunswick.

Dumont et al. (in press) similarly applied a combination of GPS data collection, with a web-based prompted recall survey. In this case, the subjects were university students, and the intention was to determine the differences between the behaviour of individuals with telephone land lines versus those that did not have land-lines.

4.4 PASSIVE DATA STREAMS

4.4.1 Methods & Applications

Passive data streams are data collected continuously without inputs from the users and usually for a purpose other than strategic or operational transportation planning. They are typically the result of a legal or administrative operation such as making a reservation, paying for something, validating a ticket, etc. In this discussion, passive data streams are divided into two categories:

- Smart Cards and Other Transaction Data Sources.
- Fixed Road-Side Technologies.

4.4.1.1 Smart Cards and Other Transaction Data Sources

The STO (Société de transport de l'Outaouais) was the first to fully implement a transit smart card system in Canada in 2004. Since then, many other transit systems have implemented or plan to implement such tools. Pelletier et al. (2011) recently published a comprehensive literature review on the use of smart cards in public transit. Usually, smart card systems are implemented for fare management reasons. Various studies, however, have demonstrated the relevance of collected data to provide interesting statistics about both supply and demand features of transit networks. Chapleau et al. (2008) suggest that "*with a well-defined methodology, the data can be used to monitor the supply and consumption of the transit system, the mobility of its users as well as the activity pattern at trip generators*". If we focus more precisely on the contributions of smart card data to the gathering of information on the movements of people, we can identify various applications. Pelletier et al. (2011) provide a list of analytical applications relevant for the context of this project:

- **Analyze travel behaviour of transit users**: Agard et al. (2006) examined travel behaviour using multiple days of observation for individual smart cards (using STO data, the Gatineau transit network). Using a similar dataset and data mining techniques, Morency et al. (2006) measured transit use variability. Instead of relying on a typical weekday, analysis conducted with smart card data can allow observation of variability among days, weeks, seasons, and years. Park and Kim (2008) also demonstrated the possibility of using these data to better understand user habits.
- **Assess turnover rates:** Analyses were also conducted with smart card data to assess how well the transit network retains its users (Bagchi and White, 2005, Trépanier and Morency, 2010).
- **Derive Origin-Destination matrices on the transit network** (Munizaga et al., 2010).
- **Forecast travel demand**: using historical data, Park and Kim (2008) created a future demand matrix.
- **Understand impacts, on demand, of various incidents /events/context:** Using one month of data Chu and Chapleau (2011) illustrated how transaction data are impacted by the occurrence of events. Descoimps et al. (2011) analyze the impact of weather on transit demand for various population segments.
- **Enrich household travel surveys:** Bayard et al. (2008) suggests that smart card data could be used to enrich classical OD survey data and that OD survey data could enrich smart card data. Trépanier et al. (2009) compared these two datasets for a specific region and have identified many issues. There are few examples of data fusion in this context but fusion techniques have been examined by various researchers (see Part III of this report).
- **Estimate transit performance indicators.** Trépanier et al. (2007) proposed a method to impute a destination point and, using this information, developed a load profile for each transit route. Trépanier et al. (2009b) and Reddy et al. (2009) also demonstrated the usability of smart card data to estimate a set of transit performance indicators (veh-km, mean speed, veh-hr, mean trip distance, schedule adherence) for various contexts. Trépanier and Vassivière (2008) propose an intranet (internet tool only accessible to the organization's employees) tool with various operational statistics.

Data are also gathered for other types of transportation services. Carsharing and bikesharing systems are two examples:

- In the case of carsharing, members have a unique identifier and every reservation and transaction is recorded in a database for administrative purposes such as billing. These transaction datasets contain the time, date and origin station of all trips made by the members using shared cars. They are very useful for developing indicators characterizing supply and demand. A portion of the shared car fleet is equipped with GPS devices so it is possible to locate the cars in real-time and to analyze their travel patterns. In addition, all transactions are known to the operators through reservation system and paper logs in each car (that members need to complete).

- In the case of bikesharing, members also have a unique identifier and their use of shared bikes can be studied. Bikesharing records include the bicycle ID, time, date, origin station and destination station of the trips. Also, administrative data include trips made by occasional users and hence contain 100% of trips made on the network.

Such data can be used to:

- **Study travel behaviour of users**. Data allow measuring regularity of use over time (longitudinal data) and the impact of various variables (weather, holidays, change in supply, etc.).
- **Estimate performance indicators**. Continuous data on the use of the system can be used to estimate performance indicators such as utilization ratio or kilometres travelled.
- **Feed demand forecasting models**.

4.4.1.2 Fixed Roadside Technologies

There are various fixed technologies that can be used to gather traffic data: counts and loops, video streams or Bluetooth devices.

Traffic counts: Many cities are equipped with vehicle counting devices, namely magnetic loop detectors embedded in roadways. The main outputs of these devices are segmented counts of vehicles (cars vs. trucks). For speed measurement, it is preferable to use data from double loops but some researchers have proposed methods to impute speed using single loops. Quality of the data and speed estimation are highly dependent on the calibration of the loops and will vary according to congestion level.

The use of loop detector data for real time travel time estimation (e.g., to provide inputs for variable message signs) is more or less standard practice in Canada and elsewhere. Robinson and Polak (2005) model travel times on road segments using loop data. They use a classification method to derive travel times. Their method is tested using London (UK) data. They suggest that these data could be combined with probe car data to obtain more precise estimations. Using loop and probe car data to obtain travel time estimations is also now quite standard practice.

It should also be noted, however, that new technologies for vehicle counting are entering operational practice. One example of these is the "Wavetronix" technology (http://www.wavetronix.com), which uses radar to detect vehicles and measure their speeds. MTQ is in the process of replacing its traditional double-loop counters with this technology.

Video streams: Cities and provinces in Canada also often have parts of their freeway and arterial networks equipped with video cameras used to monitor or detect incidents and congestion levels. Cameras are also installed at many intersections to observe interactions between users. Various algorithms have been developed over time to extract data from traffic video cameras, by registration plate recognition or by following moving objects.

Friedrich et al. (2008) present five applications of an Automated Number Plate Recognition (ANPR) system in the context of transport planning and engineering: vehicle classification, travel time measurements, through traffic surveys, route choice observations and estimation of

Origin-Destination matrices. In the case of ANPR, specific systems need to be installed typically including: infrared detecting camera, optical colour detecting camera and infrared light emitting array of LEDs.

Saunier and Sayed (2006) propose a feature-based tracking algorithm that allows the extraction of information on individual traces of moving objects from videos. After correct calibration of the video, movements of moving objects and attributes of these movements (speed, travel times) can be derived. This algorithm was used in various settings:

- Safety analysis: pedestrian behaviour at intersections (Ismail et al., 2009), conflicts between pedestrian and cars (Ismail et al., 2010).
- Speed / travel time estimation on some road segments (Saunier and Morency, 2011).

Bluetooth: Bluetooth is a proprietary open wireless technology standard for exchanging data over short distances. Multiple vehicles have systems installed that use this technology and many devices (computers, smart phones) are Bluetooth-ready. Bluetooth devices have a unique identifier that can be captured using a proper antenna. Typically, antenna will be placed along a corridor and record time stamped unit identification number. Using records from multiple antennas makes it possible to derive travel times between antenna locations. Hence, depending on the setting, data could be used to derive O-D matrix and partial route choice of a sample of vehicles (cordon setting).

While the available data have usually been used to provide information on vehicle movements, it is also becoming possible to study pedestrian behaviour. Malinovskiy et al. (2012) investigate the feasibility of using Bluetooth for pedestrian studies using two separate sites. Their results suggest that "*given sufficient populations, high-level trend analysis can provide insights into pedestrian travel behaviour*."

4.4.1.3 Automatic Passenger Counters and Onboard GPS Units

Automatic passenger counters (APCs) are increasingly being used to provide automated counts of passengers boarding (and possibly alighting) from transit vehicles. The major challenge with such devices is to accurately differentiate between boardings and alightings when the same door is used for both movements. Similarly, the installation of GPS units onboard transit vehicles to accurately track these vehicles in time and space is increasingly common in Canadian transit systems. Onboard GPS serves many purposes (facilitate onboard stop annunciators, real-time tracking and operational control of vehicles, etc.), including coordinating with APCs to provide accurate spatial-temporal locations for the recorded boardings and alightings.

4.4.2 Strengths & Weaknesses; Issues & Opportunities

4.4.2.1 Smart Cards and Other Transaction Data

The first obvious limitation of smart card and other transaction datasets is that they only cover a portion of urban mobility: the trips made on the transportation networks covered by the transaction dataset.

A second challenge is that transaction data systems generate very large datasets. With the increasing proportion of systems and system users that use smart cards, the data sets will tend to represent 100% of the behaviour on this network. While this is valuable information, it is necessary to develop tools for the validation, imputation and processing of these data on a continuous basis. This requires significant investment of money and time.

Pelletier et al. (2011) identify additional strengths and weaknesses of smart card use in public transit. We have extracted the following advantages and disadvantages of the system with respect to data collection on people movements:

- Advantages:
  - The user role in data collection previously achieved by the survey process is minimized (Bagchi and White, 2004). This can be generalized to all passive data streams. There is no input from the traveler.
  - Trip data combined with personal data improves data quality and increases the number of statistics available compared to using solely smart card data (Bagchi and White, 2005).
- Disadvantages:
  - No information on trip purpose or on user assessment of service can be obtained (Bagchi and White, 2005).
  - In tap-on only systems, the user's ultimate destination is not obtained but imputation methods can be applied to determine alighting stop and destination (Bagchi and White, 2005). Tap-on/tap-off systems[35] obviously provide both boarding and alighting stops.
  - The research and development cost of smart card systems that can be used for travel analysis is high (Deakin and Kim, 2001).
  - There is a need for service providers to undertake surveys to validate and calibrate imputation procedures (Bagchi and White, 2005).
  - Market penetration needs to be sufficient to assure a representative sample of the population of interest (Utsunomiya et al., 2006).
  - Not all systems currently require the card to be read upon entry and/or exit. Such "honour" systems involve only random onboard checks for valid cards.

4.4.4.2 Fixed Roadside Technologies

Saunier and Morency (2010) compare four traffic data sources with respect to cost, spatial and temporal coverage, type of output data and identify the main challenges associated with each source.

According to Friedrich et al. (2008), "c*lassical stationary detectors such as induction loops can only measure volumes and local speed. They do not permit to measure travel times for longer distances or to survey the route choice behaviour of drivers*." They also indicate that methods such as Floating Car Data

---

[35] Such as the new Translink system which is scheduled for launch in 2013.

(FCD) allow these measurements but that a critical sample level cannot be reached due to high installation and data transfer costs. It is also important to note that conventional loop detectors can provide unreliable readings, are prone to failure and are expensive to install and to replace.

Table 4.2: Comparison of Traffic Data Sources

	Cost	Spatial coverage	Temporal coverage	Data types	Challenges
Floating Car	High	moderate	small	speed, travel times	small sample size
GPS	Low	Large	large	speed, travel times	matching GPS positions to the network, varying sample size, no control
Bluetooth	Low	Small	continuous	travel times	ongoing development with some detection issues
Video	moderate	Small	continuous	speed, counts, density, behaviour	ongoing development, varying accuracy

4.4.3 Applications in Canada

Various demonstrations of the potential of smart cards to develop supply and demand indicators were done using data from a small transit authority in the Gatineau region (Pelletier et al., 2011 summarizes the various studies). New studies are currently being conducted with the Montreal OPUS data (smart card for the Montreal transit system).

Bracewell (2008) used Ticketmaster data to identify the origins of the ticket purchasers for the 2010 Olympic Games in Vancouver Whistler. He found differences between the origins of ticket purchasers and the origins of those attending events (using on-site market research), presumably because people were purchasing tickets for other people, or were entering their work address, versus their home address.

A recent survey of provincial and municipal organizations in Canada indicated that over half engaged in an in-house traffic counting program (TAC, 2009). Approximately 80% of organizations that collect traffic counts are able to obtain vehicle classifications of those vehicles, thus making it possible to distinguish trucks from cars. Traffic counts are only one element of roadway network performance. The traffic counting systems seldom obtained information about other attributes such as vehicle speed (47%), vehicle length (17%) and vehicle weight (24%). Methods included automatic vehicle classification recorders, manual vehicle classification recorders, video classification counts, and electronic sensors (loop inductors, WIM piezoelectric, radar (RTMS), etc.).

4.4.3.1 Bluetooth

Roorda et al. (2009) describe the outcomes of a Bluetooth system of data collection that has been installed on Highway 401 in Toronto. While the system is preliminary, the project demonstrated that reasonable travel times could be attained from an inexpensive Bluetooth monitoring system. Bachmann et al. (in press), then used a data fusion algorithm to combine Bluetooth travel time information with information from loop detectors. They found that significant improvements in travel time estimation could be attained by fusing data from multiple sources using a variety of fusion methods.

4.5 SOCIAL NETWORKS

4.5.1 Methods & Applications

Recently, social networks such as Facebook or Twitter have changed the way people interact and access information as well as the speed at which this is done. In the transportation field, there have been few attempts to use these media to interact with travellers. Many transit authorities are providing information to their users through text messages, tweets or Facebook status but it has not really been tested to recruit survey respondents. According to a survey conducted by 6S Marketing (2009), 70% of Canadians say they use social media and Facebook is the most popular social networking site. A 2009 study among Canadians aged 18 years and older by Ipsos Reid showed that women (59%) are more likely to have a social network profile than men (52%) (Dewing, 2010). This study also confirmed that Facebook was the dominant social networking site among Canadian adults. It also confirms that the use of internet and social networks is not uniformly distributed across the population, differences being observed between ages, dwelling locations (higher in urban areas), linguistic group (higher among Anglophones in the older population).

Social networks can be used to recruit respondents for web-based transport surveys. In this method, however, the sampling frame cannot be controlled. The sampling method is snowball sampling (also known as chain referral sampling or respondent driven sampling) (Gile and Handcock, 2010). Invitation to participate is first sent to a convenience sample that acts as seed for recruitment. The process is hence non-probabilistic and depends on the initialization process (selection of seeds).

There are few documented examples of using these media to construct a sample. Benfield et al. (2006) discuss the use of various media to recruit respondents. According to these authors, "*different recruitment procedures can have different effects on the resulting sample and (b) the right recruitment procedure, with some luck, can yield interestingly large samples for the study*". In one of their surveys, the authors used a snowball sampling technique and sent an initial recruitment e-mail to 60 friends, colleagues and family. They managed to yield 189 responses in one month. They have estimated that the snowballing effect had stopped after the third or fourth iteration.

Social networks also have the potential to be mined for planning relevant information. One can often find locational / trip-making information on people's blogs or tweets that, for example, describe where they went for lunch or a comment/complaint about traffic/transit. While perhaps qualitative in nature and derived from ill-defined populations, such data can be obtained at fairly low cost and may prove useful for certain planning purposes.

4.5.2 Strengths & Weaknesses, Issues & Opportunities

The main advantage of using social networks to recruit respondents is its obvious low cost. Disadvantages include:

- Sampling bias: only people connected to one of the network used to disseminate seeds are reached
- No sampling frame (no control over temporal distribution of collected data, across days of the week for instance)
- There are some security issues with respect to the transmission of data or links using the web or networks: risks of hacking, data corruption, email blockages are still possible.

Social networks are also a subject of interest for travel surveys since they have an impact on the travel associated with social behaviour. Research by Carrasco et al. (2008, 2011) has developed methods for determining a person's social network using a variety of survey techniques. Crowdsourcing as an emerging technology may also provide opportunities for the collection of travel behaviour data, although formal applications of these very new methods for transportation planning purposes do not seem to yet exist.

4.5.3 Applications in Canada

Few, if any, applications of using social networking applications (e.g., Facebook, Twitter) to assess transportation behaviour have been found in Canada. However, combinations of paper/pencil surveys and technology recruitment have been used to identify the nature of social networks. Carrasco et al. (2008) assessed the composition of social networks and their impacts

on activity travel behaviour. Carrasco et al. (2008) used a variety of technologies and contact methods including mobile telephone, email, text-message and face to face meetings.

SECTION 4 REFERENCES

6S Marketing (2009). Social Media Survey Results, http://www.6smarketing.com/social-media-survey-results/ (page consulted December 15th 2011)

Abt SRBI (2011) Household Travel Survey with GPS data loggers,

Agard, B., Morency, C., Trépanier, M., 2006. Mining public transport user behaviour from smart card data. In: 12th IFAC Symposium on Information Control Problems in Manufacturing – INCOM 2006, Saint-Etienne, France, May 17–19.

Alsnih, R. (2007). Characteristics of web based surveys and applications in travel research. Travel Survey Methods. Quality and Future Directions. Elsevier. pp.569-592

Armoogum, J., W. Axhausen, K., Madre, J.-L. (2009) Lessons from an Overview of National Transport Surveys, from Working Group 3 of COST 355: “ Changing Behavior Toward a More Sustainable Transport System”, in Transport Survey Methods: Keeping Up with a Changing World, Emerald. pp.621-634.

Bachmann, C., Roorda, M.J., Abdulhai, B. and B. Moshiri (In Press). Fusing a Bluetooth Traffic Monitoring System with Loop Detector Data for Improved Freeway Traffic Speed Estimation. Accepted for publication in *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* (accepted (conditionally) November 29, 2011)

Bagchi, M., White, P.R., 2004. What role for smart-card data from bus system? Municipal Engineer 157, 39–46.

Bagchi, M., White, P.R., 2005. The potential of public transport smart card data. Transport Policy 12, 464–474.

Bayart, C., Bonnel, P. (2008). Comparison of web and face-to-face household travel survey- application to Lyon case, Proceedings of the 2008 European Transport Conference, http://www.etcproceedings.org/paper/download/3405, page consulted December 12th, 2011.

Bayart, C., Bonnel, P., Morency, C., 2008. Survey mode integration and data fusion: Methods and challenges. Resource paper for Workshop on Best Practices in Data Fusion, 8th International Conference on Survey Methods in Transport Annecy, France, May 25-31, 2008

Benfield, J. A., & Szlemko, W. J. (2006). Internet-based data collection: Promises and realities. Journal of Research Practice, 2(2), Article D1. Retrieved [December 1st 2011] from, http://jrp.icaap.org/index.php/jrp/article/view/30/51

Bonnel, P., Lee-Gosselin, M., Madre, J.-L., Zmud, J. (2009). Keeping Up with a Changing World: Challenges in the Design of Transport Survey Methods, in Transport Survey Methods: Keeping Up with a Changing World, Emerald. pp.3-13.

Bonnel, P., Morency, C., Bayart, C. (2009). Survey mode integration and Data fusion: Methods and Challenges, in Transport Survey Methods: Keeping Up with a Changing World, Emerald. 587-612.

Bourbonnais, Pierre-Leo, Morency, Catherine (2011). Web-based travel survey: a demo, presented at the 9th International Conference on Transport Survey Methods, Chile, Nov. 2011.

Bracewell D. (2008) The Road to the 2010 Olympic and Paralympic Games -Transportation Findings of Large Public Events in Downtown Vancouver, 2008 Annual Conference and Exhibition of the Transportation Association of Canada: Transportation - A Key To A Sustainable Future

Braunsberger, K., Wybenga, H., Gates, R. (2007). A comparison of reliability between telephone and web-based surveys, Journal of Business Research, Volume 60, 2007, pp.758-764.

Carrasco J. A., B. Hogan, B. Wellman and E. J. Miller (2008) Collecting social network data to study social activity-travel behaviour: an egocentric approach, *Environment and Planning B: Planning and Design*, 35: 961-980

Carrasco, J.A., C. Bustos, and B.C. Aguayo (2011). Affective personal networks versus daily contacts: Analyzing different name generators in an activity-travel behavior context, 9th international conference on Survey methods in Transport, Termas de Puyehue, Chile.

Chapleau R., M. Trepanier and K.K. Chu (2008) The Ultimate Survey for Transit Planning: Complete Information with Smart Card Data and GIS, 8th International Conference on Survey Methods in Transport, Harmonization and data Comparability, Annecy, France

Charlton B., J Hood, E. Sall, M.A. Schwartz (2011) Bicycle Route Choice Data Collection Using GPS-Enabled Smartphones, *Transportation Research Board 90th Annual Meeting*, Papers DVD

Chu, A., Chapleau, R. (2011). Smart Card Validation Data as a Multi-Day Transit Panel Survey to Investigate Individual and Aggregate Variation in Travel Behaviour, Presented at the 9th International Conference on Transport Survey Methods, Chile.

Chung, E., & Shalaby, A. S. (2005). A trip reconstruction tool for GPS-based personal travel surveys. *Transportation Planning and Technology*, 28(No. 5), 381-401.

Clark A. C. And S. T. Doherty (2010) A multi-instrumented approach to observing the activity rescheduling decision process, *Transportation Research Part D: Transport and Environment*37: 165-181

Dalumpines, R. and D.M. Scott (2011) GIS-based map-matching: development and demonstration of a post-processing map-matching algorithm for transportation research. *Lecture Notes in GeoInformation and Cartography,* 1(2):101-120.

Deakin, E., Kim, S., 2001. Transportation Technologies: Implications for Planning. University of California Transportation Center, Paper # 536, 27 p.

de Fabritiis C., R. Ragona and G. Valenti (2008). Traffic Estimation And Prediction Based On Real Time Floating Car Data, *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*: 197-203.

Descoimps, E., Agard, B., Trépanier, M. (2011). Comment les conditions climatiques influencent-elles l'utilisation du transport collectif ? Normalité des déplacements et impacts météorologiques, Dixièmes rencontres francophones Est-Ouest de socio-économie des transports, Montréal, Québec, Canada, 2-3 juin.

Dewing, M. (2010). Social Media 2.0. Who uses them? Library of Parliament, Background paper, Publication no 2010-05-E, Ottawa, Canada (http://www.parl.gc.ca/Content/LOP/ResearchPublications/2010-05-e.pdf, page consulted December 14th 2011).

Dumont, J., A. Shalaby and M.J. Roorda (In Press). GPS-Aided Survey for Assessing Trip Reporting Accuracy and Travel of Students Without Land Lines. Accepted for publication in *Transportation Planning and Technology*.

Duncan, S.J., H.M. Badland and G. Schofield (2009) Combining GPS with heart rate monitoring to measure physical activity in children: A feasibility study. *Journal of Science in Medicine in Sport*. 12(5): 583-585.

Exner, J.-P., Zelle, P., and B. Streich (2011). Monitoring laboratory spatial planning: New benefits and potentials for urban planning through the use of urban sensing, geo and mobile web. Paper presented at the 2011 Computers in Urban Planning and Urban Management (CUPUM) Conference, Lake Louise, July 5-8.

Friedrich, M., Jehlicka, P., Schlaich, J. (2008): Automatic number plate recognition for the observance of travel behavior, 8th International Conference on Survey Methods in Transport: Harmonisation and Data Comparability, Mai 2008, Annecy, Frankreich.

Frignani, M., J. Auld, A. Mohammadian, C. Williams and P. Nelson (2010). Urban travel route and activity choice survey (UTRACS): An internet-based prompted recall activity travel survey using GPS data. *Transportation Research Record 2183*. pp. 19-28.

Garcia C., A. Silva, B. Garcia, M. Abou-Zeid, M. Ben-Akiva, C. Choudhury, F. Pereira, and M. Silva (2011) Integrated transportation and energy activity/travel web-based survey. Presented at the 9th International Conference on Survey Methods in Transport. Termas de Puyehue, Chile.

Geostats (2011). GeoStats (2011 Household Travel Surveys for Jerusalem, New York/New Jersey and Cleveland USA, http://www.geostats.com/service_travel.htm (page consulted Dec. 19, 2011).

Gile, K. J. and M. S. Handcock (2010). Respondent-driven sampling: an assessment of current methodology. Sociological Methodology 40(1): 285-327.

Gonzalez, M.C., Hidalgo, C.A., Barabasi, A-L (2008). Understanding individual human mobility patterns, Nature, 453:5, doi:10.1038/nature06958.

Goulias, K. (2000).Resource paper for the Workshop on multi-method and multi-instrument surveys , TRB Transportation Research Circular E-C008: Transport Surveys: Raising the Standard

Greaves S. And R. Ellison (2011) A GPS/Web-based Solution for Multi-day Travel Surveys:

Processing Requirements and Participant Reaction, 9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile

Greaves, S., S.S. Fifer, R. Ellison, G. Germanos (2010). Development of a global positioning system web-based prompted recall solution for longitudinal travel surveys.. *Transportation Research Record, 2183*. pp. 69-77.

Ismail, K, Sayed, T., Saunier, N. (2010). Automated Analysis of Pedestrian–Vehicle Conflicts: Context for Before-and-After Studies, Transportation Research Record: Journal of the Transportation Research Board, Issue Number: 2198, pp 52-64

Ismail, K., T. Sayed and N. Saunier. (2009). Automated Collection Of Pedestrian Data Using Computer Vision Techniques. In Transportation Research Board Annual Meeting Compendium of Papers, Washington D.C., January 2009. 09-1122

Kagerbauer M., W. Manz and D. Zumkeller (2011) Methodological Analysis of Different Methods within one Multi Day Household Travel Survey – PAPI, CATI and CAWI in Comparison, 9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile

Kelly P., A. Doherty, E. Berry, S. Hodges, A. Betterham and C. Foster (2011) Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study, International Journal of Behavioural Nutrition and Physical Activity, 2011, 8:44

Kestens, Y. (2011). Panelist , Obesity, Diet, and Physical Activity session, 2011 mHealth Summit,Washington, D.C. (http://www.mhealthsummit.org/program_speakers_ykestens.php).

Kohla B., G. Sammer, R. Wally, M. Herry and R. Tomschy (2011). Comparing Trip Diaries with GPS-Tracking Results of a Comprehensive Austrian Study, 9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile

Lee-Gosselin, M.E.H., Doherty, S.T. & Shalaby, A. (2010). Data collection on personal movement using mobile ICTs : old wine in new bottles?, in: Wachowitz, M. (Ed) : Movement-Aware Applications for Sustainable Mobility, Information Science Reference, IGI Global, pp 1-15

Li, Z. and A. Shalaby, (2008). Web-based GIS System for Prompted Recall of GPS-assisted Personal Travel Surveys: System Development and Experimental Study, CD Proceedings of the 87th Annual Transportation Research Board Meeting, Washington D.C.

Loustau, P., Grasset, V., Morency, C., Trépanier, M. (2010) Comparing floating car data and carsharing GPS data for travel time assessment, presented at the 12th World Conference on Transport Research, Lisbon, Portugal.

Malinovskiy, Y, Saunier, N, Wang. Y. (2012). Pedestrian travel analysis using static bluetooth sensors. In Transportation Research Board Annual Meeting Compendium of Papers, 2012. 12-3270 Considered for publication in Transportation Research Record: Journal of the Transportation Research Board

Morency, C., Trépanier, M., Agard, B., 2007. Measuring transit use variability with smart-card data. Transport Policy 14 (3), 193–203.

MTO (2009) City of Toronto and Regions of Durham, Peel and York 2008 Travel Time Study. Final Report prepared by IBI Group. August.

Munizaga, M., Palma, C., Mora, P., 2010. Public transport OD matrix estimation from smart card payment system data. In: Presented at the 12th World Conference on Transport Research, Lisbon, Paper No. 2988

Noureldin, A., Karamat, T., Eberts, M. and El-Shafie, A. (2009). Performance Enhancement of MEMS Based INS/GPS Integration for Low Cost Navigation Applications, IEEE Transactions on Vehicular Technology, V58 (3), pp: 1077 – 1096, March 2009

Park, J.Y., Kim, D.J., 2008. The Potential of Using the Smart Card Data to Define the Use of Public Transit in Seoul. Transportation Research Record: Journal of the Transportation Research Board, No. 2063, Transportation Research Board of the National Academies, Washington, DC, pp. 3–9

Pelletier, M.P, Trépanier, M., Morency, C. (2011). Smart card data use in public transit: A literature review, Transportation Research Part C 19 (2011) 557–568.

Potoglou, D., Kanaroglou, P.S. (2008). Comparison of phone and web-based surveys for collecting household background information, Paper presented at the 8th International Conference on Survey Methods in Transport, France, May 25-31, 2008.

Reddy, A., Lu, A., Kumar, S., Bashmakov, V., Rudenko, S., 2009. Application of Entry-Only Automated Fare Collection (AFC) System Data to Infer Ridership, Rider Destinations, Unlinked Trips, and Passenger Miles. 88th Annual Meeting of the Transportation Research Board, Washington, 21 p. (CD-ROM)

Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., Srivastava, M. 2010. Using mobile phones to determine transportation modes. ACM Trans. Sensor Netw. 6, 2, Article 13 (February 2010), 27 pages. DOI = 10.1145/1689239.1689243 http://doi.acm.org/10.1145/1689239.1689243

Riandey, B., Quaglia, M. (2009) Surveying Hard-to-Reach Groups, in Transport Survey Methods: Keeping Up with a Changing World, Emerald. pp.127-144.

Robinson, Steve, Polak, John W. (2005). Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the k-NN Method, *Transportation Research Record: Journal of the Transportation Research Board,* No. 1935, pp. 47–56.

Roorda, M., B. Sharman, C. Sekula and P. Masters (2009). Preliminary Analysis of a System for Real-time Monitoring of Bluetooth Device Data on an Urban Freeway. Paper presented at the Translog 2009 conference. Hamilton, June 17-18.

Saunier, N., Morency, C. (2011). Comparing data from mobile and static traffic sensors for travel time assessment, ASCE Conf. Proc. T&DI Congress 2011: Integrated Transportation and Development for a Better Tomorrow Proceedings of the First T&DI Congress 2011

Saunier, N., Sayed, T. (2006). A feature-based tracking algorithm for vehicles in intersections, Proceeding CRV '06 Proceedings of the 3rd Canadian Conference on Computer and Robot Vision IEEE Computer Society Washington, DC.

Schüssler, N. and K. W. Axhausen (2008) Identifying trips and activities and their characteristics from raw GPS data without further information. *8th International Conference on Survey Methods in Transport.* Annecy France.

Schüssler N., L. Montini and C. Dobler (2011) Improving Post-Processing Routines for GPS Observations Using Prompted-Recall Data, 9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile

Sharman, B. and M.J. Roorda (2011). Analysis of Freight GPS Data: A Clustering Approach for Identifying Trip Destinations. *Transportation Research Record No. 2246*. 83-91.

Sharman, B., Roorda, M.J. and K.M.N. Habib (In Press). Comparison of Parametric and Non-Parametric Hazard Models of Stop Durations on Urban Commercial Vehicle Tours. Paper accepted for publication in *Transportation Research Record (subject to revisions)*.

Sharp J. and E. Murakami (2005). Travel Surveys: Methodological and Technology-Related Considerations, *Journal of Transportation and Statistics*, 8(3) 97-113.

Sioui, L., Morency, C., Trépanier, M., Viviani, M., Robert, B. (2009). A Web-based Travel Survey for Carsharing Members, presented at TRANSLOG 2009, Hamilton.

SmartCard Alliance (2006) Smart Cards and Parking, A Smart Card Alliance Transportation Council White Paper, Publication Number TC-06001

Statistics Canada (2010). 2009 Canadian Internet Use Survey, http://www.statcan.gc.ca/daily-quotidien/100510/dq100510a-eng.htm, page consulted November 10th, 2010.

Statistics Canada (2010). 2010 Residential Telephone Service Survey, http://www.statcan.gc.ca/daily-quotidien/110405/dq110405a-eng.htm, page consulted December 11h, 2011.

Steuart, G. (2011) personal communication.

Stopher, P., Clifford, E., Swann, N., Zhang, Y.(2009). Evaluating voluntary travel behaviour change: Suggested guidelines and case studies. Transport Policy, 16, pp. 315-324 (2009)

Stopher P. R., C Prasad, L. Wargelin and J Minser (2011) Conducting a GPS-Only

Household Travel Survey, 9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile.

TAC (2009). Framework for High Quality Data Collection of Urban Goods Movement, Phase 2 - Final Report. Report Prepared for Transportation Association of Canada. October.

TCRP Synthesis 69 (2006) Web-Based Survey Techniques A Synthesis of Transit Practice, *Research Sponsored by the Federal Transit Administration in Cooperation with the Transit Development Corporation*, Project J-7, Topic SH-07

Teeuw, W.B., Koolwaaij, J., Peddemors, A. (2011). User behaviour captured by mobile phones, Workshop on Interactive Human Behavior Analysis in Open or Public Spaces, Amsterdam, http://www.idiap.ch/workshop/interhub2011/papers/Teeuw/at_download/file, page consulted December 12th, 2011.

Thériault, M., Lee-Gosselin, M., Alexandre, L., Théberge, François Dieumegarde (2011). Pencil-and-paper versus Web Surveys of Weekly Mobility Comparisons on Conviviality, Technical and Privacy Issues. 9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile

Timmermans, H.J.P., Hato, (2009) E. Electronic Instrument Design and User Interfaces for Activity-Based Modeling, in Transport Survey Methods: Keeping Up with a Changing World, Emerald. pp.437-461.

Trépanier, M., Chapleau, R., Tranchant, N., 2007. Individual trip destination estimation in transit smart card automated fare collection system. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 11 (1), 1–15 (Taylor & Francis).

Trépanier, M., Morency, C., 2010. Assessing transit loyalty with smart card data. Presented at the 12th World Conference on Transport Research, Lisbon, Paper No. 2341.

Trépanier, M., Morency, C., Agard, B., 2009b. Calculation of transit performance measures using smartcard data. Journal of Public Transportation 12 (1), 79–96.

Trépanier, M., Morency, C., Blanchette, C., 2009. Enhancing Household Travel Surveys Using Smart Card Data? 88th Annual Meeting of the Transportation Research Board, Washington, 15 p. (CD-ROM).

Trépanier, M., Vassivière, F., 2008. Democratized smart card data for transit operators. In: 15th World Congress on Intelligent Transport Systems, New York, USA, 12 p

Trevor H. And Eric Hildebrand (2011) Experiences with GPS Travel Diaries in Rural Older Driver Research, *TRB 90th Annual Meeting, Transportation Research Board Annual Meeting 2011 Paper #11-4258*

Tsui, A. and A. Shalaby (2006). An Enhanced System for Link and Mode Identification for GPS-based Personal Travel Surveys, *Transportation Research Record: Journal of the Transportation Research Board,* 1972:38-45.

Utsunomiya, M., Attanucci, J., Wilson, N., 2006. Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning, Transportation Research Record: Journal of the Transportation Research Board, No. 1971, Transportation Research Board of the National Academies, Washington, DC, pp. 119–126.

Wolf, J. J. Wilhelm, J. Casas and S. Sen (2011) A case study: Multiple data collection methods and the NY/NJ/CT regional travel survey. *9th International Conference on Survey Methods in Transport.* Termas de Puyehue, Chile.

Yang, J., Lu, H., Liu, Z., Boda, P.P.(2010). Physical Activity Recognition with Mobile Phones: Challenges, Methods and Applications. In: L. Shao et al. (Eds.): Multimedia Interaction and Intelligent User Interfaces: Principles, Methods and Applications, Springer-Verlag, London, pp. 185-213 (2010).

APPENDIX: ANNOTATED BIBLIOGRAPHY

This annotated bibliography focuses on the recent literature, 2007-2011 inclusive, with selected earlier documents included.

Documents are classified using the category codes and sub-codes listed below. Sub-codes are shown in brackets in the categorizations; e.g., HI(WB) indicates a web-based household interview survey.

Category Codes:

Household Interviews: HI
  Face to face: FtF
  Telephone: Tel
  Mail-back: MB
  Web-based: WB

Choice-based surveys: CB
  Roadside: RS
  On-board: OB

Standard technology-based methods: STB
  Geographic positioning system: GPS
  Roadside detectors: RSD
  Smart-card: SC
  Mobile phone: MP

Emerging technology-based methods: ETB
  Remote sensing (satellite/aerial): RSSA
  Web apps: WA
  Social network software: SNS
  Smart phone: SP
  Accelerometers: Acc
  Personal health sensor: PHS
  Environmental sensors: ES
  Bluetooth: Blu

Methodological advances implemented or recommended in the study: MAIR

All links were accessed between February 4, 2012 and February 8, 2012

Andrews D., B. Nonnecke and J. Preece (2003). Electronic Survey Methodology: A Case Study in Reaching Hard-to-Involve Internet Users, *International Journal of Human-Computer Interaction*, 16(2): 185-210

Categories:	HI (WB)
Applications:	Online
Methods Used and/or Summary:	"In this article, quality criteria for electronic survey design and use based on an investigation of recent electronic survey literature are presented. The application of these criteria to reach a hard-to-involve online population—nonpublic participants of online communities (also known as "lurkers")—and survey them on their community participation, a topic not salient to the purpose of their online communities is demonstrated in a case study."
Strengths:	"[I]t is possible to coax non-public participants into publicly participating in an online survey." Piloting is essential in online surveys.
Weaknesses:	The authors did not implement all the recommended criteria for quality electronic surveys but did implement those that made sense for the research objectives, context, and content of this study. The survey server was infected with a virus and was also hacked, it is important to provide strong firewall security.
MAIR:	
URL:	http://www.tandfonline.com/doi/abs/10.1207/S15327590IJHC1602_04

Arentze T. and H. Timmermans (2008). Social Networks, Social Interactions, and Activity-Travel Behaviour: A Framework for Micro-Simulation, *Environment and Planning B: Planning and Design,* 35:1012-1027

Categories:	HI (FtF, WB, MB), and ETB (SNS)
Applications:	None
Methods Used and/or Summary:	"This paper introduced a framework for incorporating social networks in the dynamic microsimulation of activity-travel patterns. A core assumption of our theory is that similarity between persons in terms of attributes, preference, and action space increases the probability that a link between them is created and sustained in time. Once a link has been created, social influence leads to knowledge exchange and the adaptation of preferences." The paper does describe the required data collection method for developing the proposed framework. However, the framework requires emerging technologies to collect the required, detailed social network interaction between individuals. Such a framework and similar approaches are being more discussed as data collection becomes more feasible.

Strengths:	Although no data collection method is discussed in the paper, the proposed activity-travel pattern framework seems intuitive and behavioural.
Weaknesses:	The theory is still at its early stages and requires further attempts to become well developed and applicable.
MAIR:	
URL:	http://www.envplan.com/abstract.cgi?id=b3319t

Baltes M. A. (2002). Customer Surveying For Public Transit: A Design Manual for Customer On-Board Surveys, *Final report for National Center for Transit Research*, NCTR - 416 - 083

Categories:	CB (OB)
Applications:	None
Methods Used and/or Summary:	“Presented is a best practices manual that describes the necessary steps in conducting a successful on-board survey of public transit customers. It was specifically developed for the public transit professional that has at least a rudimentary understanding of the purposes and procedures in survey research and is searching for specific guidance on how to “best” conduct an on-board survey of its customers. This how-to manual provides public transit professionals with a much better understanding of the total customer surveying process and its importance in planning and ultimately the highest quality service to the riding public.”
Strengths:	“It describes the various components or steps of the on-board transit customer surveying process from specifying and clearly defined objectives, various methods of data collection, questionnaire construction, sample size, appropriate level(s) of analysis, accurate and truthful reporting of results, data entry, report writing, and data archiving.”
Weaknesses:	
MAIR:	
URL:	http://www.nctr.usf.edu/pdf/On-Board%20Survey%20Manual.pdf

Barcelo J., L. Montero, L. Marques and C. Carmona (2010). Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic, *Transportation Research Record, Journal of the Transportation Research Board*, 2175:19-27

Categories:	ETB (Blu)
Applications:	Barcelona Spain
Methods Used and/or Summary:	“Traditional technologies, such as inductive loop detectors, do not usually produce measurements of the quality required by real-time applications. Therefore, one wonders what could be expected from newer information and communication technologies, such as automatic vehicle location, license plate recognition, and detection of mobile devices. The main objectives of this paper are to explore the quality of the data produced by Bluetooth detection of mobile devices that equip vehicles for travel time forecasting and its use in estimating time-dependent origin–destination matrices. Ad hoc procedures based on Kalman filtering have been designed and implemented successfully, and the numerical results of the computational experiments are presented and discussed.” Travel times of links in a network are approximated and forecast by detecting vehicles equipped with Bluetooth mobile devices, hands free phones, Tom Tom or Parrot devices, etc.
Strengths:	This method can be also used to estimate dynamic origin to destination matrices on motorways.
Weaknesses:	The impact of weather and congestion on the accuracy of estimation was not studied.
MAIR:	
URL:	http://trb.metapress.com/content/17x1v88n6p8363p1/

Bechx C., L. I. Panis, D. Jansseens and G. Wets (2010). Applying activity-travel data for the assessment of vehicle exhaust emissions: Application of a GPS-enhanced data collection tool, *Transportation Research Part D: Transport and Environment,* 15(2): 117-122

Categories:	STB (GPS), ETB(WA, SP)
Applications:	Belgium
Methods Used and/or Summary:	The methodology is not innovative compared to other studies of GPS enabled estimation of travel attributes. However, the application of a built-in GPS logger for estimating vehicle emissions is noteworthy. “During the survey period, the activity-based survey GUI, which is the major interface of the application, is used to register the activity-travel diary data.”... “The built-in GPS logger is used to trace physical travel paths and the travel times.”... “Speed profiles, based on instantaneous (1 Hz) speed data from the GPS receiver in the data collection device, are composed for every detected vehicle trip in the travel survey and used as input for a vehicle simulation tool. This simulation tool, VeTESS (Vehicle

	Transient Emissions simulation Software), developed within the EU 5th framework project DECADE, calculates emissions and fuel consumption made by a vehicle during a defined ‘drive-cycle, then the emissions per second for CO2, CO, NOx, HC and PM are calculated.
Strengths:	
Weaknesses:	Only 32 respondents were studied. No discussion was provided regarding the algorithm(s) used for processing the GPS log data.
MAIR:	
URL:	http://www.sciencedirect.com/science/article/pii/S1361920909001278

Bohte W. and K. Maat (2009). Deriving and Validating Trip Purposes and Travel Modes for Multi-Day GPS-Based Travel Surveys: A Large-Scale Application in the Netherlands, *Transportation Research Part C: Emerging Technologies,* 17(3): 285-297

Categories:	STB (GPS)
Applications:	Netherlands
Methods Used and/or Summary:	“This article presents an innovative method that combines GPS logs, Geographic Information System (GIS) technology and an interactive web-based validation application. In particular, this approach concentrates on the issue of deriving and validating trip purposes and travel modes, as well as allowing for reliable multi-day data collection”
Strengths:	The sample was large enough (1200 respondents) to compare the performance of their model against the Dutch Travel Survey.
Weaknesses:	“It does not accurately reflect the feasibility of mode detection using GPS data from mobile phones. Since characteristics of assisted GPS technology used in GPS-enabled mobile phones include increased sensitivity and a reduced time-to-first-fix, mobile phones can yield location data that are significantly different from data generated by traditional stand-alone GPS devices.” They developed their own rule-based system to determine the mode and purpose of the trip.
MAIR:	
URL:	http://www.sciencedirect.com/science/article/pii/S0968090X08000909

Bonnel P., M. Lee-Gosselin, J. P Zmud and J. L. Madre (2009). *Transport Survey Methods: Keeping Up with a Changing World, Emerald Group Publishing Limited, United Kingdom*

Categories:	HI (FtF, Tel, MB and WB), STB (GPS, MP) and ETB (SP, ES, PHS, Acc, SNS RSSA, WA)
Applications:	Various
Methods Used and/or Summary:	This book provides an overview of the State of Art and Practice for transport survey methods. It starts from challenges of surveying hard to reach groups. Then it discusses how to improve the capture of urban goods movement data. In a series of chapters, the authors discuss GPS-based surveys and the impact of electronic instrument design and user interface on survey quality. Finally, the last part of the book is allocated to emerging or persistent survey issues and data harmonization with a specific focus on vehicle-based surveys and data fusion approaches.
Strengths:	Strengths and weaknesses of some of the methods discussed in chapters of the book are presented separately in the report.
Weaknesses:	
MAIR:	
URL:	http://books.emeraldinsight.com/display.asp?K=9781848558441&cur=EUR

Bracewell D. (2008). The Road to the 2010 Olympic and Paralympic Games -Transportation Findings of Large Public Events in Downtown Vancouver, *Annual Conference and Exhibition of the Transportation Association of Canada: Transportation - A Key to a Sustainable Future*

Categories:	CB (), ETB ()
Applications:	Vancouver, BC
Methods Used and/or Summary:	The authors used Ticketmaster data to collect information about event participants including postal codes and addresses of the original ticket purchasers.
Strengths:	The authors create indirect links from collected data to socio-demographic attributes of the participants.
Weaknesses:	"When compared to the stated origins of event attendees from the on-site market research, the number of Vancouver residents who actually attend the events is less than what the Ticketmaster data findings suggest." Ticket purchasers may have used work address or purchased tickets for friends.

MAIR:	The authors conducted a set of surveys using typical survey methods to validate the progress of the city in supporting more sustainable transportation modes. Similar methods can be used for other Canadian cities hosting large events.
URL:	http://www.tac-atc.ca/english/resourcecentre/readingroom/conference/conf2008/docs/k2/racewell.pdf

Carrasco J. A., B. Hogan, B. Wellman and E. J. Miller (2008). Collecting Social Network Data to Study Social Activity-Travel Behaviour: an Egocentric Approach, *Environment and Planning B: Planning and Design,* 35: 961-980

Categories:	ETB (SNS)
Applications:	East York area of Toronto, Canada,
Methods Used and/or Summary:	"This paper presents a data collection effort designed to incorporate the social dimension in social activity-travel behaviour by explicitly studying the link between individuals' social activities and their social networks. The main hypothesis of the data collection effort is that individuals' travel behaviour is conditional upon their social networks; that is, a key cause of travel behaviour is the social dimension represented by social networks. With this hypothesis in mind, and using survey and interview instruments, the respondents' social networks are collected using an egocentric approach that is constituted by the interplay between their individual social structures and their social activity behaviour. More explicitly, individuals' networks are a context within which to elicit social activity- travel generation, spatial distribution, and information communication and technology use. The resultant dataset links aspects, in novel ways, that have been rarely studied together, and provides a sound base of theory and method to study and potentially give new insights about social activity-travel behaviour. Social network data can show how aggregated measures of the individuals' social networks and interaction patterns can provide insights about travel behaviour, especially regarding the propensity to perform face-to-face social activities."
Strengths:	The authors create indirect links from collected data to socio-demographic attributes of the participants. Data collection methods used: call by cell phone, call by regular phone, e-mail, instant message, talk with face-to-face, meet at restaurants or bars and visit or host as visitors.
Weaknesses:	"Network boundaries are difficult to define. People do not easily recall their network members, and need appropriate `prompts' to elicit them; in addition, networks are very large in general, and different social network members may have different importance, depending on the phenomenon studied. Information

	about the network members needs to balance detail and the interviewee's burden.”
MAIR:	The authors seek to link social network data to trip-making decisions.
URL:	http://www.envplan.com/abstract.cgi?id=b3317t

Carrasco, J. A., C Bustos and B C. Aguayo (2011) Affective Personal Networks versus Daily Contacts: Analyzing Different Name Generators in an Activity-Travel Behaviour Context, *9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile*

Categories:	STB (MP), ETB (SP)
Applications:	Greater Concepción Area, Chile,
Methods Used and/or Summary:	This paper presents a new data collection effort, in a social, urban, and temporal context where social activity influences-travel behaviour. To study the social dimension of travel, the authors emphasize studying the social network. To do so, they ask individuals to generate names of contacts with whom they have connections. Four name generating methods were used: Emotional Closeness, Network Capital, Social Activities and Time Use. “Using personal network data, this paper has reviewed how four name generators capture the participants’ social context in relation with their spatial and temporal patterns of social interaction. A special focus was put on how these techniques help to understand the role of income and access to amenities on those spatial and temporal patterns.”
Strengths	The application of a new data collection method in transportation is explored in this paper.
Weaknesses	
MAIR:	
URL:	

Chapleau R., M. Trepanier and K.K. Chu (2008). The Ultimate Survey for Transit Planning: Complete Information with Smart Card Data and GIS, *8th International Conference on Survey Methods in Transport, Harmonization and data Comparability, Annecy, France*

Categories:	STB (SC)
Applications:	Montreal Canada

Methods Used and/or Summary:	"The paper argues that with a well-defined methodology, the data can be used to monitor the supply and consumption of the transit system, the mobility of its users as well as the activity pattern at trip generators. Descriptions of the validation and enrichment procedures as well as their logic are given so that they can be adapted for use in other systems."
Strengths:	"The numerous advantages of the smart AFC system over traditional survey, the high resolution of data and the completeness of information make it the ultimate survey for transit planning." A simple heuristic was used to locate the residence of participants.
Weaknesses:	Attributes of card holders were not collected or inferred. Most of the travel and personal characteristics of participants remained unknown and no specific method was recommended for modeling them. Therefore, the developed methodology is not yet capable of substituting a travel survey.
MAIR:	"As more and more transit agencies adopt the smart card AFC technology and more data become available, there are opportunities and needs to explore innovative ways to extract and analyse data for transit planning and modelling purposes."
URL:	http://www.isctsc.cl/archivos/2008/13%20B1%20chapleau%20et%20al.doc

Charlton B., J. Hood, E. Sall, M.A. Schwartz (2011). Bicycle Route Choice Data Collection Using GPS-Enabled Smartphones, *Transportation Research Board, 90th Annual Meeting,* 11: 2652

Categories:	STB (MP), ETB (SP)
Applications:	San Francisco, California USA
Methods Used and/or Summary:	"A freely downloadable iPhone/Android smartphone "app" called CycleTracks to collect actual bicycle routes traversed by city cyclists."... "Once installed on a user's smartphone, a single "tap" would start and stop recording a bicycle trip; after completing a trip, the app automatically uploaded the track to a central database/web server, via the phone's built-in data plan. Approximately 5,000 usable bicycle trips were collected from hundreds of users in the region."... "A bicycle route choice model developed using the data revealed sensitivity to slope, presence of bike lanes and/or bike route designations, trip purpose, and gender. The bike route choice model is now being integrated into San Francisco's regional travel model."
Strengths:	The cost of the survey is very low while the quality of data is desirable.

Weaknesses:	“Demographic data was optionally provided by some users, and showed a bias toward frequent cyclists, and toward male users”
MAIR:	
URL:	http://amonline.trb.org/12kktc/12kktc/1

Cherubini M., R. De Oliviera, A. Hiltunen and N. Oliver (2011). Barriers and Bridges in Adoption of Today’s Mobile Phone Contextual Services, *13th International Conference on Human Computer Interaction with Mobile Devices and Services, Stockholm, Sweden*

Categories:	STB (MP), ETB (SP)
Applications:	
Methods Used and/or Summary:	Three research questions are discussed in this paper: What are the human needs that support the adoption of contextual services and applications? How do contextual needs find their correspondence to general human needs? What are the most/least relevant barriers that people face to adopt mobile phone contextual services? “This paper presents ethnographic observations, a diary study and a large-scale quantitative questionnaire (n=395) designed to study the reasons for adoption and refusal of context aware mobile applications. Through a qualitative study we identify 24 user needs that these applications fulfill and 9 barriers for adoption. We found that for many of the identified needs the end-goal is not that of receiving information, thus complementing work on mobile information needs. Also, this work offers an actionable list of obstacles that prevent contextual services to reach a larger audience. Finally, our findings suggest the opportunity to develop novel mobile applications that fulfill needs in the activity and personal contextual dimensions, and that of developing an application store for feature phones.
Strengths:	The authors provided an actionable list of obstacles that prevent contextual services from reaching a larger audience
Weaknesses:	
MAIR:	
URL:	http://dl.acm.org/citation.cfm?doid=2037373.2037400

Clark A. C. and S. T. Doherty (2010). A Multi-Instrumented Approach to Observing the Activity Rescheduling Decision Process, *Transportation Research Part D: Transport and Environment,* 37: 165-181

Categories:	STB (GPS), ETB(WA, SP)
Applications:	Waterloo, Canada
Methods Used and/or Summary:	The data collection methodology consists in six stages. A preplan schedule is collected by open-ended personal interview. Responses are codified and stored in tables. Passive GPS tracking occurs. An internet-based prompted recall diary is completed with assistance from the researcher. The preplan is compared with the executed schedule and the researcher identifies rescheduling scenarios. An in-depth rescheduling interview is conducted in person. The GPS tracking system consisted of a wearable SiRF Star III GPS receiver (smaller than a deck of cards) linked wirelessly to a smartphone. The key methodological advances and suggestions: "preplanned activity scheduling surveys should pay careful attention to the interface/report-format, as it was discovered in this paper that subjects prefer a variety of mediums including point form, calendar, and verbal only." "Much care is needed in designing planning time horizon survey questions to avoid forcing subjects to generalize, misinterpret, or provide erroneous responses." "Subjects were not always able to specifically quantify the limits of partial elaboration"
Strengths:	"This new methodology proved valuable as a pretense from which to engage subjects in a discussion on the *why* and *how* of rescheduling decisions."
Weaknesses:	"The hardware and prediction algorithm had occasional flaws that should be the focus of continued improvement to ensure less need for respondent interaction and transferability to other settings and cultures."
MAIR:	
URL:	http://www.springerlink.com/content/k3837551030n1754/

Dalumpines R. and D. M. Scott, (2011). GIS-based Map-matching: Development and Demonstration of a Postprocessing Map-matching Algorithm for Transportation Research, *Lecture Notes in Geoinformation and Cartography,* 1(2): 101-120

Categories:	HI (Tel), ETB (GIS-based map matching)
Applications:	Halifax, Nova Scotia

Methods Used and/or Summary:	“This paper presents a GIS-based map-matching algorithm that makes use of geometric, buffer, and network functions in a GIS to illustrate the suitability of a GIS platform in developing a post-processing map-matching algorithm for transportation research applications such as route choice analysis. This algorithm was tested using a GPS-assisted time-use survey that involved nearly 2,000 households in Halifax, Nova Scotia, Canada. Actual routes taken by household members who travelled to work by car were extracted using the GPS data and the GIS-based map-matching algorithm.”
Strengths:	“The algorithm produced accurate results in a reasonable amount of time. The algorithm also generated relevant route attributes such as travel time, travel distance, and number of left and right turns that serve as explanatory variables in route choice models.”
Weaknesses:	
MAIR:	This post processing method can be used in other applications to improve the quality of the processed results. Post-processing is a staple task in GPS-based prompted recall survey.
URL:	http://www.springerlink.com/content/p056441nnw15nj2m/

de Fabritiis C., R. Ragona and G. Valenti (2008). Traffic Estimation And Prediction Based On Real Time Floating Car Data, *11th International IEEE Conference on Intelligent Transportation Systems, Beijing, China*, 197-203

Categories:	CB (OB), STB (GPS) and ETB (floating-car data)
Applications:	Rome, Italy
Methods Used and/or Summary:	This paper discusses “basic component of ATIS (advanced traveler information systems) and ATMS (advanced traffic management system) applications. In this view the use of real-time floating-car data (FCD), based on traces of GPS positions, is emerging as a reliable and cost-effective way to gather accurate travel times/speeds in a road network and to improve short-term predictions of travel conditions.” ... “Traffic speed estimates are deduced at an interval of 3 minutes from GPS traces transmitted in real-time from a large number (and still growing) of privately owned cars (about 600.000) equipped with a specific device covering a range of insurance-related applications.”
Strengths:	Unlike previously proposed FCD techniques (mostly using data from taxi or bus fleets), this system exploits data from a large number of privately owned cars, to deliver real-time traffic speed information throughout the Italian motorway

	network and along some important arterial streets located in major Italian metropolitan areas.
Weaknesses:	The disadvantages of this method are high installation and data transfer costs.
MAIR:	
URL:	http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4732534

Dillman, D. (2010). New Data Collection Modes and the Challenge of Making them Effective. *Presentation for the Workshop on the Future of Federal Household Surveys, National Research Council, Washington, DC*

Categories:	HI (FtF, Tel, MB, WB)
Applications:	USA
Methods Used and/or Summary:	Prof. Dillman discusses the major four household interview methods and how the resultant data quality from these methods can be improved. A web mode encourages required answers and fewer “don’t know” options. Fill-ins are possible from previous answers. Audio, video, and other add-ons are possible, and typically there are no hidden categories. He mentioned that “Still, if people are given a choice of responding by either mail or Internet, most will choose mail. And, if mail is withheld to encourage respondents to use the web, research has shown that the respondents who end up participating during follow-up are very different from one mode to another.”
Strengths:	
Weaknesses:	
MAIR:	
URL:	http://www.nap.edu/openbook.php?record_id=13174&page=35

Duncan S. J., H. M. Badland and G. Schofield (2009). Combining GPS with Heart Rate Monitoring to Measure Physical Activity in Children: A Feasibility Study, *Journal of Science and Medicine in Sport,* 12(5): 583-585

Categories:	STB (GPS), ETB (PHS)
Applications:	New Zealand

Methods Used and/or Summary:	“The recent development of global positioning system (GPS) receivers with integrated heart rate (HR) monitoring has provided a new method for estimating the energy expenditure associated with children's movement. The purpose of this feasibility study was to trial a combination of GPS surveillance and HR monitoring in 39 primary-aged children from New Zealand.” “Each participant and their legal guardian provided written informed consent. Spatial location was assessed to the nearest metre using a 12-channel F500 GPS receiver (FRWD Technologies Ltd., Oulu, Finland), while a coded transmitter belt (Polar Electro, Kempele, Finland) attached to the chest of each participant enabled HR monitoring. The F500 model is well suited for child monitoring given the water resistant construction and the lack of external buttons or controls. Location, distance, speed, and HR data were collected during school lunch periods using a 1-s recording interval.”
Strengths:	“Combining GPS with HR monitoring is a promising new method for investigating both the spatial location and energy expenditure associated with children's physical activity.”
Weaknesses:	“Data collection took place in a restricted outdoor area over a relatively short timeframe.”
MAIR:	
URL:	http://www.sciencedirect.com/science/article/pii/S1440244008001977

El Esawey M. and T. Sayed (2010). Travel Time Estimation in Urban Networks Using Buses as Probes, *Annual Conference of the Transportation Association of Canada Halifax, Nova Scotia*

Categories:	ETB (RSSA)
Applications:	Vancouver, BC
Methods Used and/or Summary:	“Using transit vehicles as probes offers a number of advantages as they cover a large portion of urban networks and the equipment required for data collection is usually already installed by transit operators. Despite the fact that transit vehicles and automobiles have different running behaviours, a relationship can be developed to estimate auto travel times using transit data. Travel time estimation using buses as probes is usually limited to their travel routes. This research investigates the potential of using bus travel time data to estimate general link travel times of neighbour (nearby) links. The main research hypothesis is that travel times of nearby links have strong correlation as these links are subject to similar traffic conditions. A general methodology is presented for travel time estimation using historical travel time data of the link itself and real-time bus data from neighbour links. A case study was undertaken using a VISSIM microsimulation

	model of downtown Vancouver. The model was calibrated and validated using real-life traffic volumes and travel time data. Travel time estimation accuracy was assessed using the Mean Absolute Percentage Error (MAPE), the value of which was 17.6%. The method was proven to be useful to estimate travel time on links that do not have real-time travel time data while having strong travel time correlation with neighbouring links.”
Strengths:	“In general, the travel time estimation accuracy was below 10% when using buses to estimate auto travel times of the same link, while it was about 17.6% when using buses for neighbour links travel time estimation. This accuracy level was considered acceptable considering the high travel time fluctuations in the study area and the complex traffic pattern that includes pedestrians, shared lanes, signalized intersections, etc.”
Weaknesses:	There many other advanced ways to estimate travel time with better level of accuracy such as estimating travel time with Bluetooth, GPS and cell phone devices.
MAIR:	None
URL:	http://www.tac-atc.ca/english/resourcecentre/readingroom/conference/conf2010/docs/t1/esawey.pdf

Frignani M., J. Auld, A. Mohammadian, C. Williams, and P Nelson (2010). Urban Travel Route and Activity Choice Survey (UTRACS): An Internet-Based Prompted Recall Activity Travel Survey using GPS Data, *Transportation Research Record, Journal of the Transportation Research Board*, 2183: 19-28

Categories:	HI (FtF and WB), STB (GPS)
Applications:	Chicago, Illinois
Methods Used and/or Summary:	Internet-based prompted-recall activity travel survey using Global Positioning System (GPS) data collection combined with a short activity preplanning and scheduling survey.
Strengths:	Data collection takes place over 14 days. Information about activity planning and scheduling collected.
Weaknesses:	Small sample size due to hardship of survey completion.
MAIR:	
URL:	http://trb.metapress.com/content/h00815k07142j685/?p=3f6f9b79d9c1455e99c0

	ce748f9a76de&pi=2

Garcia C., J. A. Silva, B. Garcia, M. Abou-Zeid, M. Ben-Akiva, C.Coudhury, F. Pereira and M Silva (2011). Integrated Transportation and Energy Activity-Travel Web-Based Survey, *9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile*

Categories:	HI (WB), ETB (SP)
Applications:	Lisbon Metropolitan Area
Methods Used and/or Summary:	The authors present an agent based model of urban activity choice and energy use from the perspective of both energy flows at the aggregate urban level and household choice behaviours. One ultimate purpose of the model is to provide policy-makers with a tool for comparing green-energy policy and initiatives. Currently in the planning stage, the model will be supported by three types of data collection pertaining to household behaviour: web-based, smart phone, and household energy telemetering The paper focuses on the design and benefits of the web-based survey which includes collection of household characteristics such as socio-economic information and vehicle characteristics, an activity diary from each member of the household over the age of 12, and household equipment and energy consumption.
Strengths:	Comprehensive model of urban energy consumption and related flows from both aggregate and household perspective.
Weaknesses:	Early stage.
MAIR:	
URL:	

GeoStats, (2011). Household Travel Surveys for Jerusalem, New York/New Jersey and Cleveland USA, *GeoStats Case Studies*

Categories:	HI (FtF, Tel, MB, WB), STB(MP)
Applications:	Jerusalem, NY/NJ and Cleveland
Methods Used and/or Summary:	“The first survey was the 100% GPS household travel survey conducted in Jerusalem. In this survey, GPS devices were provided to more than 6000 households during face-to-face recruitment interviews. Participants used these devices on the following day, and subsequently the GPS devices were downloaded and a GPS-based prompted recall was conducted on a laptop. This study was led by

	the Jerusalem Transportation Masterplan Team and was conducted using GeoStats' TripBuilder survey software system, which was designed to collect travel details in GPS and diary based travel surveys." "The second survey is the NYMTC / NJTPA regional travel survey, which includes 30 counties in NYC/ New York, New Jersey, and Connecticut. In this study, 10% of all households across the study area (more than 3000 households) received GPS devices for all adult household members to wear for multiple days. Once the equipment was returned, the GPS data were processed into trips and then loaded into TripBuilder, where telephone interviewers or participants themselves completed the GPS-based prompted recall interview via the project website. The remaining 90% of households participated in a traditional diary-based travel survey, with travel details also collected via TripBuilder." Most recently, GeoStats has begun work on the Cleveland 100% GPS household travel survey, which is just the second 100% GPS study to be conducted in the United States. In this survey, at least 4,250 households will use GPS devices to collect travel details for a three-day period. These GPS data will then be used to impute travel details, eliminating the need for travel diaries and travel reporting interviews.
Strengths:	Comparison of survey techniques. Single system ensures consistency of question flow/branching, logic checks / validation, geocoding method, and data formats.
Weaknesses:	
MAIR:	
URL:	http://www.geostats.com/service_travel.htm http://www.travelsurveymethods.org/pdfs/GPS-Based_Prompted_Recall_Studies.pdf

Giamio G., R. Anderson, L. Wargelin and P. Stopher (2010). Will It Work? Pilot Results from First Large-Scale Global Positioning System–Based Household Travel Survey in the United States, *Transportation Research Record, Journal of the Transportation Research Board,* 2175: 26-34

Categories:	HI (WB), STB (GPS)
Applications:	Cincinnati, USA
Methods Used and/or Summary:	"The Greater Cincinnati Household Travel Survey (HTS) is a proof-of-concept study for replacing travel diaries with a large-scale multiday Global Positioning System (GPS) survey. The objectives are to collect multiple-day data from more than 3,000 households with portable GPS devices and improve existing processing software to provide data that support modeling approaches in Ohio. No diaries are collected

	for household members younger than 12 years old. A subsample of follow-up prompted recall surveys allow respondents to review GPS interpreted travel information for verification. This paper, with data from the spring 2009 pilot, describes the survey process developed for this HTS. It documents that with an address-based sample frame, advance letters, and Internet and phone recruiting, a significant subsample of cell phone–only households can be recruited and surveyed with GPS; a representative sample of households can be recruited for a GPS-based survey, based on a comparison of pilot sample household characteristics with available Public Use Microdata Samples data; and response rates for difficult-to-reach households such as cell phone–only, lower income, and zero-vehicle households can be improved with a cash incentive ($25). The paper provides principles and describes the prompted recall survey developed to obtain additional data from a subset of respondents beyond the GPS recorded travel for improving imputation software.”
Strengths:	
Weaknesses:	A GPS device alone cannot accurately locate the respondents. Cell phones provide better results.
MAIR:	
URL:	http://trb.metapress.com/content/n62820jp38363x21/?p=61bb15bf959c46a18ace664b28fc7eaa&pi=2

Gonzalez P.A., J.S. Weinstein, S.J. Barbeau, M.A. Labrador, P.L. Winters, N.L. Georggi and R. Perez (2010). Automating mode detection for travel behaviour analysis by using global positioning systems enabled mobile phones and neural networks, *The Institution of Engineering and Technology, Intelligent Transport Systems,* 4(1): 37–49

Categories:	STB (GPS) and ETB (WA, SP)
Applications:	Tampa, Florida
Methods Used and/or Summary:	“As demonstrated in this research paper, automatic mode detection is feasible when utilising a neural network, and assisted GPS data collected via a mobile application such as TRAC-IT for GPS-enabled mobile phones. Furthermore, mode detection accuracy was actually improved when only a small subset of GPS coordinates required to re-create the user’s path”. Assisted GPS data were gathered for 38 car, 38 bus and 38 walking trips.
Strengths:	The “neural network correctly predicted 92.11% of the car trips, 81.58% of the bus trips and 100% of the walking trips for this test series.”

Weaknesses:	"One limitation to the research presented in this paper is that the GPS data used to train and test the neural network were manually segmented by the cell phone user into trips, each of which contained a single mode of transportation. In other words, the user of the cell phone who was being surveyed indicated via input to the active diary portion of the TRAC-IT Java ME mobile application where the trip started"
MAIR:	
URL:	http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05409621

Goulias, K. (2000). Resource Paper for the Workshop on Multi-Method and Multi-Instrument Surveys, *Transportation Research Board, Transportation Research Circular E-C008: Transport Surveys: Raising the Standard*

Categories:	HI (FtF, Tel, MB)
Applications:	Various
Methods Used and/or Summary:	Goulias' literature review opens with a summary of differences between surveys including methods of approaching potential respondents, survey design and methodology, and linking results with extant data. This serves as an extensive background for the introduction of the titular topic of multi-method, multi-instrument surveys. The author discusses data combination, referring to some existing endeavours to harmonize data formats and collection techniques in major surveys, as well as statistical methods for combining datasets. Major points regarding combination: analysis should be undertaken in two stages. Exploratory, to identify and correct introduced biases, and combinatorial when the process of linking and "enriching" takes place. Disaggregation of data from large scale surveys is possible by matching key indicators with those of the desired sample to create subsets or synthetic baselines. The author refers to the concept of collecting both qualitative and quantitative data for the same project as "triangulation" and lists a wide variety of arguments why it is desirable including increasing researcher confidence in results and allowing for testing amongst competing theories. The most important point made by the author is the need to identify different sources of variation in responses when using multi-method, multi-instrument surveys. These sources are summarized as: "respondent characteristics and/or trip making circumstances, differences in responses that are due to survey method(s) and instrument(s) used, differences in responses that are due to the time period of

	the survey (time of day, day of week, season, year, and so forth), and unobserved time-varying and time-invariant factors." The magnitude and direction of these sources of variation should be extensively explored through experimental pre-testing before finalizing a multi-method, multi-instrument survey.
Strengths:	Strong recommendations for best-practices multi-method, multi-instrument surveys.
Weaknesses:	Brief. Lack of practical examples.
MAIR:	Use of longitudinal surveys for pre-testing to identify direction and magnitudes of "components of variation" in multi-method, multi-instrument surveys.
URL	http://onlinepubs.trb.org/onlinepubs/circulars/ec008/workshop_a.pdf

Greaves S., S.S. Fifer, R. Ellison and G. Germanos (2010). Development of a Global Positioning System Web-Based Prompted Recall Solution for Longitudinal Travel Surveys, *Transportation Research Record, Journal of the Transportation Research Board,* 2183:69-77

Categories:	HI (FtF and WB), STB (GPS)
Applications:	Sydney, Australia
Methods Used and/or Summary:	Internet-based prompted-recall activity travel survey using Global Positioning System (GPS) data collection combined with a short activity preplanning and scheduling survey.
Strengths:	Data collection occurred over 10 weeks. Information on activity planning and scheduling was collected. The authors note: "highly accurate data of this nature can be collected for several weeks with little respondent burden."
Weaknesses:	Small sample size (30 motorists as pilot).
MAIR:	
URL:	http://trb.metapress.com/content/48k1685p08r73h66/?p=ad16242d2c8a4a9bbf1cb89c1d169bf4&pi=7

Harvey, A. (2009). Halifax Regional Space-Time Activity Research (STAR) Project - A GPS-Assisted Household Time-Use Survey, *St. Mary's University, Halifax, Canada*

Categories:	HI (Tel), STB (GPS)

Applications:	Halifax, Nova Scotia
Methods Used and/or Summary:	The Halifax Space-Time Activity Research (STAR) Project was claimed to be the world's first largest GPS-assisted prompted-recall time diary survey. "The survey was conducted for a 2-day period covering approximately 2,000 households in Halifax, Nova Scotia, Canada from 2007 to 2008. Person-based GPS devices were used." "The STAR project sampled 1971 households or about one household in 78 within HRM, during the period April, 2007–May, 2008. Primary respondents over the age of 15 completed 48-h time diaries and detailed questionnaire surveys." "After carrying the device for 48 hours and after the travel days are complete, a member of our research team will call the respondent to complete a questionnaire."
Strengths:	Like other GPS prompted recall surveys, the STAR data are particularly rich, since the project couples time diaries with GPS tracking of travel, and thus provides a complete picture of space–time activity.
Weaknesses:	
MAIR:	"A license agreement with a local IT solutions provider with national and international connections enabled commercialization of the software developed by the project. The suite of software products is now being marketed in the courier, freight, field service and public transit sectors."
URL:	http://www.smu.ca/partners/turp/pages/projects/STAR/STAR_Main.htm

Hato E. and H. J. P. Timmermans (2008). Electronic Instrument Design and User Interfaces for Activity Based Modeling, *8th International Conference on Survey Methods in Transport, Harmonization and Data Comparability, Annecy, France*

Categories:	HI (MB, WB), STB (GPS), ETB (RSSA, WA, SP)
Applications:	
Methods Used and/or Summary:	"The focus of this resource paper will be primarily on active data collection. However, because some interesting hybrid data collection instruments have been developed lately, we will also discuss such progress." The paper discusses four data collection methods, their strengths and weaknesses: Computer–assisted activity diaries: "It does not require much effort to appreciate that the use of these computer-assisted instruments results in less errors compared to the traditional paper and pencil methods However, the use of computers also presents a challenge to some respondents. Moreover, it requires extra time to learn how to use the software. Furthermore, in many cases, there is less flexibility. Of course, with mobile technology, this may change. In any case, it means that there is potential sampling bias as not everyone will have a computer,

	or cellular phone. The software also needs to be activated, which does not only take time, but could also involve more of a mental threshold than simply completing a form, for example during the commercial breaks of a favourite television show." Re-enactment sessions: Unlike the computer-assisted data collection tool, these methods provide virtual reality re-enactment session to trigger environmental cues, perceived atmosphere and sense of a place or experience. It performed better in reporting the number of stops and the number of activities in the schedules but most of the duration aspects were reported better by the traditional paper and pencil diary. The technology cannot yet be readily applied to city-wide data collection. Interactive computer experiments: In this method, subjects express their behaviour or some other response measure in reaction to some change generated by the computer. There is the potential issue that they view the interactive experiment as a game as opposed to a serious task. Sample size would be small because subjects need to come to the location of the system and it is costly. Stated preference methods: Data collection involves presenting respondents with hypothetical questions related to transport choice. Research indicates that the formulation and presentation of the question(s) may have large impacts on responses.
Strengths:	Comprehensive review of historic and novel data collection techniques.
Weaknesses:	
MAIR:	
URL:	http://www.isctsc.cl/archivos/2008/A3%20RP%20Hato%20and%20Timmermans.doc

Herrera J.-C., D. Work, X. Ban, R. Herring, Q. Jacobson and A. Bayen (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment, *Transportation Research C,* 18: 568–583

Categories:	STB (GPS), ETB (SP, WA)
Applications:	San Francisco, USA
Methods Used and/or Summary:	"Mobile Millennium is a research project that includes a pilot traffic-monitoring system that uses the GPS in cellular phones to gather traffic information, process it, and distribute it back to the phones in real time. The public-private research partnership - UC Berkeley, Nokia Research Center, and NAVTEQ, with sponsorship from the California Department of Transportation - launched the pilot program from the Berkeley campus on November 10, 2008. It ran for exactly 12 months. During that time, more than 5,000 users downloaded the Mobile Millennium traffic software onto their phones." "Mobile Millennium also highlighted some future challenges that need to be

	addressed by transportation agencies and businesses before similar systems become more commonplace. These challenges include new procurement approaches that are focused on purchasing information rather than equipment, defining the respective roles (and business models) of the public and private sectors in provided traffic information to consumers, and trade-offs between individualized information delivered to a smart phone and distracted driving."
Strengths:	"Mobile Century was intended as a proof of concept to test traffic data collection from GPS-equipped cell phones in one hundred vehicles driven on a 10-mile stretch of a highway located in the San Francisco Bay Area. The phones, which effectively served as vehicle probes, stored vehicle speed and position information every three seconds. "
Weaknesses:	
MAIR:	
URL:	http://www.sciencedirect.com/science/article/pii/S0968090X09001430

Hong F., J. A. Prozzi and A. Leung (2008). Sampling Schemes for Weigh-in-Motion Traffic Data Collection, *Transportation Research Record, Journal of the Transportation Research Board,* 2049: 38-44

Categories:	CB (RS), ETB (weigh-in-motion)
Applications:	Austin Texas
Methods Used and/or Summary:	"This study proposed a methodology for the comprehensive investigation of the effect of different sampling schemes on WIM data accuracy. The selected typical schemes involved three frequencies (monthly, quarterly, and yearly) and three lengths of time at each data collection frequency (1 day, 2 continuous days, and 1 week)."
Strengths:	"In summary, the methods developed by this study can be used by state highway agencies for the more cost-effective collection and processing of WIM data."
Weaknesses:	The disadvantages of this method are high installation and data transfer costs.
MAIR:	
URL:	http://trb.metapress.com/content/82r127457773534l/?p=3779e2a9fefb40c4983e86eb8264d1b2&pi=4

Inbakaran C. and A. Kroen (2011) Travel Surveys – Review of international survey methods, *Australasian Transport Research Forum, Adelaide, Australia*

Categories:	HI (FtF, Tel, MB, WB), STB (GPS, MP)
Applications:	Melbourne, Australia
Methods Used and/or Summary:	"The paper investigates travel survey methods in Australia, and in particular Melbourne, and considers the applicability of methods used elsewhere as well as their potential to improve response rates and data quality. In Australian capital cities travel surveys are either conducted as face-to-face interviews with travel diaries or as self-completion questionnaires which are personally dropped off. As there are different advantages and disadvantages with different methods of conducting survey methods and new technologies have developed in recent years, this paper explores the experiences in other cities and countries and the lessons learned in these areas." A brief review of methods in the following categories is provided: Face-to-face interview, Telephone surveys, Internet Survey, GPS devices and mobile phones. Further, the authors describe the methodology of several specific transport surveys: The Victorian Integrated Survey of Travel and Activity (VISTA) – Melbourne which is mail out mail back survey with a small pilot survey of wearable GOS devices with the travel diary (190 respondents), Sydney Household Travel Survey (HTS) continuous since 1997, South East Queensland Travel Survey (SEQTS) 2003,2004, 2006-2008 and 2009, Metropolitan Adelaide Household Travel Survey (MAHTS) 1999, Greater Hobart Household Travel Survey, 2008-2009, Transportation Tomorrow Survey, Toronto, Chicago Regional Household Travel Inventory – also known as Travel Tracker Survey – has been conducted between January 2007 and February 2008, which uses a mix of telephone and mail contact, Mobility in Germany' (Mobilität in Deutschland) is a national travel survey of about 50,000 households.
Strengths:	"The comparison of the Victorian travel survey to the other surveys has demonstrated that the sampling method in Melbourne is quite effective, and unless it is possible to use data similar to municipal registers relatively easily, should be continued."
Weaknesses:	
MAIR:	
URL:	http://www.atrf11.unisa.edu.au/Assets/Papers/ATRF11_0106_final.pdf

Janssens, D., T. Bellmans, E. Moons and G. Wets (2008). Simulating Emergent Behaviour and Evolution of Activity-Travel patterns: Data Collection Challenges, *8th International Conference on Survey Methods in Transport, Harmonization and data Comparability, Annecy, France*

Categories:	STB (GPS, MP) and ETB (SP, WA, PDA)
Applications:	Flanders
Methods Used and/or Summary:	This paper provides detailed data on activity-travel patterns collected using a combination of paper-and-pencil and GPS/PDA devices. In comparison with other activity-based studies or that time (2008), the survey period is long (7 days). As an interesting contribution, they undertake a data fusion exercise to combine travel survey data with time use survey data by using an iterative proportional fitting method. "The automated activity-travel diary survey tool has been called PARROTS, which stands for PDA (Personal Digital Assistant) system for Activity Registration and Recording of Travel Scheduling. PARROTS runs on a PDA and uses the Global Positioning System (GPS) to automatically record location data."
Strengths:	Long survey period.
Weaknesses:	
MAIR:	
URL:	http://www.isctsc.cl/archivos/2008/31%20B3%20Janssens%20et%20al.doc

Kagerbauer M., W. Manz and D. Zumkeller (2011). Methodological Analysis of Different Methods within one Multi Day Household Travel Survey - PAPI, CATI and CAWI in Comparison, *9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile*

Categories:	HI (PAPI, CATI and CAWI)
Applications:	Stuttgart, Germany
Methods Used and/or Summary:	A combination of three survey methods, PAPI, CATI and CAWI, were used in one survey setup to collect data from the Region of Stuttgart over seven days. The survey showed, that in general the PAPI method works very well and is accepted in all parts of the population. PAPI can still be a useful data collection method. The authors conclude that: "Therefore by using the PAPI design it was possible to obtain reliable results without unexpected non-response effects."
Strengths:	The comparison between the three methods showed that: "Younger people have a strong preference of joining the CAWI survey while elderly or retired people clearly favour the telephone interview."

Weaknesses:	They did not consider the GPS-based prompted recall method in their comparison.
MAIR:	
URL:	

Kelly P., A. Doherty, E. Berry, S. Hodges, A. Betterham and C. Foster (2011). Can We Use Digital Life-Log Images To Investigate Active and Sedentary Travel Behaviour? Results From a Pilot Study, *International Journal of Behavioural Nutrition and Physical Activity,* 8:44

Categories:	ETB (digital camera)
Applications:	
Methods Used and/or Summary:	“Active travel such as walking and cycling has potential to increase physical activity levels in sedentary individuals. Motorised car travel is a sedentary behaviour that contributes to carbon emissions. There have been recent calls for technology that will improve our ability to measure these travel behaviours, and in particular evaluate modes and volumes of active versus sedentary travel. The purpose of this pilot study is to investigate the potential efficacy of a new electronic measurement device, a wearable digital camera called SenseCam, in travel research. Participants (n = 20) were required to wear the SenseCam device for one full day of travel. The device automatically records approximately 3,600 time-stamped, first-person point-of-view images per day, without any action required by the wearer. Participants also completed a self-report travel diary over the same period for comparison, and were interviewed afterwards to assess user burden and experience.”
Strengths:	“Direct observation of travel behaviour from time-stamped images shows considerable potential in the field of travel research. Journey duration derived from direct observation of travel behaviour from time-stamped images appears to suggest over-reporting of self-reported journey duration.”
Weaknesses:	“The device has certain limitations; there are particular settings where participants are not comfortable to wear it and in certain situations such as very low light the images do not always show the journey clearly.”
MAIR:	
URL:	http://www.ijbnpa.org/content/8/1/44

Kochan B., T. Bellemans, D. Janssens, G. Wets and H.J.P. Timmermans (2010) Quality assessment of location data obtained by the GPS-enabled PARROTS survey tool, *Journal of Location Based Services,* 4(2): 93-104

Categories:	HI (WB), STB (GPS), ETB (SP)
Applications:	Flanders, Belgium
Methods Used and/or Summary:	"A custom tool, PARROTS (PDA system for Activity Registration and Recording of Travel Scheduling) was developed to collect activity-travel diary data and global positioning system (GPS) based location data during trips."
Strengths	If the PDA is switched on, PARROTS and the main GUI start automatically. Data is immediately available in an electronic format. The authors find a lower rate of attrition in the PDA sample than that of paper and pencil.
Weaknesses	There are some errors in reporting start and end of trips for which GPS logs were unavailable.
MAIR:	
URL:	http://www.tandfonline.com/doi/abs/10.1080/17489725.2010.506662

Kohla B., G Sammer, R. Wally, M. Herry and R. Tomschy (2011). Comparing Trip Diaries with GPS-Tracking: Results of a Comprehensive Austrian Study, *9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile*

Categories:	HI (WB), STB (GPS)
Applications:	Austria
Methods Used and/or Summary:	"In an Austrian travel survey (MobiFIT) different techniques were used over three days on a total of 235 participants to assess the pros and cons of both survey approaches. The mobility parameters show that self-reported paper-pencil surveys yield accurate social-demographic information about the respondents as well as trip purposes and modes of transportation, whereas too few trips are reported. Automated GPS-based methods have weaknesses identifying trips, mode and purpose. With GPS the length and duration of trips – when identified correctly can be determined more accurately. Comparisons have shown that respondents both over and underestimate their self-reported trip distances, so that the average daily trip distance amounted to almost the same from both approaches."
Strengths:	Participants were split into three groups and surveyed in two methods: GPS with prompted recall, and trip diaries. The results were compared.

Weaknesses:	The automated data processing method is still incomplete and requires more research to improve the quality of results. This study mainly emphasizes the potential and usefulness of large scale GPS-based surveys such as the next Austrian national travel survey.
MAIR:	
URL:	

Krygsman S. and J. Nel (2008). Deriving Transport Data with Cell Phones: Methodological Lessons from South Africa, *8th International Conference on Survey Methods in Transport, Harmonization and data Comparability, Annecy, France*

Categories:	STB (GPS, MP) and ETB (SP, RSSA, WA)
Applications:	Cape Town, South Africa
Methods Used and/or Summary:	"Cellphone data expressly allow for the extraction of home and work locations as well as an indication of intermediate locations. The construction of OD matrices is therefore possible. Possible routes taken can be identified with some methodological innovation. Cellphones also allow for deriving a richer description of individuals' entire daily activity and travel patterns than what is normally available from standard OD-surveys." They used *labelling process* (?) to construct a 48-hour activity and trip record of individuals. Then the data is overlaid with other data layers such as satellite land use classification data, transport zones information and land use zonal data.
Strengths:	Supplementing conventional travel surveys notably OD-surveys
Weaknesses:	"A more in-depth understanding of the accuracy and reliability of cellphone data is required before the technique can be applied in large scale surveys. Specifically, cell sizes and the relationship with transport zones should be considered. Cellphone boundaries seldom 'match' transport zones."
MAIR:	
URL:	http://www.isctsc.cl/archivos/2008/63%20B4%20Krygsman%20and%20DeJong.pdf

Lane N. L., E. Miluzzo, H. Lu, D. Peebles, T. Choudhury and A T. Campbell (2010). A Survey of Mobile Phone Sensing, *IEEE Communications Magazine,* 48(9): 140-150

Categories:	ETB (SP)
Applications:	
Methods Used and/or Summary:	This paper discusses several emerging technologies for data collection using smart phones. "Mobile phones or smartphones are rapidly becoming the central computer and communication device in people's lives. Application delivery channels such as the Apple AppStore are transforming mobile phones into App Phones, capable of downloading a myriad of applications in an instant. Importantly, today's smartphones are programmable and come with a growing set of cheap powerful embedded sensors, such as an accelerometer, digital compass, gyroscope, GPS, microphone, and camera, which are enabling the emergence of personal, group, and community scale sensing applications. We believe that sensor-equipped mobile phones will revolutionize many sectors of our economy, including business, healthcare, social networks, environmental monitoring, and transportation. In this article we survey existing mobile phone sensing algorithms, applications, and systems. We discuss the emerging sensing paradigms, and formulate an architectural framework for discussing a number of the open issues and challenges emerging in the new area of mobile phone sensing research." The authors list different applications of low-level sensor data and high-level events, context, and activities inferred from mobile phone sensor data in transportation, social networking, environmental monitoring, and health and well-being. They identify three building blocks in the mobile phone sensing architecture: sense, learn, and inform, share and persuasion. The paper concludes: "The primary obstacle to this new field is not a lack of infrastructure; millions of people already carry phones with rich sensing capabilities. Rather, the technical barriers are related to performing privacy-sensitive and resource-sensitive reasoning with noisy data and noisy labels, and providing useful and effective feedback to users."
Strengths:	
Weaknesses:	
MAIR:	
URL:	http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5560598

Lawson, C. (2008). Data Visualization Techniques, 8^{th} *International Conference on Survey Methods in Transport, Harmonization and data Comparability, Annecy, France*

Categories:	HI(), STB(), ETB ()
Applications:	
Methods Used and/or Summary:	The author summarizes a wide variety of data visualization software and its applications to the field of transportation planning. Applications discussed include computer assisted drawing and design, animation and simulation, intelligent transportation systems including freight and passenger data, dynamic interactions, GIS, and space-time path analysis. Further, the author discusses the incorporation of visualization techniques into travel surveys. Topics discussed include hardware and software formats, confidentiality and data archiving issues. The paper closes with a list of topics for discussion including data asset management strategies, participatory research, the open-source environment and sources of support.
Strengths:	
Weaknesses:	
MAIR:	
URL:	http://www.isctsc.cl/archivos/2008/B5%20RP%20LAWSON.doc

Licoppe C., D. Diminescu, Z. Smoreda and C. Ziemicki (2009). Using Mobile Phone Geolocalisation for 'Socio-Geographical' Analysis of Co-ordination, Urban Mobilities, and Social Integration Patterns, *Journal of Economic and Social Geograph*y, 99(5): 584-601

Categories:	ETB (SP, Blu, WA)
Applications:	Paris, France
Methods Used and/or Summary:	This research aimed to reconstruct urban mobilities and communication practices through mobile phone base data. A software probe that can be implemented on a user's mobile phone, and which allows the joint recording and collection of the successive locations experienced by the user was developed (through the identification of the cell in which the mobile phone is located) and all types of communicative acts performed through the mobile phone. This has been combined with in-depth interviews with subjects over one week of their mobility and mobile communication behaviour. The precision of these data is more limited

	than that of GPS systems and does not show a person's movements within a short distance (a 100 metre radius in town). However, with the combination of location and mobile communication data, individuals' movements as well as the spatiotemporal distribution of their mobile communication practices can be mapped.
Strengths	Because the methodology allows for observation over many months, the authors have been able to analyse entanglements between mobility and mobile communication practices leading to very different patterns.
Weaknesses	
MAIR:	
URL:	http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9663.2008.00493.x/full

Ling X. (2011). Understanding Travel Behaviour and Vehicle Emissions from GPS and Diary Data: An Application to Southern California, *Master of Science Thesis, University of Irvine*

Categories:	STB (GPS)
Applications:	Southern California, USA
Methods Used and/or Summary:	"The purpose of this thesis is to explore the impact of socio-economic characteristics of drivers on travel behavior and on vehicular emissions of various air pollutants using microscopic data. My starting dataset was collected by SCAG in 2001 and 2002 during their post 2000 Census Regional Travel Survey. Of the 16,939 households who answered the survey, 297 provided self-reported 24-hour travel diary data and detailed GPS data for their vehicles, which was instrumented for SCAG's survey. After selecting 100 out of these 297 households based on their socio-economic characteristics and the completeness of their answers, I relied on 2003 imagery in Google Earth to match diary and GPS data. An extensive clean-up of this dataset yielded a sample of 701 trips, for which I estimated emissions of CO, CO2, NOx, HC, PM10, and PM2.5 using OpMode in EPA's MOVES2010 (Motor Vehicle Emissions Simulator) from second-by-second GPS travel data. A statistical analysis of the results reveals that men make longer trips than women, although the difference in their emission rates is not statistically significant. Moreover, people 60 or older are the greenest drivers: their driving patterns are more environmentally benign because they accelerate/decelerate less than younger people. Finally, I found significant differences in emission rates based on different household income levels."
Strengths:	By using GPS and travel diaries, participants' driving behaviour can be considered in emission estimations.

Weaknesses:	The author uses MOVES which limits consideration of the impact of disaggregate behaviour of individuals on the level of emission.
MAIR:	
URL:	http://gradworks.umi.com/14/97/1497074.html

Lucas K. (2011) Qualitative Methods in Transport Research: Is 'Action Research' a Methodology too Far?, *9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile*

Categories:	ETB (action research)
Applications:	London, Leicester, UK
Methods Used and/or Summary:	This paper discusses 'action research' as an emergent transport survey method. "Reason and Bradbury's handbook of action research identifies the method as *'a participatory and democratic process concerned with developing practical knowing in the pursuit of worthwhile human purposes, grounded in a participatory worldview'."..."* It is a method which is increasingly being brought into play to help local communities understand the environmental and social consequences of their travel behaviours and to encourage them to seek alternative travel options. The paper identifies that these micro-level transport projects present a significant challenge for travel survey methodology and suggests that transport researchers should be seeking to use the best of technology together with more collaborative data collection methods to capture the impacts of such initiatives." For example in the case of a GPS prompted recall survey being done in the UK (i-Connect), the authors mention that the use of community researchers in the data collection process can considerably improve the quality of the survey.
Strengths:	"[A]ction research with communities is also a tool for empowerment and the new-found power to act it gives to its participants may not always encourage them to do so in the ways in which policymakers wish them to."
Weaknesses:	"The question remains as to how far transport researchers should be intervening to drive or guide such projects in the 'right' direction of change and whether, if they do so, they undermine the objectivity of their research observations."
MAIR:	
URL:	

Manfredini F., P. Tagliolato and C. Di Rosa (2011). Monitoring Temporary Populations Through Cellular Core Network Data, *Lecture Notes in Computer Science,* 6783: 151-161

Categories:	STB (MP), ETB (SP)
Applications:	Lombardia, Italy
Methods Used and/or Summary:	Collecting data about tourists attending a big event is not an easy task. This study attempts to study the behaviour of visitors during the period of the Milan International Design week. Having limited demographic data of the mobile users and linking it to the data provided by the mobile switching centre (MSC), a simple descriptive analysis is provided.
Strengths:	The idea of studying visitors travel behaviour using MSC data is remarkable.
Weaknesses:	The scope of the analysis is very limited. A very simple analytical method is used to study the behaviour of cell phone users.
MAIR:	
URL:	http://www.springerlink.com/content/l51l87187750p68h/

Messelodi A., C. M. Modena, M. Zanin, F. G. B. De Natale, F. Granelli, E. Betterle and A. Guarise (2009) Intelligent Extended Floating Car Data Collection, *Expert Systems With Applications,* 36(1): 4213-4227

Categories:	CB (OB), STB (GPS) and ETB (floating-car)
Applications:	Rome Italy
Methods Used and/or Summary:	“An on-board car-data analyzer hypothesizes continuously the traffic status along the trip. … [T]he hypothesis of traffic queue or jam is discarded by a vision-based module that looks ahead for the presence of vehicles: if there is no vehicle ahead, the low speed of the host vehicle is not due to traffic. A second on-board vision module is able to react on temporary danger warning signs, typically placed on road-works sites. This real-time algorithm labels images as containing, or not, the specific road-works pattern.”
Strengths:	“An extended FCD packet is prepared containing vehicle identifier, time, position, vehicle data, estimated traffic level and images of the scene ahead. The xFCD packets collected by the moving platforms are transmitted to the Main Center where they are merged by proper algorithms to maintain an updated map of the traffic conditions and where image content can be analyzed by specialized algorithms.”

Weaknesses:	The disadvantages of this method are high installation and data transfer costs.
MAIR:	
URL:	http://www.sciencedirect.com/science/article/pii/S0957417408002042

NCHRP (2011). State DOT Public Transportation Performance Measures: State of the Practice and Future, *Transportation Research Board, Report to the National Cooperative Highway Research Program on Project 20-65. Washington, DC*

Categories:	HI ()
Applications:	
Methods Used and/or Summary:	"The research finds that many state DOTs are tracking public transportation performance measures to increase accountability to stakeholders, improve management and decision-making, and comply with state mandates and federal data requirements. Most of these performance measures focus on ridership and internal factors (e.g., cost, efficiency), though quality and asset management are becoming more widespread. States with the most advanced public transportation performance measurement were notable for the linkages they made between their goals, performance measures, and funding decisions; their data collection efforts; collaboration with public transportation providers; and reporting methods. A number of challenges remain, however, for advancing public transportation performance measures at state DOTs. Collecting data and connecting performance to funding decisions are two key challenges. Many state DOTs pointed to a need to find ways to compare disparate public transportation systems and to collect accurate and relevant data from their public transportation providers. Moreover, developing appropriate performance measures is often challenging, given the disparate nature of different types of public transportation services, particularly in rural areas."
Strengths:	
Weaknesses:	
MAIR:	
URL:	http://www.trb.org/Main/Blurbs/166065.aspx

Ortuzar J. D., J. Armoogum, J. L. Madre and F. Potier (2010). Continuous Mobility Surveys: The State of Practice, *Transport Reviews,* 31(3): 293-312

Categories:	HI (FtF, Tel, MB, WB), SBT (GPS)
Applications:	Various
Methods Used and/or Summary:	The authors summarize the current state of practice in continuous data collection for transportation planning. Endeavours are examined that occur at the national and regional levels. The authors go on to contrast longitudinal (panel) and cross-sectional data collection. Interrupted continuous surveys such as the National Transport Survey in Sweden were treated with other surveys, cross-sectional by design. The authors discuss sampling issues specific to continuous data collection endeavours. Conclusions drawn heavily favour the collection of further continuous data throughout the year, from a geographically and chronologically controlled sample, supported by permanent and motivated staff.
Strengths:	
Weaknesses:	The paper does not focus sufficiently on emerging technology based surveys.
MAIR:	
URL:	http://www.tandfonline.com/doi/abs/10.1080/01441647.2010.510224

Pelletier M. P., M. Trepanier and C. Morency (2011). Smart card data use in public transit: A literature review, *Transportation Research Part C: Emerging Technologies,* 19(4): 557-568

Categories:	ETB (SC)
Applications:	Gatineau, Quebec, Canada
Methods Used and/or Summary:	The authors describe several studies using smart card data that have been undertaken. They go on to classify 3 types of studies. Strategic-level studies have as aim long term planning endeavours which may link usage data, geographic data, and socio-economic data either directly, or as an enrichment from household surveys. Care must be taken to expand the collected data to represent all users in the system, not just those with smart cards. Tactical level studies seek to improve day to day functioning of the transit system by identifying route load-profiles and boarding and alighting points to identify efficiency adjustments. Operational-level studies seek to create system-level indicators such as adherence to schedule, vehicle/person kilometres per run, route or day, and correct functioning of the smart-card fare payment system itself.

	The authors go on to discuss commercialization of smart card systems by various means and conclude with directions for future research.
Strengths:	Comprehensive introduction to smart card technology. Strong framework for identifying the scope and aims of smart card studies., tactical-level studies, operational-level studies and commercialization.
Weaknesses:	The authors only discuss smart card data use in transit and do not provide the advantages and disadvantages of using such data to complement household travel surveys.
MAIR:	
URL:	http://www.sciencedirect.com/science/article/pii/S0968090X1000166X

Roorda, M. and A. Shalaby (2008). Transportation Data Collection in the Greater Golden Horseshoe: A Framework and Priorities for Improvement, *Final Report for the Data Management Group, UTRAC, University of Toronto*

Categories:	HI(Tel, WB)
Applications:	Greater Golden Horseshoe (GGH)
Methods Used and/or Summary:	The report is prepared for the Data Management Group (DMG) which undertakes the Transportation Tomorrow Survey (TTS) in the GGH every 5 years. The authors seek to answer two key questions. Will data from the TTS, in conjunction with other sources, be sufficient to answer transport questions in the near future? How should the methodology of the TTS be updated to deal with technological and societal changes currently underway? The authors define the “universe of travel” that makes up all travel in the GGH, list sources of data pertinent to this universe, and identify elements about which little data is collected e.g. non-systematically or not at all. The report goes on to prioritize which under- reported submarkets are most in need of increased data collection. These priorities are based on the amount of traffic, its impacts (“environmental, social, economic, health and safety, operational, etc.”, and its relevance to policy interventions. The merit and failings of current data collection for the forum of policy intervention are examined. The new data collection priorities are listed in descending order as follows, while the report itself includes discussion of relevant precedents in Canada and the United States. 1: “Goods movement, service provision and other business travel, by automobile and truck, within the GGH.” 2:” Economic elements of personal travel, including income and

	costs of travel and parking." 3: "Information about the types and ages of vehicles owned by the household." 4: "Non-motorized personal trips to/from non-work/non-school trip purposes." 5: "Travel of children under the age of 11." 6: "Weekend travel, and seasonal variation in travel." 7: "Tourism and intercity business travel." Issues of data harmonization are discussed briefly. The authors go on to identify challenges facing the TTS which can be generalized to any travel survey such as sample selection and recruitment, as well as specific and general strategies to address these challenges including dual-frame sampling and new survey instruments. Finally, specific recommendations are made to address the initial key questions of the TTS.
Strengths:	Axiomatic definitions of travel data availability and deficiencies. Specific recommendations to address deficiencies.
Weaknesses:	
MAIR:	
URL	

Sanggu L. And M. D. Heckman (2011). Travel Pattern Analysis Using Smart Card Data of Regular Users, *Transportation Research Board, 90th Annual Meeting,* 11: 4258

Categories:	ETB (SC)
Applications:	
Methods Used and/or Summary:	"Smart card fare collection systems are no longer a new trend: this will be the inevitable fare payment media in future transit networks. This paper addresses the use of smart card data to understand transit passenger behavior. Specifically, this paper (1) proposes a method of inferring boarding stops using Geographic Information Systems (GIS) and database management using Structured Query Language (SQL), and (2) analyzes the travel patterns of regular transit users overall and by fare type. The experiment of matching Automatic Fare Collection (AFC) data to Automatic Passenger Count with Vehicle Location (APC/VL) data clearly shows that intersection-level identification of passengers' boarding stops can be successfully inferred. Through structured queries of the AFC data, regular users of a typical weekday are defined. The need for multi-day analysis to examine the travel and activity behaviour of regular users is emphasized. Spatial-temporal characteristics then are analyzed by fare type and duration between transactions. In addition, the variability and regularity of transit use is investigated using data from the first transaction during each weekday."

Strengths:	
Weaknesses:	
MAIR:	
URL:	http://amonline.trb.org/13icu6/13icu6/1

Schüssler N., L. Montini and C. Dobler (2011). Improving Post-Processing Routines for GPS Observations Using Prompted-Recall Data, *9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile*

Categories:	HI (WB), STB(GPS), ETB(Acc)
Applications:	Zurich, Switzerland
Methods Used and/or Summary:	Using a GPS device with a built-in accelerometer, the authors collected data for a pilot of 15 individuals over four weeks. Participants carry the device during the day. Overnight when the device is charged, the GPS data is transmitted to a central SQL database. It is then post processed to find mode and stop points. The coordinates of the public transport stops are inputted to data processing software. The authors use a fuzzy logic method to determine the mode of transportation. As data was processed, participants can see the results via the internet to confirm or correct them.
Strengths:	The post processing software detects the actual mode in most of the cases.
Weaknesses:	“The stop point detection seems to detect slightly more stop points than there were mode transfers and activities resulting in a higher share of short stages. Moreover, several of the long duration stages are missing.” Stop points are defined as: zero speed longer than a certain time threshold, high density clusters, signal loss for longer than a certain time threshold, changes from and to walk based on speed and acceleration and no movement recorded by accelerometer. Therefore, due to technological limitations of the GPS device (such as cold start) around 30 percent of stop points are not detected accurately.
MAIR:	
URL:	

Schüssler, N. and K. W. Axhausen (2008). Identifying Trips and Activities and their Characteristics from GPS Raw Data without Further Information, *8th International Conference on Survey Methods in Transport, Harmonization and data Comparability, Annecy, France*

Categories:	STB (GPS)
Applications:	Zurich, Winterthur and Geneva
Methods Used and/or Summary:	Several post-processing procedures for data cleaning, trip and activity detection and mode identification in JAVA are explored. Not only are these methods able to handle the huge amount of data efficiently, it is also independent of the quality of the network and the spatial information available. Modules developed: data filtering, detection of trips and activities, mode stage determination, mode identification, map-matching
Strengths:	Mode of transportation was identified using a fuzzy logic approach.
Weaknesses:	Mode detection was not validated. Several assumptions were made about the behaviour of individuals.
MAIR:	
URL:	http://www.isctsc.cl/archivos/2008/51%20B4%20Schuessler%20and%20Axhausen.pdf

Seidl, D. (2012). Spatial Analysis of GPS Unit Returns in the Greater Minneapolis Region, *American Association of Geographers, Annual Meeting, Modelling and Visualizing Travel Behavior II*

Categories:	HI(FtF, Tel) STB(GPS),
Applications:	Minneapolis, USA and Southern California, USA
Methods Used and/or Summary:	The authors use a prompted recall survey of Minneapolis residents where diary data is collected via the web. GPS is used for validation of the diary survey. A pilot of 250 households (550 individuals) were surveyed in 2010. The final target is to collect data on 5000 households in the Metropolitan Area by 2012. This project is sponsored by the Metropolitan Council of Governments. Abt SRBI is conducting a similar survey in the Greater Los Angeles Area which is sponsored by the Southern California Association of Governments (SCAG). This data will be collected between 2011 and 2013. Around 6000 households will be interviewed among which it is planned to recruit 1750 household to carry GPS loggers during the day. Data will be collected for 3-7 days .
Strengths:	The device and data processing procedure are very advanced and therefore capable of determining spatial and temporal activity attributes.
Weaknesses:	

MAIR:	
URL:	http://meridian.aag.org/callforpapers/program/AbstractDetail.cfm?AbstractID=41416

Sharp J. and E. Murakami (2005). Travel Surveys: Methodological and Technology-Related Considerations, *Journal of Transportation and Statistics,* 8(3): 97-113

<table>
<tr><td>Categories:</td><td>HI (Tel, FtF, MB and WB)</td></tr>
<tr><td>Applications:</td><td>None</td></tr>
<tr><td>Methods Used and/or Summary:</td><td>"The primary objective of this paper is to introduce and discuss survey methodology considerations for the next series of personal travel surveys conducted by the U.S. Department of Transportation"

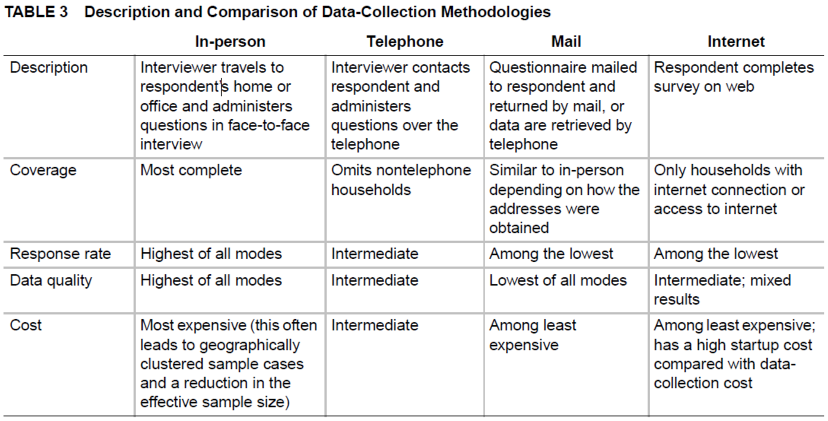



TABLE 3 Description and Comparison of Data-Collection Methodologies
<table>
<tr><th></th><th>In-person</th><th>Telephone</th><th>Mail</th><th>Internet</th></tr>
<tr><td>Description</td><td>Interviewer travels to respondent's home or office and administers questions in face-to-face interview</td><td>Interviewer contacts respondent and administers questions over the telephone</td><td>Questionnaire mailed to respondent and returned by mail, or data are retrieved by telephone</td><td>Respondent completes survey on web</td></tr>
<tr><td>Coverage</td><td>Most complete</td><td>Omits nontelephone households</td><td>Similar to in-person depending on how the addresses were obtained</td><td>Only households with internet connection or access to internet</td></tr>
<tr><td>Response rate</td><td>Highest of all modes</td><td>Intermediate</td><td>Among the lowest</td><td>Among the lowest</td></tr>
<tr><td>Data quality</td><td>Highest of all modes</td><td>Intermediate</td><td>Lowest of all modes</td><td>Intermediate; mixed results</td></tr>
<tr><td>Cost</td><td>Most expensive (this often leads to geographically clustered sample cases and a reduction in the effective sample size)</td><td>Intermediate</td><td>Among least expensive</td><td>Among least expensive; has a high startup cost compared with data-collection cost</td></tr>
</table>
Table 3 provides a good summary of the four methods of household interview methods.
Recommendations for better data quality: improving response rate, non-response follow-up study and bias analysis, modified sample design/frame, improved coverage of rarer important modes of travel, reduce trip underreporting.</td></tr>
<tr><td>Strengths:</td><td></td></tr>
<tr><td>Weaknesses:</td><td></td></tr>
<tr><td>MAIR:</td><td></td></tr>
<tr><td>URL:</td><td>https://2bts.rita.dot.gov/publications/journal_of_transportation_and_statistics/volume_08_number_03/pdf/entire.pdf#page=107</td></tr>
</table>

Simas Oliviera M. G. and J. Casas (2010) Improving Data Quality, Accuracy, and Response in On-Board Surveys: Application of Innovative Technologies, *Transportation Research Record, Journal of the Transportation Research Board,* 2183: 41-48

Categories:	HI (WB), STB(GPS)
Applications:	United States
Methods Used and/or Summary:	"Transit agencies conduct origin–destination on-board surveys periodically to gather information regarding travel patterns and demographic data of their users and to collect customer satisfaction information. These surveys constitute a highly valuable means of obtaining important information on an agency's customers to provide a basis for effective transit planning and for regional travel demand modeling efforts. This paper describes the application of innovative technologies in the data collection process to improve data quality, data completeness, and data collection management. This includes the simultaneous collection of boarding and alighting count data at the stop level using Global Positioning System–enabled personal digital assistants, association of distributed surveys to boarding locations, and a web-based sample and productivity management system. These technologies allow for automatic collection of boarding location, arrival and departure times, and transit trip times. An imputation procedure was developed to derive the most likely alighting location of each collected sample. Joint application of these technologies reduces survey length and thereby minimizes respondent burden."
Strengths:	
Weaknesses:	
MAIR:	
URL:	http://trb.metapress.com/content/jt010h52497k026l/?p=794a1b52c06a4d18b46f46ad6977bf7c&pi=4

SmartCard Alliance (2006). Smart Cards and Parking, *Smart Card Alliance Transportation Council, White Paper TC-06001*

Categories:	STB (SC), ETB (Radio frequency - based transponders)
Applications:	USA

Methods Used and/or Summary:	Smart cards play an integral role in new payment strategies by providing increased security, enabling more distributed processing and providing a variety of communications options. In 18 US cities smart card transit payment systems have been established with a budget of over $1 billion. Similar developments have occurred in the toll collection industry, using long range radio frequency-based transponders. These technologies enable researchers to collect data without contacting respondents.
Strengths:	
Weaknesses:	The type of data collected by these technologies is very limited and privacy issues decelerate the process of linking these data to other data sources with socio-demographic information.
MAIR:	
URL:	http://www.smartcardalliance.org/resources/lib/SmartCards_Parking_FINAL_123005.pdf

Smith K. and G. Spitz (2010). Internet Access: Is Everyone Online Yet and Can We Survey Them There?, *Transportation Research Record, Journal of the Transportation Research Board,* 2175: 35-41

Categories:	HI (WB)
Applications:	United States
Methods Used and/or Summary:	“This paper presents two case studies that review data collected in travel surveys in metropolitan areas in the United States to identify coverage error, specifically whether lack of Internet access by a segment of the population leads to coverage error. The first case study analyzes data from several road pricing surveys to quantify differences between those who have and do not have Internet access. The second case study analyzes data from two transit origin–destination surveys in which respondents were invited to provide contact details for additional market research. This paper compares the overall sample with those willing to be surveyed in the future by telephone and those willing to receive a future survey invitation by e-mail. Both case studies find that the samples with access to the Internet are similar to the larger full samples that include those without Internet access and therefore that the coverage error found in the Internet-only samples is small. The results suggest that, for surveys of general populations of drivers or transit riders, surveying only those with Internet access does not introduce significant coverage error into travel survey samples.”
Strengths:	The final conclusion is interesting that: “for surveys of general populations of

	drivers or transit riders, surveying only those with Internet access does not introduce significant coverage error into travel survey samples."
Weaknesses:	The conclusion cannot be generalized because it is only based on two case studies.
MAIR:	
URL:	http://trb.metapress.com/content/fk03517828081400/?p=ff12c5c8414540dbb9fdd55d9e3e0f29&pi=3

Smith, P. (2009). Survey Harmonisation in Official Household Surveys in the United Kingdom, *International Statistical Institute, Invited Paper Proceedings, Durban, South Africa*

Categories:	HI ()
Applications:	United Kingdom
Methods Used and/or Summary:	It is explained that the focus of the presentation is on the original design of the U.K. Integrated Household Survey (IHS) but includes a discussion of the challenges the United Kingdom has faced related to the design over the years. The design of the IHS relies on the use of modules formed from four existing continuous household surveys: the Labour Force Survey (LFS) (including some regional supplementary surveys), which serves as the IHS survey core and provides the majority of sample cases (200,000 households); the General Lifestyle Survey (formerly the General Household Survey); the Living Cost and Food Survey (formerly the Expenditure and Food Survey); and the Opinions Survey (formerly the Omnibus Survey). After the original modular design incorporated these four surveys, others, such as the English Household Survey, were added. The idea behind the modules was to standardize concepts and questions across the surveys. In its current form, the survey sample includes 265,000 households and uses a staged approach.
Strengths:	The expected benefits of the proposed improvement in the survey design include reduced sampling variance due to increased sample size, cost savings associated with the "un-clustering" of the sample designs, and two-phase calibration, which will enable the use of the estimates from the core in calibration for components.
Weaknesses:	The survey core of IHS ended up being too long to be practically administered in the field.
MAIR:	
URL:	http://www.statssa.gov.za/isi2009/ScientificProgramme/IPMS/0238.pdf

Stopher P. R., C. Prasad, L. Wargelin and J. Minser (2011). Conducting a GPS-Only Household travel Survey, *9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile*

Categories:	STB (GPS), ETB (WA)
Applications:	Ohio-Kentucky-Indiana Metropolitan Area
Methods Used and/or Summary:	Using a GPS device, second by second travel data was collected. The data is processed to impute trip ends, mode occupancy and purpose. The respondents were also asked to complete an internet PR survey for one of the three days surveyed. The results of this PR survey are used to validate the results of processing and to improve the software. "The target sample size was originally set at over 3,500 households to be collected steadily over a twelve-month period, although this target was reduced downwards during the course of the survey."
Strengths:	The processing software seems to work effectively because for most of the cases it accurately determines the mode or purpose of the trip. Determining mode of transportation by using internal accelerometers, gyroscopes, barometers, and magnetic compasses built in PDAs and smart phones is a way of accurately imputing travel attributes. Therefore, accurately imputing travel mode, purpose and destination by only using GPS data is a remarkable achievement.
Weaknesses:	The survey is not a prompted recall survey. Therefore the respondents may forget details of the trip. Limited travel attributes can be imputed by only using GPS data. There is no discussion about how this GPS –based survey can substitute household travel surveys because it still requires an additional PR survey supplement.
MAIR:	
URL:	

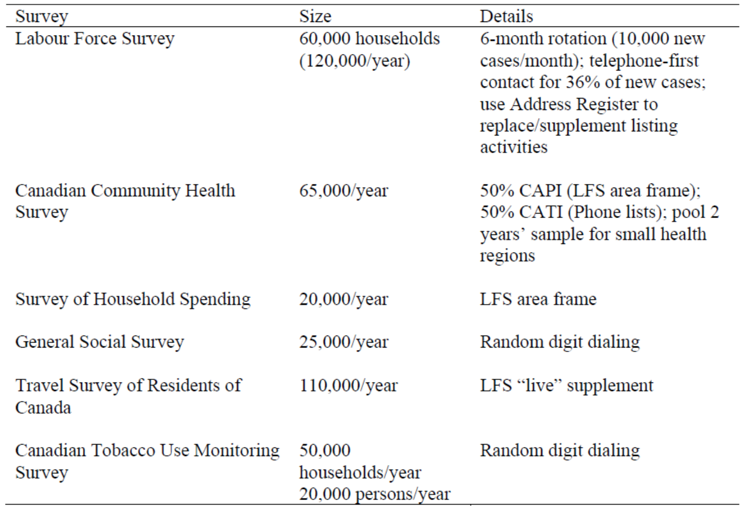
Stopher, P.(2009). Collecting and Processing Data from Mobile Technologies, *8th International Conference on Survey Methods in Transport, Harmonization and data Comparability, Annecy, France*

Categories:	HI (), STB (GPS, MP) and ETB (SP, RSSA, WA, Blu)
Applications:	Ohio, United States

Methods Used and/or Summary:	This paper attempts to answer whether one of the following technologies – GPS, GSM, RFID, Wi-Fi, or combinations thereof – are ready to replace conventional travel surveys. It concludes that "The fact that no serious efforts have yet been made to build demand models from mobile technology data is probably a further element that must be tested before there is more widespread use of these technologies." As there is no standard software for HTS with mobile technologies therefore, "the heuristics that are applied, or the use of Artificial Intelligence or fuzzy logic programming can enhance the results."
Strengths:	"From the standpoint of the sample, it appears that the samples from mobile technology surveys will be slightly less biased than those from conventional surveys. While similar biases are likely to exist relating to one-person households and non-car-owning households, there appears to be some evidence that there will be less bias against large households."
Weaknesses:	"[L]imitations of the use of mobile technologies by age may be an issue for such data"
MAIR:	
URL:	http://www.isctsc.cl/archivos/2008/B4%20RP%20Stopher.doc

Tambay, J.-L. (2010). Statistics Canada's Household Survey Strategy, *National Research Council of the National Academies, Presentation for the Workshop on the Future of Federal Household Surveys, November 4-5, Washington, DC*

Categories:	HI (FtF, Tel, MB, WB)
Applications:	Canada
Methods Used and/or Summary:	"Currently, Statistics Canada has three major sampling vehicles for household surveys: (1) the Labour Force Survey (LFS) area frame design, (2) random digit dialling (RDD), and (3) a census of population, conducted every five years." "Statistics Canada has developed several strategies grouped under the term "New Household Survey Strategy," including survey integration, spreading interviewer and response burden, development of a master sample, creation of a population frame, and integration of listing activities. The process of survey integration includes using a common core of questions for all surveys, harmonizing content modules, creation of a master sample, and integrating survey and census listing activities." This table summarized the major Canadian surveys with monthly data collection



Survey	Size	Details
Labour Force Survey	60,000 households (120,000/year)	6-month rotation (10,000 new cases/month); telephone-first contact for 36% of new cases; use Address Register to replace/supplement listing activities
Canadian Community Health Survey	65,000/year	50% CAPI (LFS area frame); 50% CATI (Phone lists); pool 2 years' sample for small health regions
Survey of Household Spending	20,000/year	LFS area frame
General Social Survey	25,000/year	Random digit dialing
Travel Survey of Residents of Canada	110,000/year	LFS "live" supplement
Canadian Tobacco Use Monitoring Survey	50,000 households/year 20,000 persons/year	Random digit dialing

Strengths:	
Weaknesses:	
MAIR:	"Other indirect methods of obtaining more complete information are also under consideration, such as matching tax records to Infodirect phone numbers to add apartment numbers that are missing on Infodirect. Exploring a cell phone billing file was also attempted. An application to sample from this frame has yet to be developed. A consequence of trying to add additional phone numbers to the frame is that regional offices are communicating that their telephone centers are already operating at capacity with the phone numbers that are currently in certain frames. Statistics Canada is also attempting to expand its address resources using such tools as municipal lists and tax forms."
URL:	http://www7.nationalacademies.org/cnstat/Tambay%20Slides.pdf

TCRP Synthesis 63 (2005). On-Board and Intercept Transit Survey Techniques, *Federal Transit Administration in Cooperation with the Transit Development Corporation,* Project J-7, Topic SH-05

Categories:	CB (OB)
Applications:	USA
Methods Used and/or Summary:	"This report documents and summarizes transit agencies' experience with planning and implementing on-board and intercept surveys. This information can help transit staff responsible for market research in their agency. A survey of 52 transit agencies found that 96% conducted on-board surveys between 2002 and 2004, with most of this group also having conducted intercept surveys. Large agencies typically conduct five or more on-board/intercept surveys annually, primarily

	focused on specific routes or geographic areas. Small agencies typically conduct surveys every 1 to 3 years, often involving the entire transit system. On-board and intercept surveys generally address two to four of the following research topics: Where and when do customers use transit? Who uses transit? How satisfied are customers? Why do customers use transit? How could the agency attract increased ridership? How effective are agency communications and information? Response rates reported by transit agencies varied widely, from 13% to 90% of riders who were asked to participate in a given survey. Within this very broad range, response rates for the majority of on-board and intercept surveys ranged from 33% to 67%, with one-half of the agencies reporting response rates in this range. As with response rates, costs also vary widely. Overall costs for surveys reported by transit agencies ranged as high as $350,000, although approximately one-fifth of surveys with cost data reported by transit agencies cost less than $10,000. Even surveys employing similar methodologies show widely different costs. Factors affecting costs include the number of completed surveys to be obtained, whether personal interviews or self-administered questionnaires are used, whether survey staff dedicated to the task or bus operators are used to distribute surveys, the density of riders at survey locations, survey length and complexity, response rates, and labour costs.”
Strengths:	
Weaknesses:	“In on-board and intercept surveys, the devil is in the details. Although transit agencies have developed many effective practices through experience, there is insufficient methodologically sound research to guide decisions in two key areas: (1) impact of design and layout of questionnaires and (2) impact of the use of incentives.”
MAIR:	
URL:	http://www.trb.org/Main/Blurbs/156542.aspx

TCRP Synthesis 69 (2006). Web-Based Survey Techniques A Synthesis of Transit Practice, *Federal Transit Administration in Cooperation with the Transit Development Corporation, Project J-7, Topic SH-07*

Categories:	STB(WB)
Applications:	USA
Methods Used and/or Summary:	“This synthesis of transit research describes how web-based surveys are being used by transit agencies and other transit researchers and documents the experiences of web-based survey research as applied to transit. In addition, this study documents not only the current state of the practice, but also provides a

	resource for successful practices in web-based surveying, discusses the technologies necessary to conduct web-based surveys, and presents some specific case studies of transit agency use of web-based survey techniques." "The synthesis ends with the conclusions including the following aspects of web-based surveys that transit agencies have found to be successful: Start simply with web-based surveys to learn the differences between web-based surveys and other survey methods. Attempt to collect databases of e-mails from customers and potential customers to use as a sampling source for research. Apply web-based survey methods in a multi-method survey environment to improve response rates by providing response alternatives and to enable the transit researcher to gain the benefits of web-based survey data and techniques. This finding is made knowing that measurement error is an issue with multi-method surveys; therefore, this must be balanced against the benefits. Research the issue of coverage error and try to minimize sampling bias. Remain cautious but optimistic about including web-based surveys in research programs as the survey methods and the Internet mature."
Strengths:	"There were 175 researchers at both transit agencies and in the private sector in the selected/convenience sample of potential respondents that the topic panel and consultant chose to survey. Of those, 25 responses were received from researchers at transit agencies and 11 from researchers in the private sector of the transit industry."
Weaknesses:	
MAIR:	
URL:	http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_syn_69.pdf

Thiagarajan A., L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo and J. Erriksson (2009). VTrack: Accurate, Energy-Aware Traffic Delay Estimation Using Mobile Phones, *7th International Conference on Embedded Networked Sensor Systems (ACM SenSys '09), Berkeley, California*

Categories:	ETB (SP, WiFi, WA)
Applications:	Boston, USA
Methods Used and/or Summary:	"Traffic delays and congestion are a major source of inefficiency, wasted fuel, and commuter frustration. Measuring and localizing these delays, and routing users around them, is an important step towards reducing the time people spend stuck in traffic. As others have noted, the proliferation of commodity smartphones that can provide location estimates using a variety of sensors—GPS, WiFi, and/or cellular triangulation opens up the attractive possibility of using position samples

	from drivers' phones to monitor traffic delays at a fine spatiotemporal granularity. This paper presents VTrack, a system for travel time estimation using this sensor data that addresses two key challenges: energy consumption and sensor unreliability. While GPS provides highly accurate location estimates, it has several limitations: some phones don't have GPS at all, the GPS sensor doesn't work in "urban canyons" (tall buildings and tunnels) or when the phone is inside a pocket, and the GPS on many phones is power-hungry and drains the battery quickly. In these cases, VTrack can use alternative, less energy-hungry but noisier sensors like WiFi to estimate both a user's trajectory and travel time along the route. VTrack uses a hidden Markov model (HMM)-based map matching scheme and travel time estimation method that interpolates sparse data to identify the most probable road segments driven by the user and to attribute travel times to those segments. We present experimental results from real drive data and WiFi access point sightings gathered from a deployment on several cars. We show that VTrack can tolerate significant noise and outages in these location estimates, and still successfully identify delay-prone segments, and provide accurate enough delays for delay-aware routing algorithms. We also study the best sampling strategies for WiFi and GPS sensors for different energy cost regimes."
Strengths:	"1) Reducing energy consumption using inaccurate position sensors (WiFi rather than GPS), and 2) obtaining accurate travel time estimates from these inaccurate positions.
Weaknesses:	"There are a couple of interesting directions for future work. We plan to develop an online, adaptive algorithm that dynamically selects the best sensor to sample taking available energy and the current uncertainty of the node's position and trajectory into account. Second, improving the quality of today's algorithms to predict future travel times on segments, using historical travel times and sparse amounts of real-time data, would be useful for traffic-aware routing.
MAIR:	
URL:	http://nms.lcs.mit.edu/papers/vtrack-cr-submission.pdf

Tomlinson M., W. Solomon, Y. Singh, T. Doherty, M. Chopra, P. Ijumba, A. C. Tsai and D. Jackson (2009). The Use of Mobile Phones as a Data Collection Tool: A Report from a Household Survey in South Africa, *BMC Medical Informatics and Decision Making,* 9(51): 1-8

Categories:	ETB (SP, WA)
Applications:	Durban, South Africa

Methods Used and/or Summary:	The authors seek to take advantage of mobile technology, combined with the improvement that mobile phones offer over PDAs in terms of data loss and uploading difficulties. "A web-based system was developed to allow electronic surveys or questionnaires to be designed on a word processor, sent to, and conducted on standard entry level mobile phones." General conclusion: The experience with a large scale baseline survey suggests that real-time quality control and supervision by data collectors make the use of a mobile phone based survey system an attractive management option and preferable to a paper based approach.
Strengths:	The cost of the device is low ($40 per person).
Weaknesses:	Requires an extensive training procedure. "Training for the data collection protocol was conducted for all the CHW's over a two day period with the data quality control officer present." The software used was not open source.
MAIR:	
URL:	http://www.biomedcentral.com/1472-6947/9/51

Trevor H. and E. Hildebrand (2011). Experiences with GPS Travel Diaries in Rural Older Driver Research, *Transportation Research Board, 90th Annual Meeting,* 11: 4258

Categories:	STB(GPS) ETB (GIS)
Applications:	New Brunswick Canada
Methods Used and/or Summary:	"This paper describes using passive Global Positioning Systems (GPS) data collection and Geographic Information System (GIS) with participant prompted recall to study the travel habits of rural older drivers. It is based upon the research of a convenience sample of 60 rural drivers (29 men, 31 women, average age 69.6 years) in New Brunswick, Canada. The transportation needs of a growing population of older rural residents, many who face the risk of not being able to meet their needs if they can no longer drive, are not well understood and represent an immediate and future policy need. GPS-based travel diaries are a useful method to obtain origin/destination and other contextual information in support of rural transportation planning. A total of 1,649 "stops" (periods of non-movement lasting 1 minute or more) by participant vehicles were recorded with the GPS units."
Strengths:	"Approximately 8% of all "stops" were due to stoplights or traffic delay. The remaining "stops" were organized into 1,494 trips (one origin with one destination), with participants supplying purposes, who was driving, and passenger

	details for 99.1% of recorded trips. Travel data were collected on average for 5.3 days per participant. An external battery for the GPS unit minimized the typical satellite acquisition but was exhausted in 10% of cases. Only 2.2% of recorded trip ends were due to lost reception or acquisition delay and in each case the missing distance data were interpolated. Service clubs and snowball sampling were the most effective means of recruiting rural participants.”
Weaknesses:	“The use of a battery pack to complement a vehicle’s power supply to the GPS unit was highly effective. The external battery was exhausted in only six of the 60 surveys which minimized the typical acquisition delay of 1 – 2 minutes and potential missed trips that could have occurred had the GPS unit power been tied to the vehicle ignition exclusively.”
MAIR:	“Jurisdictions with substantive rural populations should consider this method to complement their rural transportation planning efforts, but should involve a third party for the data collection and analysis to avoid perceptions of it being a government driving assessment tool.”
URL:	http://amonline.trb.org/165poe/165poe/1

Wiehe S. E., A. E. Carroll, G. C. Liu, K. L. Haberkon, S. C. Hoch, J. S. Wilson and D. Fortenberry (2008). Using GPS-enabled cell phones to track the travel patterns of adolescents, *International Journal of Health Geographics,* 7:22

Categories:	STB (GPS, MP), ETB (SM, PDA)
Applications:	Indiana, USA
Methods Used and/or Summary:	Participants carried Blackberry 7520 GPS-enabled cell phones (cost $150/phone). “Using the BlackBerry Java Development Environment, a background module was created that ran on startup and in the background. It was therefore invisible to the phone user and could not easily be turned off. At 5 minute intervals, the program obtained a "fix" on the participant's location and, through an http call, transmitted the location coordinates, battery level, device ID, and timestamp to a java servlet running on a server dedicated to the study. All data were stored in a MySQL database.” The participant (15 adolescent women from a clinic-based setting) answered several questions each evening before going to bed using a program developed for the Blackberry platform.
Strengths:	
Weaknesses:	Device is costly

MAIR:	“Next steps include monitoring larger numbers of adolescent women, repeatedly over a year, to investigate differences in travel patterns by whether they engage in various health-risk behaviours.”
URL:	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2426678/pdf/1476-072X-7-22.pdf

Wolf J., J. Wilhelm, J. Casas and S. Sen (2011). A Case Study: Multiple Data Collection Methods and the NY/NJ/CT Regional Travel, *9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track, Termas de Puyehue, Chile*

Categories:	HI (FtF, Tel, MB and WB) STB (GPS)
Applications:	New York, North New Jersey, and Connecticut
Methods Used and/or Summary:	“This survey is using a combination of web, telephone, and mail-out/mail-back methods to recruit and retrieve household, person, vehicle, and travel information from approximately 18,800 households. Ten percent of the sampled households are participating by using wearable Global Positioning System (GPS) devices that collect detailed travel data which, in turn, is processed and presented back to the households in a GPS-based prompted recall interview administered by web or telephone.”
Strengths:	By providing multiple methods of data collection they maximized survey participation. They collected data from 83% of the targeted population.
Weaknesses:	It does not suggest anything new to each data collection method.
MAIR:	
URL:	

Yen K. S., T. B. Swanston, P.W. Wong, A. Adamu, M. Assadi, A. Benouar, B. Ravani and T. A. Lasky (2009). Support for Business Case Development for the GPS-Automated Travel Diary (GPS-ATD) in Preparation for the 2010 State-Wide Travel Behaviour- Phase 1, *Final Report of Contract IA 65A0210*

Categories:	ETB (PDA, SP)
Applications:	California, USA

Methods Used and/or Summary:	A complete comparison of the advantages and disadvantages of using passive GPS loggers, PDA with built-in GPS and GPS-ATD devices is provided in the report. Furthermore, the authors evaluate the cons and pros of different user interfaces that are available for each of the technologies. The authors conclude: the GPS-ATD provides increased data quality and information with simultaneous decreased user burden. They developed a prototype software for a generic GPS-enabled android-based device.
Strengths:	
Weaknesses:	
MAIR:	
URL:	http://ahmct.ucdavis.edu/?projects=gps-automated-travel-diary

Yu P., M. De Courten, E. Pan, G. Galea and J. Pryor (2009). The Development and Evaluation of a PDA-Based Method for Public Health Surveillance Data Collection in Developing Countries, *International Journal of Medical Informatics,* 78(8): 532-542

Categories:	ETB (PDA, SP)
Applications:	Fiji
Methods Used and/or Summary:	"EpiData and Epi Info are often used together by public health agencies around the world, particularly in developing countries, to meet their needs of low-cost public health data management; however, the current open source data management technology lacks a mobile component to meet the needs of mobile public health data collectors. The goal of this project is to explore the opportunity of filling this gap through developing and trial of a personal digital assistant (PDA) based data collection/entry system. It evaluated whether such a system could increase efficiency and reduce data transcription errors for public surveillance data collection in developing countries represented by Fiji. A generic PDA-based data collection software eSTEPS was developed. The software and the data collected using it directly interfaces with EpiData. A field trial was conducted to test the viability of public health surveillance data collection using eSTEPS. The design was a randomised, controlled trial with cross-over design. 120 participants recruited from the Fiji School of Medicine were randomly assigned to be interviewed by one of six interviewers in one of the two ways: (1) paper-based survey followed by PDA survey and (2) PDA survey followed by paper-based survey. Data quality was measured by error rates (logical range errors/inconsistencies, skip errors, missing values, date or time field errors and incorrect data type). Work flow and cost were evaluated in three stages of the survey process: (1) preparation of data collection

	instrument, (2) data collection and (3) data entry, validation and cleaning. User acceptance was also evaluated in the two groups of participants: (1) data collectors and (2) survey participants.”
Strengths:	“None of the errors presented in 20.8% of the paper questionnaires was found in the data set collected using PDA. 62% of the participants perceived that the PDA-based questionnaire took less time to complete.” ... “The eSTEPS field trial proves that PDA was more efficient than paper for public health survey data collection.”
Weaknesses:	
MAIR:	
URL:	http://www.sciencedirect.com/science/article/pii/S1386505609000392

Yuan Y., M Raubal and Y. Liu (2011) Correlating Mobile Phone Usage and Travel Behaviour - A Case Study of Harbin, China, *Computers, Environment and Urban Systems, In Press*

Categories:	ETB (SP)
Applications:	Harbin, China
Methods Used and/or Summary:	The correlation between mobile phone call frequencies and three indicators of travel behaviour: (1) radius, (2) eccentricity, and (3) entropy is explored. In addition, explanatory factors, such as age, gender, social temporal orders and characteristics of the built environment as they impact the relationship between mobile phone usage and individual activity behaviour are examined. Some travel attributes can be inferred using phone call frequencies.
Strengths:	This research is a good start point to extract travel behaviour of individuals by exploring their phone call pattern.
Weaknesses:	The presented method can only poorly simulate travel attributes.
MAIR:	
URL:	http://www.sciencedirect.com/science/article/pii/S0198971511000652

PART III – DATA INTEGRATION/FUSION METHODS

1. INTRODUCTION

Many transportation datasets provide useful information about travel behaviour or transportation system performance, but they often contain data gaps that limit modelling or analysis capabilities. Chalasani and Axhausen (2005) classify transportation data into the following categories: travel survey data, transportation infrastructure data, transportation system performance data and geographic data. Each of these categories may contain data of different types and contexts. However, in almost all transportation analysis, we need data belonging to more than one type. Moreover, with increasing difficulty in constructing appropriate sampling frames and reaching diverse target populations,[36] as well as the increasing amount and quality of data required for advanced transportation modelling (among other planning applications), it is becoming impossible to collect all the data required during a single survey or with a single method (Bayart et al. 2008). Given this, integrating or *fusing* multiple data sources together to create unified databases is increasingly important in urban transportation planning analysis and modelling. The purpose of this part of the report is to discuss the data fusion problem and the variety of methods available for integrating/fusing individual datasets to create new, more comprehensive datasets.

The concept of data integration/fusion involves creating a new, more comprehensive dataset by joining/combining multiple datasets to fill information gaps (Gautier, 1999; Saporta, 2002). A common example of this is the use of Census data to add socio-economic variables (income is a typical example) to travel survey records that are missing these variables. Compilation of comprehensive databases by fusing multiple sources can enhance the useful information available to transportation system users, transportation system operators/service providers and system planners (Amey et al. 2009), and, in some cases may even eliminate the need to undertake new surveys or data collection efforts by using available data to construct the needed information. It also may be possible to combine datasets collected in different time periods and in different spatial contexts, as long as there are some common elements among the individual datasets.

Section 2 of this part provides an overview description of the data fusion problem and introduces several key issues that generally need to be addressed in data fusion exercises. Section 3 establishes a typology of data fusion contexts. Section 4 then provides brief descriptions of the various data fusion techniques currently available.

---

[36] See Part II: *Literature Review* for a discussion of these issues.

2. DATA FUSION: DEFINITIONS & ISSUES

Figure 2.1 illustrates the typical data fusion problem in which two datasets (A and B) exist that share a certain number of common variables (X_A and X_B). One of the datasets contains additional variables (Y_A) that are not found in the second dataset. The problem is to impute the values of these missing variables (Y_B) given the relationships (correlations) between the X and Y variables observed in dataset A. In this problem, the X variables are called *common variables*, the Y variables are *target variables*, and datasets A and B are referred to as the *donor* and *receptor* datasets, respectively.

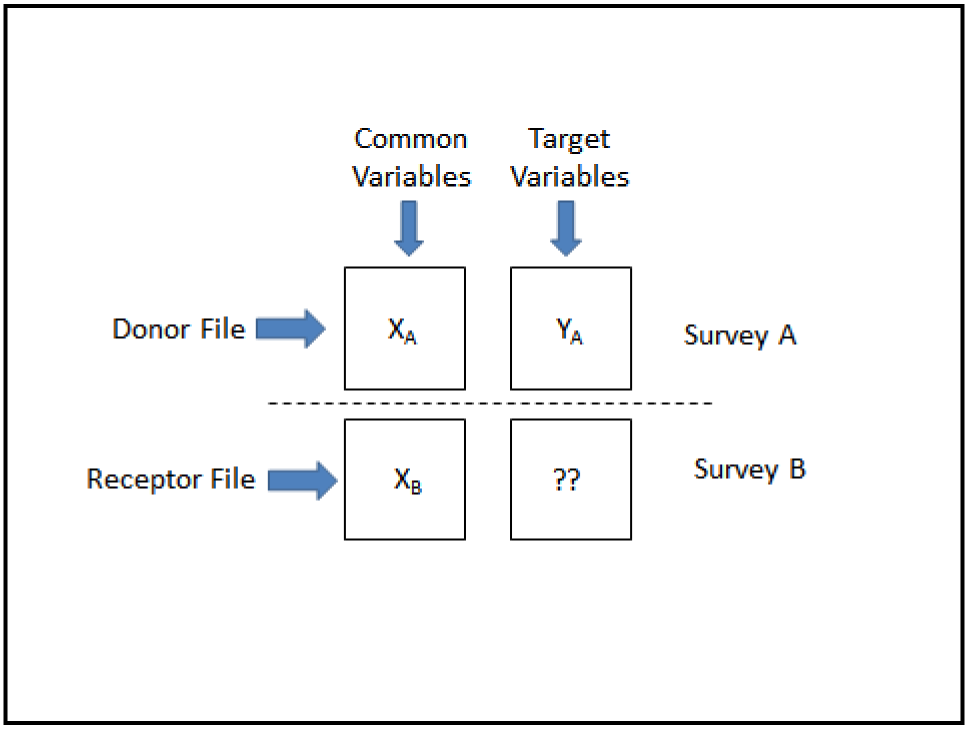


Figure 2.1: The Data Fusion Problem (D'Ambrosio, 2007**)**

All data fusion problems one way or another involve using the observed relationships (joint distribution) among the X_A and Y_A variables to *impute* appropriate values of the Y_B variables, given the observed X_B variables (Gilula et al. 2006). Building on this basic task, several major types of data fusion applications exist, including:

- Using the combined A and B datasets once the Y_B variables are known for modelling or analysis purposes. Using a combined revealed preference/stated preference (RP/SP)[37] dataset for mode choice model estimation (see Sections 3 and 4) is an example of this type of application.

---

[37] A revealed preference survey gathers information on actual travel choices made by a respondent. A stated preference survey asks respondents to state what they would do in a given hypothetical situation. See Part II*: Literature Review* for further discussion of RP and SP surveys. A combined RP/SP survey augments RP questions

- Using the updated B (receptor) dataset for modelling or other analysis purposes. In this case, the B dataset is typically a file of travel survey records and the A dataset contains information that is missing in the original B dataset but that is needed for the given application. Imputing incomes for trip-makers in a travel survey dataset from census data is a typical, simple example of this type of application.
- Creating a *synthetic population* for use in microsimulation models. In this case, the original B dataset typically consists of a set of aggregate tables of data (household size by income by census tract; person age by gender by labour force status; etc.) for the entire population being modelled. The dataset A contains much more detailed information for a sample drawn from this population (e.g., a set of individual person or household records, in which each record corresponds to a specific person or household and contains the actual attributes for this agent – age, gender, labour force status, etc.). A Public Use Microdata (PUM) file of actual individual census records (with spatial and individual identifiers deleted) provided by most census agencies,[38] or other types of surveys are typical examples of such a donor dataset. This dataset provides a sample estimate of the joint distribution of the full set of variables (X and Y) of interest and can be used in a variety of ways with the aggregate "marginal" population tables in the receptor file to synthesis the desired set of disaggregate agents for this population (Pritchard, 2008).
- If the A and B datasets are repeated cross-section surveys[39] with the same sets of variables, then these datasets can be combined to create a *pseudo-panel*. In this pseudo-panel observations in one dataset are linked to observations in the second dataset that match sufficiently closely to be treated as if they were the "same" respondents in the two datasets. Thus, the linked observations can be treated as if they were observations from a panel survey. Such pseudo-panel data can provide the basis for dynamic, process-based modelling (Bernard et al. 2012). Construction and use of pseudo-panel data is not yet generally used in operational planning practice, but is receiving increasing attention in research applications. Pseudo-panels have the potential to provide deeper insights into changes in travel behaviour over time than might be obtained from simple analysis of repeated cross-section data while avoiding the cost and difficulties of panel surveys.

Thus, data fusion requires:

- Two or more datasets that collectively contain the required information.
- At least one common variable among the datasets.
- A suitable statistical method for undertaking the desired fusion.

---

about actual choices with a set of SP experiments that are linked in way or another to the RP choices that the respondents make.

[38] See, for example, http://datalib.chass.utoronto.ca/major/canpumf.htm.

[39] See Part II***: Literature Review*** for a discussion of repeated cross-section surveys.

A major challenge in data integration is the different levels of aggregation that often exists in different datasets. Differences in aggregation may be spatial (e.g., different zone systems), temporal (e.g., different time periods) or semantic, i.e., the categorization of the variables themselves (e.g., use of different categories for age, income, occupation/industry definitions, etc.). These differences need to be reconciled if the datasets are to be correctly fused. Ever-increasing planning requirements to address complex policy issues, combined with the availability of more detailed datasets and increased computational power leads to the need/opportunity for using high resolution/microscopic travel behaviour data across spatial, temporal and semantic dimensions to develop advanced analysis/modelling methods. This further accentuates both the need for data fusion techniques to construct the required complex, detailed databases and the technical challenges involved. The aggregation levels of different datasets play a crucial role in success and failure of such data fusion efforts (Polak, 2006; D'Orazio et al. 2006).[40]

---

[40] Whenever possible, care should be taken in designing data collection efforts to gather data at as disaggregate level as possible so as to facilitate matching of these data at various levels of aggregation (i.e., data can always be aggregated but it cannot be disaggregated with certainty). Similarly, accessing disaggregate base data (as opposed to published more aggregate results), should obviously be pursued whenever possible.

3. DATA FUSION CONTEXTS

Data fusion for transportation analysis can be broadly classified into the following exhaustive set of categories:

- ***Spatial context***: Data fusion based on common spatial contexts (neighbourhood features, regions, cities, neighbourhoods, census tracts, traffic analysis zones, etc.). That is, the common variable between the two datasets is the zone or other spatial indicator that allows one set of zonal variables to be merged with a second set.
- ***Temporal context***: Data fusion based on a common temporal context (same year, month of the year, week of the month, day of the week, time of the day, etc.).
- ***Semantic context***: Data fusion based on common semantic contexts (same age groups, same gender groups, same household size groups, same auto ownership levels, same occupation groups, same income groups, etc.)
- ***Mixed spatial, temporal and/or semantic context***: Data fusion based on combinations of spatial, temporal and semantic contexts.
- ***Survey mode context***: Data fusion for datasets collected through different modes of surveys (face-to-face, telephone interview, web-based, etc.).
- ***Data type context***: Data fusion based on data types for advanced econometric modelling (e.g., combination of revealed reference (RP) and stated preference (SP) data).

Each of these contexts is briefly discussed below.

Spatial context: Expanding sample survey data to represent the whole population of a study area is the simplest and most widely used example of data fusion based on a common spatial context. Calculation of weighting factors for the survey observations based on census data of the study area allows the sample dataset to be expanded to represent the whole population. In such cases, calculation of the expansion weighting factors is very straightforward, since it is simply the inverse of the selection probability of the individual in the sample from the sample frame.

Matching datasets via postal codes, census tracts and traffic analysis zones to supplement household travel survey data with land use data, aggregate census information, etc. is another common example of data fusion used in transportation analysis. The major challenge in these applications occurs when the two zone systems (spatial aggregations) involve overlapping boundaries. For example, census tracts and traffic zones often are not simple aggregations of one another, so that a given traffic zone may intersect with multiple census tracts, or *vice versa*. In such cases, at least three options for reconciling the spatial aggregations exist:

- Aggregate both zone systems up to a new, common super-zone system in which each of the original zones is wholly contained within one and only one super-zone.
- Use GIS-based calculations to split the units of observation (persons, households, etc.) in one zone and allocate these units to the portions of the other zone with which the first zone intersects. This allocation of units is often done on the basis of the fraction of the zone's area that lies in each of the overlapping zones. This implies a uniform distribution of the units within the zone, a reasonable assumption in the absence of additional information. If information exists to alter this assumption it should be used. For example, the available of detailed

information concerning dwelling unit distributions within the zone (from Google Earth©, DMTI Spatial datasets[41] or other similar sources) may permit non-uniform allocations to be made with confidence.

- Use GIS-based calculations to split both zone systems into a common set of fine-grained grid cells to which the zone unit/attributes can be allocated. Again, either a uniform distribution assumption or data-driven allocations from Google Earth/DMTI type datasets can be used to perform these calculations.

Temporal context: Attaching transportation level-of-service variables (travel times, costs, etc.) for a given time period that have been computed using road and transit network assignment models to travel survey records is a very simple example of temporal context fusing, in which one of the common variables is the time period within which the trip occurred. Combining different count datasets by time period is another common and important temporal context fusing.[42]

Although not often discussed, temporal aggregation is always an issue in any transportation analysis, given that trips occur continuously over time and travel times are continuous variables. When analysing trip-making by time period one is always making the assumption of uniform trip rates, mode splits, etc. throughout the time period. With respect to the fusion problem, inconsistent time period definitions obviously pose problems similar to those caused by inconsistent zone systems (e.g., it may be the case that two sets of traffic counts use different peak-period definitions).

Semantic context: Creating income variables to attach to travel survey data based on census information is a common example of data fusion within a semantic context. Income variables (average income, median income, etc.) can be generated from census data according to occupation types, job categories, neighbourhood or traffic analysis zone, etc. that are common between the Census and travel survey datasets. In such cases, income variables for each observation in the travel survey data can be directly imputed by assigning Census incomes to persons or households in the survey that have the same values of the common variables (same household size, occupation type, etc.). Direct imputation and more complex semantic fusion techniques are discussed further in Section 4.

Aggregation issues also exist in semantic fusion contexts. A very typical case involves two data tables that share a common variable (e.g., age) but use different aggregation categories. If the two categorizations map consistently into one another (e.g., one uses 5-year intervals and one uses 10-year intervals) then the finer categorization can simply be aggregated to the coarser categories. If the two categorizations overlap with one another, however, methods similar to the spatial aggregation case must be used, i.e., one of:

- Aggregate both categorizations to a common set of super-categories that both original categorizations unambiguously map into.

---

[41] http://www.dmtispatial.com/en.aspx?gclid=CP2H1p78vq0CFY0BQAoddnZxAg.

[42] If data from different seasons is being combined, then, obviously, possible seasonal effects need to be accounted for,

- Split one of the categorizations and allocate portions to the other categorization using appropriate assumptions (again often a uniform distribution assumption is made).
- Break down both categorizations to a finer, common system and statistically allocate data from both categorizations to the common finer system. For example, in the case of age categories, both categorizations could be disaggregated down to single years.

Survey mode context: Tradition survey methods of face-to-face, mail-back and telephone interviews are no longer the only ways by which we collect data, with innovative survey methods such as computer- and web-based surveys and other new technologies (GPS, mobile phones, etc.) being used for collecting travel behaviour data.[43] In many cases, multiple surveys are used to collect data from the same spatial and/or time contexts. Integrating data collected by different survey modes raises a number of statistical and practical issues. In particular, each survey method has its own set of instrument biases and other data quality issues. Thus, even in cases in which the fusion of data collected in a multi-mode framework may appear to be straightforward, analysis and modelling using the fused dataset should recognize the mode-specific randomness and errors existing within this dataset.

Data type context: Similarly, stated preference (SP) data are increasingly being used in place of or to supplement traditional revealed preference (RP) data. In particular, it is very common to combine RP and SP data in the joint estimation of travel demand models (typically disaggregate mode choice models) in order to take advantage of the strengths of both types of data. This is a particular form of data fusion in which two datasets are both used to determine the parameters of a single model, thereby incorporating information from both datasets within the same model. Not only does this represent a multi-mode survey application, which, as discussed above, implies varying survey instrument biases across the two dataset, but SP data are hypothetical responses gathered in an experimental setting, as opposed to RP data which are observations of actual choices made by respondents in real-world contexts. Thus, unobserved differences can be expected between the nature of the variables in the two datasets (stated versus actual choices; hypothetical versus real-world explanatory variables) that need to be accounted for in the statistical estimation process, as well as in the interpretation of the survey results themselves.

Mixed contexts: Data fusion problems more often than not involve multiple contexts. The example used above concerning matching trip travel times to travel survey records actually involves temporal (trip start time), spatial (trip origin and destination) and semantic (trip mode) components. Mixed contexts are not necessarily more difficult to deal with than single contexts (as the trip ravel time example illustrates), although the combinatorics of dealing with aggregation issues across multiple dimensions may in some cases be challenging. A simple example of the latter situation might be two datasets in which both age categories and trip mode choice definitions both involve different aggregations that make fusing the two datasets more problematic.

---

[43] See this project's Part II*: Literature Review* for further discussion of these trends.

4. DATA FUSION TECHNIQUES

4.1 INTRODUCTION

As described in Section 2, all data fusion problems involve inferring missing values of the target (Y) variables in the receptor dataset by making use of the observed relationships (joint distribution) among the common and target variables in the donor dataset. Data fusion techniques can be broadly classified into the following categories (Bayart et al, 2008; D'Ambrosio et al, 2007, Gilula, 2006):

Matching with certainty: If common variables can be matched with certainty, then no statistical procedure for record fusing is needed. Such approaches are also called exact matching or record linkage (Johansen, 2005). Common examples include:

- Matching travel survey records with aggregate demographic, land use or other zone-based attributes based on census tracts or traffic analysis zones. In this case, the common variable is the zone which is common between the donor and receptor datasets and which allows the donor target variables to simply be copied to the receptor record. Additional common variables, of course, may also exist (same income group, etc.).
- Adding level of service variables to travel survey records from network assignment models outputs by matching time of the day, origin zone, destination zone and mode of transportation.
- Joining RP and SP data collected from the same individual by matching individual-specific identifiers.

Matching with uncertainty: Statistical techniques are needed whenever the datasets' common variables do not permit exact matching to occur. All statistical procedures dealing with such cases can be termed as matching with uncertainty. This refers to that fact that, since the common variables among the datasets do not match exactly, the analyst needs to use the joint distribution among the common and target variables estimated from the donor dataset to statistically impute values for the receptor target variables.

Two broad classes of approaches exist for matching with uncertainty:

- ***Classical or explicit method***: This approach involves econometric estimation using the donor dataset to develop a mathematical model that predicts a target variable value as a function of a set of common variables; i.e., a model of the form Y = f(X) is developed using the observed data in the donor dataset. This model is then used to impute the same variable of interest that is missing in the other dataset as a function of the common variables in that dataset. Depending of the type of the variable of interest different statistical techniques can be used. If the variable of interest is a continuous variable, then linear regression, non-linear regression, non-parametric regression, etc. are all possible techniques that might be used. If the variable of interest is a discrete variable, then logistic regression, binomial regression, etc. are candidate procedures. As one example of this approach, Hain et al. (2011) developed a wage rate model using Statistics Canada Survey of Labour and Income Dynamics data that is then used to estimate wages in a population synthesis procedure.

- ***Implicit method***: Implicit methods of data fusion involve matching records of one dataset with the records of another where the values of the overlapping variables are not exactly the same, but are "as similar as possible". Section 4.2 discusses implicit methods for matching with uncertainty in detail.

The statistical reliability of matching with uncertainty obviously depends on the statistical properties of the methods used to synthesize the target variable(s). Using such synthesized variables obvious introduces some modelling error into the data set (which will then be transmitted to any analyses or models making use of these data). The assumption, obviously, is that the additional explanatory power provided by the synthesized variables exceeds the "noise" introduced through the synthesis process. While this will generally be the case (e.g., introducing a robust estimate of income into a model is typically preferred to leaving income effects out of the model all-together), the assessment of the impact of the introduced uncertainty on the analysis/model should always be considered. In the case of explicit methods, this will be indicated by the statistical properties of the econometric model used. Section 4.2 discusses statistical issues associated with implicit methods. In all cases, care should be taken that the fused dataset is validated as best as possible in terms of:

- Consistency with any available "external" datasets not used in the fusion process.
- Lack of bias relative to the original datasets.

4.2 IMPLICIT METHODS FOR MATCHING WITH UNCERTAINTY

4.2.1 Definitions

To understand implicit matching, first define the following terms:

I	Set of receptor observations, i=1,…,I
J	Set of donor observations, j=1,…,J
$weight_i$	The sampling weight for observation i in the receptor dataset[44]
$weight_j$	The sampling weight for observation j in the donor dataset
X_i	Vector of common variable values for observation i in the receptor set
X_j	Vector of common variable values for observation j in the donor set
Y_i	Vector of target variable values to be determined through the matching process for receptor observation i
Y_j	Vector of known target variable values for donor observation j
$distance_{ij}$	Some measure of the "distance" or "difference" or "mismatch" between observation i in the receptor set and observation j in the donor set. A simple such metric would be the sum of absolute differences between the X_i and X_j variables; i.e.,:

$$distance_{ij} = \sum_{k=1}^{K} |X_{ik} - X_{jk}| \qquad \text{where K = the number of common variables} \qquad (4.1)$$

---

[44] If the sample is unweighted, then all weights equal 1 in value. In this discussion, the general case of weighted observations is considered throughout.

$weight_{ij}$ The weight attached to donor observation j's contribution to the calculation of receptor observation i's Y_i values; $weight_{ij} \geq 0$. That is:

$$Y_{it} = \sum_{j=1}^{J} weight_{ij} * Y_{jt} \qquad t=1,\ldots,T = \text{number of target variables} \qquad (4.2)$$

where the $weight_{ij}$ variables are chosen so as to minimize the total distance ("mismatch") among the target and donor common variables:

$$\underset{\{weight_{ij}\}}{\text{MIN}} \quad \sum_{i=1}^{I} \sum_{j=1}^{J} weight_{ij} * distance_{ij} \qquad (4.3)$$

4.2.2 Unconstrained versus Constrained Implicit Matching Approaches

Rodgers (1984) classifies the implicit method of data fusion into two broad categories (as explained by Bayart et al. 2008): unconstrained and constrained methods, each of which is briefly described below.

Unconstrained method: In this approach, the Y_j values are simply attached to receptor observation i for the donor observation j whose X_j common variables are closest in value to i's X_i values. Observation i retains its original weight and so:

$$\begin{aligned} weight_{ij} &= weight_i \text{ if donor observation j is assigned to receptor observation i} \\ &= 0 \text{ otherwise} \end{aligned} \qquad (4.4)$$

This approach is labelled "unconstrained" since donor records may be allocated to as many receptor records as required, and, as a result, the original weights (and, hence distribution of values) assigned to the Y_j target variables in the donor set are not, in general, maintained in the imputed receptor Y values.

Constrained method: In a constrained approach, multiple donor records can be used to compute weighted values for the target Y_i variables (equation 4.2), where these weights are computed so that they reproduce both the original receptor observation weights ($weight_i$) and the original donor observation weights ($weight_j$). That is, the weights chosen to minimize the objective function (4.3) must simultaneously satisfy the constraints:

$$\begin{aligned} &\sum_{j=1}^{J} weight_{ij} = weight_i \quad where \quad i = 1,2,3......I \\ &and \\ &\sum_{i=1}^{I} weight_{ij} = weight_j \quad where \quad j = 1,2,3......J \end{aligned} \qquad (4.4)$$

Constrained methods are generally preferred to unconstrained methods since failing to constrain the weights to replicate the donor weights can lead to significant bias in the estimation of the target Y_i values (Pendyala, et al. 2010).

Figure 4.1 provides a simple illustration of unconstrained and constrained matching applied to the same dataset. In this figure, Table 1a contains “file A” a receptor dataset consisting of 8 records (A1-A8), while Table 1b displays “file B”, a donor set containing 6 records (B1-B6). Two common variables exist between the two datasets (sex and age), but the values of these variables do not exactly match between their records (e.g., record A1 is a 42 year-old male; no male of this age exists in file B). Receptor set A has two “Y” variables (birth year and “ln(E)”), while donor set B has a third “Y” variable (“ln(P)”) that we would like to “add” (impute values for) to the receptor set records. In this example, each receptor observation has a weight of 3, while each donor observation has a weight of 4.

Table 1c in the figure shows the output from an unconstrained matching procedure. In this table, each A record has had the ln(P) value of the B record attached to it for the B record that is “closest” in terms of its sex and age values as possible to the A record values. For example, record A1 (with sex=M and age=42) has the B5 record (sex=M; age=41) ln(P) value of 7.243 attached to it. Note that the number of records in output file remains the same (8), with each record having the same weight as before the matching (in this case, a weight of 3 for each record). Also note, however, that both the mean and the standard deviation of the ln(P) variable in the matched dataset are <u>not</u> the same as in the original donor file: the mean has changed from 5.55 to 6.30 and the standard deviation has changed from 1.57 to 1.06. These differences indicate that the unconstrained matching procedure has introduced bias in the imputed values of ln(P) due to the failure to maintain the donor file weights (i.e., the weighted distribution of this variable’s values in the population as defined in the file B observations).

In the constrained method’s results shown in Table 1d in Figure 4.1 this problem is corrected. Multiple donor records are used to compute the receptor records’ target variables. In this case study, for example, donor records B2 and B5 are both matched (with weights 1 and 2, respectively) to receptor record A1.[45] This results in an increased number of records in the output file (12 instead of the original 8), but the total number of weighted records remains the same, equal to the number of weighted records in both the original A and B files (24). As a result, the means and standard deviations for all variables remains the same as in the original files (i.e., no bias has been introduced through the matching process). It should be noted that the total “distance” or “mismatch” in the constrained match is larger than in the unconstrained match (6.46 versus 4.88), but this is a small price to pay to avoid bias in the imputation of the target variables.

---

[45] This is equivalent to creating adding a weighted average value of ln(P) to the A1 record that is equal to $(\text{weight}_{A1,B2} * \ln(P)_{B2}) + (\text{weight}_{A1,B5} * \ln(P)_{B5}) / (\text{weight}_{A1,B2} + \text{weight}_{A1,B5}) = (1*5.524 + 2*7.243)/(1+2) = 6.670$.

Table 1a: Simplified Example of Statistical Matching
File A

Case	$S_1=X_1^A$ Sex	S_2 Year Born	X_2^A Age	y^A ln(E)	W_i^A Weight
A1	M	1938	42	9.156	3
A2	M	1945	35	9.149	3
A3	F	1917	63	9.287	3
A4	M	1925	55	9.512	3
A5	F	1952	28	8.494	3
A6	F	1927	53	8.891	3
A7	F	1958	22	8.425	3
A8	M	1955	25	8.867	3
Mean	0.50		40.38	8.97	
SD	0.53		15.32	0.38	

Table 1b: File B

Case	$t_1=X_a^B$ Sex	$T_2=X_2^B$ Age	Z^B Ln(P)	W_j^B Weight
B1	F	33	6.932	4
B2	M	52	5.524	4
B3	M	28	4.223	4
B4	F	59	6.147	4
B5	M	41	7.243	4
B6	F	45	3.230	4
Mean	0.50	43.00	5.55	
SD	0.55	11.58	1.57	

Table 1c: Unconstrained Match

A Case	B Case	$X_1^A= X_1^B$	X_2^A Age	X_2^B Age	d_{ij}	y^A	z^B	w_{ij}
A1	B5	M	42	41	1	9.156	7.243	3
A2	B5	M	35	41	6	9.149	7.243	3
A3	B4	F	63	59	4	9.287	6.147	3
A4	B2	M	55	52	3	9.512	5.524	3
A5	B1	F	28	33	5	8.494	6.932	3
A6	B4	F	53	59	6	8.891	6.147	3
A7	B1	F	22	33	11	8.425	6.932	3
A8	B3	M	25	28	33	8.867	4.223	3
	Mean	0.50	40.38	43.25	4.88	8.97	6.30	
	SD	0.53	15.32	12.40	3.00	0.38	1.06	

Table 1d: Constrained Match

A Case	B Case	$X_1^A= X_1^B$	X_2^A Age	X_2^B Age	d_{ij}	y^a	z^B	w_{ij}
A1	B2	M	42	52	10	9.156	5.524	1
A1	B5	M	42	41	1	9.156	7.243	2
A2	B3	M	35	28	7	9.149	4.223	1
A2	B5	M	35	41	6	9.149	7.243	2
A3	B4	F	63	59	4	9.287	6.147	3
A4	B2	M	55	52	3	9.512	5.524	3
A5	B1	F	28	33	5	8.494	6.932	3
A6	B4	F	53	59	6	8.891	6.147	1
A6	B6	F	53	45	8	8.891	3.230	2
A7	B1	F	22	33	11	8.425	6.932	1

Figure 4.1: Unconstrained and Constrained Matching Examples (Rogers, 1984)

4.2.3 Implicit Matching Methods

Several methods exist for implicit matching. Pritchard (2008) provides a detailed discussion of these methods, which the most common of which is *Iterative Proportional Fitting (IPF).* As illustrated in Figure 4.2, in this approach the detailed joint sample distribution (typically obtained from a census public use microdata file or a travel survey file) is iteratively "updated" using aggregate population "marginals" until it matches the aggregate population totals (Beckman et al. 1996). Numerous variations on this approach exist, which are typically used to synthesis disaggregate populations for input into activity/travel microsimulation models (Pritchard, 2008; Pendyala et al. 2010; Mueller and Axhausen, 2010).

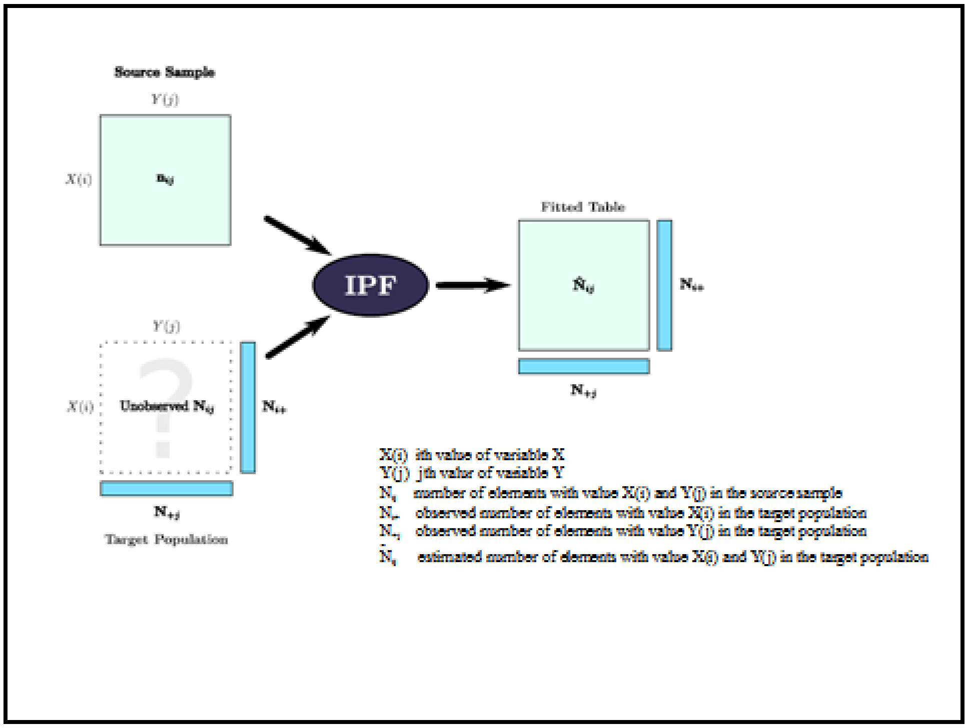


Figure 4.2: Iterative Proportional Fitting Procedure (Pritchard, 2008)

4.3 Joint RP and SP Model Estimation

The combined use of data from RP and SP surveys in the joint estimation of a common model is a special form of data fusion in that the "fusion" occurs" within the estimated parameters of the joint model, rather than a new, joint dataset *per se*. Hensher et al. (1999) and Louviere et al. (2000) present various procedures for combining RP and SP data in joint model estimations. While addressing the differences in RP and SP datasets can vary across different types of analysis and econometric models, the general approach is to recognize the differences in variances and heterogeneity among individual travellers choices expressed through RP and SP methods.

5. CONCLUSIONS

Data fusion techniques are currently used for several types of applications, including:

- “Filling in” missing variables in survey datasets (e.g., imputing missing income information in travel surveys).
- Synthesizing disaggregate agents as inputs into microsimulation models.
- Generating large, combined datasets for largely descriptive analyses.
- Combining revealed and stated preference survey data.
- (Less commonly) converting repeated cross-section survey data into a pseudo-panel dataset to support dynamic, longitudinal analysis and modelling.

Data fusion clearly can be successfully used for the applications listed above. As experience with data fusion/synthesis techniques continues to expand and as computing power continues to grow, it can be expected that such methods will see increasing application in operational transportation planning settings. This is particularly the case given steadily increasing concerns about the affordability and/or feasibility of conventional survey methods. It must be noted, however, that data fusion methods cannot, correct sampling errors, measurement errors, coverage errors and non-response errors (Venigalla 2004) inherent in the original survey data.

Stopher and Greaves (2007) raise the possibility of using fused/synthesized data derived from a variety of sources directly to build predictive models in the absence of detailed “real” data. The main challenge in this application is to build in sufficient explanatory power in the generated datasets to reflect travel behavioural processes and sensitivities to transportation and land use policies, as opposed to merely the assumptions used in generating the synthesized dataset. Nevertheless, the potential to create significantly enriched datasets, especially through the fusing of activity/travel survey data with census and/or other Statistics Canada survey datasets is very promising and deserves more investigation than has occurred to date. In particular, the potential to fuse travel survey data (with detailed spatial information concerning trip-making) with time use survey data (with detailed temporal information concerning both in- and out-of-home activities) is an interesting direction for further research.

REFERENCES

Amey, A., L. Liu, F. Pereira, C. Zegras, M. Veloso, C. Bento, and A. Biderman (2009). State of the Practice Overview of Transportation Data Fusion: Technical and Institutional Considerations. Paper# ITS-CM-09-01, Working Paper Series Working Paper Series, MIT-Portugal Program

Bayart, C., P. Bonnel, and C. Morency. (2008). Survey Mode Integration and Data Fusion: Methods and Challenges. Resource paper for Workshop on Best Practices in Data Fusion, 8th International Conference on Survey Methods in Transport Annecy, France, May 25-31.

Beckman, R.J., K.A. Baggerly, and M.D.McKay (1996). Creating Synthetic Baseline Populations. *Transportation Research A*, 30(6):415–435.

Bernard, J.-T.; Bolduc, D.; Yameogo, N.-D. (2012). A Pseudo-Panel Data Model of Household Electricity Demand , forthcoming, *Resource and Energy Economics*

Chalasani, V.S. and K.W. Axhausen (2005). Conceptual Data Model for the Integrated Travel Survey and Spatial Data, in: *Proceedings of ASC 2005 Maximising Data Value*, R. Khan, R. Banks, R. Cornelius, S. Evans and T. Manners (ed.), ASC, Chesham, pp. 123-135. (http://www.ivt.ethz.ch/vpl/publications/reports/ab302.pdf, accessed November 2011)

D'orazio, M., M. Di Zio and M. Scanu (2006). *Statistical Matching: Theory and Practice*, New York: John Wiley and Sons.

D'Ambrosio, A., Aria, M., Siciliano, R. (2007). Robust Tree-Based Data Imputation Method for Data Fusion. M.R. Berthold, J. Shawe-Taylor, and N. Lavrac (Eds.): IDA 2007, LNCS 4723, Berlin: Springer-Verlag, 174–183.

Gautier, J-M. (1999). Mégabase de consommateurs, sondages et statistique. In: Enquêtes et sondages - méthodes, modèles, applications, nouvelles approches (G. Brossier, A. M. Dussaix, ed.). Paris: Dunod, 73-80.

Gilula, Z., McCulloch, R.E., Rossi, P.E. (2006). A Direct Approach to Data Fusion. *Journal of Marketing Research*, Vol XLIII.

Hensher, D., Louviere, J., Swait, J. (1999). Combining Sources of Preference Data, *Journal of Econometrics*, 89, 197-221.

Johansen, H. (2005.) Health Studies Using Linked Administrative Hospital Data. *Statistics Canada International Symposium Series – Proceedings*, Symposium 2005: Methodological Challenges for Future Information Needs, http://www.statcan.ca/english/freepub/11-522-XIE/2005001/9482.pdf (accessed October 2011).

Louviere, J., Hensher, D., Swait, J. (2000). *Stated Choice Methods: Analysis and Application*. New York: Cambridge University Press.

Muller, K. and K.W. Axhausen (2010). Population Synthesis for Microsimulation: State of the Art, Zurich: Institute for Transport Planning and Systems, Swiss Federal Institute of Technology.

Pendyala, R., K. Konduri, K. Goulias and C. Bhat (2010). Population Synthesis Using PopGen, presentation to the SCAG Activity-Based Travel Demand Model Expert Panel Meeting, Los Angeles: June 29.

Polak, J. (2006). OPUS: Optimising the Use of Partial Information in Urban and Regional Systems, Presentation, and Research Methods Festival Programme. http://www.ccsr.ac.uk/methods/festival/programme/opus1/polak.ppt (accessed November 2011)

Pritchard, D. (2008). *Synthesizing Agents for Land use Transportation Modelling: Connecting Persons to Dwellings,* MASc thesis, Toronto: Department of Civil Engineering, University of Toronto.

Rodgers, W.L. (1984). An Evaluation of Statistical Matching*, Journal of Business and Economic Statistics*, 2(1), 91-102.

Saporta. G. (2002). Data Fusion and Data Grafting, *Computational Statistics and Data Analysis*, 38(4), 465-473.

Stopher, P., Greaves, S.P. (2007). Household Travel Surveys: Where Are We Going? *Transportation Research Part A* 4, 367–381

Venigalla, M. (2004). Household Travel Survey Data Fusion Issues, paper presented at the National Household Travel Survey Conference, USA.

PART IV – SURVEY OF DATA SOURCES FOR URBAN TRANSPORTATION APPLICATIONS

1. INTRODUCTION

While the focus of this study is on data collection methods,[46] the cost and effort required in data collection, combined with the ever-increasing data requirements of complex urban planning contexts, raises the question of whether other sources of data are available that might address at least some transportation needs, thereby eliminating (or at least reducing) the need for original data collection efforts by transportation agencies. Indeed, as is discussed at greater length in Section 3, Canadian transportation agencies have always exploited a variety of data sources to support their modelling, planning, policy analysis and other analytical needs. Given the proliferation of government and private sector databases over the past number of years, available via the internet among other sources, it is important to investigate the possibility for using new data sources (or existing data sources in new ways) in urban transportation applications.

At least eight major sources of transportation-relevant data potentially exist for use by Canadian urban transportation agencies. These are:

- Canadian Census.
- Other Statistics Canada (StatCan) datasets.
- Other federal (non-StatCan) datasets.
- Data collected by professional transportation organizations (TAC, CUTA, etc.).
- Municipal datasets.
- Provincial datasets.
- Commercial/private sector datasets.
- Open source datasets (typically web-based).

It is quite impossible within the time and budget constraints of this project to generate an exhaustive enumeration of all possible datasets that are potentially useful for urban passenger transportation planning applications. Indeed such an exhaustive inventory is virtually impossible to assemble within any reasonable timeframe, particularly for municipal, provincial and private/open source datasets, given the very large number of these, as well as the extent to which these undoubtedly vary across the country. Rather, the objective of this report is to provide a representative overview of key datasets and their possible applicability to urban transportation analysis.

---

[46] See Part II***: Literature Review*** for a detailed review of current and emerging urban transportation data collection methods.

Section 2 of this part provides an overview discussion of urban datasets. Section 3 then presents summary results concerning current usage of such datasets by Canadian agencies, based on questions about such usage that were included in this study's survey of Canadian transportation agencies.[47] Section 4 concludes the main body of the report with a brief discussion of implications for the development of a framework for urban passenger transportation data collection and management.

An annotated list of individual datasets identified during the study team's scan of the literature and the web is provided in the Appendix to this part. This appendix contains an index of datasets reviewed and a brief description of each dataset using a standard template providing the following information for each dataset:

- Name.
- Source.
- Description.
- Date(s) collected.
- Typical application(s).
- Source of documentation.

Datasets are listed alphabetically by name within each of the following group:

- Federal.
- Provincial.
- Municipal
- Professional transportation associations.
- Private sector.

Again, it should be stressed that the datasets documented in the Appendix are far from exhaustive in nature, especially at the provincial, municipal and private sector levels.[48] It is suggested that if TAC finds this list useful it may want to take on the task of maintaining and growing this list over time. Possible mechanisms for doing might include:

- Maintain open wiki site where interested parties can upload documentation of datasets (their own or others') in a standardized format.
- Require researchers working on TAC-sponsored projects to add information concerning the datasets that they use within these projects to the wiki.

---

[47] See Part V***: A Survey of Canadian Urban Transportation Data Collection Methods and Usage***.

[48] The Appendix is also somewhat eclectic in that it includes datasets pertaining to freight transportation and air travel, which may be tangential (at best" to most urban passenger transportation analyses. These have been included as a contribution to a possible future compendium of datasets covering a broader range of transportation issues than just urban passenger travel.

2. OVERVIEW DISCUSSION OF CANADIAN DATASETS

2.1 INTRODUCTION

This section provides a brief overview of major types of data potentially available for use in Canadian urban passenger transportation analysis, modelling and planning. These datasets are divided into five broad categories:

- Canadian Census and other data collected by Statistics Canada (StatCan).
- Provincial and municipal datasets.
- Data collected and/or maintained by professional transportation associations.
- Commercial (private sector) datasets.
- Open-source datasets.

It should be noted that, with the exception of open-source data, these datasets are not always provided free of charge – including many StatCan products and some other public sector datasets. This Canadian policy of charging for the use of data that have been collected using public funds can be contrasted with that of other countries such as the United States, where publicly funded data are generally provided free of charge to the public. Thus, while the use of data that have been collected by others (centrally by higher-levels of government or otherwise) is often cost-effective, the costs associated with purchasing these data can still be non-trivial for government agencies, university researchers, etc. operating on limited budgets.

2.2 CANADIAN CENSUS AND OTHER STATCAN DATASETS

Since 1981 a full census of the Canadian population has been undertaken every five years.[49] The Census provides a wealth of socio-economic information that is of potential relevance to urban passenger transportation analysis, modelling and planning. A full discussion of all Census data and their possible applications is well beyond the scope of this report. Key points to note concerning the Census include:

- Prior to 2011, a “short” form was completed by all Canadians, while a supplementary, much more detailed “long” form was completed by a subset of households. Much of the information of great relevance to urban planning was gathered in the long form. In 2011 the Federal Government cancelled the mandatory long form and replaced it with a voluntary survey. Results from the 2011 voluntary survey have not yet been released at the time of writing this report. Significant concerns, however, exist, among most census data users that the statistical representativeness of the new survey will be compromised by its voluntary nature.
- Census data typically comes in various formats:
    - Standard tabulations/reports.
    - Special/custom tabulations.

---

[49] Prior to 1981, the Census was conducted every 10 years, with “mini” Censuses being undertaken in 1966 and 1976.

  - Public Use Microdata, in which actual disaggregate Census records are made available for use, but with all identifying geography removed so that the identity of the respondent remains confidential.
- Census data are published at various levels of geography (Canada, Provinces/Territories, Census Metropolitan Areas, Census Agglomerations, Census Divisions, Census Subdivisions, Census Tracts, Forward Sortation Areas / Dissemination Areas), depending on the data table involved. If special geographies are required special tabulations can be requested from Statistics Canada on a cost-recovery basis.

While the full set of Census data of relevance to urban transportation cannot be readily enumerated, important components include:

- Place of work data (location, occupation, industry, income).
- Usual mode of transportation for the journey to work.
- Commuting distance to work.
- Household data (household composition, income, etc.).
- Demographic data (age, gender, household structure, etc.)
- Place of residence data (location, dwelling unit characteristics, price/rent, operating costs)
- Ethnic origin, immigration status and years since immigration.
- Education status, level of education achieved, etc.

For further details concerning the Canadian Census, see the Statistics Canada website: http://www12.statcan.gc.ca/census-recensement/index-eng.cfm

In addition to the Census, a major database on Canadian socio-economics is the **CANSIM database**: http://www5.statcan.gc.ca/cansim/home-accueil?lang=eng&p2=50. This online database provides access to a wealth of data, including some 235 tables in the transportation category, of which 70% relate to road transportation. Examples of relevant data include: car sales, fuel sales, car registrations, and passenger-km of travel (see the Appendix for example CANSIM datasets).

A third major source of StatCan data is the monthly Labour Force Survey (http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3701&lang=en&db=imdb&adm=8&dis=2), which provides a wealth of data on employed and unemployed persons using a national rolling panel survey. An important limitation of this dataset for transportation analysis and modelling purposes is the lack of geographic detail below the Census Metropolitan Area (CMA) level – a problem which is common among many StatCan datasets.

An extensive list of current and past StatCan surveys is found under the heading “Transportation” at: http://www23.statcan.gc.ca/imdb-bmdi/pub/indexth-eng.htm

while a guide to StatCan transportation data is provide at:

http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?catno=50F0001G&lang=eng

2.3 PROVINCIAL & MUNICIPAL DATASETS

A large amount of municipal and provincial data of various kinds, derived in a variety of ways (surveys, counts, inventories, etc.) are used in urban passenger transportation planning applications. At the municipal level these typically include:

- Land use/zoning data.
- Transit ridership counts. Most urban transit agencies gather passenger counts on a regular basis, using manual methods (ride counts and point counts) and/or automated systems (automated passenger counters and smart cards). The count data are typically used for schedule updating and general route planning purposes. The data are also used for validating transit data collected in urban travel surveys.
- Transit customer satisfaction surveys. Many transit agencies across Canada conduct these surveys to collect data on preferences and satisfaction of transit riders with various transit service characteristics such fare, speed, reliability, accessibility, frequency, etc.
- Road traffic count data; screenline counts; roadway speed-time studies.
- GIS databases:
    - Street network files.
    - Traffic signal locations; parking regulations.
    - Aerial photos.
- Parcel/building footprints.
- Assessment data (usually provincially controlled), and/or other sources of data concerning housing and commercial real estate building stock.
- Population and employment forecasts.
- Social services data.
- Custom tabulations of Census data.
- Parking supply, price and usage data.
- Employment surveys; business registers.

With respect to transit data, it is important to note that public transit operators are increasingly providing schedule and ridership information in interesting electronic formats. Much of the data compiled by the Canadian Urban Transit Association (CUTA), the American Public Transit Association (APTA), etc. can be obtained directly from transit operators. Some transit operators in Canada also provide real-time operational data (usually vehicle locations).

Provincial data can include:

- Land use data.
- GIS databases.

- Assessment data.
- Motor vehicle and driver license registration data.
- Population and employment forecasts.
- Other economic data/forecasts.
- Other?

Several typical barriers exist to the use of municipal or provincial data gathered by other agencies for transportation planning purposes. These include:

- Data are often not stored and maintained in a format that is readily accessible/usable for planning analysis purposes. That is, the data are stored in databases structures that support the immediate "business" of the agency collecting the data but that do not easily support other uses. Agencies often lack the resources (particularly staff time) to provide the data to other users in usable formats.
- Data are often stored for current conditions only. That is, the database is continuously updated, but historical data are not archived. This represents a serious loss of information over time which, if properly maintained and archived, could prove extremely valuable for planning purposes.
- Agencies are often reluctant to share their data with other agencies, for a variety of reasons.

These barriers can be overcome in two ways. First, a cooperative, inter-agency approach to data management and sharing needs to be developed, in which each agencies data needs are clearly articulated and data collection and management procedures can be implemented so as to improve the usability of the data by all parties. This may involve cost-sharing arrangements among the participating agencies to cover the additional resources required by the agency that collects the data.

Second, going beyond simple data sharing, the *open data* concept is gaining traction in many Canadian governments, particularly within many municipalities. Open data involves literally opening up one's dataset for access and use by both other agencies and the general public. This requires the datasets to be organized and documented in a manner that supports general usage by a variety of users, commitment to maintaining these datasets over time, and a change in attitude from thinking in terms of "my data" to "our data". Open data initiatives hold great promise for considerably increasing data availability, quality and usability.

2.4 PROFESSIONAL TRANSPORTATION ASSOCIATIONS

Both the Transportation Association of Canada (TAC) and the Canadian Urban Transit Association (CUTA) collect and maintain data concerning the Canadian transportation system. These datasets are described in the Appendix. The Canadian Automobile Association (CAA) maintains information on the cost of driving in Canada (http://caa.ca/docs/eng/CAA_Driving_Costs_English.pdf).

At the international level, the International Association of Public Transport (UITP) maintains statistics on transit systems around the world (http://www.uitp.org/knowledge/Statistics.cfm), while the International Road Federation (IRF) similarly collects data on road systems (http://www.irfnet.org/statistics.php),

2.5 COMMERCIAL (PRIVATE SECTOR) DATASETS

Many commercial data providers exist that, for a price, provide data that may be of use for urban passenger transportation analysis, modelling and planning purposes. Types of data include:

- Repackaged census data.
- Real estate market data.
- On-route navigation system data (real-time travel times, congestion levels, etc.).
- Employment and firm data.
- Land use data.
- GIS maps and datasets.
- Remote sensing data.
- Data from car-sharing and bike-sharing programs.
- Credit/debit card transaction data.
- Fuel price and consumption data (e.g., http://www.kentmarketingservices.com/dnn/).

Issues in the use of private sector data include:

- Cost.
- On-going availability.
- Completeness/representativeness of the data.

The explosion of web-based data services and passive data collection method means that increasing opportunities for use of commercial datasets will exist and possibly provide alternatives to traditional data collection methods. One example of this is the use of increasingly-universal availability of GPS-based "probes" provided by cell phone users, car navigation systems, etc. to replace loop detector data as a primary means of monitoring roadway performance.

2.6 OPEN-SOURCE DATASETS

In addition to commercial datasets, the increasing availability of the web and of people who are connected using a variety of devices opens new opportunities for data diffusion but also for data gathering. The last few years have seen the increasing importance of tools such as Wikipedia that gather knowledge through the free contributions of people all around the world. Similar tools and website are emerging and offer the possibility for people to feed systems with data that are then freely available to the community. One relevant example for transportation is Open-street map, a cooperative map that offers many layers that are being continuously developed and enriched by the community (http://www.openstreetmap.org/). According to our understanding, there will be increasing opportunities offered by tools and systems that benefit from the contributions of the community. The use of the information from these wiki-sites still requires exogenous validation but already provides good grounds for moving forward. With the increasing involvement of a wider community, it may become one of the main sources of information to rely on.

In addition to these individual and networked initiatives, a very promising trend in Canada and elsewhere is the implementation of *open data* initiatives, in which governments at all levels are opening up their datasets for free and unconstrained use by the public. Many motivations seem to underlie this trend, including:

- Addressing the issue raised in the introduction to this section; i.e., a recognition that data collected by the public sector should be made openly available to the public.
- "Outsourcing" data application development. The provision of data freely and conveniently encourages individuals and companies to develop a wide variety of "apps" for using these data, in ways that that the agency could not afford (or have the expertise) to do. As a result, the data become useful in a variety of previously unheard of ways. This benefits both the public and the agencies involved.
- Cost reductions. Given that public agencies inevitably are faced with numerous requests for data, for which they rarely have the staff and resources to address effectively, simply providing open access to the data reduces demand on staff time and resources while improving service to the data-requesting public.

Many Canadian cities are in various stages of open data initiatives. Toronto (www.toronto.ca/open) and Vancouver (http://vancouver.ca/your-government/open-data-catalogue.aspx) provide especially interesting and voluminous information concerning many aspects of municipal life, including transportation. An exhaustive list of open data sites is almost impossible to assemble and is well beyond the scope of this project, but a few other Canadian municipal open data initiatives include:

- Calgary: http://calgaryonlinestore.com/publicdata.asp
- Edmonton: http://data.edmonton.ca
- Guelph: http://guelph.ca/services.cfm?itemid=78870&smocid=1550
- Halton Region (Ontario): http://openhalton.ca/projects/
- Mississauga: http://www.mississauga.ca/portal/residents/mississaugadata

Transit agencies are also releasing large data sets free of charge to the general public. A particularly important trend is the sharing of transit network and service data in a standard format (Google Transit Feed Specification – GTFS) with Google Maps. GTFS files provide a high-quality, standardized basis for the development of transit network models for transportation modelling purposes.[50]

At the federal level, the GEOBASE program (www.geobase.ca) offers free access to very detailed road and rail network data for the entire country. In addition, the Canadian open data portal (http://www.data.gc.ca) provides access to a wide variety of Canadian geospatial data

Finally, among undoubtedly many others, the Canadian GIS web site (http://canadiangis.com/) provides information concerning a wide variety of GIS datasets, while ESRI Canada has a "Community Maps Program", which involves the collaborative development of common maps and GIS content for Canadian communities (http://www.esri.com/software/arcgis/community-maps-program).

---

[50] For documentation on GTFS see https://developers.google.com/transit/gtfs/reference.

3. CURRENT PRACTICE: RESULTS OF A NATIONAL SURVEY

In January 2012 a web-based survey of Canadian transportation agencies was undertaken by the project team to gather information concerning current Canadian urban passenger transportation data needs and data collection practices.[51] Included in this survey were questions concerning use of data source other than the data that the agency collects itself.

Figure 3.1 summarizes the types of data sources currently used by the agencies who responded to the survey. As can be seen from this figure Statistics Canada datasets (Census, Labour Force Survey, other datasets) are used very extensively, illustrating the important role that StatCan data play in urban passenger transportation planning in the country. Not unexpectedly, a variety of provincial and municipal datasets are also routinely used. Commercial and open-source data are also used, although not to the same extent as government data sources.

Table 3.1 elaborates on Figure 3.1 by showing a more detailed list of specific datasets and types of data, by primary source (federal, etc.) that are used by Canadian agencies, as reported in the survey.

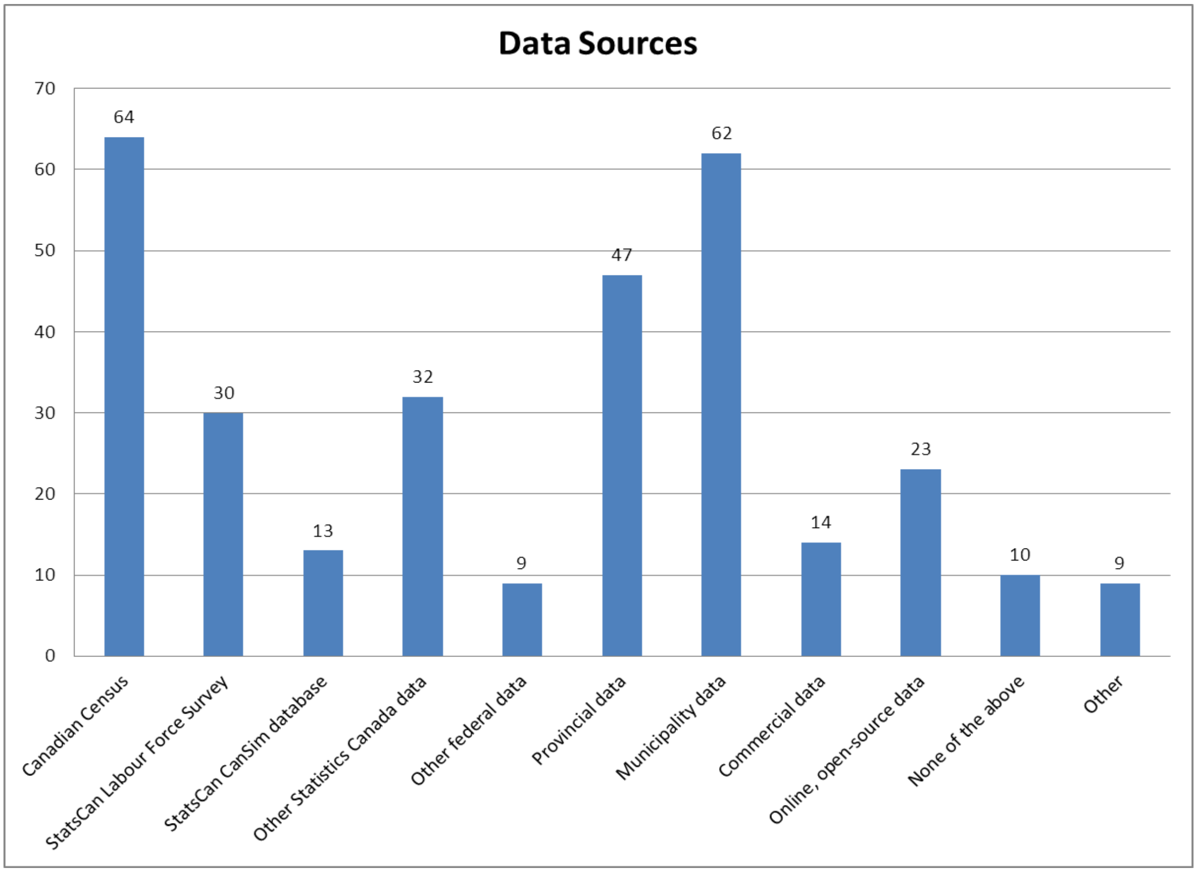


Figure 3.1: Types of Data Sources Used by Canadian Transportation Agencies

---

[51] For full documentation of this survey, see Part V*: A Survey of Canadian Urban Transportation Data Collection Methods and Usage.*

Table 3.1: Datasets Used by Canadian Transportation Agencies

Statistics Canada Data
Census Boundary Files
Census Data
Employment Data
Fuel Price Data
Journey to Work Data
Passenger Bus Data
Population Data
Public Use Microdata Files
Transportation Data
Various
Other Federal Data
Canada Mortgage and Housing Corporation Data
Canada Post Geographic Data
Economic Data
Employment Data
Household Expenditures
Non-Census Population and Demographic Data
Natural Resources Canada Data
Place of Work
Transit
Transport Canada Data
Transportation Data
Provincial Data
Agence Metropolitaine de Transport (Metropolitan Transportation Agency, Québec)
Alberta Transportation Traffic Counts
Collision Data
Commercial Vehicle Survey
Demographic and Employment Data, BC Stats
Driver's License Data

Employment Data
Health Card Data
Highway Traffic Volume Data
Institue de la Statistique du Québec (Quebec Institute of Statistics)
Land Information Ontario Data
Ontario Ministry of Infrastructure Data
Ontario Ministry of Municipal Affairs and Housing Data
Ontario Ministry of Transportation Data
Population Data
Property Assessment Files
Répertoire Adresses Québec (Québec Address Register)
School Enrolment Data
Société de l'Assurance Automobile du Québec (Automotive Insurance Society of Québec
Traffic Volume Data
Transportation Tomorrow Survey (CUTA)
Vehicle Registration Data
Yukon Bureau of Statistics
Municipal Data
Assessment Data
Building Permit Data
Business Licensing Data
Census Data
Citizen Satisfaction Survey
Community Needs Assessment
Community Profile Data
CUTA Data
Economic Development Data
Employment Data
GIS Data
Health Data
Household Surveys
Land Index

Map Data
Municipal Water Data
Ontario Municipal Benchmarking Initiative Data
Orthophotography
Parking Data
Pedestrian Counts
Planning Estimates
Property Register Data
Road Inventories
Standard Labelled Road Network Data
Tax Data
Traffic Count Data
Transit Ridership Counts
Commercial Data
Address Listings
Business Directories
Commercial Location Data
Commercial Square Footage Data
Commercial Vehicle Counts
Development Data
Economic Forecasts
Employer Directories
Employment Data
Environics Analytics Demographic Trends
Fuel Sale Data
Municipal Property Assessment Corporation Data
Road Data
Online, Open-Source Data
Google Earth Data
Google Maps Data
United States Census Data

United States National Household Travel Survey
Wikipedia
Data from Other Sources
CUTA Data
Transportation Tomorrow Survey (CUTA)
University of Toronto Data Management Group

4. IMPLICATIONS FOR A NATIONAL DATA COLLECTION FRAMEWORK

Given the expense involved in collecting and maintaining data, as well as the expertise required to undertake these tasks correctly, any opportunity to use data that are collected for other purposes and to make multiple uses of datasets should be exploited, providing, of course, that these data are, indeed, suitable for the task at hand. The rapid proliferation of web-based datasets, both commercial and open-source, is a particularly promising opportunity in this regard. Steps that transportation agencies at all levels of government (municipal, provincial and federal) can take to facilitate and encourage the use of third-party data include:

- Reiterate at every opportunity the critical role that Statistics Canada datasets (most notably but certainly not exclusively the Census) play in urban passenger transportation analysis and planning. The loss or degradation of these datasets would constitute a serious blow to urban planning and would be one that simply could not be cost-effectively compensated for by local initiatives.
- Actively promote open data initiatives within your agency and your level of government.
- Include consideration of multiple uses of datasets during the design of data collection and management programs.
- Set examples of leadership in the use and advancement of state-of-the-art data collection and in the operational use of high quality data in urban planning and decision-making applications.
- Develop a web-based inventory of metadata concerning datasets for agencies to reference. The Appendix to this report represents one starting point for the content for such an inventory.

APPENDIX ANNOTATED INVENTORY OF URBAN PASSENGER RELATED DATASETS

INDEX OF DATASETS

National Data Sources

NAME: AIR PASSENGER ORIGIN DESTINATION SURVEY, CANADA-U.S.A.

Source: CANSIM

Description: "The Air Passenger Origin and Destination - Canada/United States survey provides estimates of the number of passengers traveling on scheduled commercial flights between Canada and the United States by directional origin and destination. The data are used by Transport Canada and the Canadian Transportation Agency for evaluating competition in the industry, developing policies for the exchange of air services with foreign countries, for airport planning and market research. The information is also used by individual carriers for evaluating market trends, measuring their own growth and planning new services, as well as by Statistics Canada as input to the Provincial accounts." (StatCan website) Data are collected from all major Canadian air carriers. In order to obtain a complete picture of Canada-United States travel, it is necessary to include data from the United States Air Passenger Origin and Destination Survey. Canada and the United States exchange data quarterly via a formal statistical exchange agreement. Data are drawn from a 10% continuous systematic sample of flight coupons. The air carriers are instructed to report data only from those flight coupons where they are the first participating carrier in the flight itinerary shown on the coupon. Information include passenger counts, fare bases, all airports and carriers listed on the coupon, and the total ticket value.

Date(s): Annually, 2000 - 2010

Application(s): Possible applications are:

- Air transportation demand analysis.
- Air traffic flow analysis.
- Airport passenger demand analysis.
- Ground transportation demand for different airports.
- Investigating competitions among different airports.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2703&lang=en&db=imdb&adm=8&dis=2

NAME: AIR PASSENGER ORIGIN AND DESTINATION - DOMESTIC JOURNEYS

Source: CANSIM

Description: "The Air Passenger Origin and Destination - Domestic Journeys survey provides estimates of the number of passengers traveling on scheduled domestic commercial flights by directional origin and destination. The data are used by Transport Canada and the Canadian Transportation Agency for evaluating competition in the industry, developing policies for the exchange of air services with foreign countries, for airport planning and market research. The information is also used by individual carriers for evaluating market trends, measuring their own growth and planning new services, as well as by Statistics Canada as input to the Provincial accounts." (StatCan website)

Date(s): Quarterly, Q4 1999 – Q4 2011

Application(s): Possible applications are:

- Air transportation demand analysis.
- Air traffic flow analysis.
- Airport passenger demand analysis.
- Ground transportation demand for different airports.
- Investigating competitions among different airports.

Documentation: Data are not available due to confidentiality restrictions.

NAME: AIRCRAFT MOVEMENT STATISTICS

Source: CANSIM

Description: Data are collected from all Canadian airports with NAV CANADA air traffic control towers and flight service stations, and are voluntarily provided by many Canadian airports without NAV CANADA air traffic control towers and/or flight service stations. Data cover itinerant and local aircraft movements and include information on aircraft type and operation, time of day, civil or military, international or domestic movements, etc.

Date(s): Monthly, May 2001 - June 2012

Application(s): Possible applications are:

- Local air passenger demand analysis
- Air traffic movement analysis

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2715&lang=en&db=imdb&adm=8&dis=2

NAME: CANADIAN CENSUS OF POPULATION

Source: Statistics Canada

Description: Data were collected every five years from 1981 to 2006 from a representative sample comprised of one in five Canadian households. Since 2006 responses are voluntary, impacting the representative nature of the data. Data from the relevant period include place of work information, commuting distance, commuting mode, demographic and socio-economic information etc. Data are available online through the Statistics Canada website

Date(s): Every 5 years, 1981 - 2011

Application(s): Possible applications are:

- Urban transportation analysis, modelling, and policy
- Expanding transportation datasets with socio-economic data

Documentation:

http://www12.statcan.gc.ca/census-recensement/index-eng.cfm

NAME: CANADIAN CIVIL AVIATION - ANNUAL REPORT

Source: CANSIM

Description: Data are collected from all Canadian air carriers of passengers and/or goods. Data comprise financial characteristics such as sources of revenue, expense detail and employment characteristics as well as route information by area of operation.

Date(s): Annually, 1999 - 2010

Application(s): Possible applications are:

- Trends analysis
- Economic investigations

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2713&lang=en&db=imdb&adm=8&dis=2

NAME: CANADIAN MOTOR VEHICLE TRAFFIC COLLISION STATISTICS

Source: Transport Canada

Description: Transport Canada collects information on traffic collisions throughout Canada by various classifications including province, severity, user class, road user class, urban vs. rural, age group etc.

Date(s): Annual, 2004 - 2011

Application(s): Possible applications are:

- Accident investigation.
- Aggregate collision trend investigation.
- Crash rates modelling.
- Traffic safety analysis.
- Economic investigations of road traffic collisions.

Documentation:

http://www.tc.gc.ca/eng/roadsafety/resources-researchstats-menu-847.htm

NAME: CANADIAN PASSENGER BUS AND URBAN TRANSIT INDUSTRIES

Source: CANSIM

Description: Data are collected from bus companies in the categories urban transit, scheduled intercity, charter, school bus, sightseeing and shuttle as well as municipalities and government agencies that operate urban transit and commuter services. Information collected includes annual financial, operating and employment data.

Date(s): Annually, 1999 - 2010

Application(s): Possible applications are:

- Aggregate transit demand analysis.

- Economic investigations of public transport.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2798&lang=en&db=imdb&adm=8&dis=2

NAME: CANADIAN VEHICLE USE STUDY

Source: Transport Canada, in partnership with Environment Canada, Natural Resources Canada, the provinces, and the territories.

Description: The CVUS is carried out nationally and its purpose is to provide annual estimates of the amount of road travel by various vehicle types across Canada. It collects information from randomly selected light-duty vehicle drivers.

The CVUS is the current successor to a series of vehicle-based surveys formerly run by Statistics Canada for Transport Canada, and which were designed to estimate vehicle use by province/territory each quarter and year.

- *National Driving Survey* 1978/79
- *Fuel Consumption Survey* 1980-1990
- *Canadian Vehicle Survey* 1999-2009

This series of surveys, CVUS included, comprise the only Canada-wide source of data on daily travel, although only travel in light-duty motorized road vehicles is covered. (NB: during its early years, the CVS was deployed, with limited success, to heavy-duty road vehicles as well.) Collectively, these surveys are a very important repeated cross-section data source, although due attention to methodological changes is required when comparing estimates from more than one survey. The CVUS is the first in this series to employ tracking technologies with the objective of reducing respondent burden and increasing the quantity/quality of data collected. In CVUS, GPS-aided automatic trip loggers are temporarily installed on respondents' usual vehicles with their consent.

Date(s): From 2012 (with some partial results from progressive deployment in 2011, quarterly sampling

Application(s): Possible applications are monitoring and policy/program analyses in the transportation planning, road safety, energy and environmental domains, involving:

- Distance driven (vehicle–kilometres);
- Fuel consumption and efficiency; and
- Passenger–kilometres (occupancy).

The CVUS also provides estimates of the amount of road travel and breaks down these estimates by types of vehicles and characteristics such as age and sex of driver, time of day and season.

Documentation: http://www.tc.gc.ca/eng/policy/aca-cvus-project-summary-2584.htm

NAME: COUPON PASSENGER ORIGIN AND DESTINATION REPORT - OTHER UNIT TOLL SERVICES

Source: Statistics Canada

Description: Data are collected from all small and medium sized airline services in Canada, both domestically and internationally. Data include origin, destination, and number of passengers. Data are available only in part, as disclosed within the annual report “Air Carrier Traffic at Canadian Airports”.

Date(s): Annualy, 1999 - 2010

Application(s): Possible applications are:

- Developing federal policy on Unit Toll services.
- Monitoring the impact of airline deregulation.
- Airport planning and market research.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getInstanceList&SurvId=2704&SurvVer=0&InstaId=16301&SDDS=2704&lang=en&db=imdb&adm=8&dis=2

NAME: COURIERS AND MESSENGERS SERVICES PRICE INDEX (CMSPI)

Source: CANSIM

Description: For courier services, defined as establishments primarily engaged in providing air, surface or combined courier delivery services within and between Canadian cities and provinces/territories, as well as some international deliveries, data are collected from the official websites of the 5 largest companies in Canada. For messenger services, defined as establishments primarily engaged in providing messenger and delivery services of small parcels within a single urban area, data are collected from a subjective sample of establishments by telephone interview. The index is published at the national level only.

Date(s): Monthly, January 2003 - June 2012

Application(s): Possible applications are:

- For economic investigations and financial analyses.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5064&lang=en&db=imdb&adm=8&dis=2

NAME: ESTIMATES OF THE FULL COST OF TRANSPORTATION IN CANADA

Source: Transport Canada

Description: The Estimates of the Full Cost of Transportation in Canada report was released in August of 2008. Its goal was “to produce defensible estimates of the financial and social costs of transportation for Canada” in the year 2000. Spearheaded by Transport Canada and a federal-provincial taskforce, a low and a high estimate were calculated for the total cost of both passenger and freight transportation.

Costs were broken into infrastructure costs (capital assets, operating costs, and land value) as well as social costs (accidents, air pollution, congestion, green-house gas emissions, and noise pollution).

Date(s): One-time, 2000

Application(s): Possible applications are:

- Relative order of cost elements for policy-makers

Documentation:

http://www.tc.gc.ca/media/documents/policy/report-final.pdf

NAME: FOR-HIRE MOTOR CARRIER FREIGHT SERVICES PRICE INDEX (FHMCFSPI)

Source: CANSIM

Description: The FHMCFSPI is part of the larger Services Producer Price Index Program (SPPI) at Statistics Canada. Data are collected for the FHMCFSPI in two phases to capture in-quarter price changes. Respondents are categorized by local vs. long distance operation and freight type.

Date(s): Quarterly, Q1 2007 - Q1 2012

Application(s): Possible applications are:

- For economic investigations and financial analyses.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5136&lang=en&db=imdb&adm=8&dis=2

NAME: FREIGHT TRANSPORTATION EXPLANATORY VARIABLES

Source: Natural Resources Canada

Description: Data provide the annual sales, stock, distance travelled, on-road and lab-tested fuel consumption by truck size (light, medium, heavy).

Date(s): Annually, 1990 - 2008

Application(s): Possible applications are:

- Freight transportation demand analysis.
- Demand modelling.
- Investigations on freight traffic distribution.

Documentation:

http://oee.nrcan.gc.ca/statistics/energy-use-data-handbook/tran_00_11_e.xls

NAME: GENERAL SOCIAL SURVEY - TIME USE

Source: CANSIM

Description: Data are collected on time spent in various activities; e.g. work, commuting, personal care, free time, etc. at various locations; e.g. home, work, in transit, etc. and with various social contacts; e.g. alone, with spouse, etc. This survey has been conducted for 24 cycles. The latest was the fifth cycle dedicated to the collection of time use data. "Cycle 24 collected data from persons 15 years and over living in private households in Canada, excluding residents of the Yukon, Northwest Territories and Nunavut; and full-time residents of institutions. The purpose of this survey is to better understand how Canadians spent their time. Time use estimates can be produced based on information reported in the time use diary portion of the survey. This diary provides a detailed record of participation in a wide variety of daily activities, as well as the time devoted to them, where these activities took place, and the social relationships of the respondent. Also, for the first time, the 2010 GSS collected information on simultaneous activities, i.e. those that are performed at the same time as a primary activity. The questionnaire collected additional information on perceptions of time, time spent doing unpaid work, well-being, paid work and education, cultural and sports activities, transportation, and numerous socio economic characteristics" (StatCan website). The GSS on time use has been used to report on the commuting time to work for individual urban areas and cities across Canada. Canadian vehicle survey (CVS).This survey is currently inactive but used to provide quarterly and annual estimates of the amount of road vehicle activity by vehicle-kilometers and passenger-kilometers. Transport Canada has begun to develop and test a successor survey to the CVS, the Canadian Vehicle Usage Survey (CVUS), which will be based on electronic recording of VKT. However, results from the CVUS may not be available until 2014.

Date(s): Every 5 years, 1986, 1992, 1998, 2005, 2010

Application(s): Possible applications are:

- Travel demand analysis.
- Time use investigations.
- Activity-travel behavioural analyses.
- Commuting travel demand analyses.
- Activity-based travel demand modelling.
- Investigating various factors affecting activity-travel demand, e.g. telecommunication, socio-economic variables, age cohorts, etc.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SurvId=4503&SurvVer=3&InstaId=16848&InstaVer=5&SDDS=4503&lang=en&db=imdb&adm=8&dis=2

NAME: INTERNATIONAL TRAVEL SURVEY

Source: CANSIM

Description: The International Travel Survey comprises three sections: Frontier Counts, Mail-back Questionnaires and Air Exit Survey of Overseas Travellers. Frontier Counts data are collected at all ports of entry into Canada and collect information on the number of international travellers by category and transportation, as well as the number of automobiles, trucks and other vehicles (motorcycles, snowmobiles, bicycles) entering Canada. The Mail-back Questionnaires and Air Exit Survey (AES) sections are not available through CANSIM, but provide summary tables of statistics from international travellers including detailed characteristics of their trips such as expenditures, activities, places visited and length of stay.

Date(s): Monthly, July 2001 - June 2012

Application(s): Possible applications are:

- Time use investigations.
- International comparison of activity-travel behaviour.
- Investigating various factors affecting activity-travel demand, e.g. telecommunication, socio-economic variables, age cohorts, etc.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5005&lang=en&db=imdb&adm=8&dis=2

NAME: LABOUR FORCE SURVEY

Source: CANSIM

Description: Data concerning gross revenue (excluding subsidies) and passenger trip information are collected from 10 of the largest urban transit companies in Canada and three major intercity bus companies. These companies represent approximately 80% of urban transit activity revenues across Canada. These data are intended as a leading indicator of trends in the urban transit sector and are one input to monthly GDP estimates. Data are only available as aggregate gross revenue and passengers carried.

Date(s): Monthly, December 1994, November 2004 - May 2012

Application(s): Possible applications are:

- Transit demand analyses.
- Transit system performance analyses.
- Economic investigations of transit services.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2745&lang=en&db=imdb&adm=8&dis=2

NAME: LARGE URBAN TRANSIT SURVEY - MONTHLY

Source: CANSIM

Description: Data are collected monthly from approximately 54,000 households selected in a multi-stage sampling design based on geographic stratum .The survey is undertaken on the 15th of every month. Basic demographic information is collected on all household members with labour force information collected for all civilian household members aged 15 or older. Data collected cover such topics as unemployment, unions, absences, hours worked, educational attainment etc.

Date(s): Monthly, October 2000 – August 2012

Application(s): Possible applications are:

- Expanding transportation datasets with socio-economic data

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3701&lang=en&db=imdb&adm=8&dis=2

NAME: MAJOR AIR CARRIERS KEY FINANCIAL AND OPERATING STATISTICS MONTHLY SURVEY

Source: CANSIM

Description: Data are collected from major air carriers on main operating statistics (enplaned passengers, passenger-kilometers, available seat kilometers and goods tons kilometers) and financial statistics (total revenue and expenses, interest expenses) by type of service (unit toll and charter) and by sector (domestic, transborder and international).

Date(s): Monthly, September 2000 - June 2012

Application(s): Possible applications are:

- Air transportation demand analysis.
- Performance analyses of air transportation services.
- Economic investigations of air transportation services.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5026&lang=en&db=imdb&adm=8&dis=2

NAME: MARINE INTERNATIONAL FREIGHT ORIGIN AND DESTINATION SURVEY

Source: CANSIM

Description: Data are extracted from administrative files for all marine vessels engaged in international maritime shipping that enter or leave a Canadian port. Quarterly data on cargo tonnage loaded, unloaded, or handled by seaborne or coastwise shipping are collected biannually.

Date(s): Biannually, January to June 2000 - January to June 2011

Application(s): Possible applications are:

- International freight demand analyses.
- Water transportation investigations.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2791&lang=en&db=imdb&adm=8&dis=2

NAME: NORTH AMERICAN TRANSPORTATION STATISTICS DATABASE

Source: North American Transportation Statistics Database

Description: The North American Transportation Statistics Database combines transportation data from a number of public agencies in Canada, Mexico and the United States. The primary agencies include Statistics Canada and Transport Canada from Canada; the Secretaría de Comunicaciones y Transportes (SCT) (Ministry of Communications and Transportation), the Instituto Mexicano del Transporte (IMT) (Mexican Institute of Transportation) and the Instituto Nacional de Estadística y Geografía (INEGI) (National Institute of Statistics and Geography) from Mexico; and the Bureau of Transportation Statistics (BTS) and the U.S. Census Bureau from the United States. The database contains 37 tables in time-series and international perspective covering 12 topics including transportation and the economy, safety, energy, freight, merchandise, passengers, infrastructure and vehicles.

Date(s): Annually, 1990, 1995-2009

Application(s): Possible applications are:

- Analysis of international trade and travel in North America

Documentation:

http://nats.sct.gob.mx/sys/index.jsp?i=3

NAME: NEW MOTOR VEHICLE SALES SURVEY

Source: CANSIM

Description: Data are collected by mail from automobile manufacturers and importers based on consolidated totals from dealers' sales reports. Data consist in monthly retail sales in dollars and units by country of manufacture and vehicle type.

Date(s): Monthly, August 2000 - June 2012

Application(s): Possible applications are:

- Demand investigations of automobiles.
- Market research for automobiles.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2402&lang=en&db=imdb&adm=8&dis=2

NAME: PASSENGER TRANSPORTATION EXPLANATORY VARIABLES

Source: Natural Resources Canada

Description: Data provide the annual sales, stock, distance travelled, on-road and lab-tested fuel consumption by vehicle type (cars, light trucks, motorcycles). Additionally, annual stock and distance travelled are provided for buses.

Date(s): Annually, 1990 - 2009

Application(s): Possible applications are:

- Travel demand analyses.
- Travel demand market segmentations.
- Economic investigations.

Documentation:

http://oee.nrcan.gc.ca/statistics/energy-use-data-handbook/tran_00_7_e.xls

NAME: RAIL COMMODITY ORIGIN AND DESTINATION STATISTICS

Source: CANSIM

Description: Data are collected from Transport Canada administrative data on Canadian National Railway (CN), Canadian Pacific Railway (CP), carriers that interline with CN and CP, as well as a number of regional and short-haul carriers that do not interline with either CN or CP. Data are collected by commodity type or commodity group, origin and destination, and container of trailer. Data are only collected for cargo originating or terminating in Canada and do not include Yukon or Nunavut due to the absence of commercial rail cargo operations.

Date(s): Annually, 1998 - 2010

Application(s): Possible applications are:

- Rail transportation demand analyses.
- Investigating rail transportation demand distributions.
- Economic investigations of rail transportation.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2736&lang=en&db=imdb&adm=8&dis=2

NAME: RAILWAY CARLOADINGS SURVEY - MONTHLY

Source: CANSIM

Description: Data are collected from approximately 27 rail carriers on number of cars and tonnes by commodity, by intermodal and non-intermodal trip, and, in the case of intermodal trip, container or trailer.

Date(s): Monthly, October 2000 - May 2012

Application(s): Possible applications are:

- Rails transportation demand investigations.
- Economic investigations of rail transportation.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2732&lang=en&db=imdb&adm=8&dis=2

NAME: RAILWAY TRANSPORT SURVEY - ANNUAL

Source: CANSIM

Description: Financial, operating and employment data are collected from all common carrier railways in Canada that provide for-hire passenger and freight services. Data are available by company type for operating and income accounts, balance sheets, property accounts, track lengths, fuel consumption, operating statistics, freight and passenger transportation, inventory of equipment, employment, and employee compensation.

Date(s): Annually, 2000 - 2010

Application(s): Possible applications are:

- Rail transportation performance analyses.
- Economic and financial investigations of rail services.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2734&lang=en&db=imdb&adm=8&dis=2

NAME: ROAD MOTOR VEHICLES - FUEL

Source: CANSIM

Description: Data are collected from provincial and territorial ministries of finance on the sales of gasoline, diesel fuels and liquefied petroleum gas (LPG) for which road taxes were paid. Gasoline data are available monthly by province, other fuel types are available annually.

Date(s): Monthly, 1999 - 2011

Application(s): Possible applications are:

- Transportation demand analysis.
- Economic investigations of motorized transportation demands.
- Transportation system analyses.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2746&lang=en&db=imdb&adm=8&dis=2

NAME: ROAD MOTOR VEHICLES - REGISTRATION

Source: CANSIM

Description: Data are extracted from administrative files received directly (since 1999) from the vehicle licensing bureaus of the provinces and territories. Data are available by vehicle type (light, heavy, buses, motorcycles, trailers, and off-road vehicles (e.g. constructions or farm vehicles).

Date(s): Annually, 2000 - 2009

Application(s): Possible applications are:

- Vehicle demand modelling.
- Vehicle demand analysis.
- Investigating existing vehicle fleet characteristics
- Travel demand investigations.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2747&lang=en&db=imdb&adm=8&dis=2

NAME: TRANSPORT CANADA ANNUAL REPORTS

Source: Transport Canada

Description: A 2007 amendment to the Canada Transportation Act (Section 52) lays out the need for an annual "brief overview" of the state of transportation in Canada, with a comprehensive review every 5 years. The first comprehensive review occurred in 2011 and covered the economic circumstances of transportation in Canada, government expenditures and revenues from transportation, air, marine, rail and road transportation, gateways and trade corridors, as well as trends and future issues. Selected data from each of these categories are available as tables within the report, as well as in their entirety within a Statistical Addendum. These same statistics are available as addenda to the more recent brief overviews, but are not discussed in such depth.

Date(s): Annually, 2009-2011 Online, 1996-2008 By Request

Application(s): Possible applications are:

- Analysis to support provincial and federal policy and funding decisions regarding transportation

Documentation:

http://www.tc.gc.ca/eng/policy/anre-menu.htm

NAME: TRAVEL SURVEY OF RESIDENTS OF CANADA

Source: CANSIM

Description: "Since the beginning of 2005, the Travel Survey of Residents of Canada (TSRC) has been conducted to measure domestic travel in Canada. It replaces the Canadian Travel Survey (CTS). Featuring several definitional changes and a new questionnaire, this survey provides estimates of domestic travel that are more in line with the international guidelines recommended by the World Tourism Organization (WTO) and the United Nations Statistical Commission. The Travel Survey of Residents of Canada is sponsored by Statistics Canada, the Canadian Tourism Commission, the provincial governments and two federal organizations. It measures the size of domestic travel in Canada from the demand side. The objectives of the survey are to provide information about the volume of trips and expenditures for Canadian residents by trip origin, destination, duration, type of accommodation used, trip reason, mode of travel, etc.; to provide information on travel incidence and to provide the socio-demographic profile of travellers and non-travellers. Estimates allow quarterly analysis at the national, provincial and tourism region level (with varying degrees of precision) on:

- Total volume of same-day and overnight trips taken by the residents of Canada with destinations in Canada.

- Same-day and overnight visits in Canada.
- Main purpose of the trip/key activities on trip.
- Spending on same-day and overnight trips taken in Canada by Canadian residents in total and by category of expenditure.
- Modes of transportation (main/other) used on the trip.
- Person-visits, household-visits, spending in total and by expense category for each location visited in Canada.
- Person- and party-nights spent in each location visited in Canada, in total and by type of accommodation used.
- Use of travel packages and associated spending and use of motor coach/other guided tours.
- Source of payment (household, government, private employer).
- Demographics of adults that took or did not take trips.
- Travel party composition." (StatCan website)

Date(s): Quarterly, Q2 2000 - Q1 2012

Application(s): Possible applications are:

- Investigating tourism demands
- Investigating long-distance travel demands.
- Inter-city passenger travel demand modelling.
- Inter-city passenger mode choice modelling.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SurvId=4503&SurvVer=3&InstaId=16848&InstaVer=5&SDDS=4503&lang=en&db=imdb&adm=8&dis=2

NAME: TRUCKING COMMODITY ORIGIN AND DESTINATION SURVEY (TCOD)

Source: CANSIM

Description: Since 2004, data are collected in a four-stage sample design from all companies in the Statistics Canada Business Register with a trucking establishment meeting a minimum annual revenue requirement. Data on shipment counts, distance, tonne-kilometers, revenue, etc. are available by shipment type, i.e. domestic, transborder, local, and long distance. Between 1999 and 2004, data are only available as quarterly aggregate estimated tonnage, tonne-kilometres and shipment counts.

Date(s): Annually, 1999 - 2010

Application(s): Possible applications are:

- Freight transportation demand analyses.
- Freight traffic flow modelling.
- Regional economic investigations.

Documentation:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=2741&lang=en&db=imdb&adm=8&dis=2

NAME: URBAN TRANSPORTATION INDICATORS - FOURTH SURVEY

Source: Transport Authority of Canada (TAC)

Description: The Urban Transportation Indicators Fourth Survey report is based on data gathered during the 2006 Canadian Census for 33 Census Metropolitan Areas (CMAs). The report also contains supplementary data from 31 CMAs that responded to a detailed TAC survey, which looked at the status of transportation and land use initiatives, transportation financing, and land use and transportation. Also included are Kent Marketing data on fuel sales and summary statistics from the Canadian Urban Transit Association. Survey questions and responses are available as a Microsoft Access database from the TAC website.

Date(s): 2006

Application(s): Possible applications are:

- Urban transportation performance analysis.
- Investigating competitions of different urban transportation modes.
- Trend analyses.

Documentation:

http://www.tac-atc.ca/english/resourcecentre/databases/TDB_UTI-2006_2010-03-30_FINAL.mdb

Provincial Data Sources

NAME: ACTIVE VEHICLE FLEET, QUEBEC (VEHICULES EN CIRCULATION SELON LE TYPE DE VEHICULES ET LEUR UTILISATION, QUEBEC)

Source: Québec database of official statistics (Banque de données des statistiques officielles sur le Québec)

Description: Data are provided on all active motor vehicles in the province of Québec by vehicle type and purpose.

Date(s): One-time, 2006

Application(s): Possible applications are:

- Transportation demand analysis.
- Vehicle ownership investigations.

Documentation:

http://www.bdso.gouv.qc.ca/pls/ken/Ken211_Selct_Theme.p_theme_tratm?p_iden_tran=REPERF904MT26175436625325x7ig7&p_modi_url=STAT_ACCU&p_id_ken03t01=57&p_id_ken03t02=244&p_id_ken03t03=718

NAME: BICYCLE-PATH LENGTHS, QUEBEC AND TOURISM REGIONS (LONGUEUR DES VOIES CYCLABLES, QUEBEC ET REGIONS TOURISTIQUES)

Source: Québec database of official statistics (Banque de données des statistiques officielles sur le Québec)

Description: Data are provided on bicycle path length by region in the province of Québec and path type. Types include paved bicycle paths, bicycle lanes, bicycle paths, and paved shoulders.

Date(s): One-time, 2006

Application(s): Possible applications are:

- Inventory for active transportation.
- Tourism traffic demand analyses.
- Investigating active transportation supply.
- Active transportation demand modelling.

Documentation:

http://www.bdso.gouv.qc.ca/pls/ken/Ken211_Selct_Theme.p_theme_tratm?p_iden_tran=REPERF904MT26175436625325x7ig7&p_modi_url=STAT_ACCU&p_id_ken03t01=57&p_id_ken03t02=244&p_id_ken03t03=718

NAME: LENGTH OF ROAD NETWORK MANAGED BY TRANSPORT QUEBEC (LONGUEURS ITINERAIRES, REELLES ET PONDEREES DU RESEAU ROUTIER SOUS LA GESTION DU MINISTRE DES TRANSPORTS)

Source: Québec database of official statistics (Banque de données des statistiques officielles sur le Québec)

Description: Highway, national- and regional-level road lengths are available by administrative region, county, and electoral district for the province of Québec. Length is calculated and provided by three different methods: itinerary (point A to point B), actual (sum of lane lengths), and weighted (length in two-lane equivalents).

Date(s): One-time, 2010

Application(s): Possible applications are:

- Transportation system capacity investigations.
- Inventory for transportation system modelling.
- Inventory for travel demand modelling.
- Inventory for transportation network analysis.

Documentation:

http://www.bdso.gouv.qc.ca/pls/ken/Ken211_Selct_Theme.p_theme_tratm?p_iden_tran=REPERF904MT26175436625325x7ig7&p_modi_url=STAT_ACCU&p_id_ken03t01=57&p_id_ken03t02=244&p_id_ken03t03=718

NAME: MODE OF COMMUTE FOR FULL-TIME EMPLOYEES OF 15 YEARS OF AGE OR OLDER (POPULATION ACTIVE EN EMPLOI DE 15 ANS ET PLUS SELON LE MODE DE TRANSPORT UTILISE POUR SE DEPLACER DU DOMICILE AU LIEU DE TRAVAIL)

Source: Québec database of official statistics (Banque de données des statistiques officielles sur le Québec)

Description: Mode of commute is provided for all full-time employees 15 years of age or older, by administrative region. Modes are car/truck/van, public transit, pedestrian, bicycle, motorcycle, taxi, other.

Date(s): One-time, 2001

Application(s): Possible applications are:

- Commuting travel demand analysis.
- Commuting mode choice modelling.
- Transportation and land use policy analyses.

Documentation:

http://www.bdso.gouv.qc.ca/pls/ken/Ken211_Selct_Theme.p_theme_tratm?p_iden_tran=REPERF904MT26175436625325x7ig7&p_modi_url=STAT_ACCU&p_id_ken03t01=57&p_id_ken03t02=244&p_id_ken03t03=718

NAME: ONTARIO ROAD NETWORK

Source: Ontario Ministry of Natural Resources - Land Information Ontario

Description: The Ontario Road Network is a geographic database available in Linear Referencing System format, Segmented format, or as a Web Map Service. Data cover 234 thousand kilometres of roads in Ontario including road characteristics such as street names, highway numbers, addresses, road classifications, surface type, direction of traffic flow, and speed limits. Data are available to municipalities, provincial ministries and federal departments, First Nations communities, conservation authorities, public health units and other public sector organizations, non-profit organizations, academic institutions, and public utilities.

Date(s): One time, current

Application(s): Possible applications are:

- Inventory for transportation network analysis.
- Transpiration network modelling.
- Input to travel demand modelling.

Documentation:

http://www.mnr.gov.on.ca/en/Business/LIO/2ColumnSubPage/STEL02_167960.html

NAME: QUEBEC MINISTRY OF TRANSPORTATION PORTAL

Source: Quebec Ministry of Transportation

Description: MTQ maintains a portal to access Quebec's urban OD survey summary results.

Date(s): Various; on-going.

Application(s): Possible applications are:

- Travel demand modelling.
- Travel demand trend analyses.
- Transit service design.
- Transit service performance analysis.
- Traffic emission investigations and modelling.

Documentation:

http://www.mtq.gouv.qc.ca/portal/page/portal/ministere/ministere/recherche_innovation/modelisation_systemes_transport/enquetes_origine_destination

Municipal and/or Cross-Regional Data Sources

NAME: MONTREAL ORIGIN-DESTINATION TRAVEL SURVEYS

Source: Technical committee on origin-destination travel surveys

Description: The Greater Montreal Area has been conducting large-scale origin-destination travel surveys since 1970, approximately every five years. The latest survey was held in 2008 and covered an area of some 8000 square kilometers. Typically, some 4-5% of the residing population provides data on all trips conducted during a specific day of the fall period. In 2008, around 70,000 households, gathering 165,000 people were surveyed. Interviews are conducted through the phone and a CATI system provides assistance to the interviewer to facilitate gathering and interactively validate provided answers. There are a lot of similarities between the TTS and the Montreal (and other regional surveys conducted in the Quebec province) as Montreal initiated the process and was at the basis of the development of the other surveys. Survey information includes: home location, number of persons living at that location, car ownership, gender and age, number of residents licensed to drive, whether they are working or attending school, location of workplace, destination, purpose and time of each trip taken the previous day, means of transportation used (car, transit, walking, bicycle, etc.), for transit trips, the route taken for travel. Aggregated data are available online through the origin-destination web site http://www.cimtu.amt.qc.ca/EnqOD/Index.asp. Hence, microdata are made available to researchers interested to conduct research on the topic. Fees have to be paid by consultants or other individuals that want to use the data for commercial purposes.

Date(s): Every 5 years, 1970-2008

Application(s): Possible applications are:

- Travel demand modelling.
- Travel demand trend analyses.
- Transit service design.
- Transit service performance analysis.
- Traffic emission investigations and modelling.

Documentation:

http://www.cimtu.amt.qc.ca/EnqOD/Index.asp

NAME: TRANSPORTATION TOMORROW SURVEY

Source: Transportation Information Steering Committee

Description: The Transportation Tomorrow Survey (TTS) is a comprehensive travel survey, originally covering only the Greater Toronto and Hamilton Area (1986), but since expanded to include the Regional Municipalities of Niagara and Waterloo, the counties of Peterborough, Simcoe, Victoria and Wellington, the Cities of Barrie, Guelph, and Peterborough and the Town of Orangeville (1996), the City of Orillia and all of the County of Simcoe (2001), the City of Brantford and the County of Dufferin (2006), and the County of Brant (2011). Greater than 150,000 interviews were conducted in recent years. The 1996, 2001 and 2006 surveys are three of the largest travel surveys ever undertaken anywhere. The

1986, 1991 and 1996 surveys each involved a major element of technology development. The use of automated geocoding was a key development in the 1986 survey. On-line Direct Data Entry (DDE) was introduced in the 1991 survey and networked computers in the 1996 survey. A telephone interview with on-line DDE and automated geocoding of all geographic information collected has proven to be the most cost effective and reliable means of collecting large quantities of travel data. Survey information includes: type of residence (house or apartment) and address, number of persons living at that location, gender and age, number of residents licensed to drive, whether they are working or attending school, location of workplace, destination, purpose and time of each trip taken the previous day, means of transportation used (car, transit, walking, bicycle, etc.), for transit trips, the route taken for travel. Data are available online through the Data Management Group of the Department of Civil Engineering at the University of Toronto. Anyone can request access.

Date(s): Every 5 years, 1986 - 2011

Application(s): Possible applications are:

- Travel demand modelling.
- Travel demand trend analyses.
- Transit service design.
- Transit service performance analysis.
- Traffic emission investigations and modelling.

Documentation:

https://www.jpint.utoronto.ca/drs/new_index.html

Professional Transportation Association Data Sources

NAME: THE CENTRE FOR DATA AND ANALYSIS IN TRANSPORTATION

Source: Laval University

Description: CDAT provides a well-structured web-based tool to search for databases on various transport-related databases.

Date(s): Various

Applications(s): Various

Documentation:

http://www.cdat.ecn.ulaval.ca/en/index.php?pid=892

NAME: URBAN TRANSPORTATION INDICATORS DATABASE

Source: Transportation Association of Canada (TAC)

Description: While the fourth iteration of the Urban Transportation Indicators Database is publicly available (see National Data Sources), the three previous databases are only available to TAC members.

Date(s): 1995, 1996, 2003

Application(s): Possible applications are:

- Urban transportation performance analysis.
- Investigating competitions of different urban transportation modes.
- Trend analyses.

Documentation:

http://tac-atc.ca/english/index.cfm

NAME: URBAN TRANSIT STATISTICS (UTS)

Source: Canadian Urban Transit Association

Description: CUTA conducts an annual survey of all public transit systems in Canada. The survey gathers a wide variety of data, including:

- General information on the service area population, fares, vehicles (active, stored, peak, and base), vehicle accessibility, employees (operators, other operations staff, mechanics, other vehicle maintenance, plant maintenance, general and administration staff, wage rates).
- Operating data including revenue vehicle kilometres, total vehicle kilometres, revenue vehicle hours, auxiliary revenue vehicle hours, total vehicle hours and staff paid hours).
- Passenger data including passenger trips, regular service passenger trips, regular service passenger kilometres, and auxiliary service passenger trips.

- Financial data including operating expenses, fuel/energy expenses for vehicles, vehicle maintenance expenses, plant maintenance expenses, general and administration expenses, debt service payment, operating revenues, capital expenditures and capital funding.

The survey data are used to calculate performance indicators which are published for each transit system in the annual *Canadian Urban Transit Fact Book*. The indicators include revenue to cost ratio, cost effectiveness and efficiency, service and vehicle utilization, and average speed.

Date(s): Annually,

Application(s):

Documentation:

Private Data Sources

NAME: ANNUAL RANKING OF TOP 100 CARRIERS (FOR-HIRE TRUCKING COMPANIES IN CANADA)

Source: Today's Trucking

Description: The largest 100 for-hire trucking companies in Canada are ranked according to a compound statistic of truck count, tractor count, trailer count, owner/operator count, and employee count. All counts are also available individually.

Date(s): Annually, 2006 - 2012

Application(s): Possible applications are:

- Freight transportation system performance analyses.

Documentation:

http://www.todaystrucking.com/top100.cfm

NAME: RETAIL TRUCK SALES, CANADA

Source: Today's Trucking

Description: Sales data are provided by class of truck, manufacturer, and province of sale.

Date(s): Monthly, January 2011 - June 2012

Application(s): Possible applications are:

- Freight transportation demand analyses.
- Freight transpiration modal share investigations.

Documentation:

http://www.todaystrucking.com/trucksales.cfm

PART V – DESIGN, CONDUCT & RESULTS OF A SURVEY OF CANADIAN TRANSPORTATION DATA COLLECTION PRACTICE

1. INTRODUCTION

This part of the report documents the design and conduct of a survey of Canadian transportation agencies, as well as the findings of this survey. The objective of the survey is to determine the current state of Canadian practice with respect to the collection and use of data pertaining to urban person travel for transportation planning purposes. The results of this survey, combined with the findings of an extensive review of urban data collection methods and issues[52] and a survey of Canadian data sources[53], provide the basis for the development of a new Canadian framework for urban transportation data collection, which is the primary objective of the overall study.

The survey is a web-based survey that was designed by the Study Team and reviewed by the TAC Project Steering Committee (PSC). It was conducted in March-April, 2012. The design and conduct of the survey are described in Sections 2-4 as follows:

- Section 2: Survey method.
- Section 3: Questionnaire design.
- Section 4: Conduct of the Survey

Section 5 then presents a detailed analysis of the information gathered within the survey.

2. SURVEY METHOD

A web-based approach was selected for this survey because of its flexibility in terms of better presentation of a long series of questions that are logically connected. Web surveys provide a better control surface and pull-down lists that can effectively reduce the apparent length of the survey to individual respondents. Compared to telephone interviews, web-based surveys provide better flexibility to the respondents, as the respondent can take time to think about specific questions or even talk to colleagues who may have better knowledge about specific issues. A web-based survey is also the quickest way of collecting data as it does not restrict the number of participants who can participate at the same time. Also, web-based surveys are self-administered and are the most cost-effective method of data collection.

For implementation of this survey, Survey Monkey© was used. Survey Monkey is a subscription-based commercial web-based survey implementation tool, which provides a wide variety of survey design

---

[52] See Part II*: Literature Reviews*.

[53] See Part IV*: Survey of Data Sources for Urban Transportation Applications*.

templates including skip logic and other options for professional survey design. It also provides secure data collection, storage and transfer systems after implementing the survey. Members of the project team also have experience in successfully using this web-survey tool for other projects.

Participants were contacted through email invitations to participate in the web-survey. Emails were distributed by TAC and the Canadian Urban Transit Association (CUTA) among their members. Two versions of survey (one in English and the other in French) were distributed to every individual so that each person could choose her/his preferred language with which to complete the survey. The English invitation letter connects to the web link of the English version of the survey, while the French invitation letter similarly connects to the web link of the French version of the survey.

3. QUESTIONNAIRE DESIGN

3.1 INTRODUCTION

The survey was designed to collect as much detailed information about the data needed by Canadian transportation agencies and their existing practices of data collection s possible, while keeping the response burden at a reasonable level. As shown in Figure 3.1, to have a consistent and logical flow of information collection the survey was divided into seven main sections.

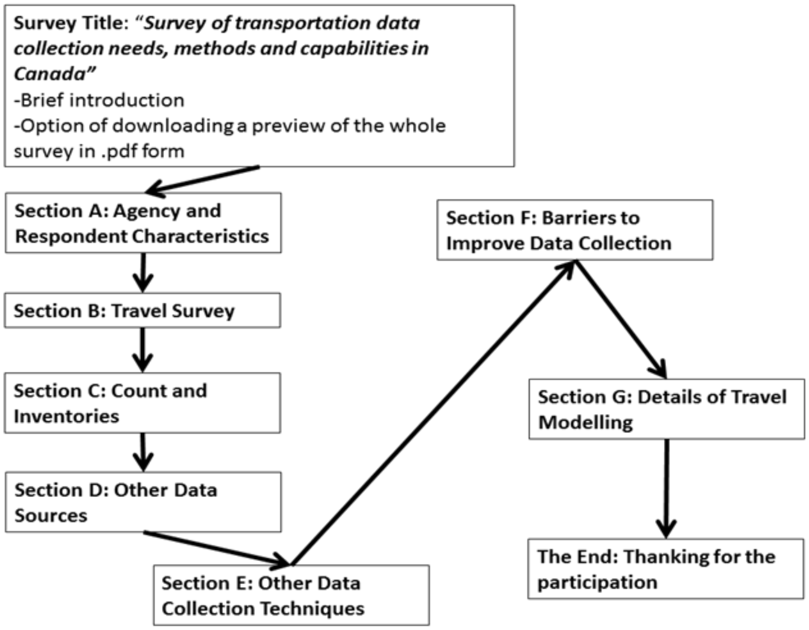


Figure 3.1: Sequence of Survey Segments

The survey starts with a brief introduction of the main objectives and purposes of the survey with a tentative estimate of the time required to complete the survey. An option of downloading a pdf copy of the complete questionnaire is also given at the beginning so that the respondent can see all the questions that will be asked during the survey. This helps the respondents in developing an expectation of the sequence of the questions as well as identifying any information that they need to get from others for proper input. This approach is unique in web-based survey practice. It was found that this reduced possible respondent confusion, and, as a result, the number of queries regarding survey completion was extremely low (only one person contacted for clarification of a particular question).

The main content of the survey is contained in Sections A-G shown in Figure 3.1. An overview of the survey structure and questions is provided in the remainder of this section. Finally, at the end of the survey, the respondent is given an option to provide email address and/or phone number if she/he agrees to be contacted later regarding clarification of any provided answer. Also, the respondent is given an option to write down the email address if the respondent would like to receive a synthesis of the responses from all the participants when it is available.

3.2 SECTION A: AGENCY & RESPONDENT INFORMATION

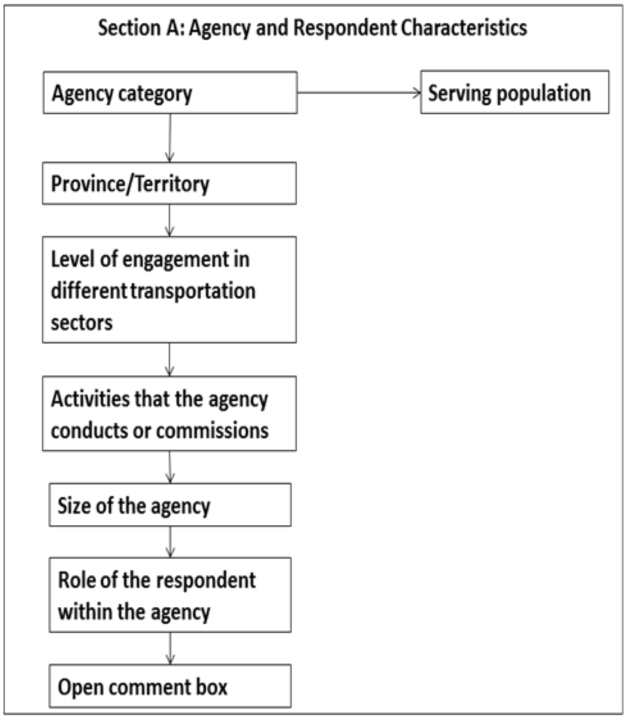


Figure 3.2: Section A - Information Collected

After the brief introduction, the first section (Section A) collects basic information about the agency and the respondent's characteristics. This section contains seven question sets and an open comment box at the end. Figure 3.2 presents a flow diagram of the sequence of the questions asked in this section. All questions except the open comment are mandatory to answer to ensure that the basic information required for proper identification of agency data needs and challenges is collected. The first question identifies the level of government of the agency in terms of local, regional, provincial, federal and others. Only local and regional agencies are then asked about the approximate population size covered by their jurisdiction. All agencies, however, are asked to identify their province or territory and organization category. The respondent is asked to choose the category that best describes his/her agency among: planning department, transportation department, public works department, transit agency, research centre and others.

Each agency's level of engagement for each of the following activities:

- Operational and planning responsibility.
- Operational responsibility only.
- Planning responsibility only.
- Not any of these responsibilities (not applicable option).

for each of the sectors following modes:

- Roads.
- Local transit.
- Commuter rail/bus.
- Intercity rail.
- Intercity bus.
- Parking.
- Bicycle.
- Pedestrian.
- Others (ferries, etc.).

Respondents are also given an open text box to mention any other responsibilities that the above mentioned discrete categories do not cover.

The next question asks about the agency's activities. A list of possible activities is presented from which the respondent may select one or more, and/or enter other activities not listed in an open text box. The activities included in the list are:

- Travel demand modelling.
- Transportation system planning.
- Transportation system regulation.
- Transportation policy analysis.
- Road design & construction.
- Road maintenance.
- Road traffic operations and control.
- Transit design & construction.
- Transit service planning.
- Transit maintenance.
- Transit operations.
- Bicycle planning, facilities and/or operations.

- Pedestrian planning, facilities and/or operations.
- Parking regulation, facilities and/or operations.
- Land use planning.
- Travel demand management.
- Intelligent transportation systems.
- Carpooling/ridesharing/carsharing.
- Research.

The next question asks about the number of full-time and part-time staff involved in data collection and/or data management activities within the agency. This question helps identify the size of the responding agency and its need for travel data. The last question of Section A asks about the position of the respondent in the organization. This question is important to evaluate the accuracy and reliability of information collected in the survey. At the end of this section the responded is given a chance to write any comment on this section in an open text box.

3.3 SECTION B: TRAVEL SURVEYS

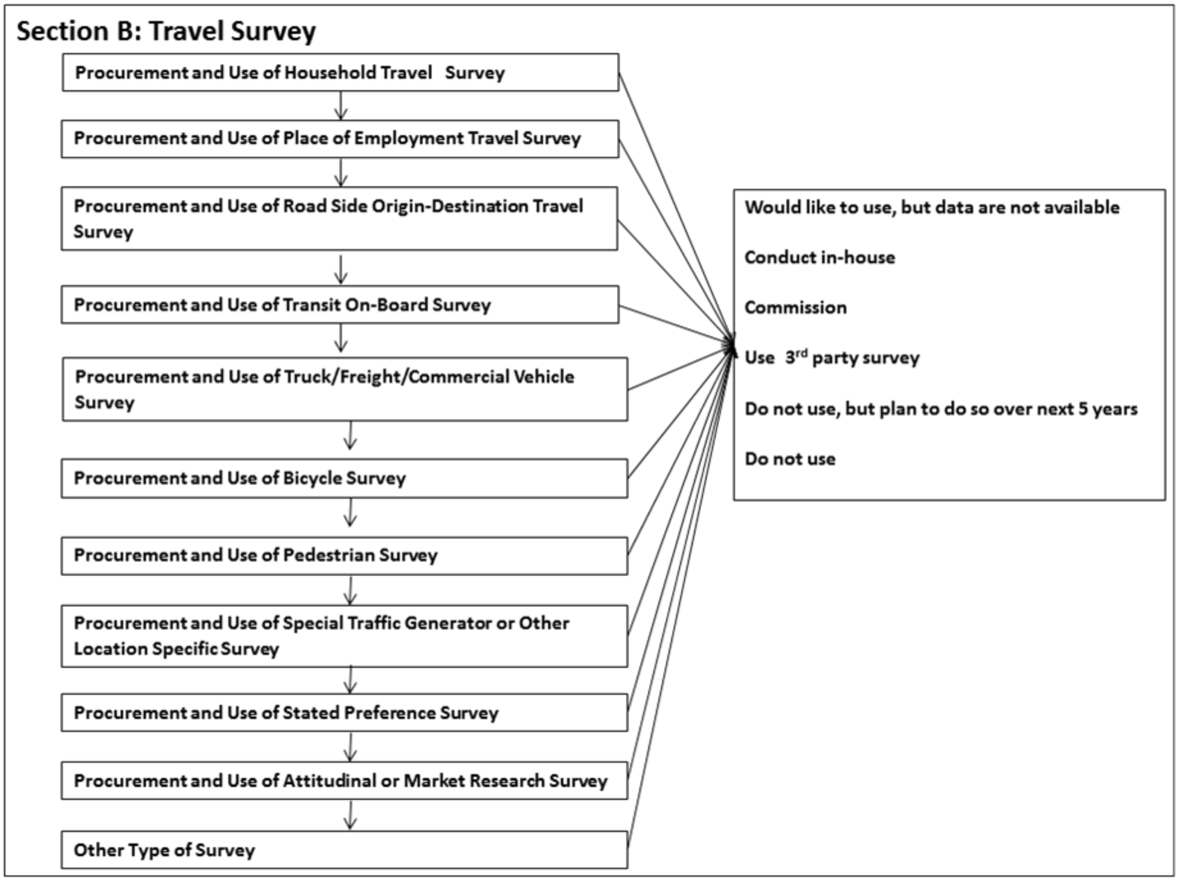


Figure 3.3: Travel Survey Section Flow Diagram

The second section (Section B) focuses on travel surveys. Figure 3.3 presents a flow diagram of the sequence of the question types used in this section. This section collects information about procurement and use of 10 different types of travel surveys. These are:

- Household travel surveys.
- Place of employment travel surveys.
- Roadside origin-destination travel surveys.
- Transit on-board surveys.
- Truck/freight/commercial vehicle surveys.
- Bicycle surveys.
- Pedestrian surveys.
- Stated preference surveys.
- Attitudinal or market research surveys.
- Other types of surveys.

For each of these 10 survey types, the respondent indicated the nature of the agency's involvement in the use and/or undertaking of surveys of this type, choosing from one of the following 6 categories:

- Would like to use, but data are not available.
- Conducts the survey in-house.
- Commissions the survey.
- Use third-party data.
- Do not use, but plan to use in next 5 years.
- Do not use.

If and only if the responding agency conducts in-house or commissions the given type of survey, are further questions asked to solicit detailed information about this survey category. For household travel surveys the following detailed information is collected:

- How many types of household travel surveys were conducted in last 10 years
- Frequency of conducting household travel surveys.
- Year the last survey was conducted.

Then for the last conducted survey, the following information is collected:

- Sample size of the survey,
- Whether all members of the household were interviewed or a single person reported on behalf of the household.
- Number of days of travel information collected per person.

- Whether the survey was on weekend or weekday travel.
- Sampling unit of the survey.
- Sampling frame.
- Survey respondent contact method.
- Mode of survey (interview type).
- Type of survey.
- Issues/concerns regarding current home-interview survey method.
- Any open comments on household travel survey.

For all other types of surveys, the detailed information collected includes:

- Frequency of the survey in last 10 years.
- Year the last survey was conducted.

Then for the last conducted survey, the following information is collected:

- Sample size.
- Whether the survey was for weekend or weekdays (not asked for place of employment surveys).
- Any open comments on place of employment travel survey.

3.4 SECTION C: COUNTS & INVENTORIES

The third section (Section C) focuses on observational data collection and usage, such as counts and inventories. Information for ten categories of counts is requested:

- Automated road traffic count.
- Manual road traffic count.
- Automated transit ridership count.
- Manual transit ridership count.
- Screenline or cordon counts.
- Truck or freight or commercial vehicle counts.
- Parking counts.
- Bicycle counts.
- Pedestrian counts.
- Other counts.

Similarly, information for six categories of inventories is requested:

- Roadways inventories (number of lanes, lane widths, etc.).
- Transit service inventories (routes, frequencies, etc.).

- Parking inventories (number of spaces, parking charges, etc.).
- Bicycle inventories (bikeways, etc.).
- Sidewalk inventories (sidewalk widths, etc.).
- Other inventories (an open text box is provided to write details about other inventories).

As for surveys, the nature of the involvement in the use and/or undertaking of each type of count and inventory are requested. Detailed follow-up questions for those who undertake counts or surveys, however, are not asked. At the end of this section, an open text box is provided to the respondent for writing any comments regarding count or inventory data or surveys that they wish to make.

3.5 SECTION D: OTHER DATA SOURCES

The fourth section (Section D) focuses on other data sources and data collection techniques that are not covered in previous sections. The Responding agency is asked to identify which of the following data sources it uses:

- Canadian census.
- Statistics Canada Labour Force Survey.
- Other federal data.
- Provincial data.
- Municipal data.
- Commercial data.
- Online, open source data.
- Other data (an open text input box is provided to write details about other data).

If the responding agency identifies any of the above-mentioned data sources, then they are asked to write the names of the specific datasets used in each of the identified categories. Results obtained from this section of the survey are presented in *Report III: Survey of Data Sources for Urban Transportation Applications*.

3.6 SECTION E: DATA COLLECTION TECHNOLOGIES

The fifth section (Section E) focuses on data collection technologies. The responding agency is asked whether they ‘use now’ or ‘plan to use in next 5 years’ or ‘do not use’ any of the following data collection technologies:

- Global positioning systems (GPS).
- Other distributed or remote sensing technologies (such as satellite imagery).
- Transit pass technologies (such as smart card).
- Debit or credit cards.
- Social media (Facebook, twitters, etc.).

- Other internet.
- Mobile devices (such as smart phone, PDAs).
- Other technologies (an open text input box is provide to write details about other technologies).

If the responding agency identifies any of the above-mentioned technologies for data collection, then are asked to write application of each of the identified categories. At the end of this section, an open text box is provided to the respondent to provide any comments that they wish to make concerning data collection technologies.

3.7 SECTION F: BARRIERS TO IMPROVED DATA COLLECTION

The sixth section (Section F) focuses on barriers to improved data collection. This section provided the respondent with a list of possible barriers and the respondent was requested to identify which, if any of these barriers the agency is currently facing that reduce their ability to collect the data that they need. The list of barriers presented to the respondents consists of:

- Competing priorities precluding allocating time to data needs
- Inadequate budget
- Lack of political support
- Lack of staff
- Union or contract rules
- Limited technical abilities of the staffs
- Lack of management support
- Lack of perceived importance of improved data
- Other reasons (an open test input box is provide to write details about other reasons).

At the end of this section, an open text box is provided to the respondent for writing any comments regarding barriers to improved data collection.

3.8 SECTION G: TRAVEL DEMAND MODELLING

A major use of urban passenger transportation data is the development and application of travel demand forecasting models. Given this, it was felt that to better understand agency data needs, it would be useful to identify whether the agency develops and/or uses travel demand models. The seventh section (Section G) focuses on this issue. The first question in this section determines whether the agency uses a travel demand model for planning and policy analysis or not. If the answer is yes, then further information is collected to identify data needs for travel demand modelling for the agency.

4. CONDUCT OF THE SURVEY

4.1 PRETESTING

The final questionnaire was developed through a series of trial runs among the project team and a group of graduate students, who had recently completed a graduate course on Travel Survey Methods. After this, a pretest of the survey was conducted among the Project Steering Committee (PSC) members, which resulted in further amendments to the survey.

The survey was originally developed in English. The final version of the survey was then translated in French by TAC, with assistance from the project team and the PSC.

4.2 SAMPLING FRAME

A possible challenge in web-based surveys is to construct an appropriate sampling frame which includes e-mail addresses. In this survey, two contact lists were used: The primary list was provided by TAC consisting of approximately 400 TAC members. This list was updated for current contact information and added to (where gaps seemed to exist) by the study team. The second list was provided by the Canadian Urban Transit Association (CUTA), consisting of about 200 of their members. This additional list was used since the TAC list did not include transit agencies, and it was felt that it was important to include transit agencies in the survey. Together, the two lists provided a representative sampling frame of urban transportation agencies across the entire country, city sizes and agency type.

4.3 SURVEY EXECUTION

Invitation e-mails were sent to the CUTA mailing list on March 19, 2012, and to the TAC mailing list on March 26, 2012, with an indicated completion deadline of April 13. Two formal reminder e-mails were sent to the full TAC list, one on April 2nd and the other on April 10. One reminder e-mail was sent on March 28 to the full CUTA mailing list.

In the period April 18 - 23, the project team, in collaboration with PSC members, identified and directly contacted a limited number of agencies to fill key gaps in the responses to the survey received to that point. The deadline for survey completion was effectively extended by individual invitation, noting that: "TAC requested a response by 13 April, but we would still very much like to hear from you". Responses received through May 4, 2012 were included in the database.

4.4 RESPONSE RATES & RESPONDENT CHARACTERISTICS

A total of 49 completed responses were received in response to the initial recruitment request by April 2, before the first reminder was sent. By April 18 (five days past the original completion deadline and after two reminder e-mails) 56 “mostly completed” and 7 “partially completed” responses had been received. After direct solicitation efforts on the part of the PSC and the project team, a final total of 124 responses were collected by April 30, 2012.

From these respondents, a total usable sample of 94 complete responses was obtained. Complete responses were those that answered all of the major branching questions (e.g.,, the first branching question asked respondents to “Please check the box that best describes your agency's procurement and use of Household Travel Surveys". If the respondent indicated that they either commission or conduct in house, then they would be asked questions about the survey).

Completed responses represent a spectrum of agencies across Canada that reflect local regional and provincial levels of government, represent population bases ranging from very small towns to heavily populated provinces, and have different agency functions. The variety represented in the sample is a good representation of the variety of users or potential users of transportation data across Canada.

Figure 4.1 shows the distribution of levels of government. Clearly, the sample is more heavily weighted toward local municipalities, followed by regional municipalities and provinces. This sample distribution reflects the distribution of such agencies that exist in Canada.

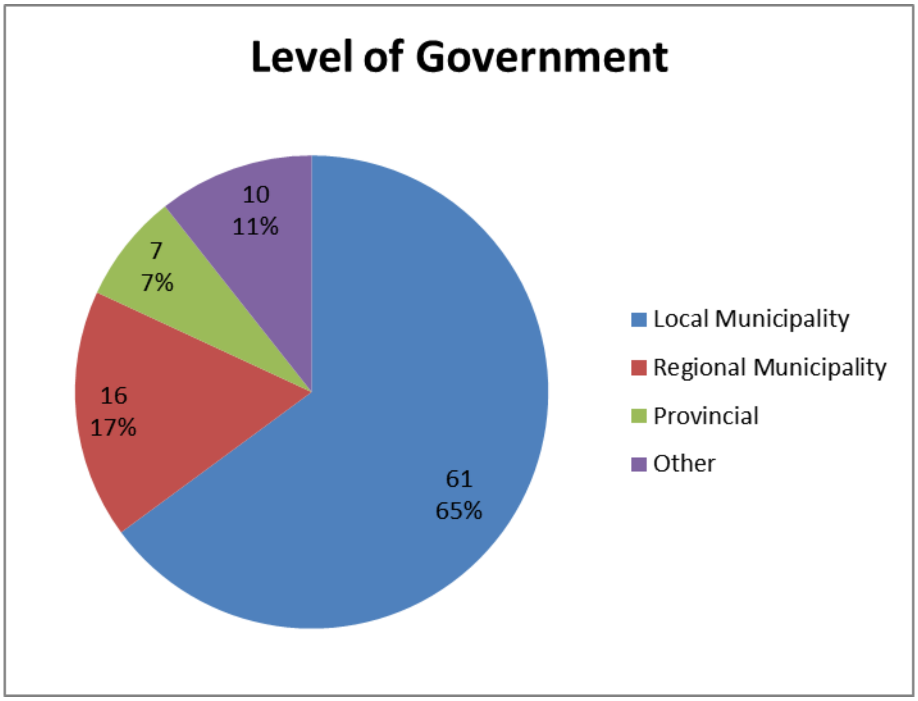


Figure 4.1: Sample distribution by level of government

Figure 4.2 shows the distribution of the population size represented by the agency. All population sizes are represented from small towns (population <10,000) to the largest population centres (population > 1M). Responses were obtained from all of the ten largest Census Metropolitan areas and the majority of all CMAs across Canada.

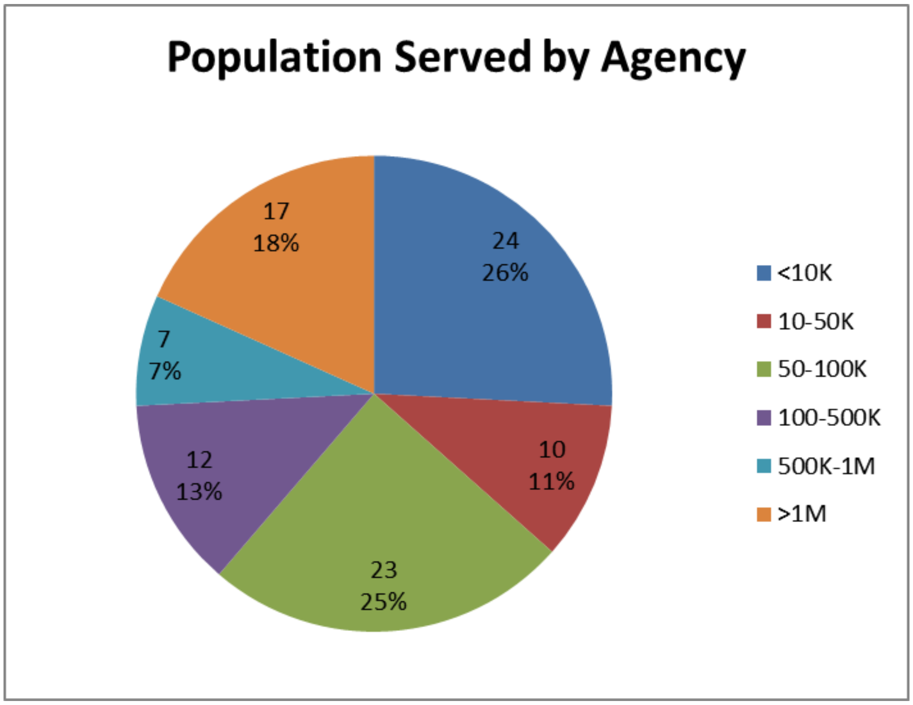


Figure 4.2: Sample distribution of population served by the agency

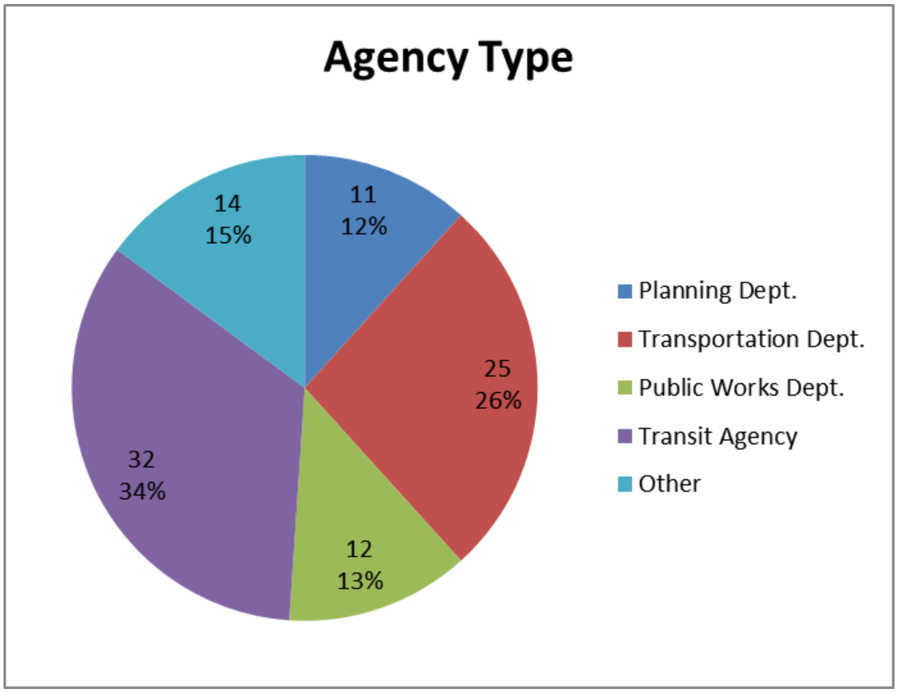


Figure 4.3: Sample distribution by agency type

Figure 4.3 shows the distribution of agency types, including planning, transportation, public works departments, as well as a significant number of transit agencies.

Figure 4.4 shows that responses were obtained from every province and territory across Canada with the exceptions of Prince Edward Island and Nunavut, which make up only 0.5% of Canada's population). Responses from each province were approximately in the same proportions as the population of each province. Notably, 73 of the 94 (78%) responses were completed in the English language, while 21 responses (22%) were completed in French.

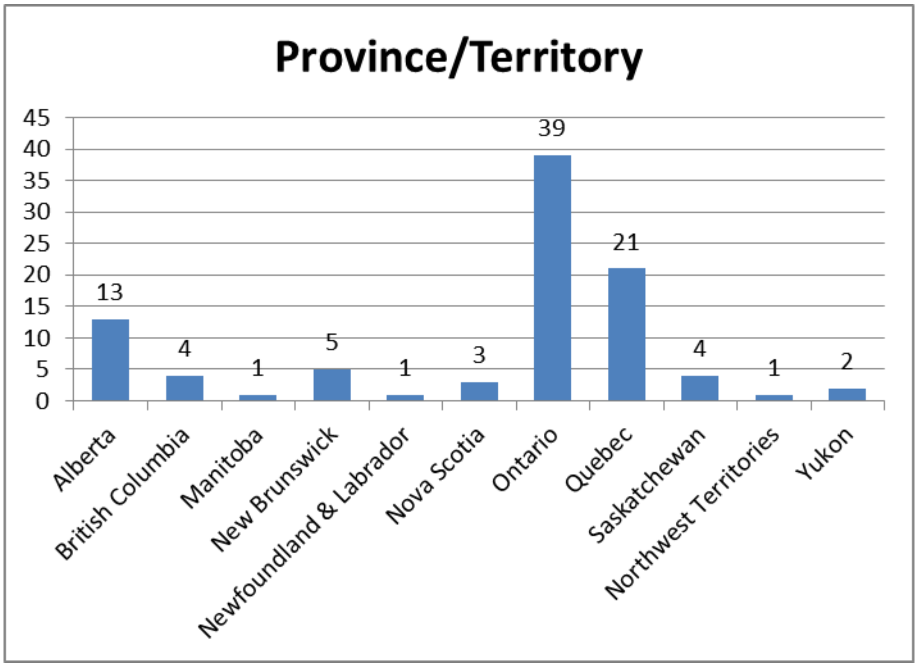


Figure 4.4: Sample distribution by province/territory

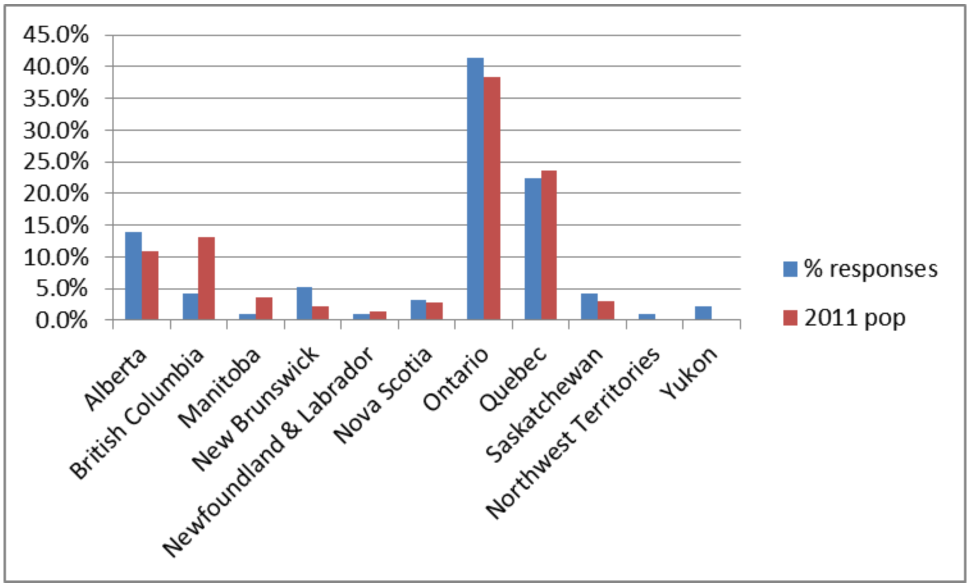


Figure 4.5: Sample distribution by province/territory in comparison with population

5. SURVEY RESULTS

5.1 INTRODUCTION

This section presents a detailed summary of key results obtained from the survey. As usual in a survey of this type, care needs to be taken in terms of generalizing from the sample results to infer population-level conclusions. As discussed above, considerable effort went into gathering as representative a sample as possible, but the net sample size (94 respondents) is still somewhat small. Nevertheless, the results obtained do provide useful insights into Canadian urban passenger transportation data collection methods and issues.

5.2 HOUSEHOLD TRAVEL SURVEYS

Almost one third of the agencies who completed our online survey either conduct household travel surveys in-house or commission them, as shown in Figure 5.1. Additionally, 27% of the participating agencies use household survey data collected by third parties, 8% plan to use this type of data in the near future, and 9% of the agencies would like to use household survey data but these data are not currently available. To sum up, nearly three quarters of the surveyed agencies either use, intend to use or would like to use household travel survey data, which speaks to the importance of this type of data in passenger transportation planning in Canada.

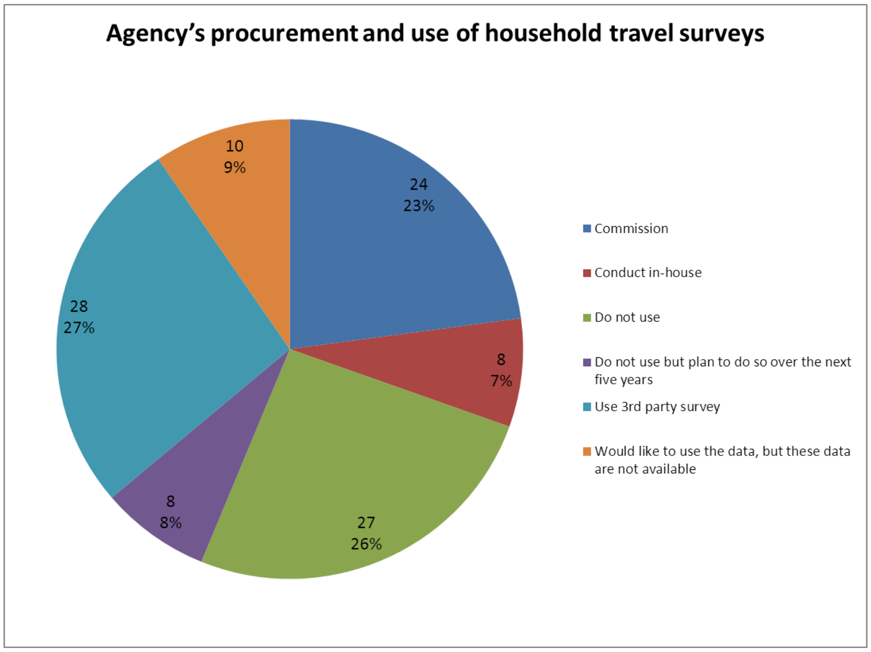


Figure 5.1: Procurement and Use of Household Travel Surveys

A further breakdown of the results by different agency characteristics (Table 5.1) provides more insight into the use and procurement of household travel surveys. A higher percentage of regional and provincial agencies conduct or commission household travel surveys relative to local/area municipalities. Urban region size seems to have less effect on household travel survey usage than one might have expected. While only 21% of very small regions (10-50 thousand population) report commissioning or using survey data, approximately 80% of all respondents in urban regions ranging in size from 50,000 to 1,000,000 across the board report usage of survey data. 71% of urban regions greater than a million in population similarly make use of survey data. Between 25% and 30% of the surveyed agencies from Alberta, British Columbia and Ontario do not use household travel data, while the corresponding percentage in Quebec is merely 5%, indicating its greater reliance on this data relative to the other large provinces. Planning departments have the lowest percentage of agencies that do not use household travel data.

Table 5.1 Procurement and Use of Household Travel Surveys by Agency Characteristics

	Commission	Conduct in-house	Do not use	Do not use but plan to do so over the next five years	Use 3rd party survey	Would like to use the data, but these data are not available	Grand Total
Local (area) municipality	11	4	16	6	17	7	61
Other	3		2		4	1	10
Provincial	3	1	3				7
Regional municipality	4	2	2	2	4	2	16
(blank)	6	1	5		4	1	17
< 10,000						1	1
10,000-50,000	1	1	13	3	4	2	24
50,000-100,000	1	1	1	2	4	1	10
100,000-500,000	4	2	1	3	9	4	23
500,000-1,000,000	5	2	1		3	1	12
> 1,000,000	4		2		1		7
Alberta	3	1	4	1	2	2	13
British Columbia	3		1				4
Manitoba		1					1
New Brunswick			2		1	2	5
Newfoundland and Labrador			1				1
Northwest Territories				1			1
Nova Scotia		1		1		1	3
Ontario	6	2	11	3	13	4	39
Quebec	9	1	1	1	9		21
Saskatchewan		1	1	1		1	4
Yukon			2				2
Other	4		6		3	1	14
Planning Department	3		1	2	2	3	11
Public Works Department	1		3	2	5	1	12
Transit Agency	6	4	7	2	9	4	32
Transportation Department	7	3	6	2	6	1	25
Grand Total	**21**	**7**	**23**	**8**	**25**	**10**	**94**

Individual agencies may use household travel surveys with different designs depending on the purposes and requirements of household travel data at different points in time. It is found that slightly above half of the agencies who conducted or commissioned household travel surveys have used two or more distinct survey designs over the past 10 years while 43% of the those agencies adhered to one distinct design over the same period, as shown in Figure 5.2. In terms of the frequency of conducting or commissioning household travel surveys, most agencies use a five-year cycle (mostly matching the Census years) or shorter, with one agency in Quebec reporting an annual household travel survey (see Figure 5.3). Other extreme cases exist, where travel surveys have been conducted on a one-time basis or every 10 years.

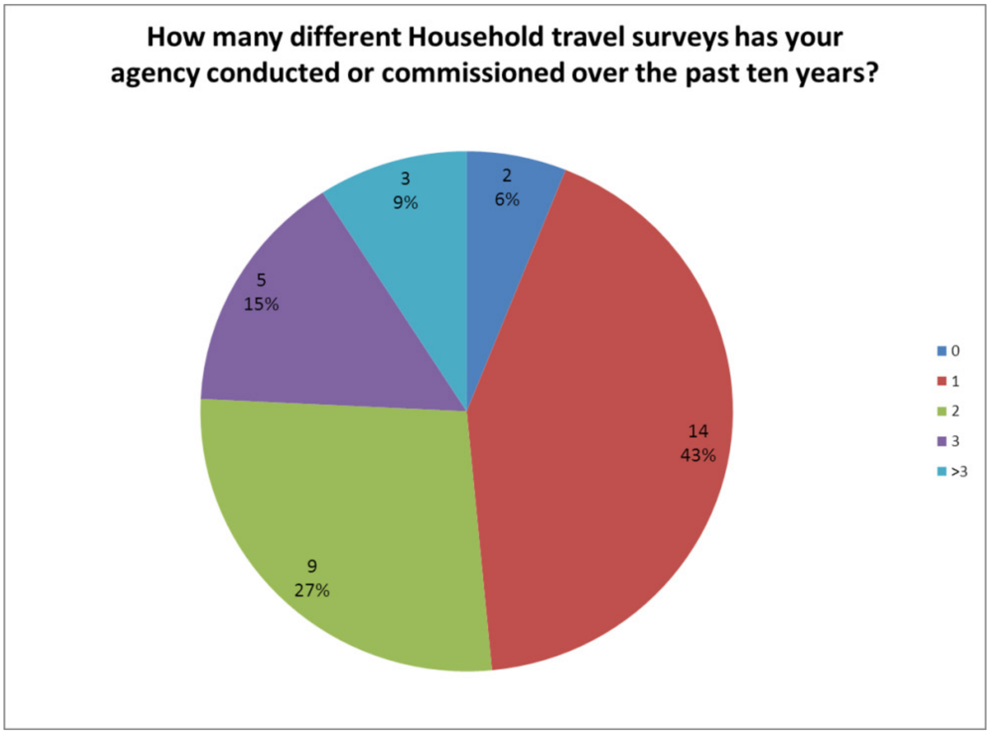


Figure 5.2: Different Designs of Household Travel Surveys

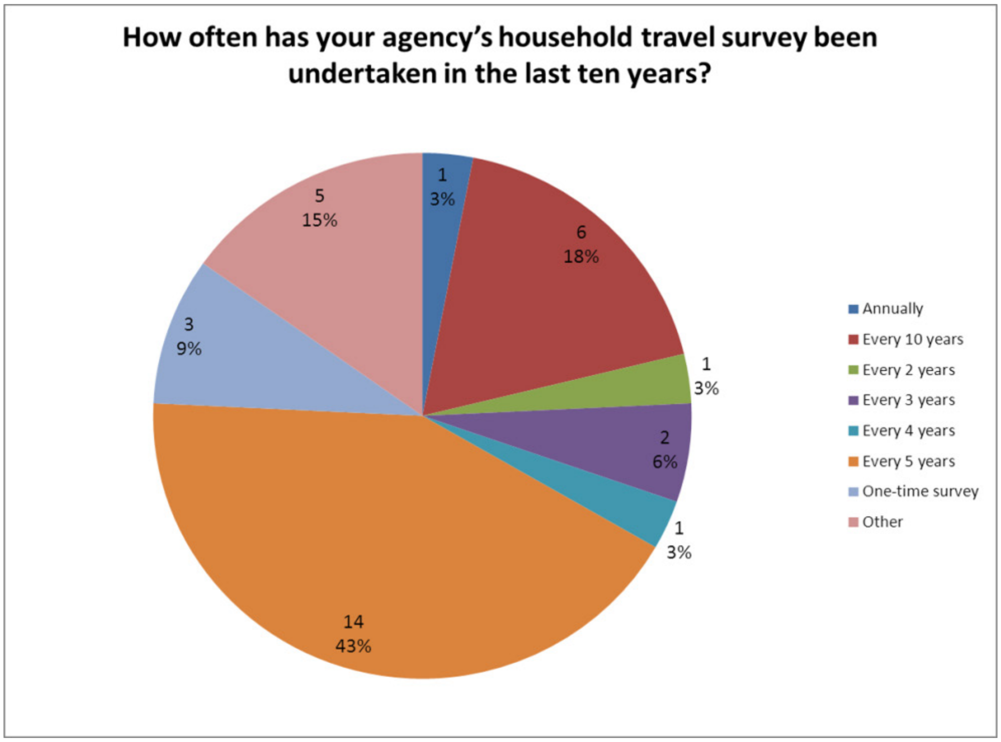


Figure 5.3: Frequency of Undertaking Household Travel Surveys

The sample size reported by participating agencies ranges from 300 subjects in a mid-size municipality in Nova Scotia to 5% of all households (about 125,000 in total) in the Greater Toronto and Hamilton Area in Ontario. While several factors could influence the selection of sample size, available budget and population size are likely the most determining factors.

The household is the most common sampling unit used in conducted or commissioned surveys, followed by person then vehicle sampling units (see Figure 5.4). As shown in Figure 5.5, survey samples were drawn mostly from telephone listings, with fewer surveys relying on other types of listings to draw their samples.

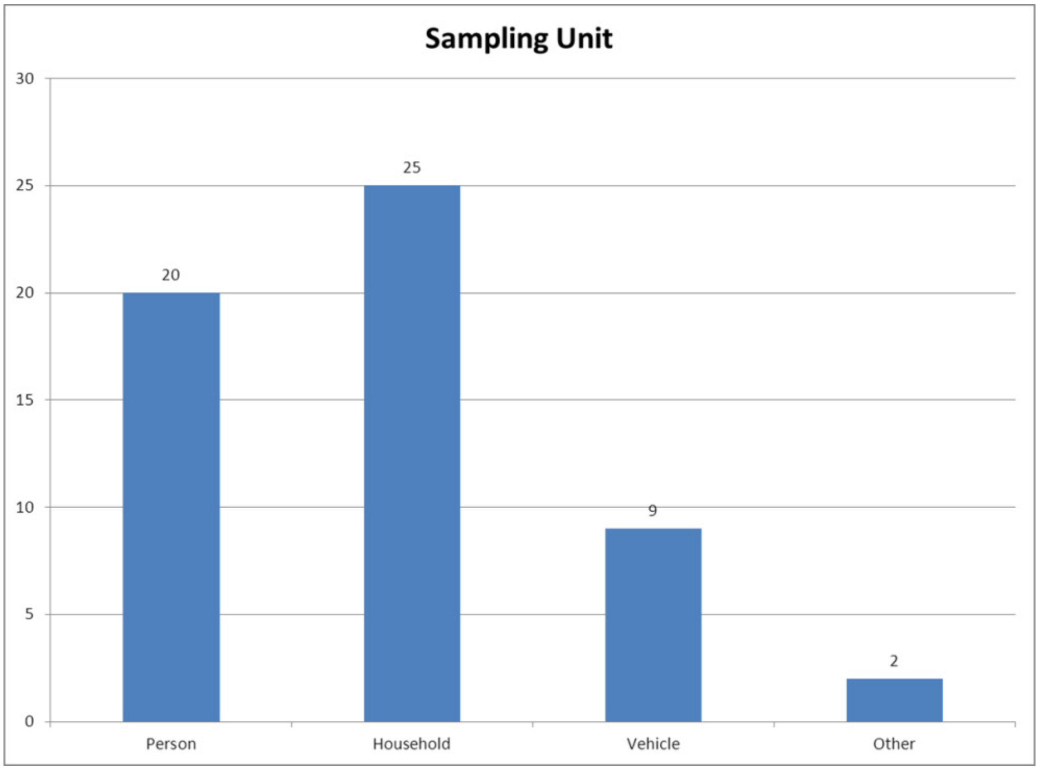


Figure 5.4: Sampling Unit of Household Travel Surveys

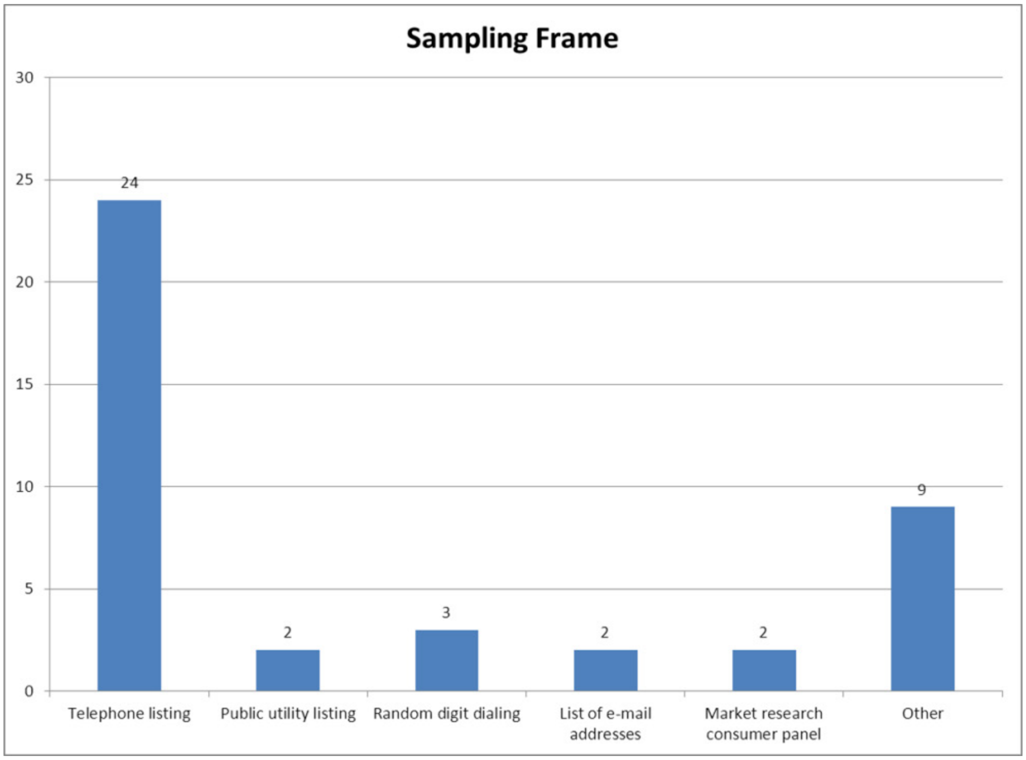


Figure 5.5: Sampling Frame of Household Travel Surveys

In a vast majority of commissioned or conducted household travel surveys (85%), interviewed subjects provided data not only on their own travel but also that of fellow members of their households (i.e. proxy reporting). The data collected in most surveys were predominantly for one day of travel per person, with two surveys collecting data on five days of travel and two other surveys collecting 7- and 20-day worth of data respectively. In one case, it was reported that travel data were collected continuously for 365 days! Nearly 60% of all reported surveys collected data on weekday travel, with the remaining surveys collecting data on travel performed on both weekdays and weekends. The household travel surveys in our study have predominantly collected trip-based travel data, with fewer surveys focusing on activity-based travel, residential location choice and household time use (see Figure 5.6).

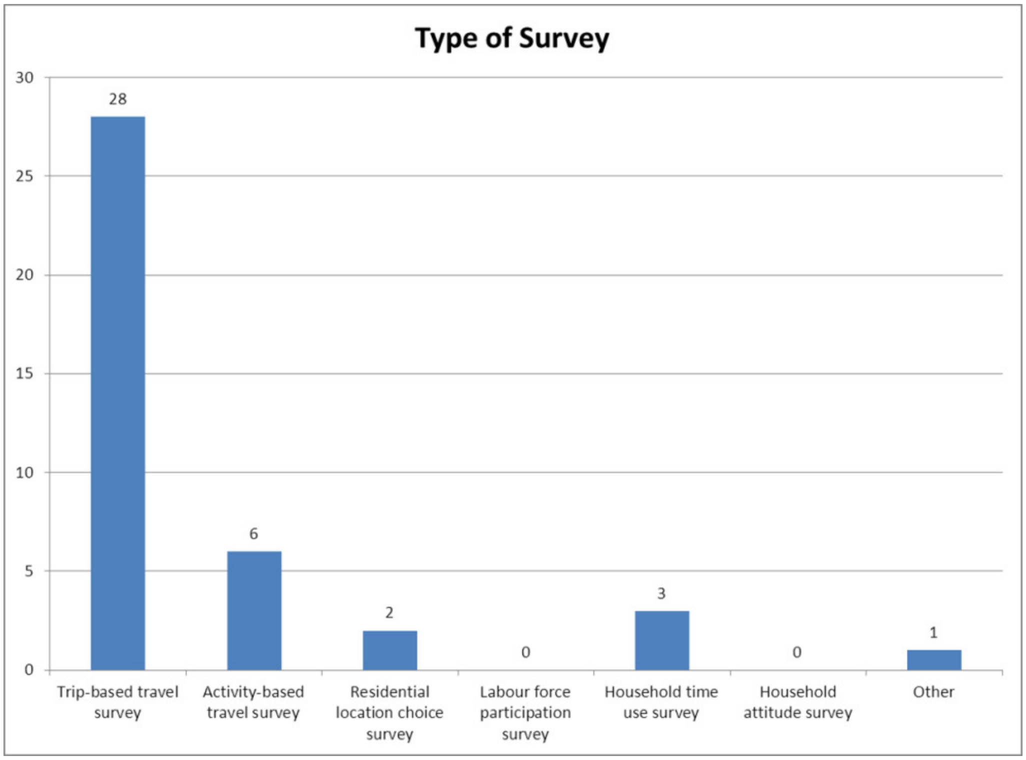


Figure 5.6: Type of Household Travel Surveys

The implementation of any household travel surveys typically starts with an initial contact of the sample subjects to inform them of the purpose of the survey and the method and timing of the interview. The next step in the process is the interview itself, sometimes accompanied with a follow-up contact as needed. While initial contacts have been made mostly either by letter or phone, the latter has predominantly been used for any follow-up contact, as shown in Figures 5.7 and 5.8. Many surveys used the telephone interview as the preferred method of data collection, with some surveys using web-based online data collection (sometimes as an alternative option for information reporting), and a fewer surveys relying on self-administered mail back questionnaires and face-to-face methods (see Figure 5.9).

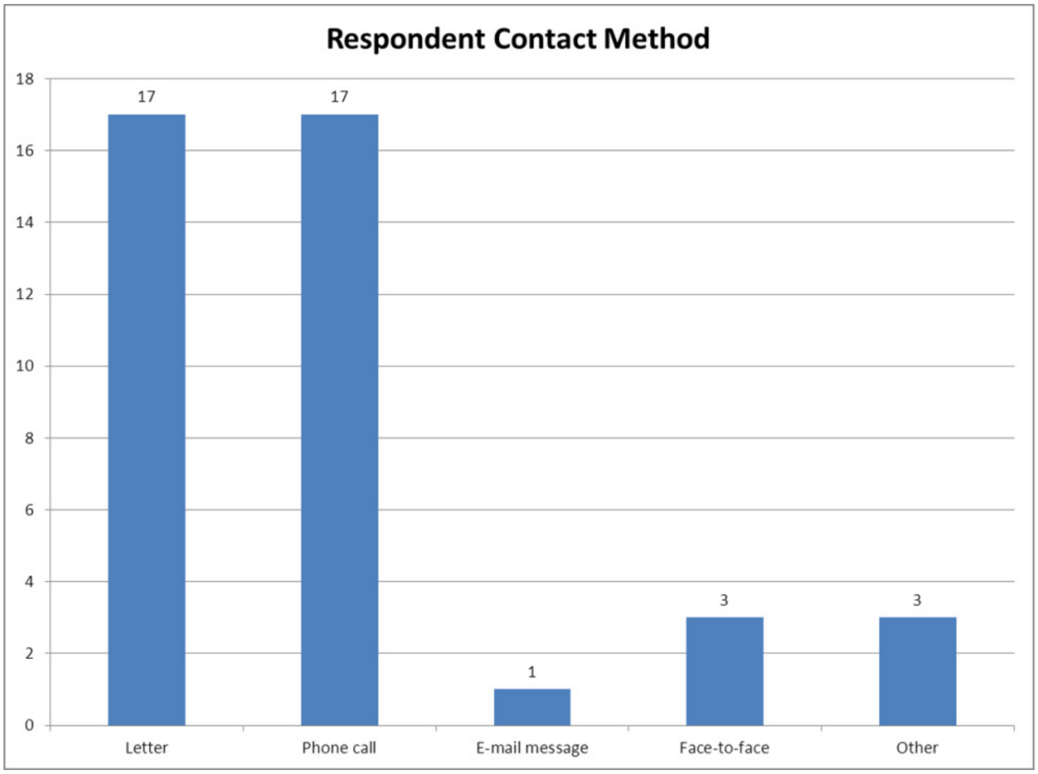


Figure 5.7: Initial Contact Method

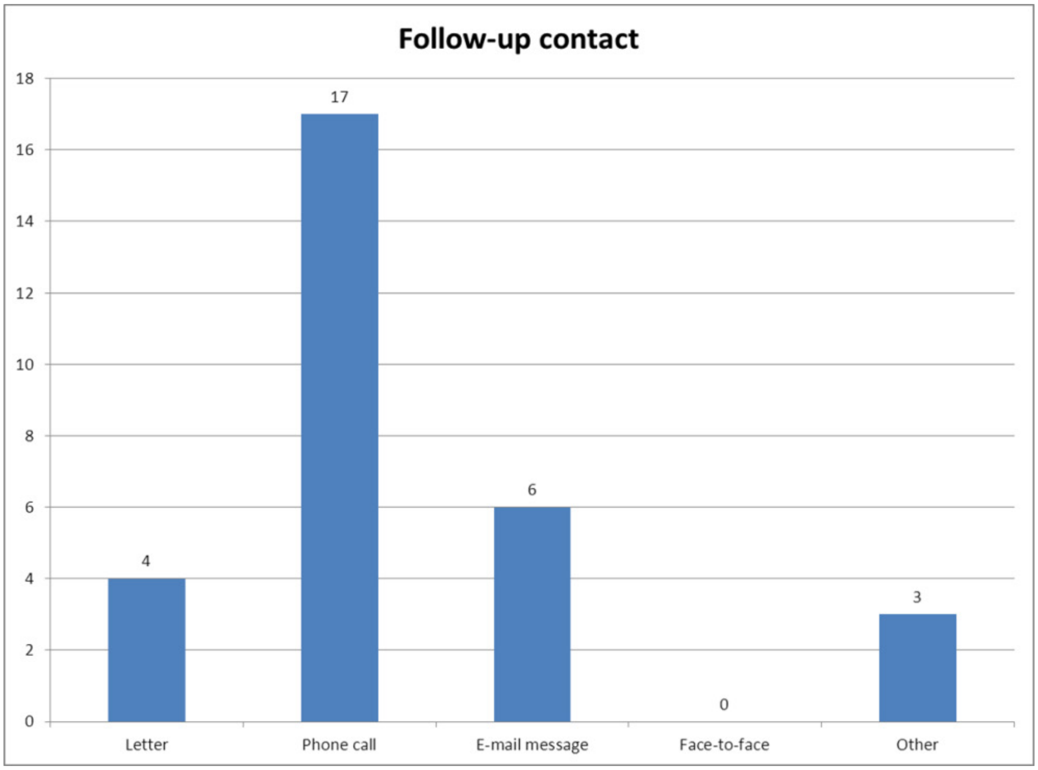


Figure 5.8: Follow-up Contact Method

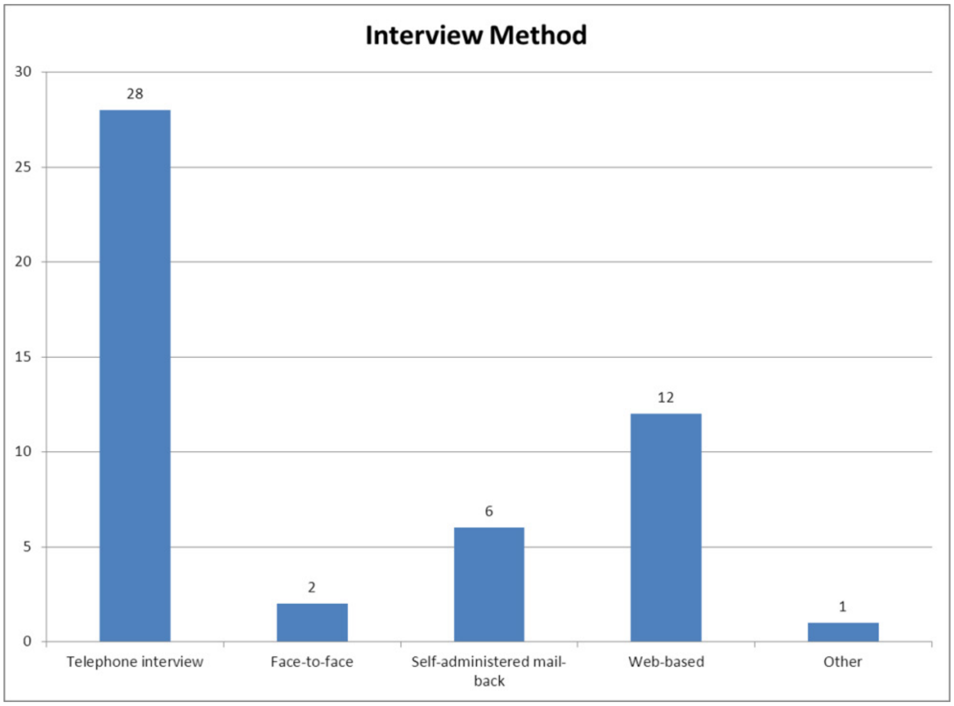


Figure 5.9: Interview Method

Figure 5.10 shows that survey data are stored mostly in a database management system which facilitates data management, retrieval and analysis by users. Other forms of data storage include spreadsheet, web-based systems and text files.

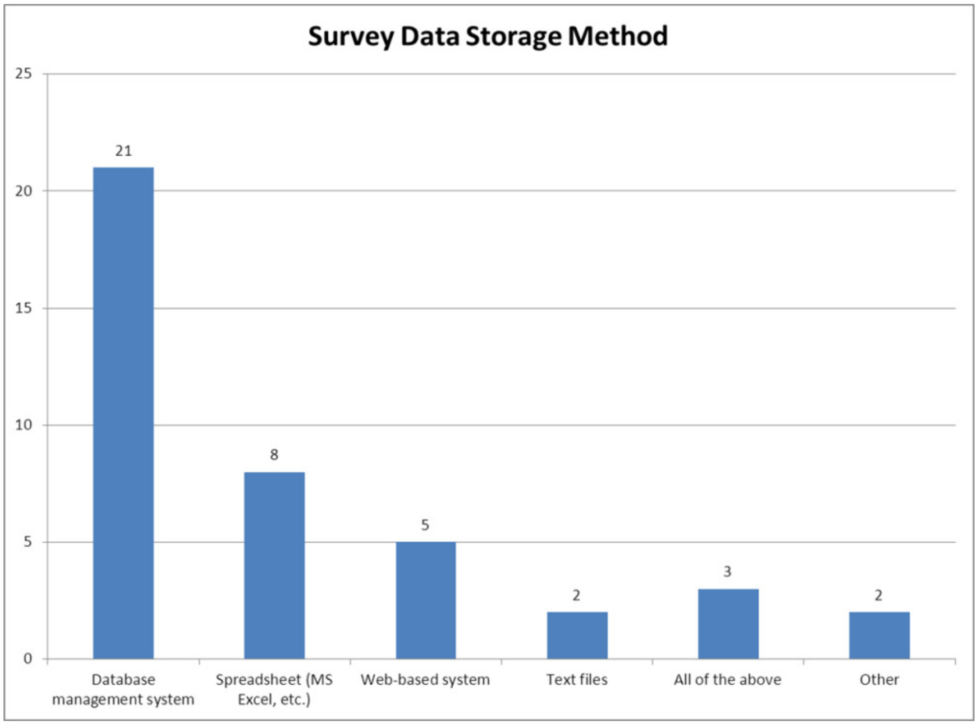


Figure 5.10: Data Storage Method

As Figure 5.11 shows, household survey data are well utilized not only by the sponsoring agencies but they are also shared with a variety of entities including other public agencies, academic researchers, consultants, private sector companies, and the general public. Only very few agencies reported that they do not share their household travel data. In most cases the data are shared free of charge, as shown in Figure 5.12.

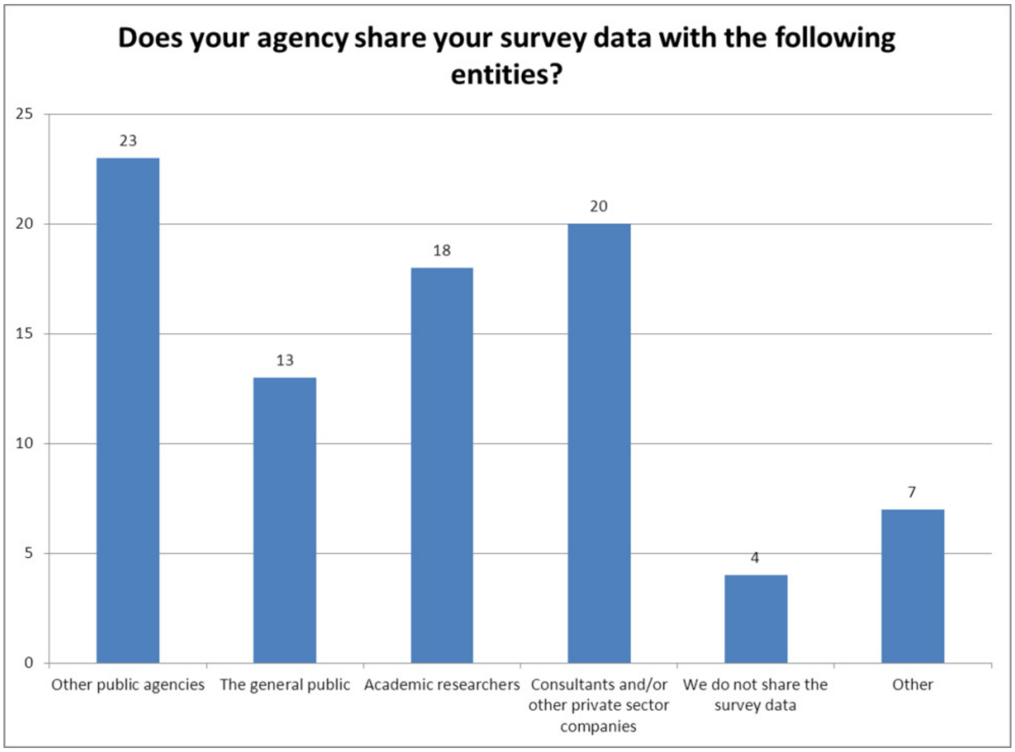


Figure 5.11: Sharing of Data

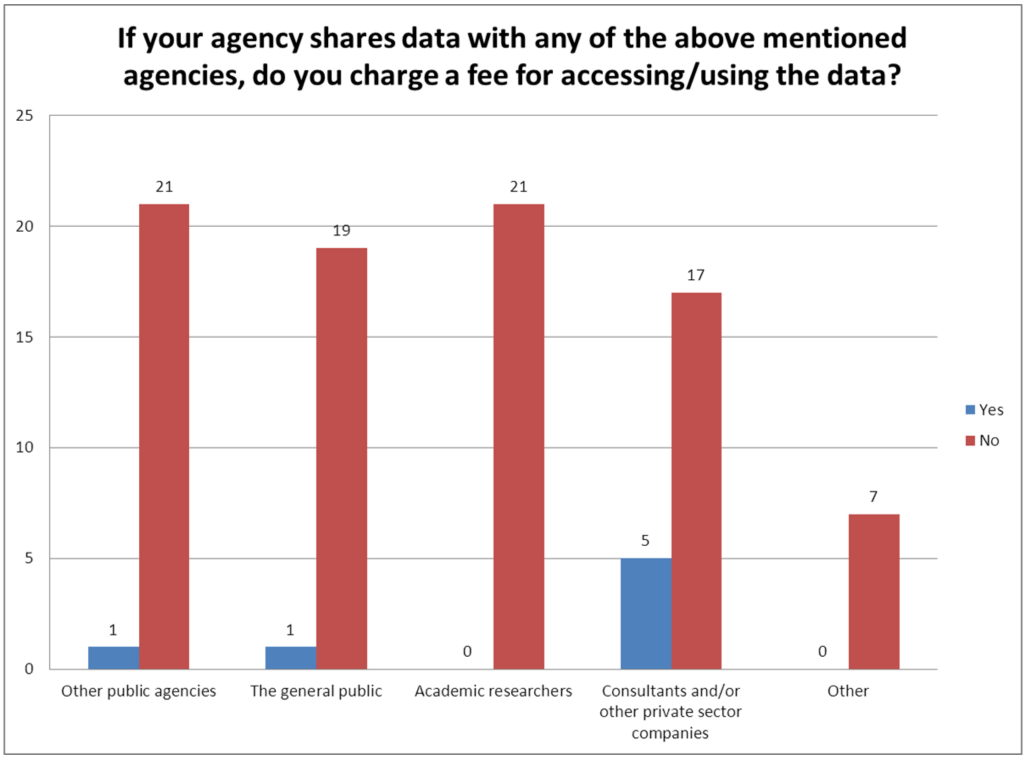


Figure 5.12: Fees for using Data

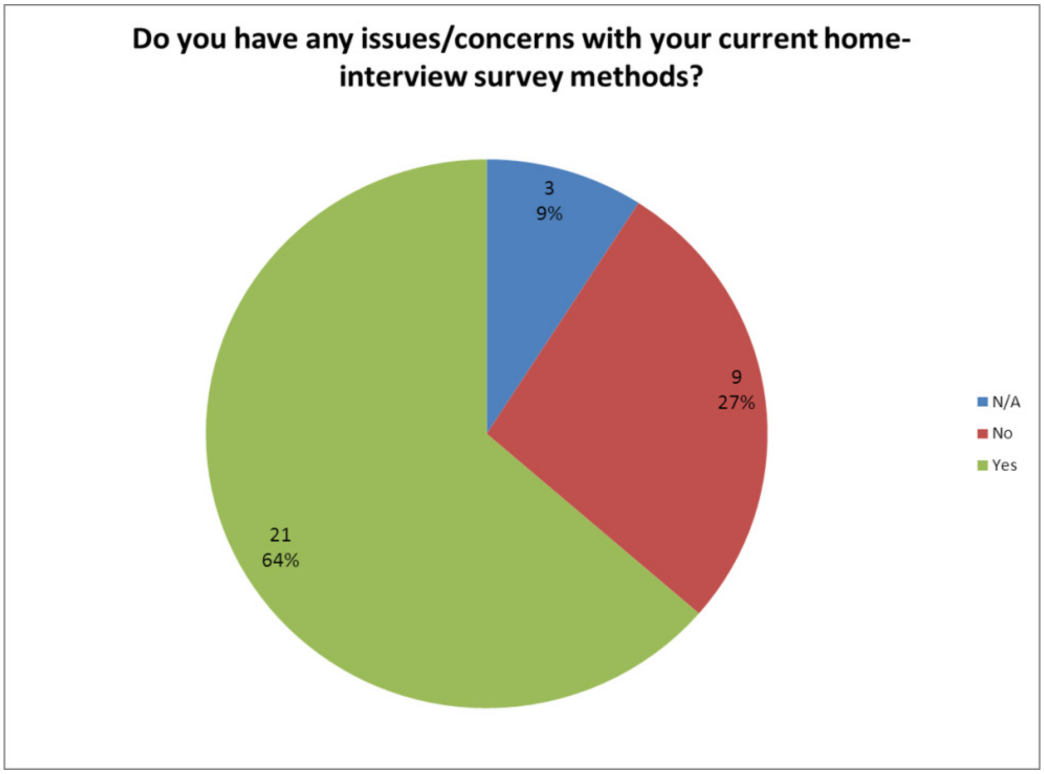


Figure 5.13: Issues and Concerns with Household Travel Surveys

Almost two thirds of the agencies who conducted or commissioned household travel surveys have expressed concerns with the survey methods used (see Figure 5.13). These concerns are paraphrased and summarized as follows (numbers in parentheses indicated the number of respondents that mentioned this specific concern).

- It is increasingly difficult to ensure a representative sample using telephone listings (14). Specific reasons mentioned include: increasing use of cell phones, and not every household has a landline) (12), young people and students in residence (4), tourists (1), business travellers (1)
- Low response rates are obtained (6). Specific populations of concern include residents of the region surrounding the city (1), low income households (1), large households (1), younger people (1), zero-trip households (1) and residents of apartments and some townhouse units that did not receive the pre-interview letter because unit numbers could not be obtained from telephone directory information (1)
- Trips are under-reported or biased (5). Specific types of under-reported trips mentioned included off-peak trips (1), discretionary trips (2), non-auto trips (1), and trips by proxy respondents (1), and there is concern about a transit bias (1).
- The survey sample size is too small for analysis of subsets of trips such as transit, cycling or pedestrian trips (4)
- The cost, time and personnel to conduct surveys is high (4)
- Web based surveys should be considered as an option for respondents (2).
- There is a lack of information on weekend travel (1) and travel variability throughout the week (1)
- Privacy issues are of concern (1)
- Hiring / retaining of bilingual surveyors (1)
- Survey expansion is problematic because it is impossible to know the true universe of the survey because of constraints / shortcomings of census data for certain populations (e.g. institutions for the elderly, residents of native reserves); inaccuracies or inconsistencies in the number of dwellings inhabited, etc. (1)
- Assessment of the individual consumption of transportation and the impact on greenhouse gases (1)
- Effectively measuring the emergence of modes such as biking, carpooling, and combined modes (1)

Many of the respondents indicated their intention to address the issues and concerns they are faced with (see Figure 5.14).

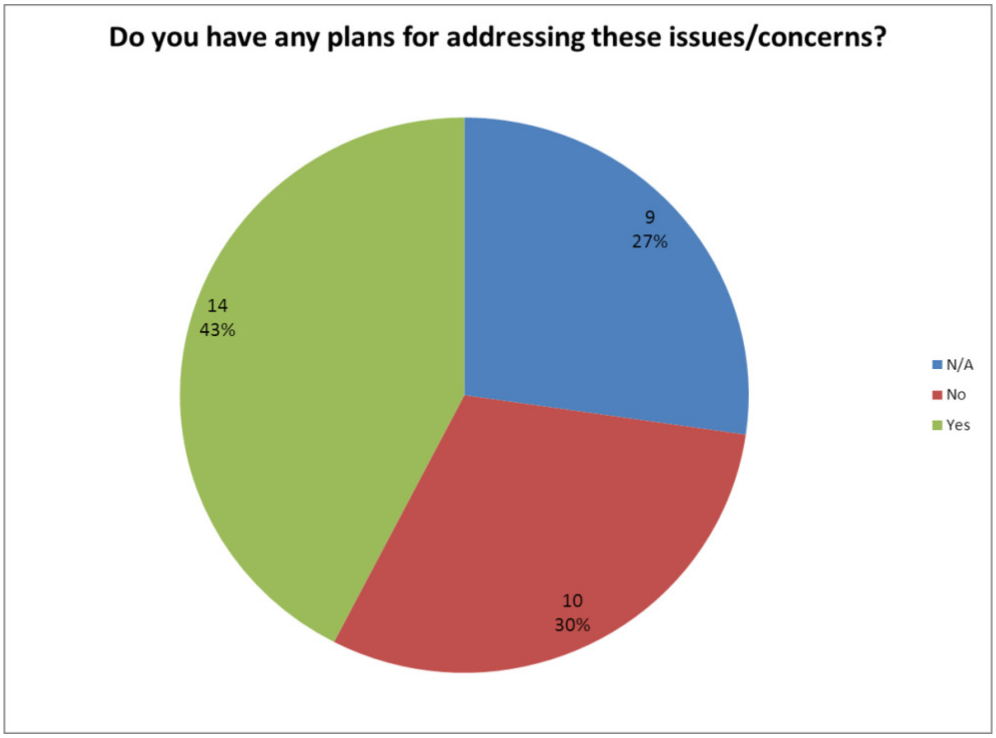


Figure 5.14: Plans to Address Issues and Concerns

5.3 OTHER TRAVEL SURVEYS

Our online survey solicited further information on other types of surveys including:

- Place of Employment Surveys.
- Roadside Origin-Destination Surveys.
- Transit Onboard/Stop Surveys.
- Truck/Freight/Commercial Vehicle Surveys.
- Bicycle Surveys.
- Pedestrian Surveys.
- Parking Usage Surveys.
- Special Traffic Generator Surveys.
- Stated Preference Surveys.
- Attitudinal or Market Research Surveys.

In general, these surveys are undertaken less widely and frequently than household travel surveys, and their specialized nature make them more specific and relevant to some agencies than others. Figure 5.15 reveals that none of the above surveys is conducted or commissioned by more than 20% of the responding agencies, with the exception of the transit onboard/stop survey (41%), parking usage survey (26%) and attitudinal or market research survey (23%).

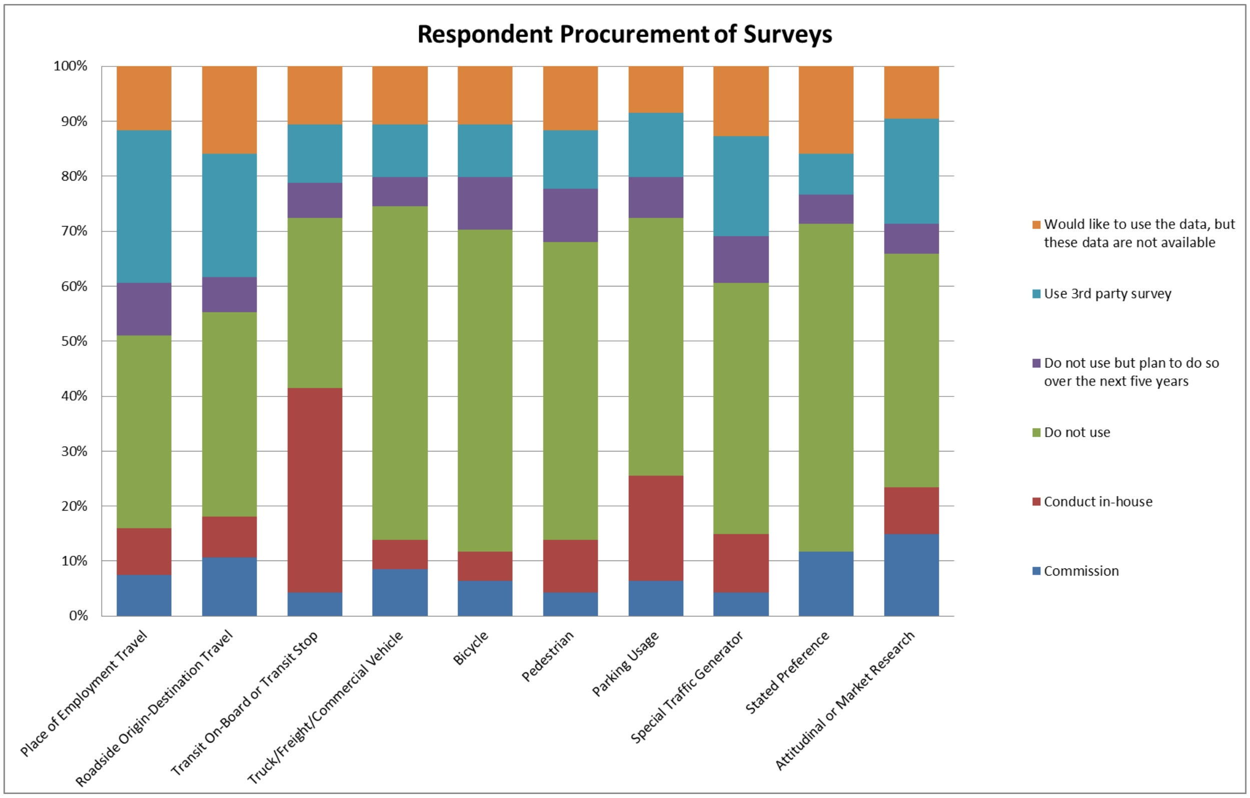


Figure 5.15: Procurement and Use of Other Surveys

Tables 5.2 through 5.11 show the level of procurement and use of each travel survey broken down by different agency characteristics. As expected, transit onboard/stop surveys are undertaken mainly by transit agencies and to a lesser extent by transportation departments (see Table 5.4). Municipalities of different sizes conduct or commission this type of survey, and a higher percentage of agencies from Alberta and Quebec undertake this survey relative to Ontario. Other surveys on active transportation (pedestrian and bicycles) and parking are undertaken mostly by transportation departments in local and regional municipalities (see Tables 5.6, 5.7 and 5.8). Roadside O-D surveys and commercial vehicle surveys are also undertaken more often by transportation departments, with Alberta showing a higher percentage of participating agencies conducting/commissioning the latter type than other large provinces (see Tables 5.3 and 5.5).

Additional survey results concerning these survey types are presented in Appendix I.

Table 5.2: Procurement & Use of Place of Employment Survey by Agency Characteristics

	Commission	Conduct in-house	Do not use	Do not use but plan to do so over the next five years	Use 3rd party survey	Would like to use the data, but these data are not available	Total
Local (area) municipality	4	5	21	8	16	7	61
Other	1		4		4	1	10
Provincial	1		5		1		7
Regional municipality	1	3	3	1	5	3	16
(blank)	2		9		5	1	17
< 10,000		1					1
10,000-50,000	1	2	12	3	4	2	24
50,000-100,000	1	1	2	2	3	1	10
100,000-500,000	1	1	3	4	10	4	23
500,000-1,000,000	2	2	4		2	2	12
> 1,000,000		1	3		2	1	7
Alberta		3	6	1	2	1	13
British Columbia			1	1	2		4
Manitoba						1	1
New Brunswick			2		1	2	5
Newfoundland and Labrador			1				1
Northwest Territories				1			1
Nova Scotia		1	1			1	3
Ontario	3	3	12	4	13	4	39
Quebec	4	1	7	1	8		21
Saskatchewan			1	1		2	4
Yukon			2				2
Other	2	1	6	2	2	1	14
Planning Department		1	2	1	5	2	11
Public Works Department	1		3	2	4	2	12
Transit Agency	3	2	10	2	10	5	32
Transportation Department	1	4	12	2	5	1	25
Grand Total	**7**	**8**	**33**	**9**	**26**	**11**	**94**

Table 5.3: Procurement & Use of Roadside Origin-Destination Survey by Agency Characteristics

Row Labels	**Commission**	**Conduct in-house**	**Do not use**	**Do not use, but Planning to do so over next 5 years**	**Use 3rd party survey**	**Would like to use the data, but these data are not available**	**Grand Total**
Local (area) municipality	6	3	26	5	11	10	61
Other	1		5		3	1	10
Provincial	1	1	3		2		7
Regional municipality	2	3	1	1	5	4	16
(blank)	2	1	8		5	1	17
< 10,000						1	1
10,000-50,000	1	1	13	3	3	3	24
50,000-100,000	1		2	1	4	2	10
100,000-500,000		1	7	2	7	6	23
500,000-1,000,000	2	3	4		2	1	12
> 1,000,000	4	1	1			1	7
Alberta		2	6		2	3	13
British Columbia	1		1		1	1	4
Manitoba			1				1
New Brunswick			1		2	2	5
Newfoundland and Labrador			1				1
Northwest Territories				1			1
Nova Scotia		1				2	3
Ontario	6	2	11	3	12	5	39
Quebec	3	2	11	1	4		21
Saskatchewan			1	1		2	4
Yukon			2				2
Other			8		4	2	14
Planning Department	2		1	1	3	4	11
Public Works Department	1		3	2	4	2	12
Transit Agency	3	1	17	1	5	5	32
Transportation Department	4	6	6	2	5	2	25
Grand Total	**10**	**7**	**35**	**6**	**21**	**15**	**94**

Table 5.4: Procurement & Use of Transit Onboard/Stop Survey by Agency Characteristics

Row Labels	**Commission**	**Conduct in-house**	**Do not use**	**Do not use but plan to do so over the next five years**	**Use 3rd party survey**	**Would like to use the data, but these data are not available**	**Grand Total**
Local (area) municipality	3	23	17	6	4	8	61
Other	1	4	2		2	1	10
Provincial		1	4		2		7
Regional municipality		7	6		2	1	16
(blank)	1	5	6		4	1	17
< 10,000		1					1
10,000-50,000	1	6	11	4	1	1	24
50,000-100,000		6	2		1	1	10
100,000-500,000	2	8	5	1	2	5	23
500,000-1,000,000		7	3		1	1	12
> 1,000,000		2	2	1	1	1	7
Alberta		7	4			2	13
British Columbia		1			2	1	4
Manitoba		1					1
New Brunswick		1	3			1	5
Newfoundland and Labrador			1				1
Northwest Territories		1					1
Nova Scotia		1	1			1	3
Ontario	2	11	15	2	5	4	39
Quebec	2	10	3	3	3		21
Saskatchewan		2	1			1	4
Yukon			1	1			2
Other		1	7	2	2	2	14
Planning Department		2	3	1	2	3	11
Public Works Department	1	2	6	1	2		12
Transit Agency	3	19	4	1	2	3	32
Transportation Department		11	9	1	2	2	25
Grand Total	**4**	**35**	**29**	**6**	**10**	**10**	**94**

Table 5.5: Procurement & Use of Truck/Freight/Commercial Vehicle Survey by Agency Characteristics

Row Labels	Commission	Conduct in-house	Do not use	Do not use but plan to do so over the next five years	Use 3rd party survey	Would like to use the data, but these data are not available	Grand Total
Local (area) municipality	5	3	39	4	3	7	61
Other		1	9				10
Provincial	1	1	3		2		7
Regional municipality	2		6	1	4	3	16
(blank)	1	2	12		2		17
< 10,000		1					1
10,000-50,000		1	16	3	2	2	24
50,000-100,000			8		1	1	10
100,000-500,000			15	2	1	5	23
500,000-1,000,000	4	1	3		3	1	12
> 1,000,000	3		3			1	7
Alberta	3	2	6		1	1	13
British Columbia	1		1		1	1	4
Manitoba			1				1
New Brunswick			2		1	2	5
Newfoundland and Labrador			1				1
Northwest Territories				1			1
Nova Scotia			2			1	3
Ontario	3	1	24	2	4	5	39
Quebec	1	1	16	1	2		21
Saskatchewan		1	2	1			4
Yukon			2				2
Other	2		7		2	3	14
Planning Department	1	2	4	1	2	1	11
Public Works Department		1	6	3	1	1	12
Transit Agency			31			1	32
Transportation Department	5	2	9	1	4	4	25
Grand Total	**8**	**5**	**57**	**5**	**9**	**10**	**94**

Table 5.6: Procurement & Use of Bicycle Survey by Agency Characteristics

Row Labels	Commission	Conduct in-house	Do not use	Do not use but plan to do so over the next five years	Use 3rd party survey	Would like to use the data, but these data are not available	Grand Total
Local (area) municipality	5	1	34	7	6	8	61
Other			7	1	1	1	10
Provincial		1	6				7
Regional municipality	1	3	8	1	2	1	16
(blank)		1	13	1	1	1	17
< 10,000						1	1
10,000-50,000			16	2	4	2	24
50,000-100,000			5	4	1		10
100,000-500,000	1	3	13	2	1	3	23
500,000-1,000,000	3	1	3		2	3	12
> 1,000,000	2		5				7
Alberta	2		6	2	1	2	13
British Columbia		1	2			1	4
Manitoba			1				1
New Brunswick		1	4				5
Newfoundland and Labrador			1				1
Northwest Territories			1				1
Nova Scotia		1	2				3
Ontario	2	1	22	4	4	6	39
Quebec	2	1	13	2	3		21
Saskatchewan			2	1		1	4
Yukon			1		1		2
Other	1		8		2	3	14
Planning Department	1	2	3	2	1	2	11
Public Works Department			8	2	1	1	12
Transit Agency	1		26	2	2	1	32
Transportation Department	3	3	10	3	3	3	25
Grand Total	**6**	**5**	**55**	**9**	**9**	**10**	**94**

Table 5.7: Procurement & Use of Pedestrian Survey by Agency Characteristics

Row Labels	**Commission**	**Conduct in-house**	**Do not use**	**Do not use but plan to do so over the next five years**	**Use 3rd party survey**	**Would like to use the data, but these data are not available**	**Grand Total**
Local (area) municipality	4	7	29	7	6	8	61
Other			6	1	1	2	10
Provincial			7				7
Regional municipality		2	9	1	3	1	16
(blank)			13	1	1	2	17
< 10,000		1					1
10,000-50,000	1	1	14	2	4	2	24
50,000-100,000		2	5	2		1	10
100,000-500,000	1	3	11	3	3	2	23
500,000-1,000,000	1	1	4		2	4	12
> 1,000,000	1	1	4	1			7
Alberta	1	3	6	1	1	1	13
British Columbia			2		1	1	4
Manitoba			1				1
New Brunswick		1	4				5
Newfoundland and Labrador			1				1
Northwest Territories			1				1
Nova Scotia		1	2				3
Ontario	2	3	19	4	3	8	39
Quebec	1	1	13	3	3		21
Saskatchewan			1	1	1	1	4
Yukon			1		1		2
Other	1		8		2	3	14
Planning Department	1	2	1	2	2	3	11
Public Works Department		2	7	1	1	1	12
Transit Agency	1		24	3	3	1	32
Transportation Department	1	5	11	3	2	3	25
Grand Total	**4**	**9**	**51**	**9**	**10**	**11**	**94**

Table 5.8: Procurement & Use of Parking Usage Survey by Agency Characteristics

Row Labels	Commission	Conduct in-house	Do not use	Do not use but plan to do so over the next five years	Use 3rd party survey	Would like to use the data, but these data are not available	Grand Total
Local (area) municipality	4	13	28	5	6	5	61
Other		1	4	1	2	2	10
Provincial	1		4	1		1	7
Regional municipality	1	4	8		3		16
(blank)	1	1	8	2	2	3	17
< 10,000	1						1
10,000-50,000		4	13	2	3	2	24
50,000-100,000	1	3	4		1	1	10
100,000-500,000	1	5	11	3	2	1	23
500,000-1,000,000		3	6		2	1	12
> 1,000,000	2	2	2		1		7
Alberta	1	3	6		2	1	13
British Columbia	1		1		1	1	4
Manitoba			1				1
New Brunswick		1	3			1	5
Newfoundland and Labrador			1				1
Northwest Territories			1				1
Nova Scotia		2	1				3
Ontario	4	8	17	4	3	3	39
Quebec		4	10	2	4	1	21
Saskatchewan			2	1		1	4
Yukon			1		1		2
Other	2	1	7		2	2	14
Planning Department	2	1	1	2	3	2	11
Public Works Department		4	6	1		1	12
Transit Agency		3	21	2	4	2	32
Transportation Department	2	9	9	2	2	1	25
Grand Total	**6**	**18**	**44**	**7**	**11**	**8**	**94**

Table 5.9: Procurement & Use of Special Traffic Generator Survey by Agency

Row Labels	Commission	Conduct in-house	Do not use	Do not use but plan to do so over the next five years	Use 3rd party survey	Would like to use the data, but these data are not available	Grand Total
Local (area) municipality	2	4	31	7	9	8	61
Other		1	2		4	3	10
Provincial		1	6				7
Regional municipality	2	4	4	1	4	1	16
(blank)		2	8		4	3	17
< 10,000					1		1
10,000-50,000	1		15	2	4	2	24
50,000-100,000		2	4	1	1	2	10
100,000-500,000		3	10	3	5	2	23
500,000-1,000,000	1	2	5	1	1	2	12
> 1,000,000	2	1	1	1	1	1	7
Alberta		3	7	1	1	1	13
British Columbia	1		1		1	1	4
Manitoba			1				1
New Brunswick		1	1	1		2	5
Newfoundland and Labrador			1				1
Northwest Territories					1		1
Nova Scotia		1	2				3
Ontario	3	2	16	3	10	5	39
Quebec		3	9	3	4	2	21
Saskatchewan			3			1	4
Yukon			2				2
Other	1		10		1	2	14
Planning Department	1	1	2	1	3	3	11
Public Works Department			5	1	4	2	12
Transit Agency		2	15	4	6	5	32
Transportation Department	2	7	11	2	3		25
Grand Total	**4**	**10**	**43**	**8**	**17**	**12**	**94**

Table 5.10: Procurement & Use of Stated Preference Surveys by Agency Characteristics

Row Labels	Commission	Do not use	Do not use but plan to do so over the next five years	Use 3rd party survey	Would like to use the data, but these data are not available	Grand Total
Local (area) municipality	5	36	3	5	12	61
Other	1	7			2	10
Provincial	3	4				7
Regional municipality	2	9	2	2	1	16
(blank)	4	11			2	17
< 10,000				1		1
10,000-50,000	1	16	2	3	2	24
50,000-100,000		5	1	1	3	10
100,000-500,000	4	12	1	1	5	23
500,000-1,000,000	1	8		1	2	12
> 1,000,000	1	4	1		1	7
Alberta	1	8		3	1	13
British Columbia	1	1	1	1		4
Manitoba		1				1
New Brunswick		3			2	5
Newfoundland and Labrador		1				1
Northwest Territories		1				1
Nova Scotia	1	2				3
Ontario	2	25	2	1	9	39
Quebec	5	10	2	2	2	21
Saskatchewan	1	2			1	4
Yukon		2				2
Other	2	9		2	1	14
Planning Department	1	3	1	3	3	11
Public Works Department		9	1		2	12
Transit Agency	2	21	1	1	7	32
Transportation Department	6	14	2	1	2	25
Grand Total	**11**	**56**	**5**	**7**	**15**	**94**

Table 5.11: Procurement & Use of Attitudinal or Market Research Survey by Agency Characteristics

Row Labels	Commission	Conduct in-house	Do not use	Do not use but plan to do so over the next five years	Use 3rd party survey	Would like to use the data, but these data are not available	Grand Total
Local (area) municipality	9	5	25	4	13	5	61
Other		1	4		3	2	10
Provincial	2		5				7
Regional municipality	3	2	6	1	2	2	16
(blank)	2	1	9		3	2	17
< 10,000				1			1
10,000-50,000	1	1	14	1	6	1	24
100,000-500,000	4		9	2	4	4	23
50,000-100,000	1	2	5		1	1	10
500,000-1,000,000	5	2	2	1	2		12
> 1,000,000	1	2	1		2	1	7
Alberta	2	4	4	1	1	1	13
British Columbia	1	1	1		1		4
Manitoba	1						1
New Brunswick			4			1	5
Newfoundland and Labrador			1				1
Northwest Territories					1		1
Nova Scotia	1		1			1	3
Ontario	5	2	15	2	10	5	39
Quebec	3	1	11	2	3	1	21
Saskatchewan			3		1		4
Yukon	1				1		2
Other	4		6		3	1	14
Planning Department	2		1	3	2	3	11
Public Works Department			8		3	1	12
Transit Agency	3	4	11	2	9	3	32
Transportation Department	5	4	14		1	1	25
Grand Total	**14**	**8**	**40**	**5**	**18**	**9**	**94**

5.4 COUNTS & INVENTORIES

Responding agencies indicated their use and procurement of various types of mode-specific counts and system inventories, as shown in Table 5.12. Road traffic counts are undertaken by many agencies, using both manual and automated methods to a similar extent. In contrast, transit ridership counts are performed more commonly using manual methods than automated ones. With varying degrees, other types of counts are performed less commonly than road traffic and transit ridership counts. In addition to counts, agencies keep inventories of transport system components and assets, such as road lanes, number of transit routes, transit fleets, parking lot capacities and charges, and bicycle and pedestrian facilities. As shown in Table 5.12, counts and inventories are undertaken more often in house than commissioned to a third party, with the exception of screen line and cordon counts.

Table 5.12: Procurement & Use of Counts and Inventories

	Commission	Conduct in-house	Do not use	Do not use, but plan to do so over next 5 years	Use 3rd party survey	Total
Road traffic counts, automated	16	31	36	0	10	93
Road traffic counts, manual	17	32	31	0	13	93
Transit ridership (boarding) counts, automated	3	37	39	7	7	93
Transit ridership (boarding) counts, manual	6	53	24	4	6	93
Screenline/cordon counts	14	12	55	1	11	93
Truck/freight/commercial vehicle counts	10	15	58	3	7	93
Parking counts	5	32	45	1	10	93
Bicycle counts	4	24	51	8	6	93
Pedestrian counts	6	26	50	4	7	93
Other counts	2	14	69	2	6	93
Roadway inventories (number of lanes, lane widths, etc.)	5	40	34	1	13	93
Transit service inventories (route, frequencies, etc.)	1	57	24	1	10	93
Parking inventories (number of spaces, parking charges, etc.)	4	33	40	1	15	93
Bicycle inventories (bikeways, etc.)	2	32	46	3	10	93
Sidewalk inventories (sidewalk widths, etc.)	2	42	36	2	11	93
Other inventories	1	11	78	0	3	93

5.5 TECHNOLOGY-BASED DATA COLLECTION

Table 5.13 reveals the extent to which different technologies are used by the participating agencies for transport data collection. Not surprisingly, GPS has found extensive use to gather transportation data, most likely on traffic and transit movements (e.g. corridor travel time studies, automated vehicle location systems). Many agencies have also reported their use of other forms of technology for data collection, such as remote sensing technologies, smart cards, transit pass technologies, debit/credit cards, social media, internet and mobile devices. According to the survey results, the near future will see a growing use of the listed technologies to collect passenger transport data.

Table 5.13: Use of Technology in Data Collection

	Use now	Plan to use in next 5 years	Do not use	Total
Global positioning systems (GPS)	57	16	18	91
Other distributed or remote sensing technologies	42	3	43	88
Smartcards	25	19	46	90
Other transit pass technologies	16	6	63	85
Debit/credit cards	23	11	52	86
Social media	29	22	35	86
Other internet	45	12	30	87
Mobile devices	27	24	38	89
Other technologies	9	5	63	77

A noteworthy finding in Table 5.13 is that 29 agencies (31% of respondents) are currently using some form of social media to collect information, while 22 (23% of respondents) indicate that they are planning to use social media within the next five years. Table 5.14 provides some additional information concerning these agencies in terms of the size of their urban areas and their geographic location within the country. As can be seen from this table, agencies using and planning on using social media are spread across the country and across city sizes. Unfortunately the survey does not provide information concerning the type of social media being used or the type of data that are collected.

Table 5.14: Distribution of Use of Social Media for Data Collection

(a) Agencies Currently Using Socia Media								
	Population							
Province	(blank)	< 10,000	10,000-50,000	50,000-100,000	100,000-500,000	500,000-1,000,000	> 1,000,000	Grand Total
Alberta	1	1	2	2				**6**
British Columbia					1		1	**2**
Manitoba						1		**1**
New Brunswick					2			**2**
Newfoundland and Labrador	1							**1**
Ontario			2	1	3	3	2	**11**
Quebec	1		1		2	1		**5**
Saskatchewan								**0**
Yukon			1					**1**
Grand Total	**3**	**1**	**6**	**3**	**8**	**5**	**3**	**29**

(b) Agencies Planning on Using Social Media in the Next Five Years								
	Population							
Province	(blank)	< 10,000	10,000-50,000	50,000-100,000	100,000-500,000	500,000-1,000,000	> 1,000,000	Grand Total
Alberta				1			1	**2**
British Columbia					1			**1**
Manitoba								**0**
New Brunswick								**0**
Newfoundland and Labrador								**0**
Ontario	1		3		4	1	1	**10**
Quebec	3		1		1	1	1	**7**
Saskatchewan			1					**1**
Yukon			1					**1**
Grand Total	**4**	**0**	**6**	**1**	**6**	**2**	**3**	**22**

The following specific applications of technology were identified by respondents in an open ended question for each form of technology (parentheses indicate the number of respondents identifying this application):

Global positioning systems (GPS)

Automatic Vehicle Location and fleet management (23) (sometimes used in conjunction with automated next stop announcement, customer notification, real time customer information, or automated passenger counters)

Asset management / data collection of infrastructure (roadways, pavement markings, road surface conditions, culverts, signs, parking, bus stops, properties, sidewalks, etc.) (11)

Travel time/speed surveys (10)

Surveys (2)

Engineering surveys and measurement (2),

Salt monitoring (2)

Real time information on network delays (2)

Ticketing system (1)

Identification of water sources (1)

Waste collection (1)

Other distributed or remote sensing technologies (e.g. satellite imagery, etc)

Technologies:

Aerial photography/ Orthophotos (9)

Satellite imagery (with GIS applications) (2)

Google maps (2)

Google Street views (1)

APC counting systems (1)

Applications:

Developing maps (for displaying route and bus location information) (5)

Salt monitoring (2)

Planning routes for new highways (1)

Identification of water sources (1)

Waste collection (1)

GIS application for asset management (1)

Identifying property lines (1)

Priority measures for buses, emergency vehicles (1)

Visualization of transportation equipment (1)

Transit services inventory (1)

Network delays (1)

Passengers counts (1)

Data retrieved from smartcards

Technologies:

Some specific brands were named, including Presto and BEA.

Applications:

Ridership (boarding data) data for transit planning (9)

Transit fares (3)

Revenue management (2)

Parking meter usage (1)

Movements of user groups (1)

Other transit pass data

Monthly sales data (4)

Ridership data (3)

Fare Revenue (3)

Transfer trace (2)

Boardings and alightings (1)

Debit/credit card data

Parking usage (2)

Transit fares, payments for passes (2)

Transaction data (2)

Ticket Payment (1)

Social media (Twitter, Facebook, etc.)

Technologies:

Twitter, Facebook, internal blogs, Web portals, Google Analytic

Applications:

Public responses, comments and feedback (3)

Customer service (2)

Public announcements (e.g. public meetings, workshop, planning/construction projects, road safety) (3)

Transit information (e.g. service changes) (2)

Travel survey (1)

Other internet

Online surveys (5)

Social media strategy (1)

Agency website (8)

Trip Planning (e.g. Google Transit, schedules, next bus) (4)

Blogs (1)

Notifications for service related announcements (e.g. disruptions) (3)

Interactive communication (1)

Mobile devices (e.g., smart phones, PDAs)

Real-time Passenger information (e.g. rider alerts, next bus) via text messaging of iPhone Apps (7)

Emails and other communications for staff (5)

Cell phone travel time /speed tracking (2)

Customer service (2)

OD survey (1)

Route choice studies (1)

Intercept surveys (1)

Enforcement (1)

Tablets for field inventories (1)

5.6 BARRIERS TO IMPROVED DATA COLLECTION

When asked about the barriers to the development of improved data collection and management methods, nearly two thirds of the responding agencies identified inadequate budget and too few staff, while more than half of the agencies attributed the lacking development of improved data collection methods to competing priorities (see Figure 5.16). The lack of perceived importance, lack of political support and limited technical capabilities of staff were also reported as barriers, albeit by a lower number of responding agencies. Union and contract rules do not seem to affect adversely the development of improved data collection and management methods.

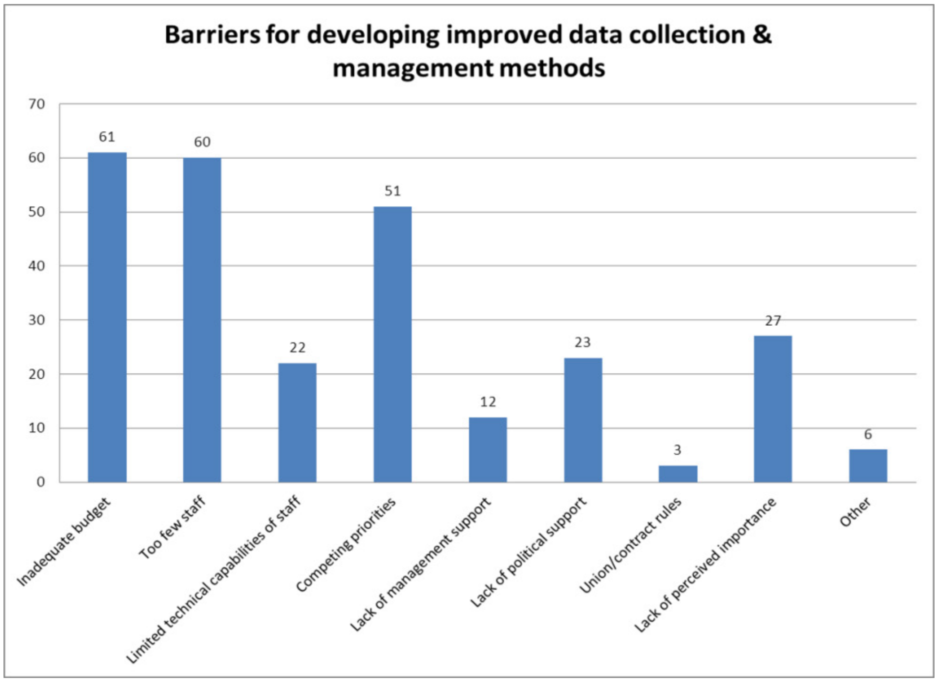


Figure 5.16: Barriers to Improved Data Collection

Few comments were provided by respondents regarding the barriers to development. They are provided as follows:

- There is a desire to collect household travel data more frequently than 10 years however costs are too high to allow for it.
- There is interest in doing research to determine if panel surveys can be conducted in between larger household travel surveys to update information and provide data needed for decision-making.
- Obtaining grants and grant processing time.
- We should increase our ability to process and analyze data

5.7 TRAVEL DEMAND MODELLING

A total of 46% of all respondents use travel demand models for planning and policy analysis. Figure 5.17 shows that a higher frequency of regional municipalities uses travel demand models for planning and policy analysis compared to provincial agencies and local municipalities. Figure 5.18 shows the frequency of agencies using models by the location of the agency's jurisdiction. Disregarding results from provinces with low response rate, more than half of the Ontario agencies participating in the survey reported their use of travel demand models, with Alberta coming second at 46% followed by Quebec at 38%. Figure 5.19 shows the level of use of travel demand models to increase with the population of the agency's jurisdiction. The percentage of agencies using travel demand models increases from a decent 50% in small jurisdictions with population between 50,000 and 100,000 to 100% in all large jurisdictions with population of one million residents or more. As expected, transportation and planning departments rely on travel demand models for planning and policy analysis than public works departments and transit agencies, as show in Figure 5.20.

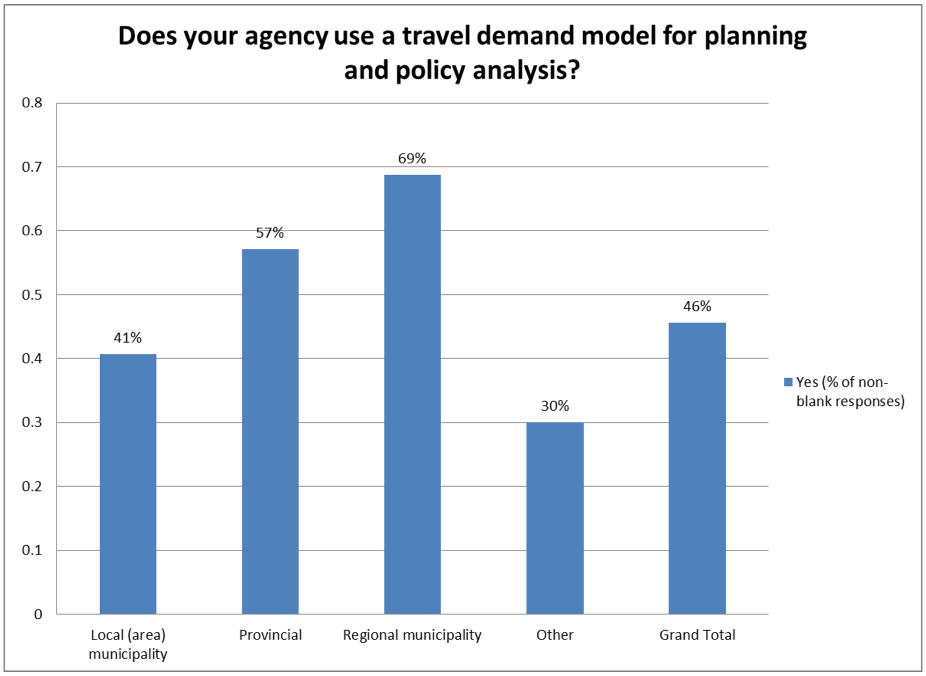


Figure 5.17: Use of Travel Demand Models by Level of Government

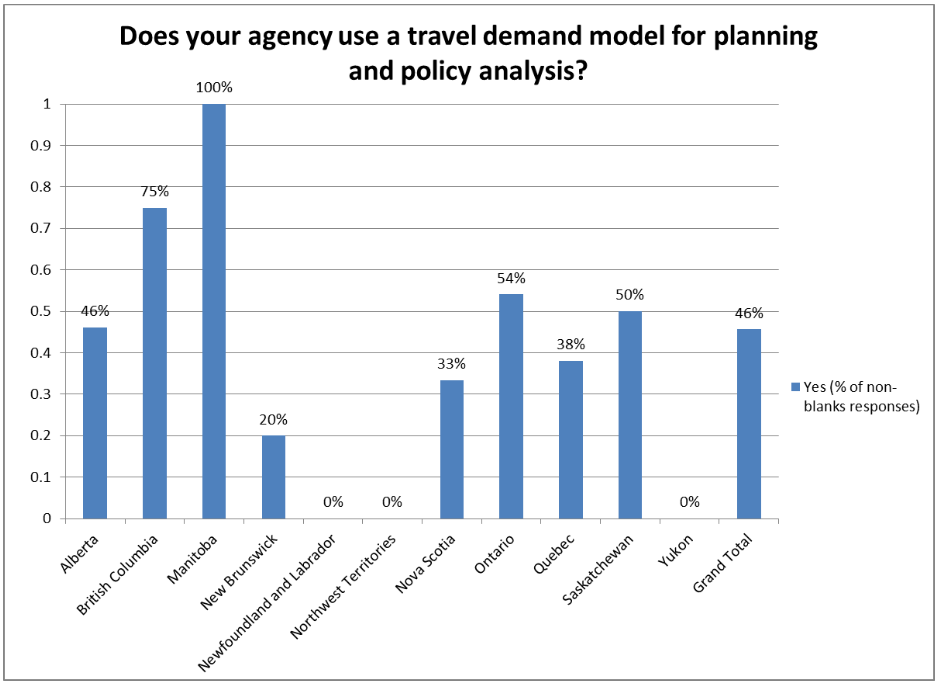


Figure 5.18: Use of Travel Demand Models by Location of Jurisdiction

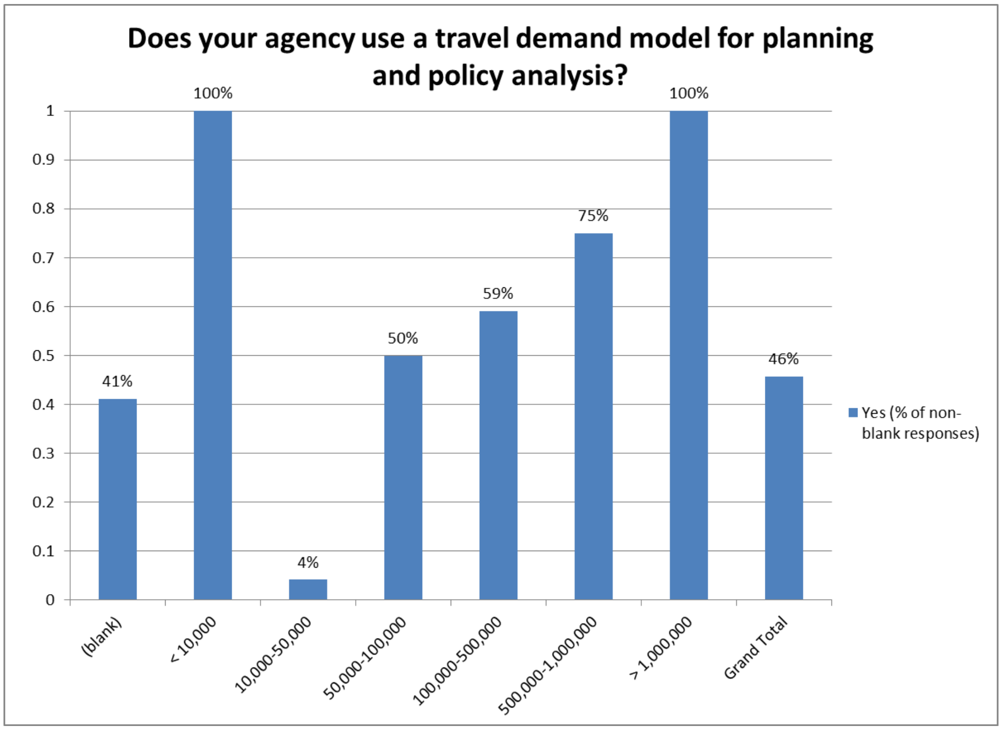


Figure 5.19: Use of Travel Demand Models by Population of Jurisdiction

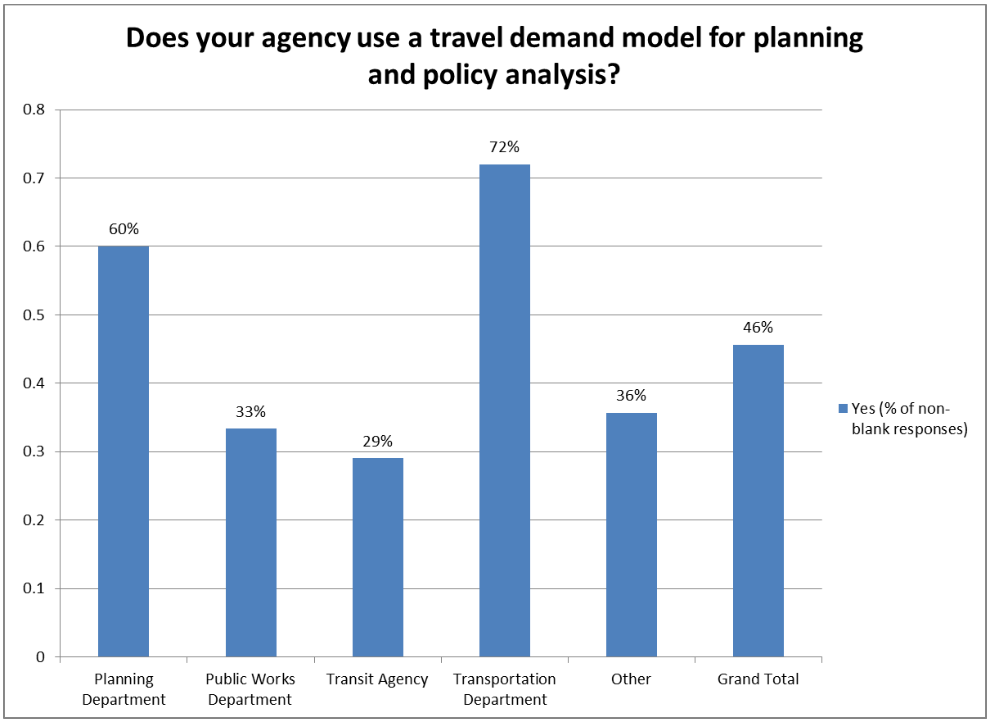


Figure 5.20: Use of Travel Demand Models by Type of Agency

Figure 5.21 shows that the responding agencies rely for the development of travel demand models on consultants, universities, or jointly by the two. However, almost one quarter of the responding agencies develop their own models in house relying on internal expertise.

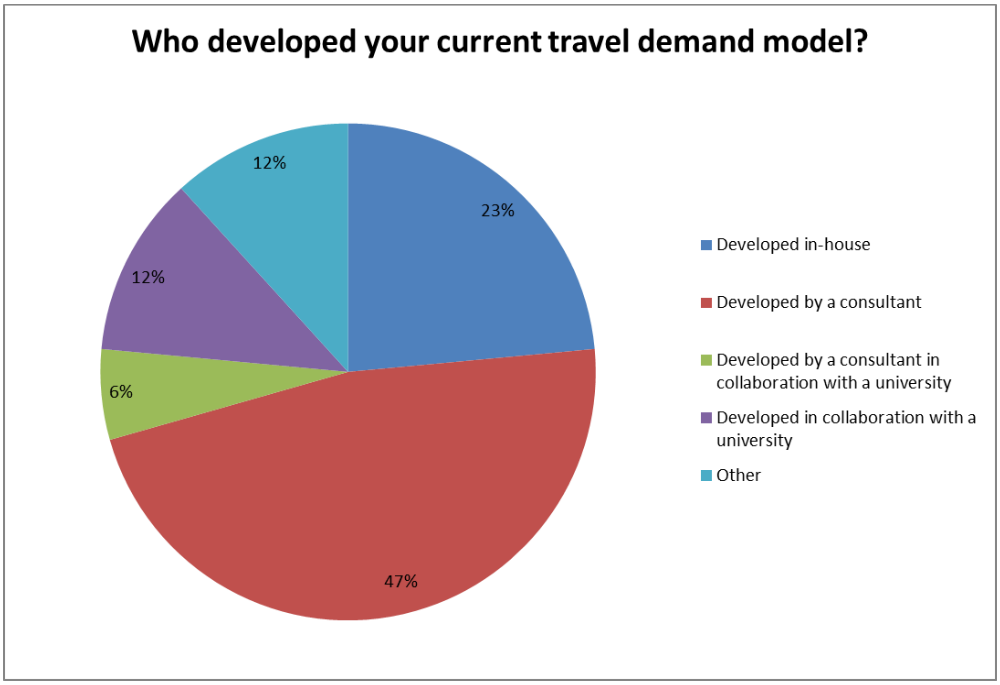


Figure 5.21: Developers of Travel Demand Models

The majority of participating agencies (80%) use a trip-based four-stage modelling system, either in a full or partial form (see Figure 5.22). A small percentage of agencies (3%) use emerging tour- or activity-based demand models, while 17% of the agencies use other forms of models. Figure 5.23 shows that the majority of models used by agencies are either for the AM peak or the PM peak, with fewer models developed for the mid-day or off peak periods.

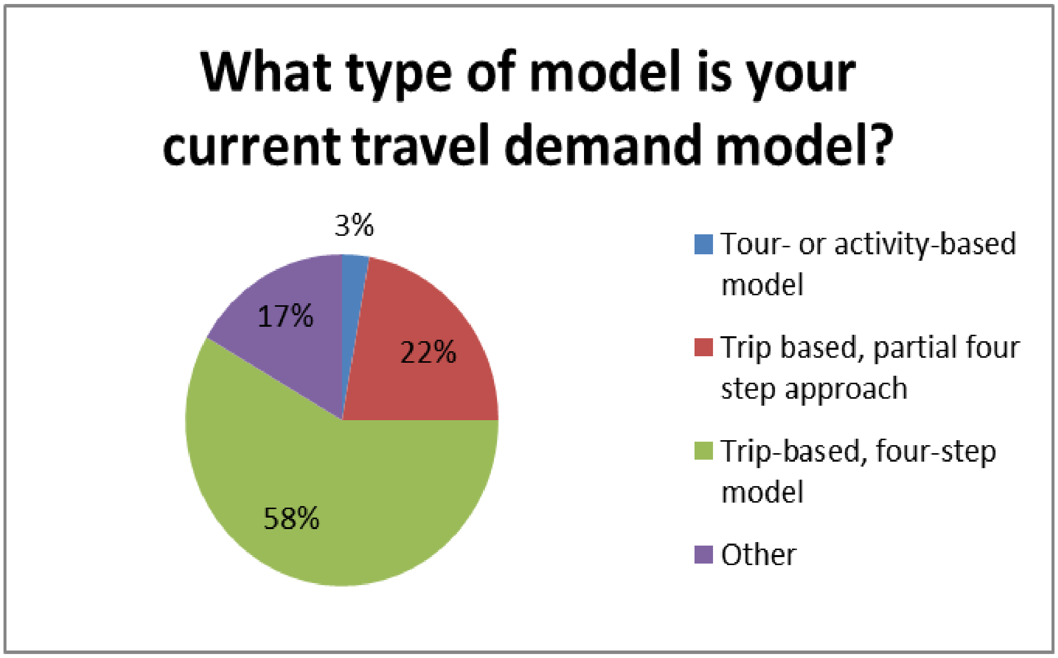


Figure 5.22: Type of Model

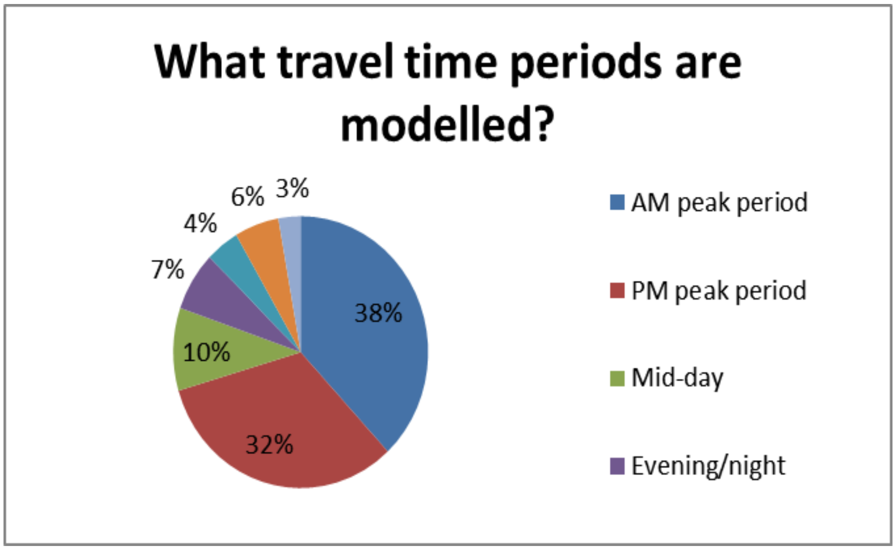


Figure 5.23: Time Period of Model

When asked about the year their model was last updated, most responding agencies have indicated this was done in the past five years and more notably in 2011 and 2012, perhaps corresponding with the last travel survey conducted (see Figure 5.24). While 43% of the responding agencies are happy with their current travel demand model, 37% indicated their plans to upgrade their models with additional 20% wishing to do the same but face constraints that will likely delay their plans (see Figure 5.25).

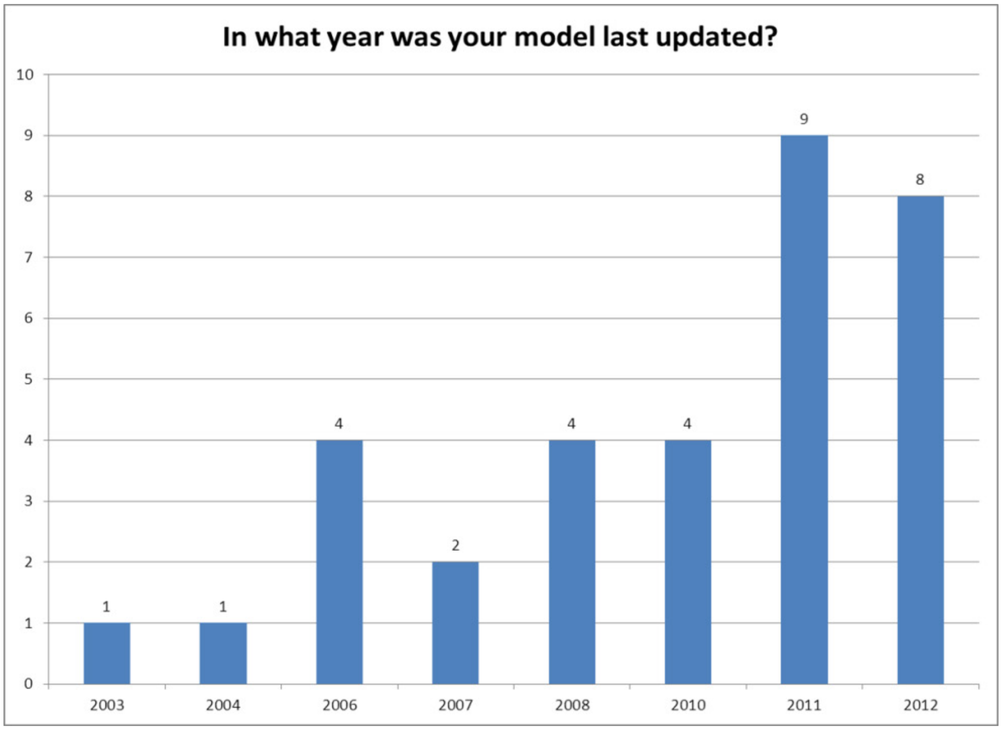


Figure 5.24: Year of Last Model Update

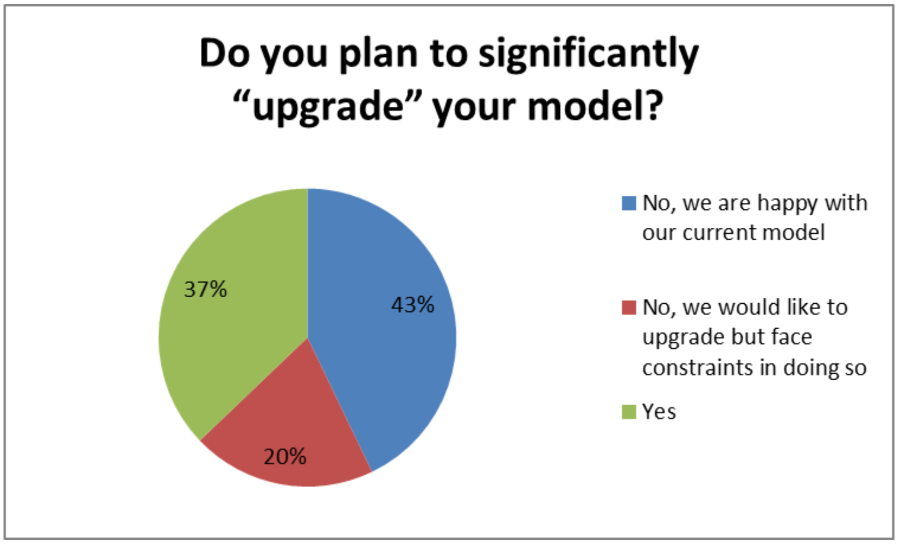


Figure 5.25: Plans to Upgrade Model

APPENDIX I – TRAVEL DEMAND MODELLING QUESTIONS

The respondent is asked whether the travel model is developed:

- In-house
- By consultants.
- By a consultant in collaboration with a university.
- Only in collaboration with a university.
- Other (with an open text box to write details about other arrangement)

The following question identifies the type of travel model. Types of travel model are classified into

- Trip-based four-stage.
- Trip-based partial four stage (e.g. traffic assignment only).
- Tour- or activity-based.
- Other types (with an open text box to write details about the type).

Then to identify data needs for model development, a question is asked to determine the level of socio-economic aggregation used in the agency's current model. Respondents are asked to identify the aggregation type from the following list:

- Total aggregation (no socio-economic categorization).
- Categorization of trip-makers with key socio-economic attributes (such as auto ownership, occupation, etc.).
- Microsimulation of individual agents.
- Others (with an open text box for explaining the other category)

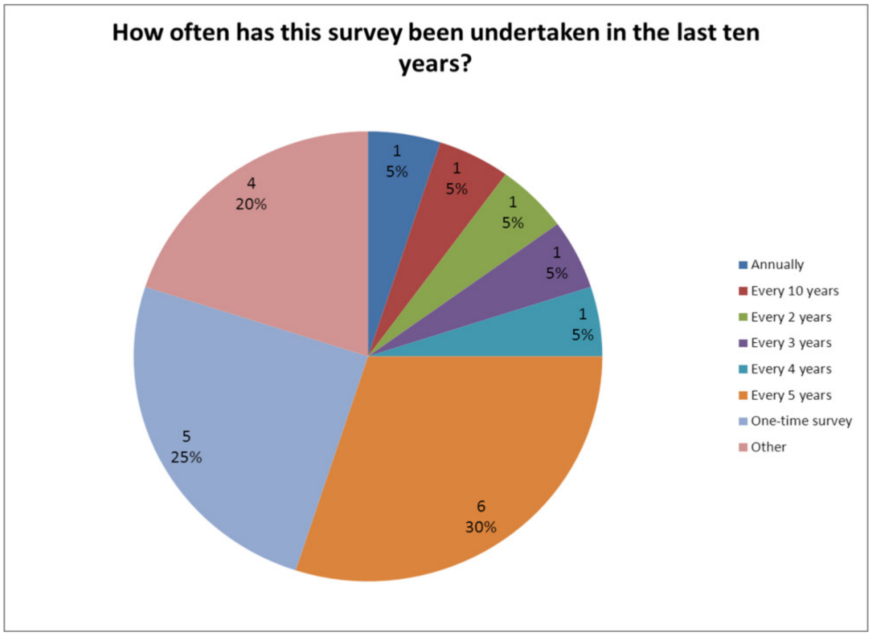
Further detail about the demand model is probed by asking for the travel time periods modelled. Travel time period is classified into the following categories:

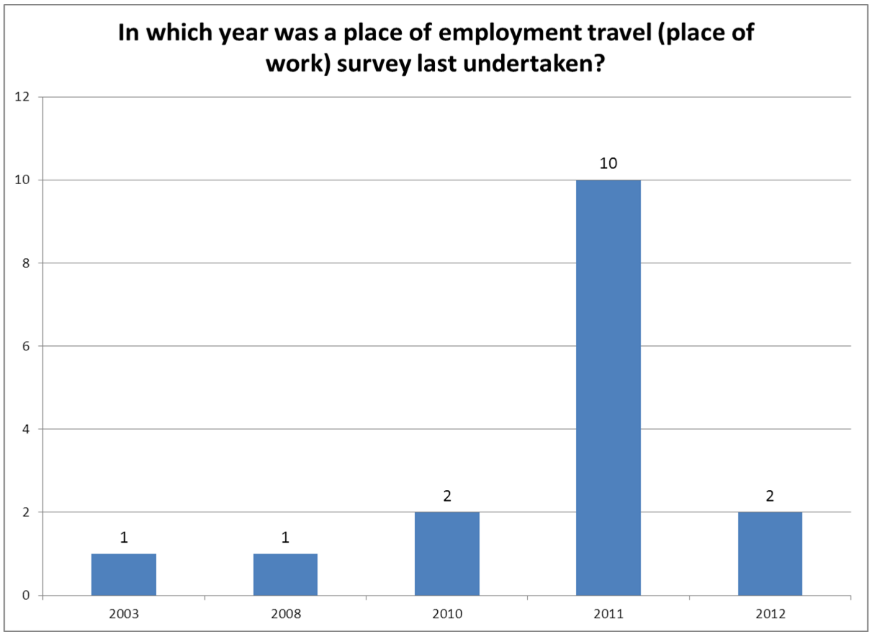
- AM peak period.
- PM peak period.
- Mid-day.
- Evening/night.
- All day (no individual time period).
- All day, continuous time representation.
- Other (with an open text box to write down in details).

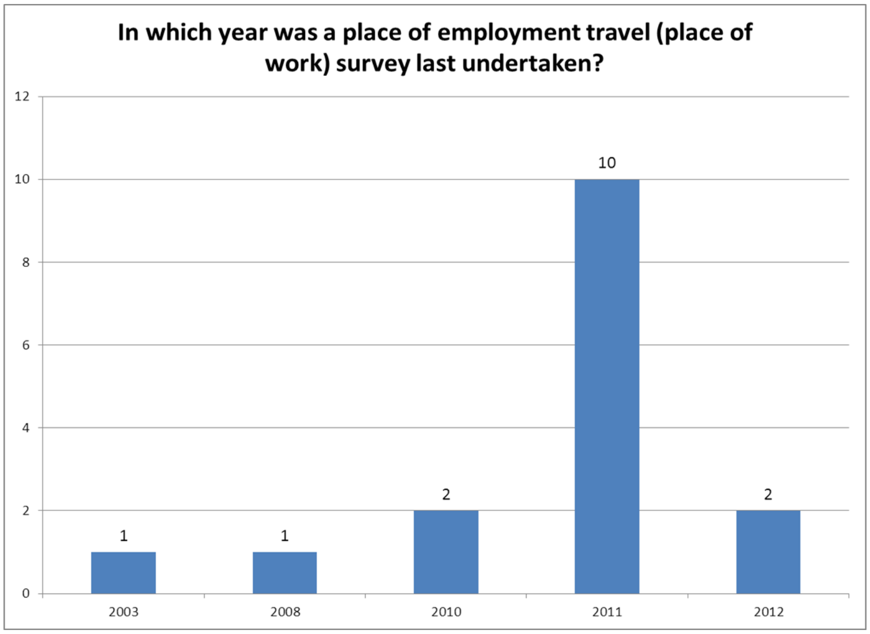
The next question asks about the frequency of updating the model system and the year when the model system was last updated. The following question asks about any plan to significantly change/upgrade the model. If the responding agency does have a plan, they are asked to write down a brief description of the plan in an open box. If the responding agency does not have any such plan, the request is made to write down details of key barriers/constraints for undertaking model upgrades. At the end of this section, an open text box is provided to the respondent for writing any other comments they wish to make regarding travel modelling.

APPENDIX II – SUMMARY RESULTS FOR OTHER TYPES OF TRAVEL SURVEYS

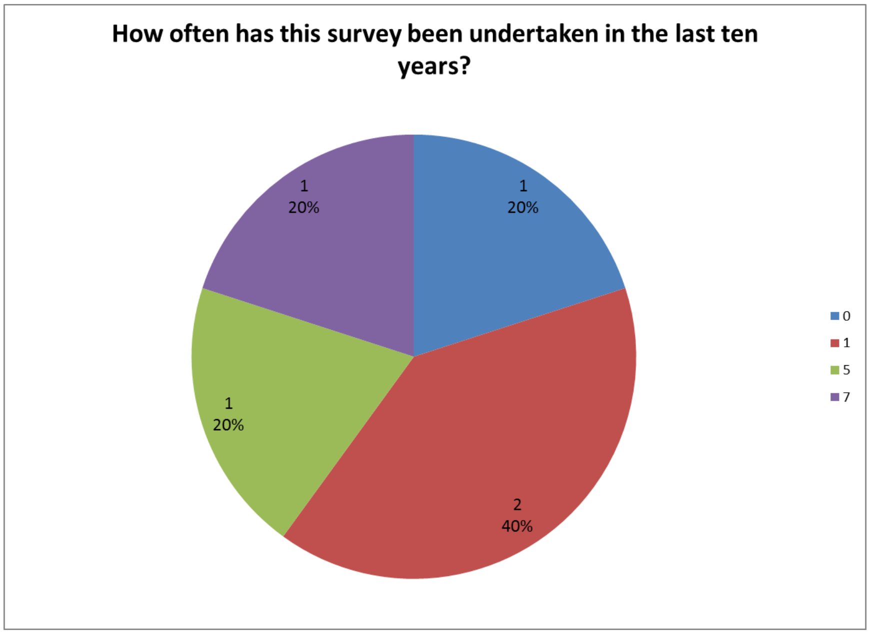
PLACE OF EMPLOYMENT SURVEY

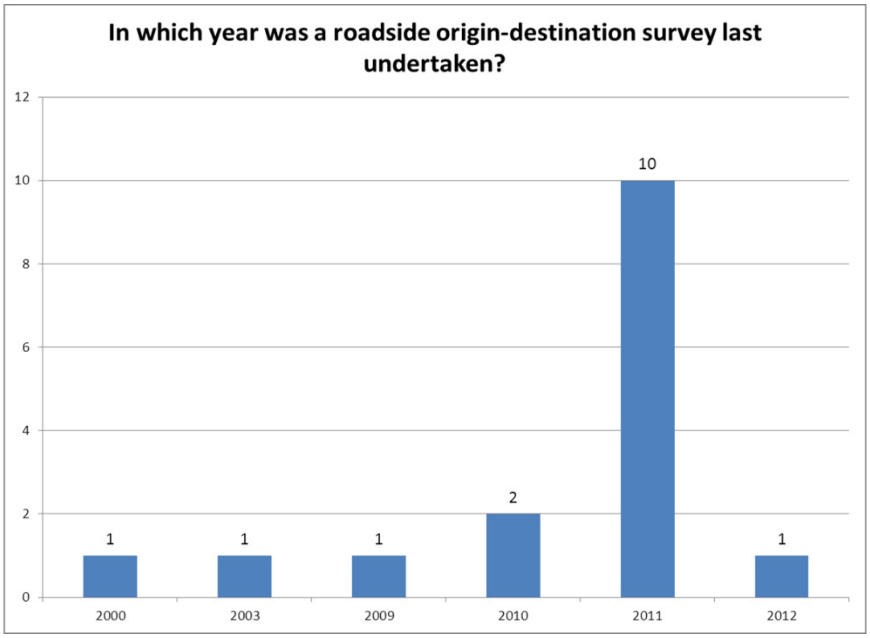


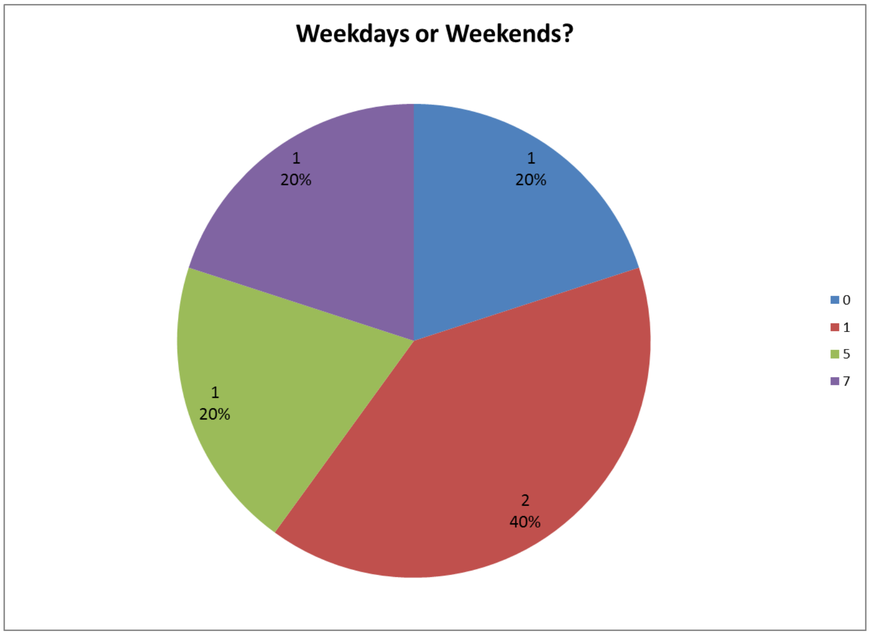




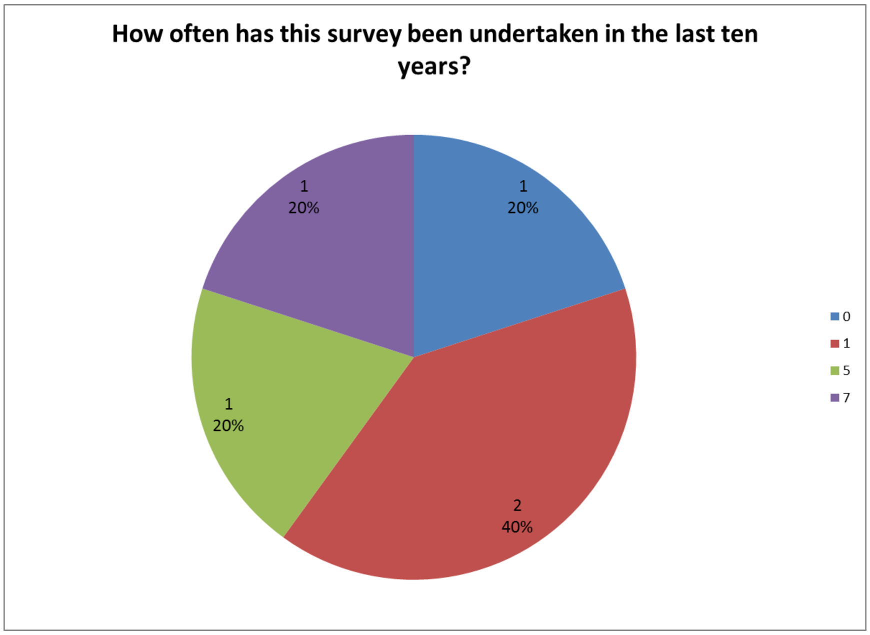
ROADSIDE ORIGIN-DESTINATION SURVEY

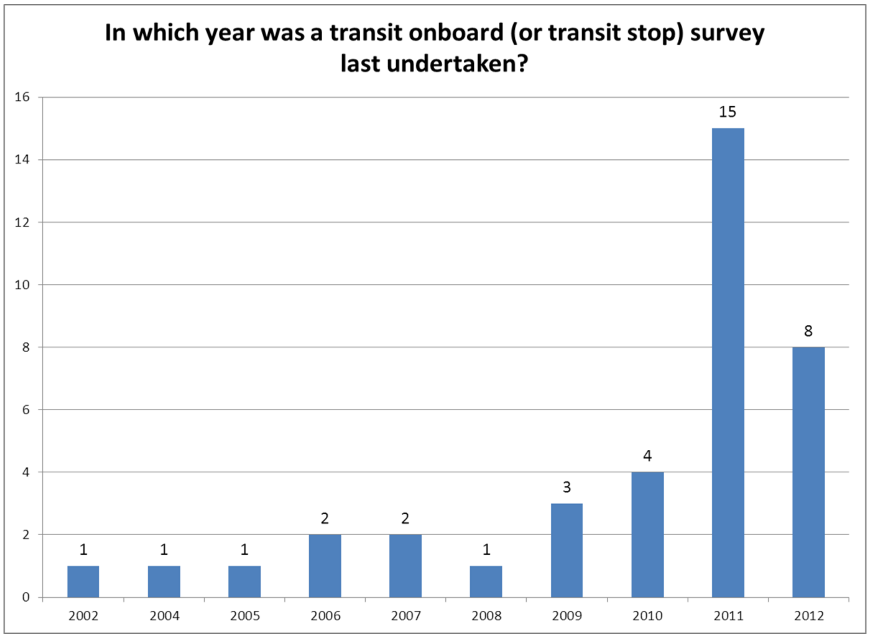


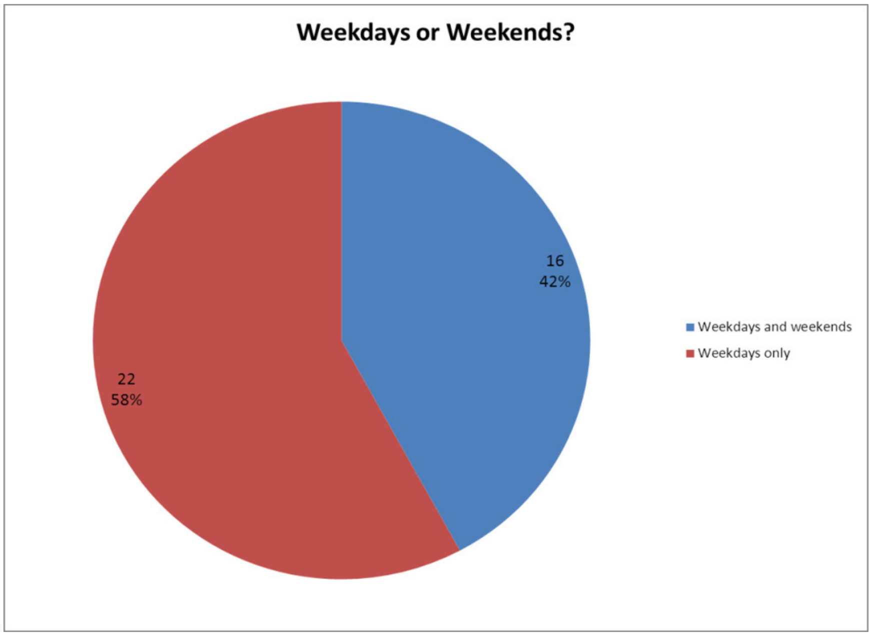

In which year was a roadside origin-destination survey last undertaken?
12
10
8
6
4
2
0
1
1
1
2
10
1
2000
2003
2009
2010
2011
2012


Weekdays or Weekends?
1
20%
1
20%
1
20%
2
40%
0
1
5
7

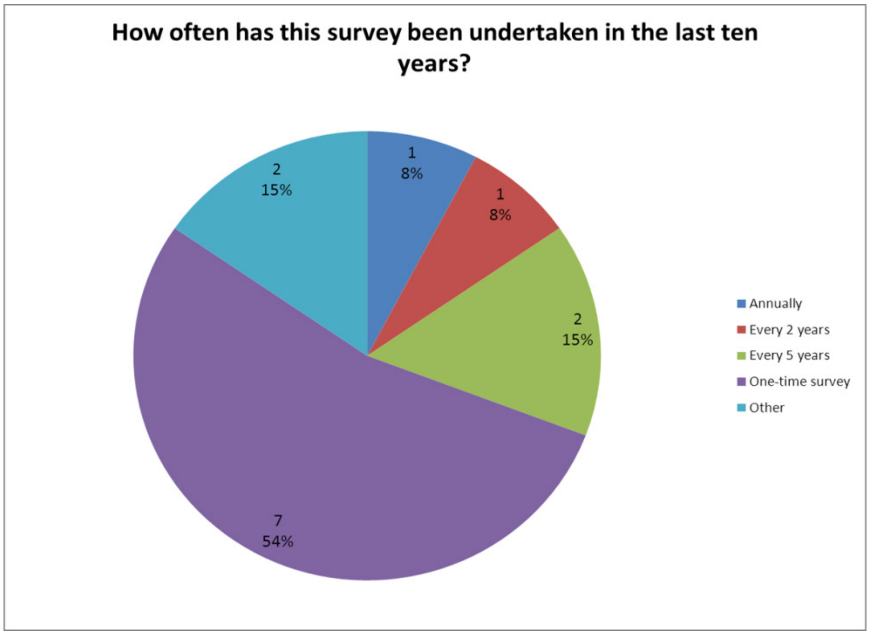
TRANSIT ONBOARD/STOP SURVEY

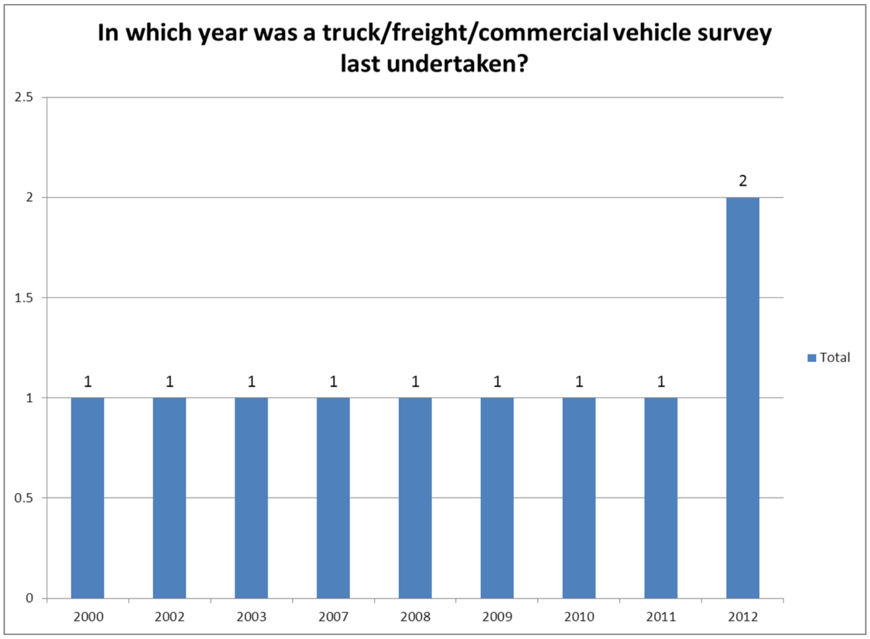


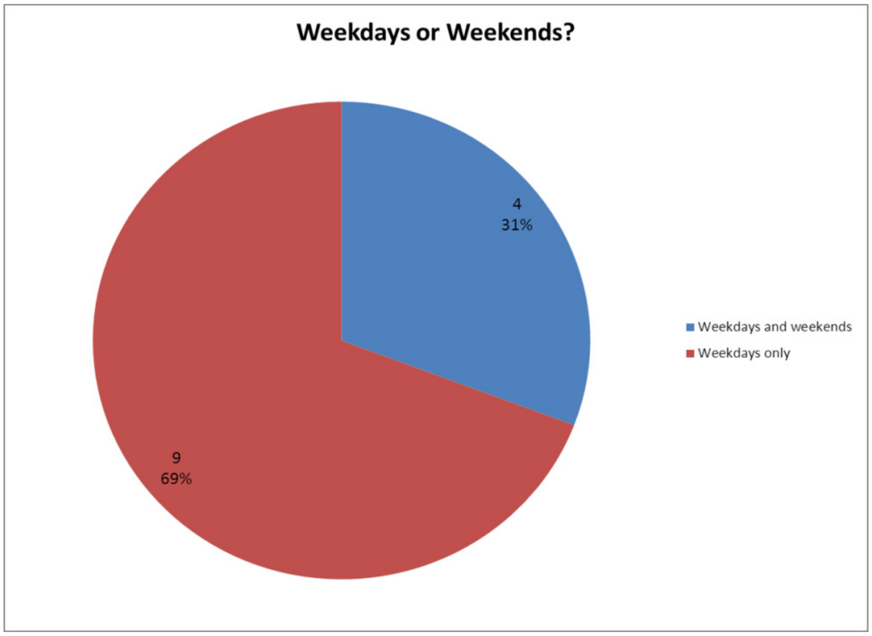




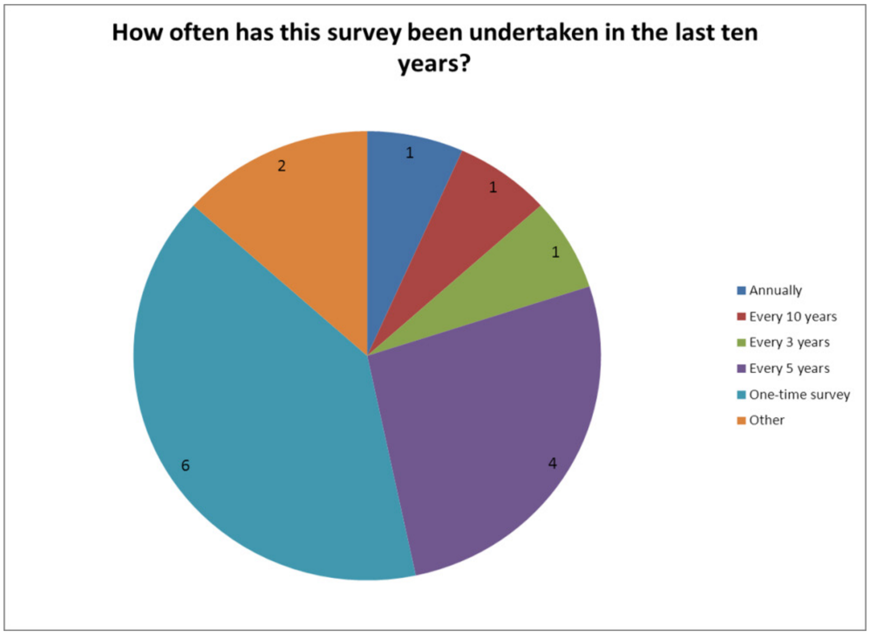
TRUCK/FREIGHT/COMMERCIAL VEHICLE SURVEY

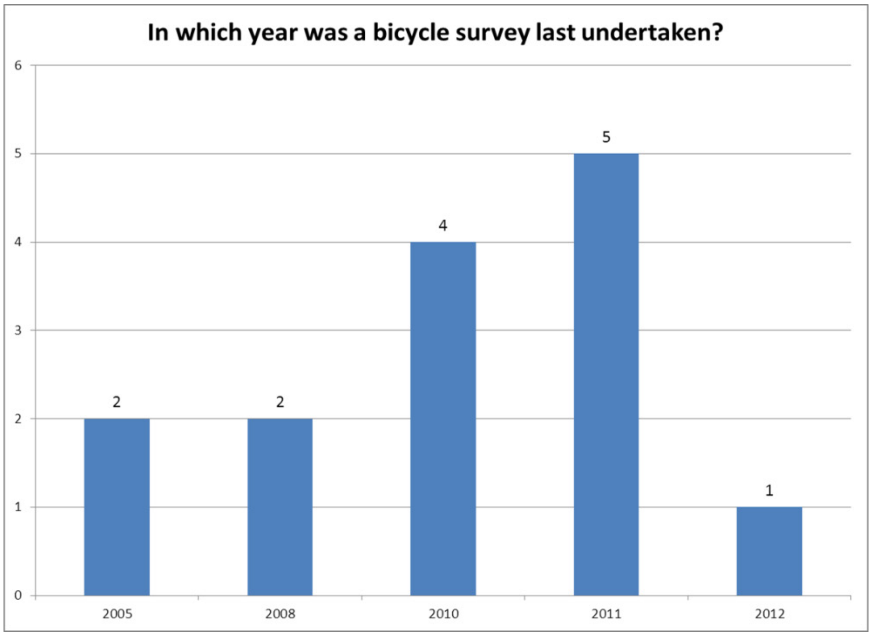


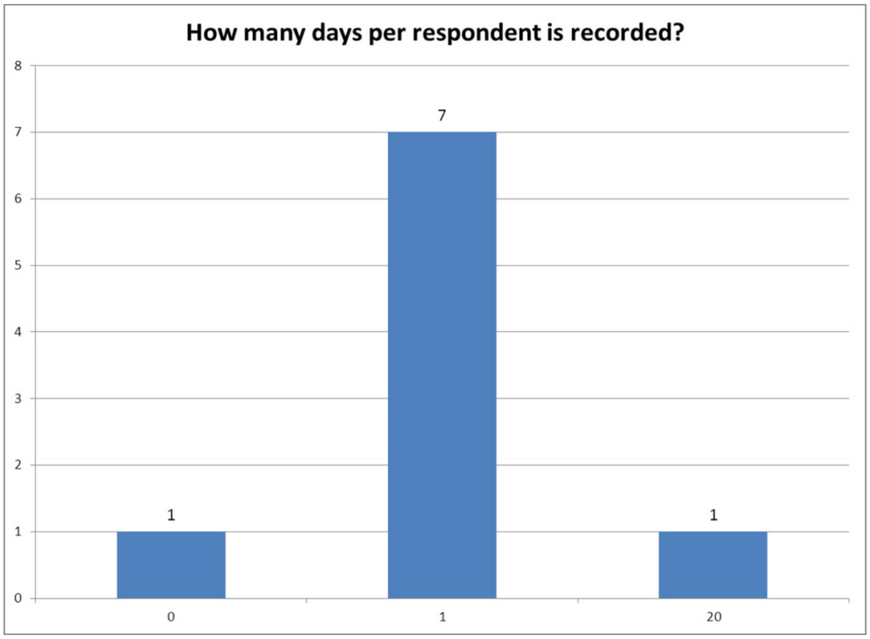

In which year was a truck/freight/commercial vehicle survey last undertaken?
2.5
2
1.5
1
0.5
0
1
1
1
1
1
1
1
1
2
2000
2002
2003
2007
2008
2009
2010
2011
2012
Total

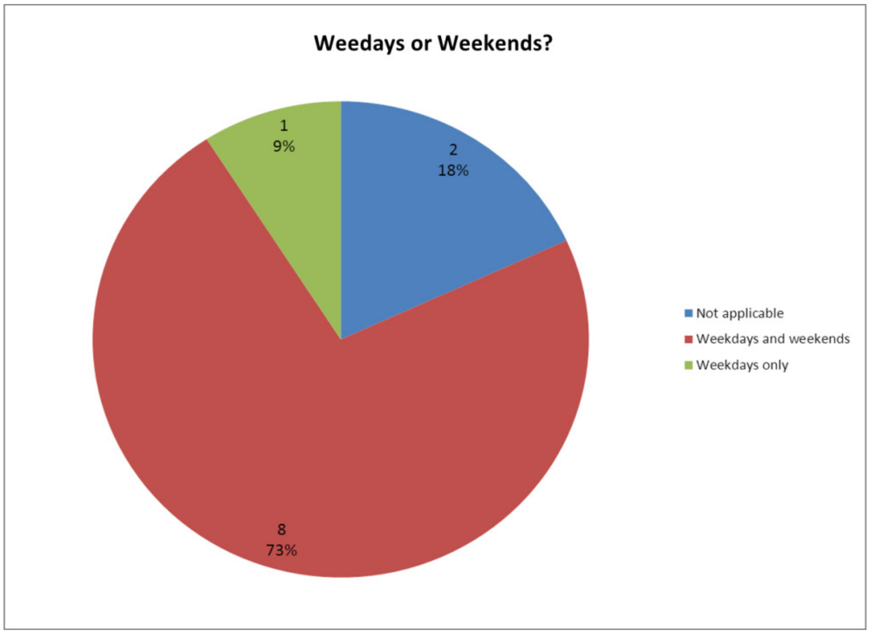

Weekdays or Weekends?
4
31%
9
69%
Weekdays and weekends
Weekdays only

BICYCLE SURVEY

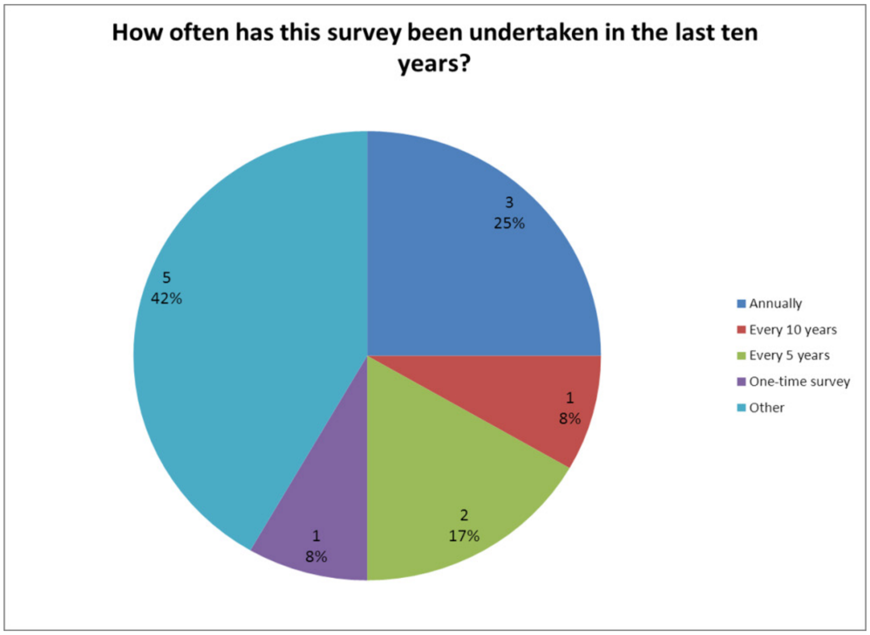


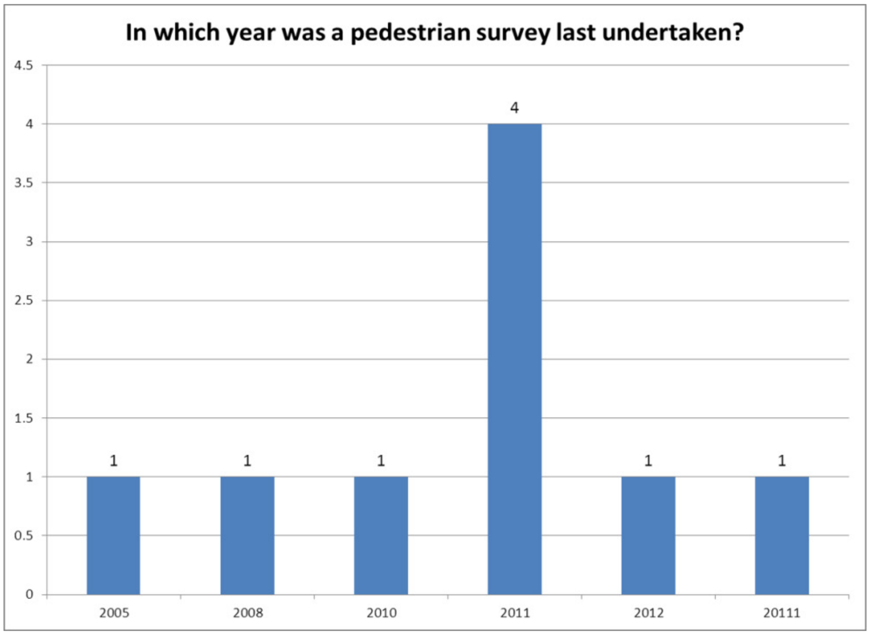


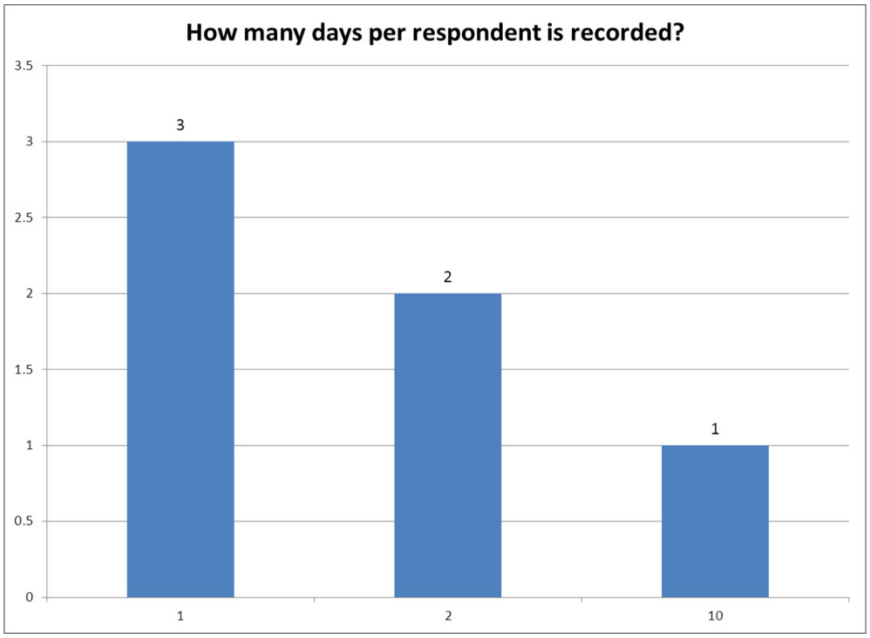

How many days per respondent is recorded?
8
7
6
5
4
3
2
1
0
1
7
1
0
1
20

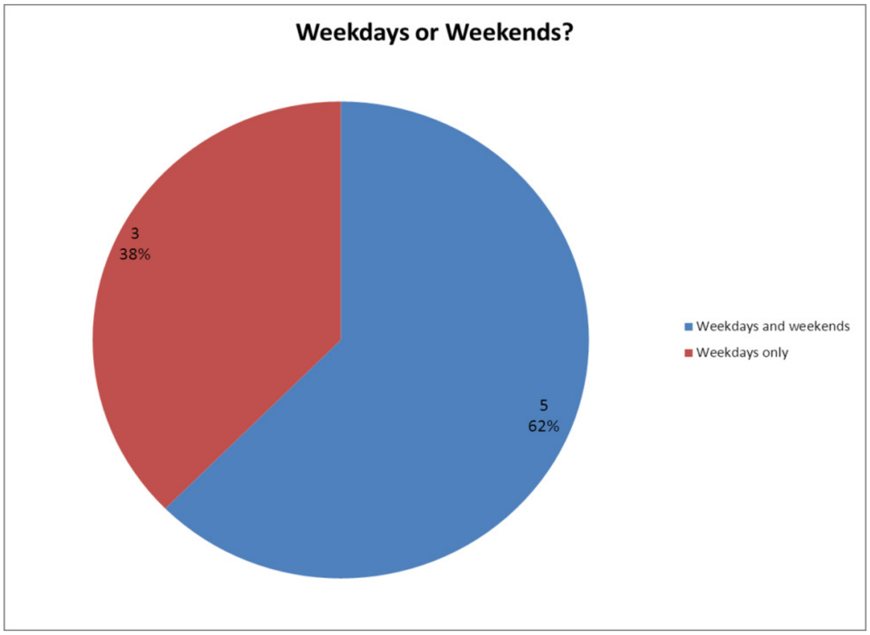

Weedays or Weekends?
1
9%
2
18%
8
73%
Not applicable
Weekdays and weekends
Weekdays only

PEDESTRIAN SURVEY

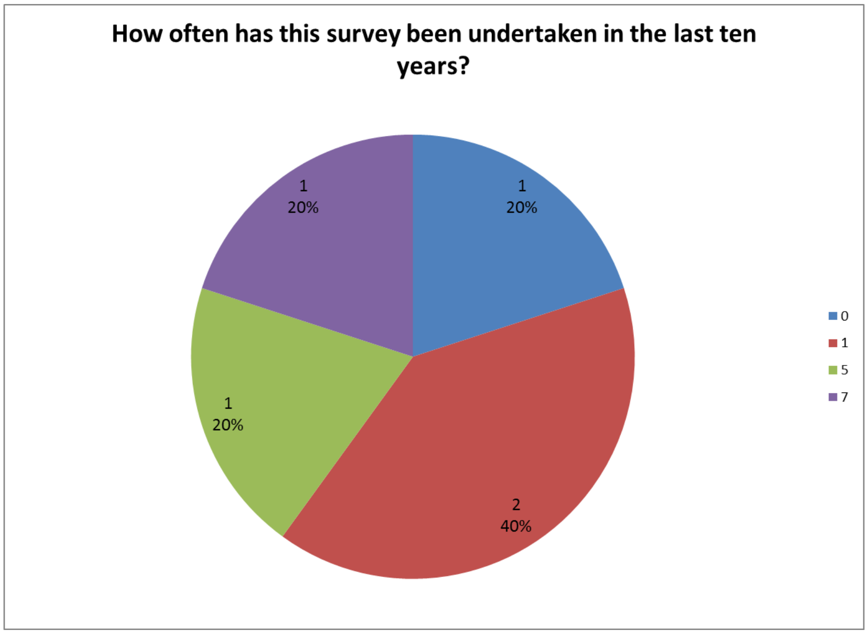


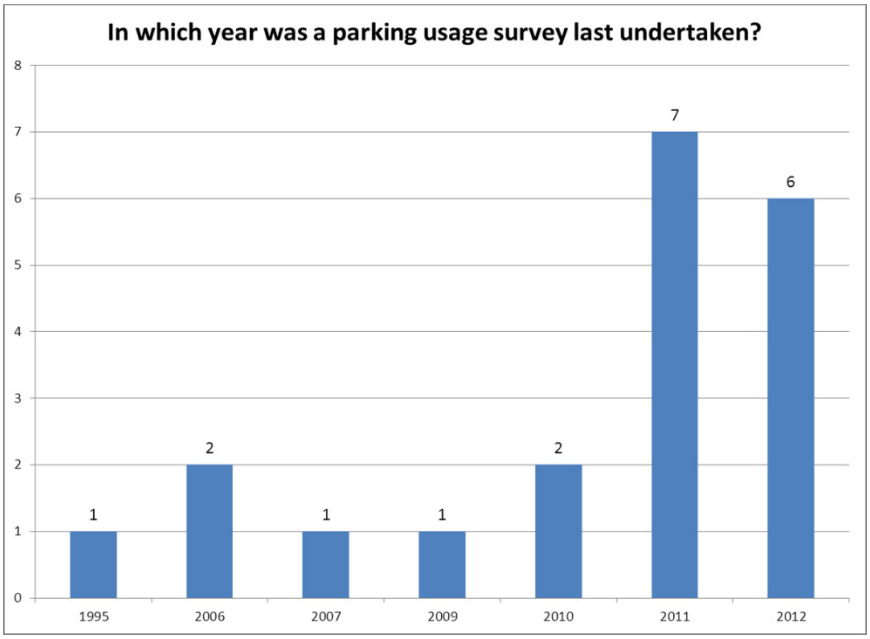


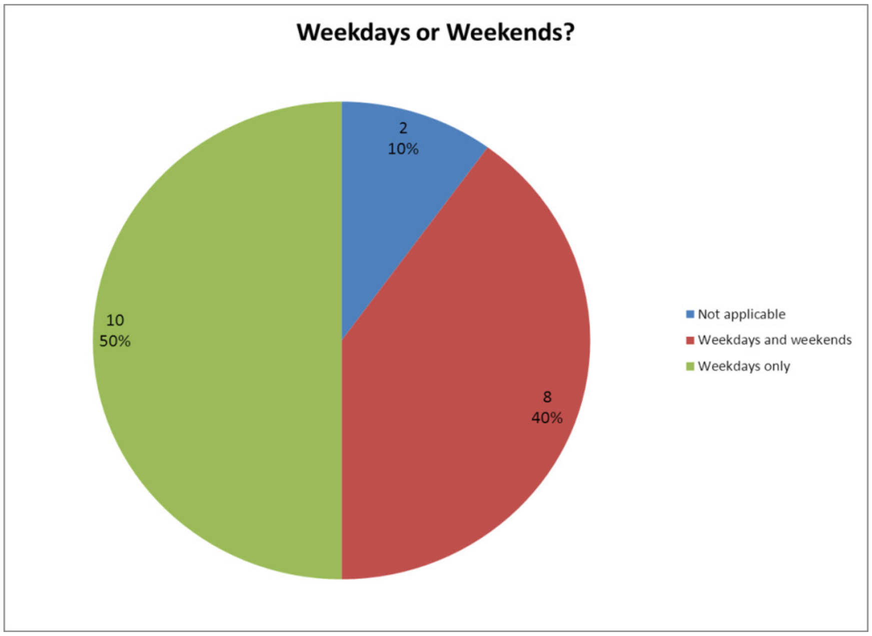

How many days per respondent is recorded?
3.5
3
2.5
2
1.5
1
0.5
0
3
2
1
1
2
10


Weekdays or Weekends?
3
38%
5
62%
Weekdays and weekends
Weekdays only

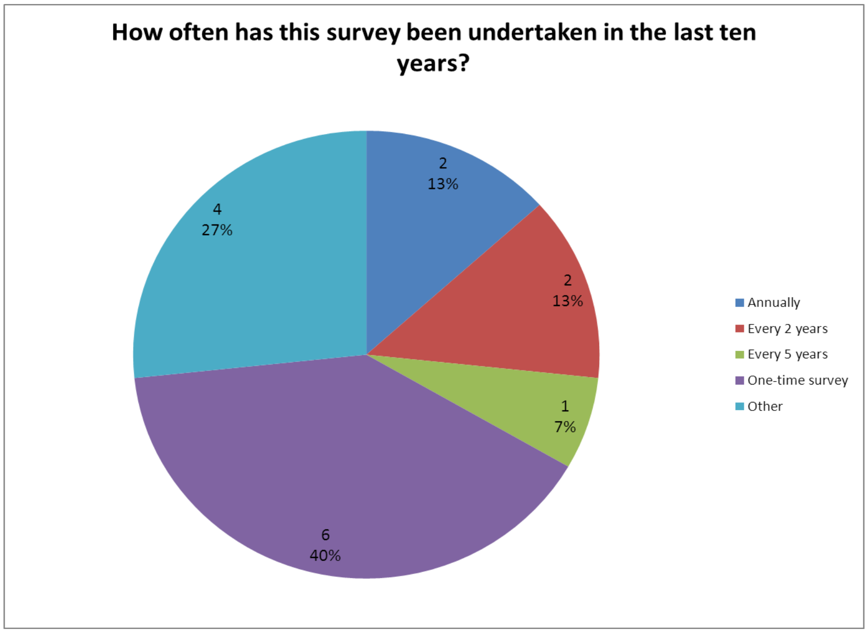
PARKING USAGE SURVEY

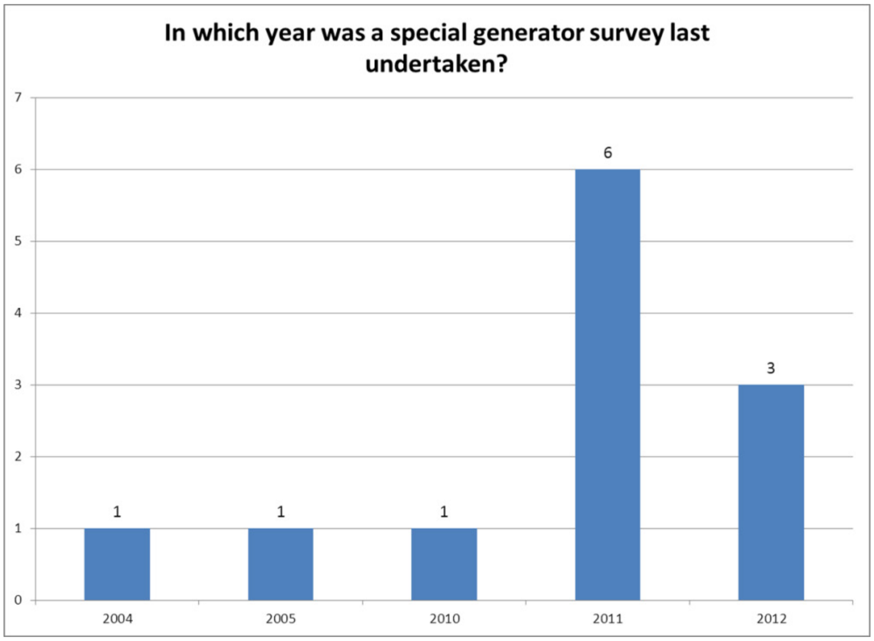


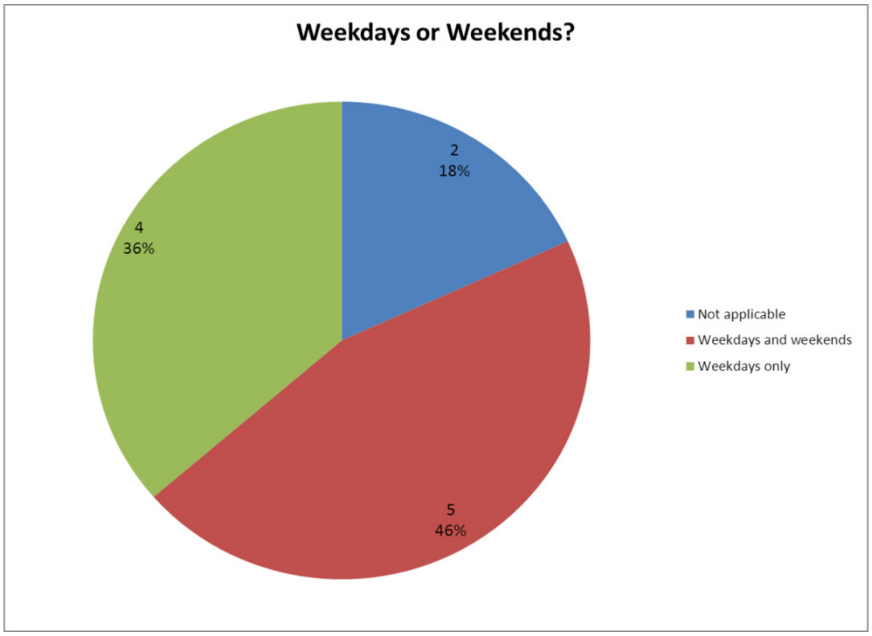




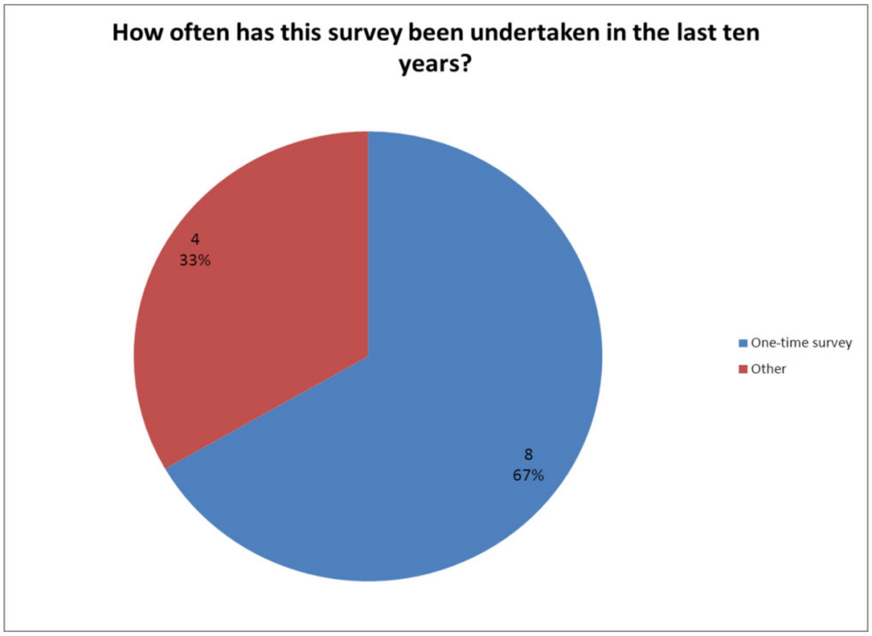
SPECIAL TRAFFIC GENERATOR SURVEY

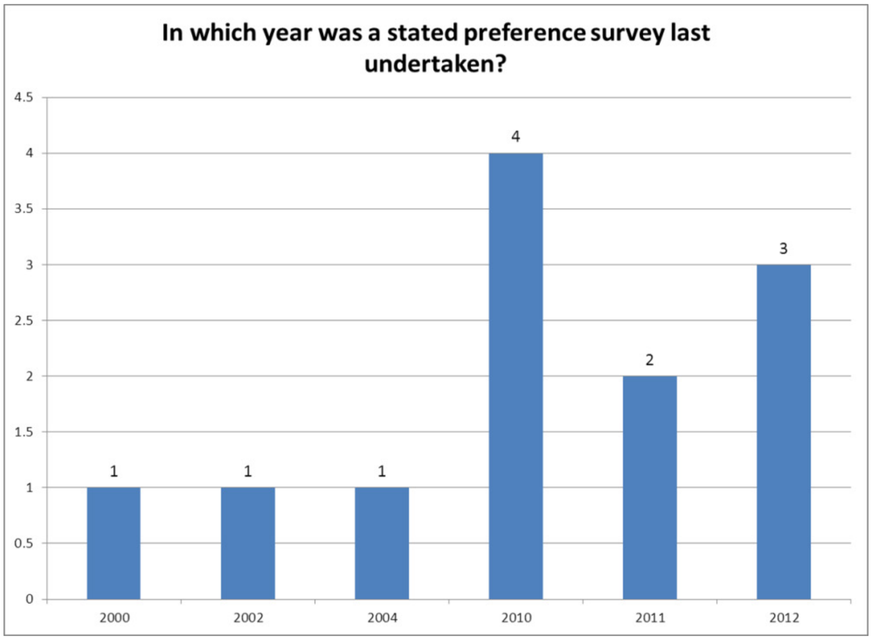



In which year was a special generator survey last undertaken?
0
1
2
3
4
5
6
7
1
1
1
6
3
2004
2005
2010
2011
2012

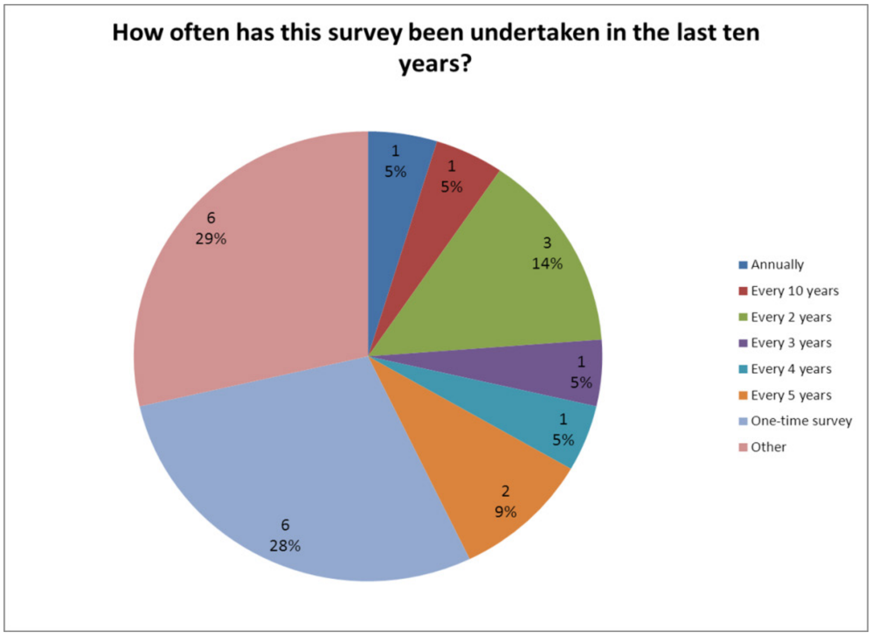

Weekdays or Weekends?
2
18%
5
46%
4
36%
Not applicable
Weekdays and weekends
Weekdays only

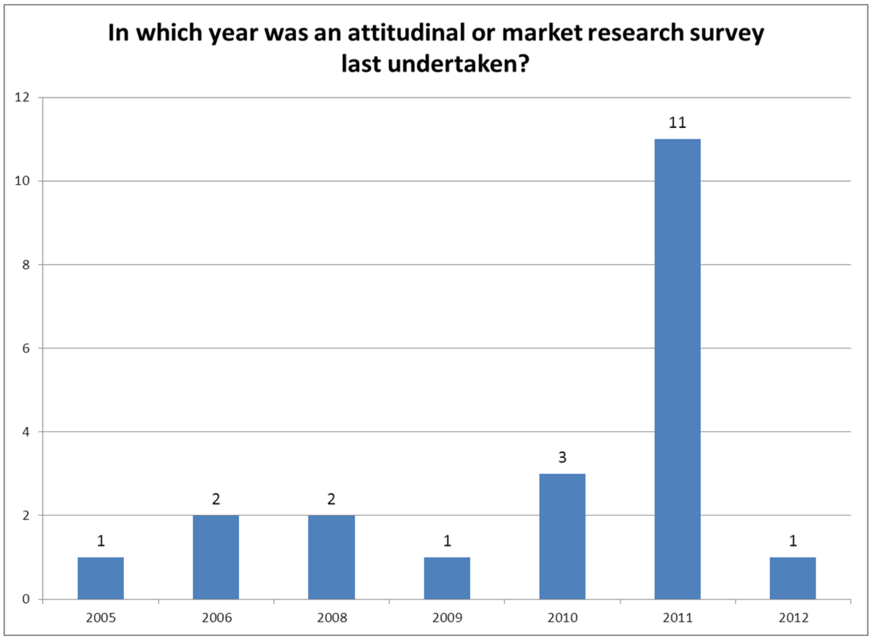
STATED PREFERENCE SURVEY





ATTITUDINAL OR MARKET RESEARCH SURVEY





PART VI – A FRAMEWORK FOR URBAN PASSENGER TRANSPORTATION DATA COLLECTION & MANAGEMENT

1. INTRODUCTION

The objective of the *Changing Practices in Data Collection on the Movement of People* project is to prepare a practical, comprehensive framework for the coordination, collection, processing and management of data on the movement of people by all modes in Canadian urban areas that is implementable and addresses data needs across the range of Canadian transportation agencies. This final part of the project report presents such a framework.

This framework is based upon several “building blocks” that have been presented in the previous parts of the project report:

- An extensive review of the survey and data collection literature that systematically identified current and emergency data collection issues, needs and methods. This review is documented in *Part II; Literature Review*.
- A review of data integration/fusion/synthesis methods that can be used to combine data from diverse sources together, thereby expanding the usefulness of available data. This review is documented in *Part III; Data Integration/Fusion Methods*.
- An inventory and discussion of a variety of sources of data pertaining to urban passenger travel (census, etc.) which Canadian transportation agencies might utilize, thereby reducing the need for original data collection efforts. This inventory is presented in *Part IV: Survey of Data Sources for Urban Transportation Applications*.
- Results from an original survey of Canadian transportation agencies undertaken by the project team to identify current Canadian data collection practice, issues and needs. The survey and its results are documented in *Part VI: Design, Conduct and Results of a Survey of Canadian Transportation Data Collection Practice*.

This part of the report is organized as follows:

- Section 2 presents the case for urban passenger data collection.
- Section 3 develops a taxonomy of urban passenger data needs.
- Section 4 provides an overview of key data collection methods.
- Section 5 discusses trends, issues and options in meeting current and emerging data needs.
- Section 6 discusses issues in data management (including privacy issues).
- Section 7 introduces a new paradigm for coordinated, cost-effective survey design (the “core-satellite” paradigm).
- Section 8 presents the recommended framework for urban transportation data collection.

2. THE CASE FOR URBAN PASSENGER DATA COLLECTION

2.1 INTRODUCTION

As is described in greater detail in Section 3 of this report, a very large amount of information of a wide variety of types is required for efficient, effective planning, design and operations of an urban transportation system. As in any other enterprise, failure to monitor and understand the characteristics and behaviour of the transportation system will lead to inefficient and unproductive designs and operations. Thus, the case for high quality data collection to support urban transportation planning and operations should be self-evident.

Despite this, data collection efforts that support urban transportation planning and operations are often under-funded. They are also increasingly in jeopardy as governments at all levels seek to “cut costs”, as well as seemingly be less interested in evidence-based planning and decision-making. The cancellation of the mandatory Census long-form is a prime example of this trend,[54] but other local and regional examples also exist of this trend towards reductions in data collection efforts and/or less usage of data and data-based models in decision-making.

Investment in adequate data collection, management and analysis “infrastructure” should be thought of in exactly these terms: an investment (as opposed to a cost) that pays significant dividends in terms of improved planning, decision-making and operational system management/control. Data collection should be an integral part of the business of transportation infrastructure and service planning, investment and operations, as, indeed, it is in any success private sector operation. That is, as is discussed in greater detail in the next sub-section, a “business case” clearly exists for urban transportation data collection that should be recognized and acted upon by all agencies concerned with urban transportation.

2.2 THE BUSINESS CASE FOR URBAN PASSENGER DATA COLLECTION

The business case for transportation data collection, management and dissemination relies on the fact that the benefits associated with a better planned, monitored and regulated transportation system outweigh the cost of data collection. The investment that Canadians make in the transportation system is enormous. Transport Canada (2010) notes the following transportation-related government expenditures in the fiscal year 2009-2010. All levels of government combined spent $39.5 billion, of which $28.9 billion was spent on roads, $5.8 billion on public transit, $1.8 billion on marine, $1.2 billion on air, $435 million on rail and $1.4 billion on multi-modal and other expenses. $33.9 billion was spent by provincial, territorial and local governments, in comparison to the $5.6 billion spent by the federal

---

[54] Although in the case of cancelling the Census long-form, cost-cutting was not the driving force but rather concerns about privacy intrusion.

government. Given such a massive level of investment in on the transportation system, it is imperative to make informed decisions based on sound knowledge that lead to the greatest social benefits.

Going beyond these aggregate numbers it is useful to compare the magnitude of costs of some key pieces of infrastructure to typical transportation data collection costs. While the cost of highway infrastructure varies across Canada, the following unit costs have been estimated (Applied Research Associates, 2008) for initial construction and maintenance of major highway infrastructure in Canada:

- Initial pavement construction for a provincial rural arterial highway in Southern Ontario is approximately $380,000 per km of two lane equivalent road.
- In Canadian provinces, average initial cost of bridge construction ranges from $1800 (British Columbia) to $4400 (Nova Scotia) per m^2 of the bridge deck area. An average size bridge ($1000m^2$) could cost $1.8 to $4.4 million.
- Annual cost of routine maintenance operations (from provincial estimates) range from $1,045 to $8,700 per km of 2 lane equivalent road.

Similarly, recent estimates of the cost of building high order exclusive right-of-way rail lines in Canada are as follows (Toronto Transit Infrastructure, 2012):

- Vancouver, Canada Line (Constructed 2005-2009) $105M/km.
- Toronto, Sheppard (Constructed 1994-2002) $170M/km.
- Toronto, Spadina (Under construction 2009-2015) $306M/km.
- Metrolinx estimates of the Sheppard Extension $177M/km.
- Laval extension of the Orange Line in Montreal (Constructed 2003-2007) $143M/km.

At the same time, the cost of maintaining Canadian transit systems in a state of good repair is equally daunting. In Montreal, for example, the Reno-System program to renovate the fixed equipment of the Montreal subway (control centre, escalators, video surveillance system, lanes, electric system, etc.) amounted to 950 million dollars during the 2004-2012 time period, with an additional 500 million dollars budgeted to complete Phase 3 of the project.

Significant costs also arise from the inability of transportation infrastructure supply to meet the demands of travellers. Congestion results in costs to society in many forms including recurrent peak period delay to drivers, increased fuel consumption and higher greenhouse gas emissions. Transport Canada (2006) has estimated the annual cost of congestion from these sources in the peak periods in major urban areas of Canada alone to be in the range from $2.3 to $3.7 billion/year. This is a conservative estimate that does not include costs of delay to freight transportation, off-peak delay, accidents, weather events, or the costs associated with other externalities like noise and vibration. Based on the 2003 Origin-Destination Montreal Survey, the MTQ estimated the cost of congestion at around 1.4 billon dollars in the Montreal region, from which 87.5% was due to travel delays (MTQ, 2009).

These very large infrastructure capital and operating costs and equally large user costs can be compared with the typical cost of undertaking a major household travel survey in Canada. In 2006 the total budget of the Transportation Tomorrow Survey (TTS) was $3.09M, inclusive of all aspects of the survey, including data collection, data management and reporting of the results (DMG, 2010). This survey involved approximately 150,000 interviews, leading to a unit cost of $20.38 per completed interview. The cost was shared by the MTO, GO Transit and the regions over which the TTS was conducted, and provided a 5% sample of 2.9 million households representing a population of 7.9 million residents. Similarly, in 2008, the total cost of the Montreal Household Travel Survey was $1.5M, with a unit cost $19.45 per household.[55]

A simple order of magnitude calculation makes obvious the relative cost-effectiveness associated with data collection versus infrastructure development and renewal. If 5% of *all Canadian households* (close to 15 million in 2011) were to be surveyed at the approximate cost of $20/interview, the total cost would be approximately $15 million. If such a survey were to be conducted once every five years (as is the current case with most large Canadian urban travel surveys), the annual cost would be $3 million/year. Using the lowest cost estimates for infrastructure and congestion cited above, this is about the same as any one of the following:

- Less than 1/100th of 1% of the total budget spent on transportation in Canada.
- Initial construction of less than 8km of 2 lane equivalent rural arterial roadway.
- Construction of less than 1700 m^2 of bridge deck (approximately 1.5 to 2 average size bridges).
- Annual maintenance of less than 3,000 km of 2 lane roadway.
- Less than 30 metres of subway line.
- Less than the cost of congestion for a single one-way commute for commuters in large urban areas of Canada.

It is not difficult to imagine that a sound data collection effort across Canada could lead to infrastructure savings far in excess of these, by directly contributing to evidence-based transportation planning. That is, such a national survey would pay for itself if it saved just 8km of two-lane roadway or 2 bridges being built somewhere in the country each year. Or taking the approximately median estimate of the cost of congestion to Canada each year of $3 billion, this national survey would pay for itself if it leads to infrastructure investments and/or operational improvements that reduced congestion costs by just 0.1%.

---

[55] http://www.enquete-od.qc.ca/docs/EnqOD08_Faitssaillants2.pdf

Explicitly quantifying the benefits of data collection within urban transportation planning is not an easy thing to do. Some indications, however, of the value that is placed on the TTS within the Greater Toronto-Hamilton Area (GTHA) include the following:

- The Data Management Group (DMG) at the University of Toronto (the keeper of the TTS database) tracks the number of online data queries that are made each year to the TTS database. During the 2006-2010 time period, over 83,000 requests for TTS information were recorded on the DMG online systems, and 80 different organizations accessed the system in 2010 alone (Data Management Group, 2011).
- Metrolinx, the GTHA regional transportation planning agency has developed a 25-year, $50 billion transportation plan for the region – "The Big Move" (Metrolinx, 2008a). This development of the overall plan and the establishment of the business cases for each individual component of the plan depends critically on travel demand forecasting results generated by the Ontario Ministry of Transportation's "GGH Model" (Metrolinx, 2008b), which was constructed using 2006 TTS data. Without TTS it would have been impossible to build GGH Model, and without its systematic, credible forecasts and analyses of alternative investment plans, evidence-based support/justification for The Big Move would not have been possible to produce.
- In the winter of 2012 an acrimonious political debate occurred within the City of Toronto concerning whether an LRT line should be built along Sheppard Avenue (at a capital cost of approximately $1 billion) or the Sheppard East subway should be extended from its current terminus at Don Mills Road to Scarborough City Centre (at a capital cost estimated to range from $2.7-3.3 billion). The City of Toronto's "GTAModel" (Miller, 2001) (also based on TTS data) was used to analyze both alternatives. Its forecasts clearly indicated that the LRT line was a more cost-effective method for attracting new riders to the TTC and reducing greenhouse gases, among other performance measures (City of Toronto, 2012). While many factors led to City Council's decision in favour of building the LRT line, to the extent that evidence from a model derived from a $2-3 million dollar survey helped save the City approximately $2 billion in an unjustifiable investment (an approximately thousand-fold difference in costs), one would surely conclude that the survey had more than paid for itself.

In Montreal, data from large-scale travel surveys play a fundamental role in the transportation planning and decision process. Examples include:

- The transportation demand forecasting model estimated by MTQ relies on the most recent detailed OD files as well as trends observed using data over a 20-year period. OD survey data define reference conditions and are updated according to observed trends in order to forecast plausible travel behaviours at 20-year horizons. The resulting forecasts, also developed similarly for other urban region in the Quebec province, feed all modelling process such as subway extension analysis or completion of the road network.
- Using 1998 OD data, MTQ has estimated congestion costs in the Montreal region; the same study was updated using 2003 OD survey to document evolution of these costs.

- In 2008, the City of Montreal adopted its 20-year transportation plan. Understanding of travel needs and description of challenges relied sets of analysis conducted using the 2003 and previous OD survey data. These data combined with MTQ forecasts were used to formulate elements of vision and strategies required to change the status quo direction.

The same is true, for example, in Edmonton, where their household travel surveys and their commodity flow survey are major data collection efforts that have been used to develop the Edmonton regional travel model (RTM). The regional travel model is used for decision support for major infrastructure decisions including LRT and roadway expansion and improvements. The LRT line from the Southeast to West in Edmonton is an example project for which the RTM was used for route selection, LRT station design and optimization of LRT scheduling/operations to minimize the number of trains required. The model also provides traffic volumes along LRT corridor and help to identify roadway/intersection configuration and right-of-way requirement. This project has an estimated cost of $3.2 Billion. Another example RTM application is the evaluation of six scenarios of the Walterdale Bridge replacement. From a traffic point of view, each of the scenarios resulted in about the same traffic patterns. From this analysis, it was decided that there is very little benefit to none to build a four lane bridge. Therefore, the recommended design that starts construction in 2013 is only three lanes. The bridge closure during construction was also evaluated. Using the RTM, it was determined that the surrounding network can still be handled the demand during construction phase. Traffic operations in the area will be worse but the longer delays will be acceptable for the short period of construction.

2.3 SUMMARY

Canadian urban areas are home to over 80% of the country's population, with this number continuing to rise in the coming decades. They face massive challenges in terms of:

- Both maintaining their existing transportation infrastructure and expanding this infrastructure to meet the needs of their growing populations.
- Supporting system operations in a fiscally sustainable fashion.
- Supporting continuing economic development of our urban regions, which must continue to be the incubators/drivers of the national economy, on-going resource sector development notwithstanding.
- Addressing pressing environment concerns (pollution, greenhouse gases, ecosystem impacts, etc.).
- Promoting more sustainable urban forms (containing urban sprawl, implementing transit-oriented development, etc.).
- Maintaining and enhancing quality of life for future generations of Canadians.

Two futures exist for Canadian cities that offer starkly different outcomes. If we fail to manage growth efficiently, effectively and sustainably, in particular in terms of providing the much expanded and improved transportation systems that our 21st Century urban regions require to be successful, then we will experience serious declines in our economic competitiveness, our environment and our quality of life. Alternatively, we can recognize the critical role that transportation plays in our economy and our society and invest in our infrastructure and services in a manner commensurate with its importance and its need, leading to a much brighter future for Canadians. Data collection in support of evidence-based planning, design and decision-making is a central element of cost-effectively building this latter path. Indeed, this path to a bright future will not be possible without adequate investment in data collection as a necessary foundation for sound decision-making.

3. URBAN PASSENGER TRANSPORTATION DATA NEEDS

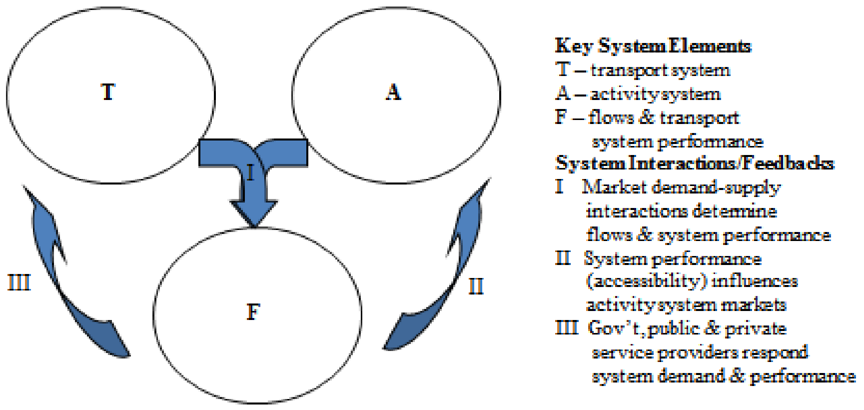
3.1 INTRODUCTION

This section provides a framework for systematically defining urban passenger transportation data needs and establishing a comprehensive taxonomy of urban passenger transportation data. It begins with a very high-level characterization of urban systems and how transportation interacts with other components of urban areas. Building upon this understanding, it then discusses how data and data collection fit within the urban transportation process, with an emphasis on a particularly critical and challenging part of data collection: travel demand data. It then concludes by presenting a comprehensive model of data of relevance to urban transportation planning that is applicable across urban scales and applications.

3.2 TRANSPORTATION WITHIN THE URBAN SYSTEM

The starting point for designing a comprehensive data collection process for urban passenger transportation planning is a clear understanding of the factors and processes that define and affect the urban transportation system. Manheim (1978) provides a very simple but powerful depiction of urban systems (Figure 3.1) as consisting of two mutually interacting systems:

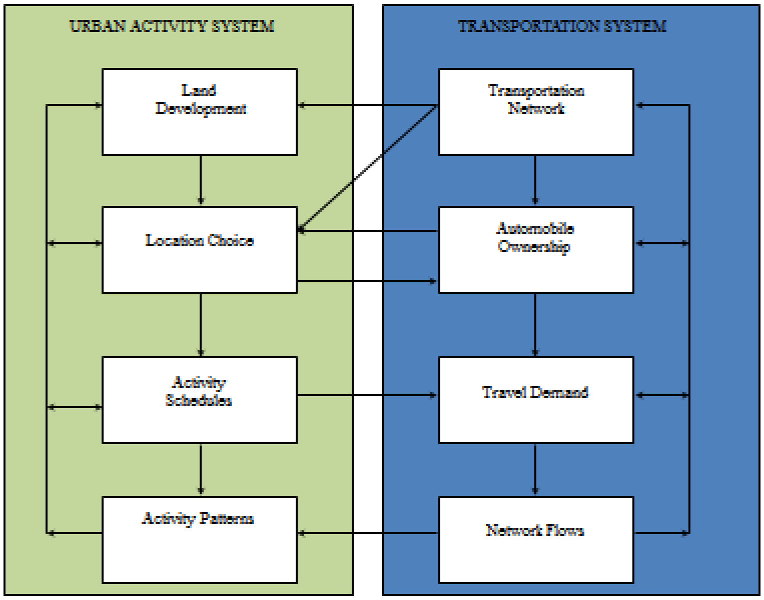
- The transportation system (T), consisting of the physical transportation networks, technologies and services, as well as supply/performance processes that define the system's operations.
- The urban activity system (A), consisting of the built urban form, the distributions of people, firms and activity generators and the economic and social processes that give rise to travel demand.



Source: Manheim (1978)

Figure 3.1: The Urban System

The interactions between the activity system (which generates the demand for travel) and the transportation system (which supplies the capability to travel) results in flows (F) of people, goods, services and money within the urban region, as well as determine the performance (travel times, congestion levels, etc.) of the transportation system as it is used by trip-makers. In Figure 3.1 these are labelled "Type I" interactions. Longer-term "feedback", however, also exists in the urban system, with the performance of the system affecting both the evolution over time of the activity system ("Type II" interactions) and the transportation system ("Type III" interactions).



Source: Meyer & Miller (2001)

Figure 3.2: Transportation – Activity System Interactions

Meyer and Miller (2001) expand Manheim's diagram (Figure 3.2) to show in greater detail the interactions between the transportation and activity systems, making explicit the fact that the *accessibility* that the provision of transportation services (including the "self-supply" of these services through possession of private automobiles) affects the evolution of urban form, while the demand for travel (and resulting network flows and performance levels) is driven by the need for people to engage in activities that are distributed in both space and time throughout the urban region. This figure also emphasizes that urban form is determined by both the supply of built form (houses, office buildings, retail space, etc.), as determined by land development processes, and the location choices of households and firms within this built form (i.e., the demand for built space).

A critical point to be drawn from Figures 3.1 and 3.2 is the need to observe and understand <u>both</u> the transportation and activity systems if transportation planning is to be successful. Given the split that exists in most municipalities between "transportation" departments (which collect and analyze transportation system data) and "planning" departments (which collect and analyse built form and associated activity system data), this often leads to myopic views of the urban system and sub-optimal analysis of urban transportation issues, such as the interconnections between urban sprawl, auto dependency and transit usage/viability.

3.3 DATA & DATA COLLECTION IN THE URBAN PASSENGER TRANSPORTATION PLANNING PROCESS

Figure 3.3 provides a high-level overview of the roles of data and data collection within the urban transportation planning process. It is through the monitoring of the transportation system (and the "external factors" that affect the transportation system) by various data collection means that we can observe, understand and analyze this system. Both primary data, directly collected from observation of the urban system, and secondary data, generated through analysis and modelling (where these models have been developed using primary data), are fundamental inputs into planning and decision-making. These data are used in myriad ways for problem diagnosis, analysis, modelling, monitoring, and information dissemination to stakeholders and decision-makers.

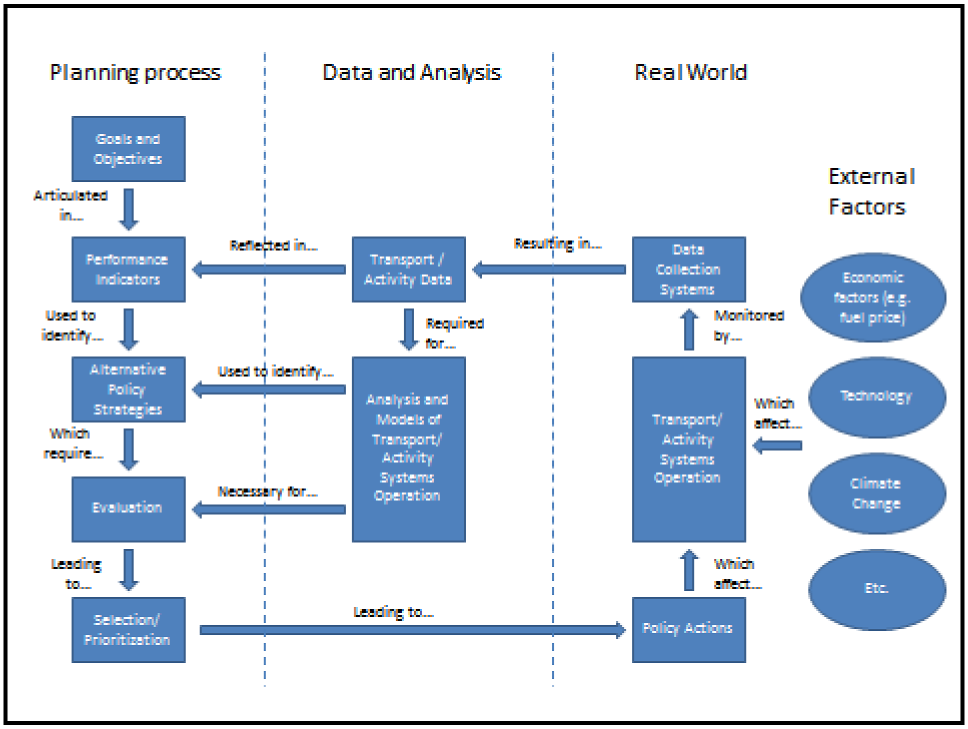


Figure 3.3: Data & Data Collection within the Transportation Planning Process

As illustrated in Figure 3.4, the data concerning the overall urban system can be categorized as pertaining to either:

- ***Characteristics of the system*** itself, which, in turn can be broken down into characteristics of the urban activity system (land use, built form, population, employment, etc.) and attributes of the transportation system (road and transit networks, schedules, fares, etc.).

- ***System behaviour***, both in terms of the demand for (usage of) transportation services and of the system's performance (speeds, congestion, travel times and costs, etc.) and impacts (pollution, greenhouse gas (GHG) emissions, noise, accidents, etc.).
- ***External factors*** affecting transportation system demand and performance (gas prices, interest rates, etc.).

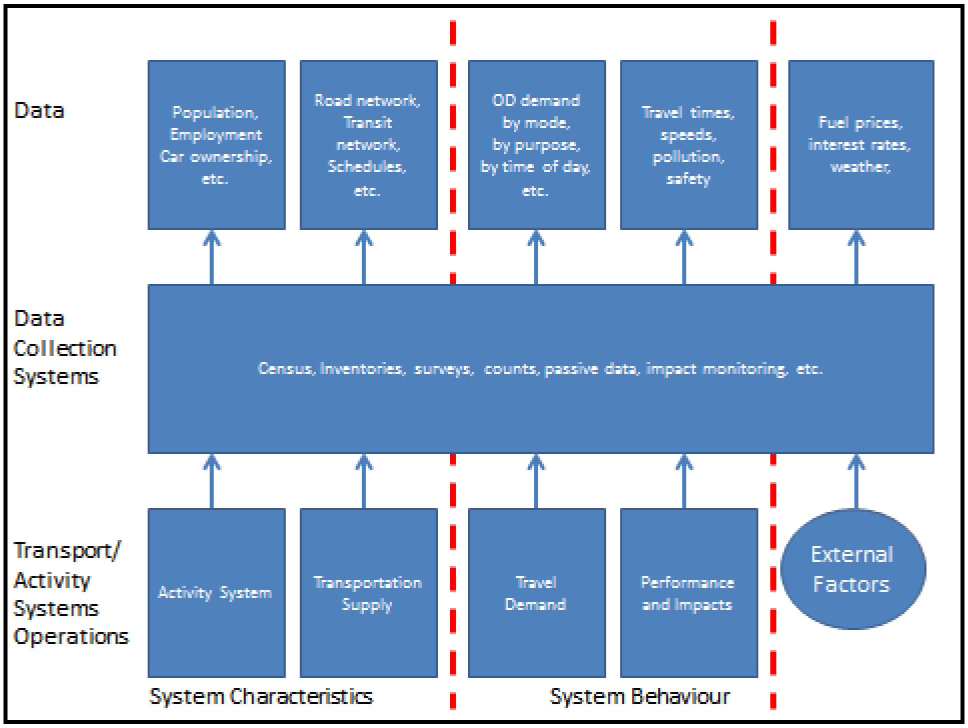


Figure 3.4: Major Transportation Data Categories

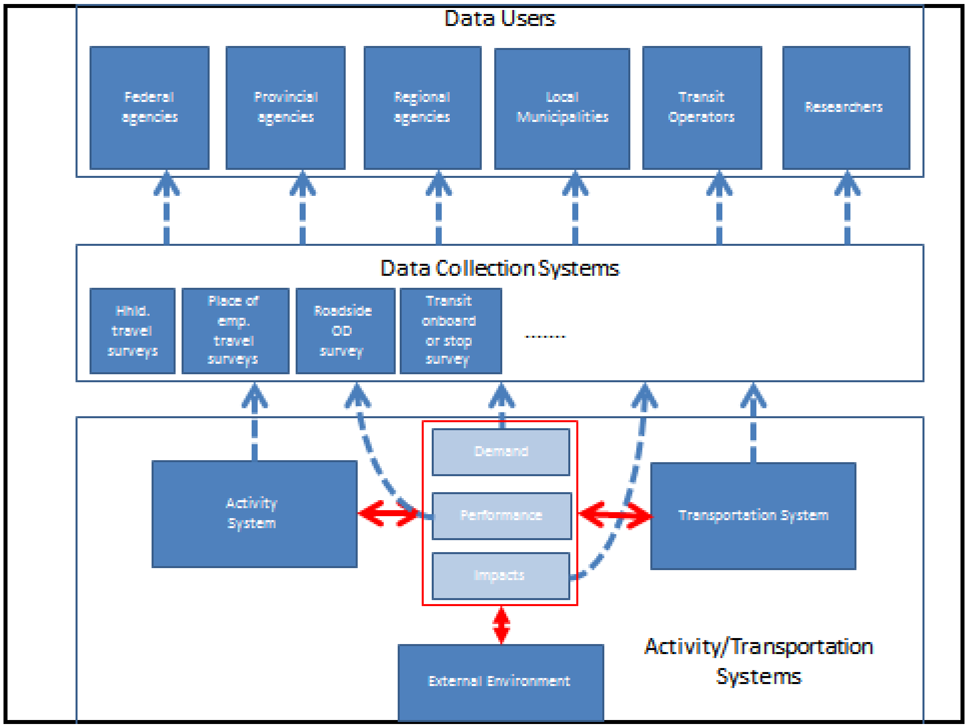


Figure 3.5: Data Users & Data Collection

As indicated in Figure 3.5, there are many users of urban transportation data, most notably government agencies at all levels, transit operators and researchers, among others.[56] They will use a variety of data collection methods to collect data relevant to their needs. A key question is the extent to which data collected by different organizations are readily available for use by other organizations, as well as the extent to which data gathered by different agencies for different purposes are sufficiently compatible (in terms of definition, data collection methods, etc.) to be combined to provide more comprehensive descriptions/analyses of the urban transportation system and the behaviours of interest within this system.

Similarly, data needs and appropriate data collection methods will vary across other dimensions that have not yet been explicitly discussed. These include:

- Jurisdictions (national, provincial, regional, municipal).
- Users/stakeholders within each jurisdiction:
    - Municipal (transit agencies, transportation planning departments, etc.).
    - Regional (regional planning agencies, conservation authorities; etc.).
    - Provincial (ministries of transportation, etc.)

---

[56] Non-government organizations (NGOs) and private sector companies also, of course, are users of urban transportation data for a variety of purposes.

  - National: environment, natural resources, transportation
  - General public (NGOs, ratepayer associations, individual citizens, etc.).
- Planning functions to which the data are to be applied:
  - Strategic (long-term).
  - Tactical/operational (short- to medium-term).
  - Emergency/event planning.
- Policy concerns to be addressed:
  - Infrastructure investment.
  - Pricing, fare policy, etc.
  - Service planning.
  - Environmental impacts (ecosystem damage, air quality pollutants, greenhouse gas emissions, noise, etc.).
  - Safety
  - Etc.
- Geographic Scale (large vs. small urban regions; regional versus local issues; etc.).

3.4 TRAVEL DEMAND DATA

The five data categories of Figure 3.4 (activity system; transportation supply; travel demand; performance and impacts; external factors) can be decomposed into a large number of individual types of information that are required in a variety of planning applications. The next sub-section deals with this decomposition in some detail. Travel demand, however, is a particularly important (and complex) component of urban transportation planning. Detailed data on current and historical travel demand are required, both for direct analysis of trip-making behaviour and trends and for constructing sophisticated travel demand forecasting models capable of credibly analysing a wide range of policies (new infrastructure investment, new pricing/fare policies, etc.) under a variety of future scenarios. Figure 3.6 illustrates the complexity of a complete representation of travel within an urban region in terms the full set of trip-makers, trip purposes, destinations, modes and time scales involved.

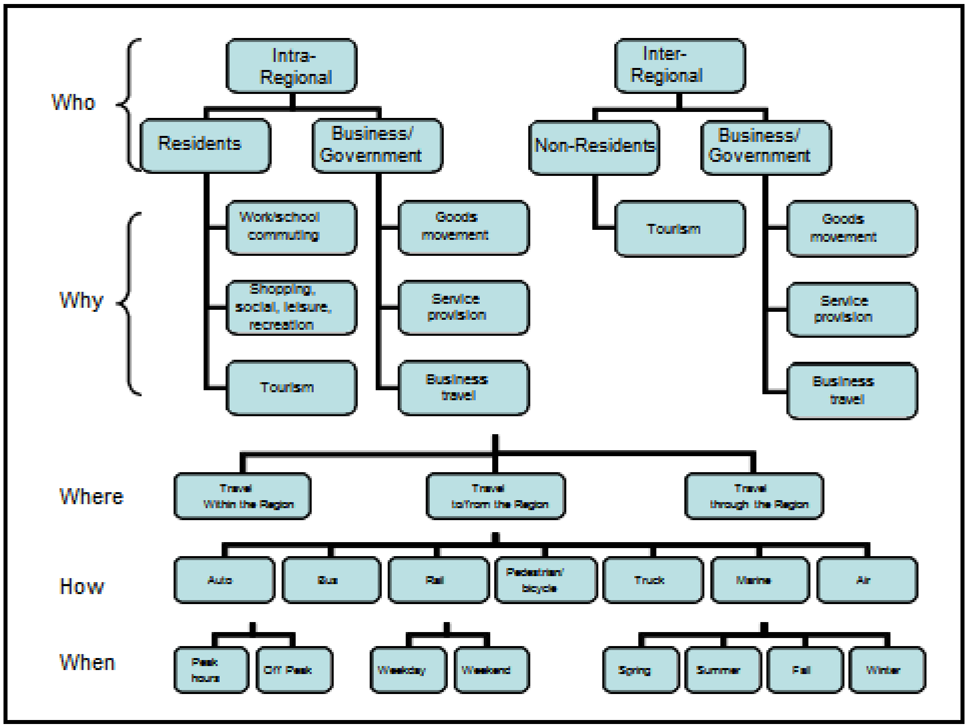


Figure 3.6: Travel Demand Components

(Source: Adapted from Roorda and Shalaby, 2008)

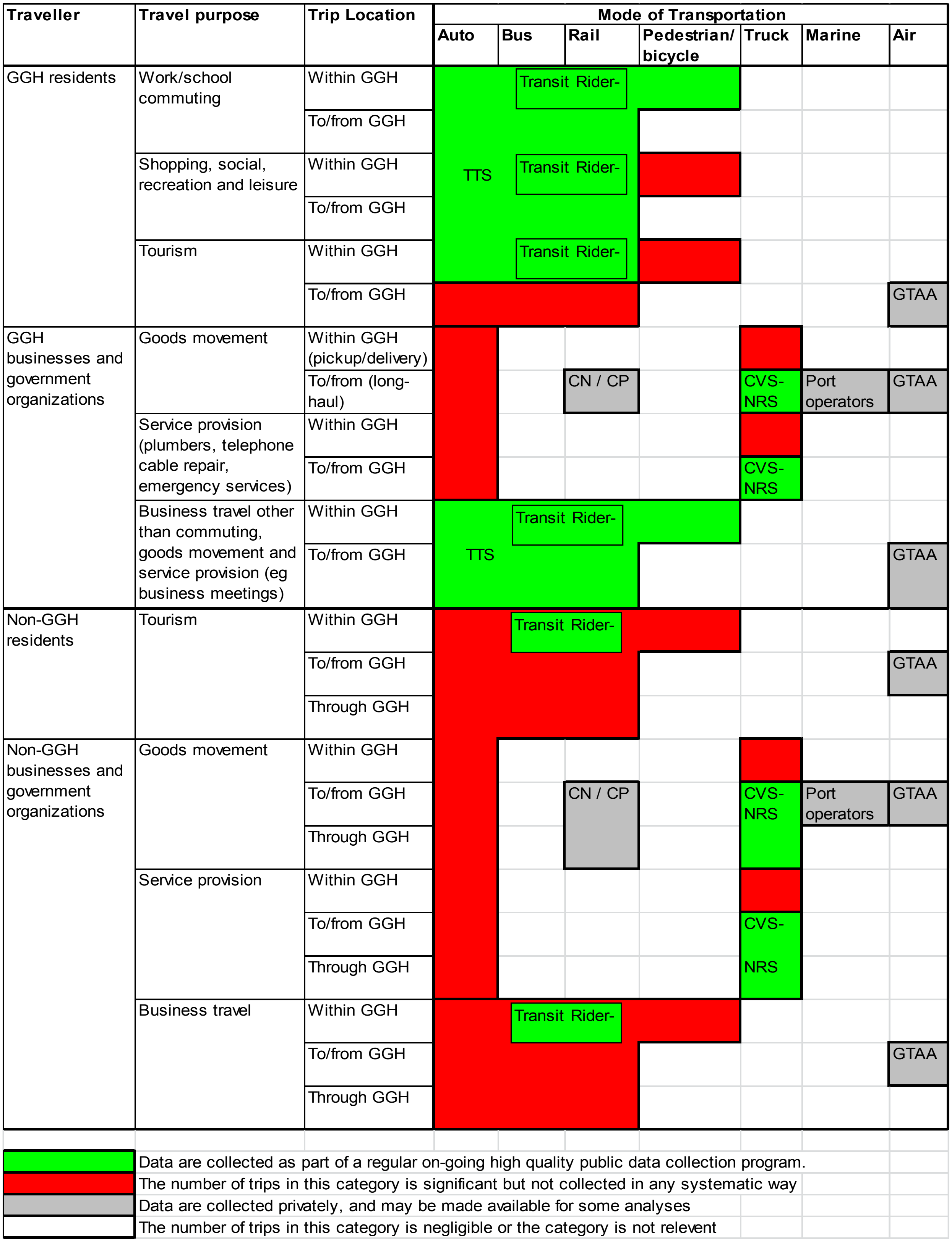
Figure 3.7 provides an overview of how the various dimensions of Figure 3.3 are currently addressed (or not)[57] in the Greater Toronto Hamilton Area (GTHA), which is reasonably representative of the current state in most of Canada's larger urban regions. As illustrated in this figure, a variety of surveys and other data collection methods, employed by a various agencies, are used to collect travel demand data in the region, with the primary methods/sources being:

- The Transportation Tomorrow Survey (TTS), which collects household-based travel information every 5 years for the region.
- On-going transit ridership surveys undertaken by transit agencies within the region.
- For truck movements, the Commercial Vehicle Survey (CVS) undertaken on a regular basis by the Ministry of Transportation, Ontario (MTO).
- Various private-sector data sources, which may or may not be readily accessible for public-sector use.[58]

---

[57] Figure 3.7 illustrates the difficulty of dealing with the full dimensionality of travel demand in a relatively simple diagram. Important dimensions that are not explicitly shown in this diagram are trip start times and durations, trip path choice information, and, in the case of goods movement, the type and quantity of goods being shipped.

[58] Another important source of data concerning travel demand that is not shown in Figure 3.7 is cordon count data, which are collected on an on-going basis by regional agencies.



Traveller	**Travel purpose**	**Trip Location**	**Mode of Transportation**						
			Auto	**Bus**	**Rail**	**Pedestrian/ bicycle**	**Truck**	**Marine**	**Air**
GGH residents	Work/school commuting	Within GGH	🟩	🟩 Transit Rider-	🟩	🟩			
		To/from GGH	🟩	🟩	🟩				
	Shopping, social, recreation and leisure	Within GGH	🟩 TTS	🟩 Transit Rider-	🟩	🟥			
		To/from GGH	🟩	🟩	🟩				
	Tourism	Within GGH	🟩	🟩 Transit Rider-	🟩	🟥			
		To/from GGH	🟥	🟥	🟥				▨ GTAA
GGH businesses and government organizations	Goods movement	Within GGH (pickup/delivery)	🟥				🟥		
		To/from (long-haul)	🟥		▨ CN / CP		🟩 CVS-NRS	▨ Port operators	▨ GTAA
	Service provision (plumbers, telephone cable repair, emergency services)	Within GGH	🟥				🟥		
		To/from GGH	🟥				🟩 CVS-NRS		
	Business travel other than commuting, goods movement and service provision (eg business meetings)	Within GGH	🟩	🟩 Transit Rider-	🟩	🟩			
		To/from GGH	🟩 TTS	🟩	🟩				▨ GTAA
Non-GGH residents	Tourism	Within GGH	🟥	🟩 Transit Rider-	🟩	🟥			
		To/from GGH	🟥	🟥	🟥				▨ GTAA
		Through GGH	🟥	🟥	🟥				
Non-GGH businesses and government organizations	Goods movement	Within GGH	🟥				🟥		
		To/from GGH	🟥		▨ CN / CP		🟩 CVS-NRS	▨ Port operators	▨ GTAA
		Through GGH	🟥		▨		🟩		
	Service provision	Within GGH	🟥				🟥		
		To/from GGH	🟥				🟩 CVS-		
		Through GGH	🟥				🟩 NRS		
	Business travel	Within GGH	🟥	🟩 Transit Rider-	🟩	🟥			
		To/from GGH	🟥	🟥	🟥				▨ GTAA
		Through GGH	🟥	🟥	🟥				

🟩	Data are collected as part of a regular on-going high quality public data collection program.
🟥	The number of trips in this category is significant but not collected in any systematic way
▨	Data are collected privately, and may be made available for some analyses
⬜	The number of trips in this category is negligible or the category is not relevent

Figure 3.7: Travel Demand Data Collection, GTHA

(Source: Adapted from Roorda and Shalaby, 2008)

Despite these quite extensive data collection efforts, it is clear from Figure 3.7 that a number of key travel demand components are not currently collected within the GTHA, notably with respect to commercial vehicle movements, business-based personal travel, non-resident/tourism travel within the region and non-motorized trip-making.

3.5 AN OBJECT-ORIENTED MODEL OF URBAN ACTIVITY/TRANSPORTATION SYSTEMS

In order to be able to deal with the complexity of data types contained within a full specification of urban passenger transportation while at the same time to provide a flexible taxonomy that readily adjusts to a wide variety of applications and contexts, an “object-oriented data model” (a “data class design”) has been developed in this study as a key component of the proposed data collection and management framework. This object oriented model is a systematic representation of the many elements of the urban system that are of relevance for transportation planning, operations and analysis. Each important element of the urban system is represented as a class (shown as boxes in the following figures) that fits within a class structure. Relevant attributes of each class are shown in the box. Hierarchical relationships between different classes are represented as lines between classes, with a numbering system to reflect the type of relationship (one to many, one to one, etc.). Each individual person, household, vehicle, trip, etc. are considered to be objects, which in turn are instantiations of a class.

There are several reasons why a formalized object-oriented model is a useful part of a framework for urban transportation data collection. First, it encourages a systematic view of what the important data elements are, and identifies ownership relationships. While much of the content of the object oriented model is intuitive and perhaps obvious, the model does highlight certain aspects that are harder to conceptualize (such as sharable transit passes, car sharing infrastructure, and the representation of mixed use developments). As such it provides a good platform for identifying data requirements for a variety of applications. Second, the class design can provide a starting point for organizing the structure of referential database systems to hold transportation data. Third, an object oriented model is an ideal tool for conceptualizing disaggregate models of travel demand, so that they may be developed in parallel with suitable datasets.

The object oriented model presented below is not intended to be fully comprehensive. Other components of the activity system (such as government entities, institutions, etc.) and other modes (such as marine, air, taxi, school-bus, skateboard, etc.) exist. Other relationships also exist beyond those elaborated in the diagrams (e.g., a person may be a shareholder of a business). The model does, however, contain what we consider to be the most important aspects of the urban activity and transportation systems as they pertain to the objectives of planning and policy analysis.

The class diagrams are hierarchical in nature. At the top of the hierarchy, shown in Figure 3.8, is the city, which is comprised of the activity and transportation systems (as also shown in Figures 3.1, 3.2, 3.4 and 3.5). Three key elements of the activity system are households, businesses and land parcels, which are elaborated in Figures 3.9, 3.10 and 3.11, respectively. These three classes are the primary generators of the activities from which demand for transportation is derived. The transportation system consists of the road system, transit system, and walk/cycle systems, which are shown in Figures 3.12, 3.13 and 3.14, respectively. These classes are intended to reflect the infrastructure used by the public, or in other words, the transportation supply.

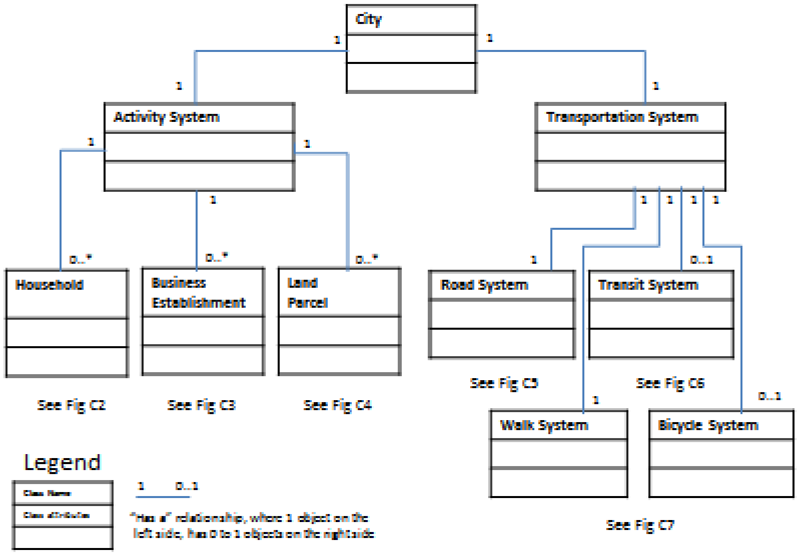


Figure 3.8: Object Oriented Model – Activity and Transportation System

Notes: Figures 3.8 to 3.14 represent a static class hierarchy. The class structure represent entities and their attributes, but not their behaviour. "Static" means that a representation of history and future are not included (e.g. Previous jobs, cars, habits, previous land uses, plans,). "Hierarchy relationships" are included, but not lateral relationships. Relationships between persons in a household, relationships between persons in different households (social networks), relationships between transport systems of different modes (e.g. Sharing a right of way), etc. are not shown.

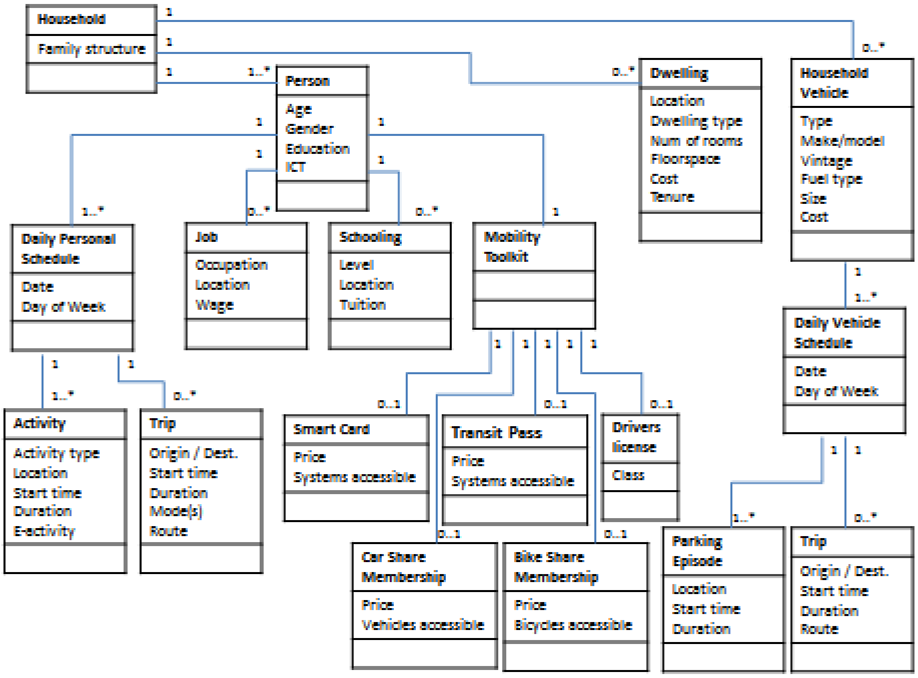


Figure 3.9: Object Oriented Model – Household

Figure 3.9 elaborates the class structure of the household. A household consists of persons, who may have one or more jobs, be enrolled in a school program and may have a driver's license. Households usually have dwellings (although the class structure is generalized to include the homeless). A household also has a mobility toolkit, which may include vehicles, transit passes, smart cards, car share memberships, etc. The mobility toolkit is considered to be a household resource, rather than a person resource because multiple people may have access to each mobility tool. However, an attribute of each mobility tool is a list of people that have access to that tool (who may drive the car, use the transit pass etc.). Household demand for travel is reflected in the schedules of people and vehicles, which consist of activities and trips (for people) and parking episodes and trips (for vehicles).

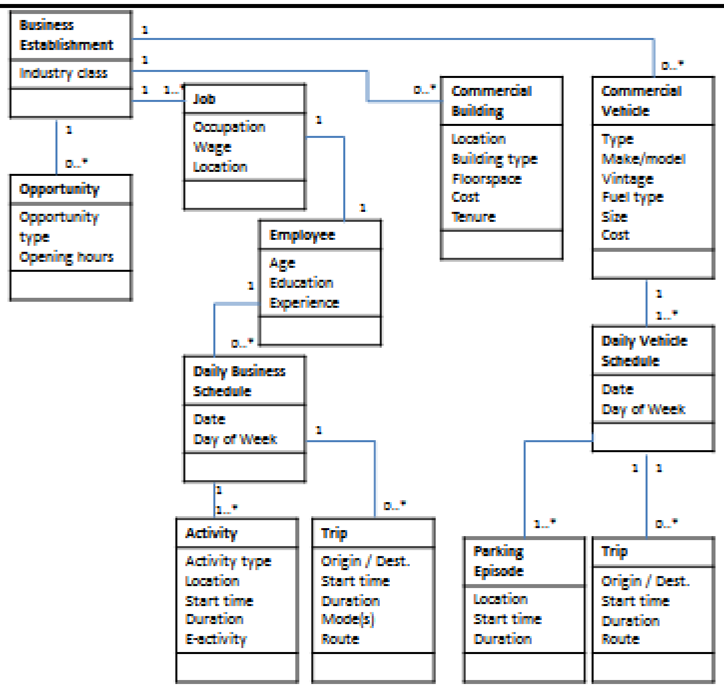


Figure 3.10: Object Oriented Model – Business Establishment

Figure 3.10 outlines a basic class structure of a business establishment. Businesses have buildings and vehicles, provide jobs, and provide opportunities (e.g., a retail establishment provides an opportunity for a person to shop, while a doctor's office provides an opportunity for people to access health care). A job is filled by an employee, who has a business schedule that may consist of business activities and trips. Notably, the same person may conduct personal travel in his/her role as a household member and also conduct business travel as an employee of a business establishment. This organization of the data, which keeps a person's dual role as household member and employee as distinct, is consistent with the usual separation of travel data collection in household and business establishment surveys. Of course personal and business travel are both important, although business travel has often been neglected in transportation planning.

Figure 3.11 shows the class structure of land parcels. A land parcel includes the buildings in which residential dwellings and non-residential floorspace exist. Land parcels also can be used for parking facilities and other facilities (parks, tennis courts, etc.). This representation of buildings allows for a representation of mixed use developments (i.e., one building consisting of residential and non-residential space), but also respects the usual way of representing space (floor area for non-residential space and dwelling units for residential space).

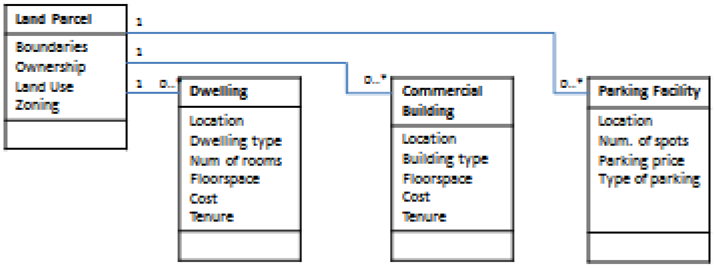


Figure 3.11: Object Oriented Model – Land Parcel

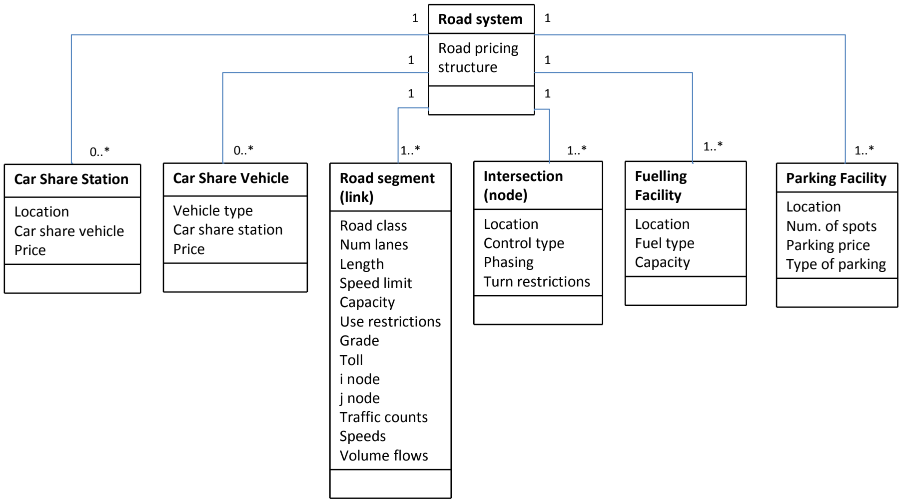


Figure 3.12: Object Oriented Model – Road System

The class representation of the road system, shown in Figure 3.12, includes the usual basic elements of road segments and intersections. Location and capacity of fuelling facilities and parking facilities is important in some policy contexts, and is likely to have impact on travel behaviour (for example, the provision of plug-in electric vehicle recharging stations may influence their use and the proportion of their operation in charge depletion mode). The control system is included as a class but not fully elaborated here because of the diversity of ITS infrastructure that is available (changeable message signs, ramp meters, coordinated traffic control systems, signals, etc.). Traveller information systems also influence travel behaviour. It is therefore essential to represent them, and ultimately collect data on the types of traveller information to which people subscribe. Finally, car share stations and vehicles have been included in the road system class, because they are available for shared public use (for those people with the required membership).

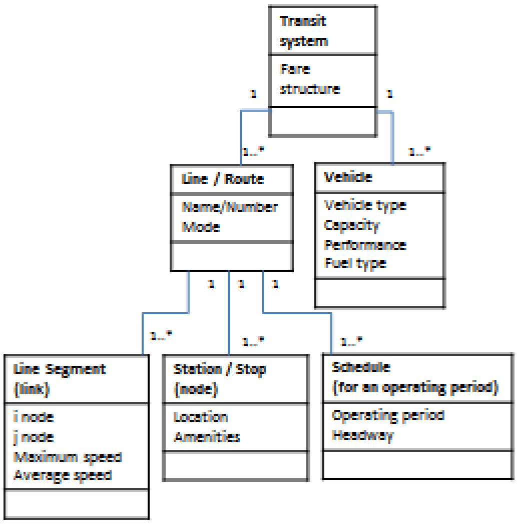


Figure 3.13: Object Oriented Model – Transit System

The transit system class structure, in Figure 3.13, describes lines/routes, the vehicles, the storage/maintenance facilities and control systems and traveller information systems that collectively define a transit system. Notably, different types of public transit can be represented with this structure, including commuter rail, bus systems, LRT, subways, ferries, etc. Only the attributes of each of these different types of public transit will differ (e.g., speed, stop locations, schedule, etc.).

The bicycle system class structure, shown in Figure 3.14, consists of bike lane segments, and the infrastructure surrounding bike share systems. Bicycles that are owned by households are considered to be vehicles that are part of the household mobility toolkit.

Finally, the walk system is represented for sake of completeness, and is also shown in Figure 3.14. It simply reflects any walk facilities (e.g., underground walkways, pedestrian bridges, sidewalks), and any pedestrian amenities (e.g. canopies, benches, drinking fountains, etc.).

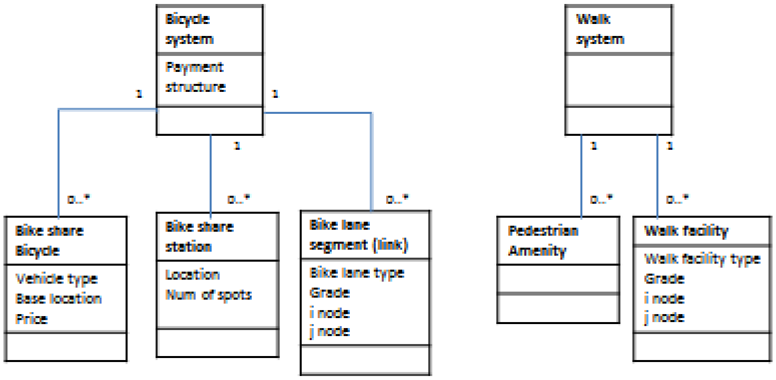


Figure 3.14: Object Oriented Model – Bicycle and Walk Systems

As noted in the figures above, a key attribute of many objects in the urban system is **location** (places of residence, work, school, trip origins and destinations, etc.). Location is the single most important variable used to link data items together (e.g., matching trip-makers to income via common places of residence), is fundamental to the spatial modelling of travel (trips by origin and destination), and the ability to map spatial data is essential to understanding both survey data and model results. Virtually all data of relevance to urban transportation planning is spatially referenced in one or more ways: by zone, by x-y geocodes, by transportation network link, etc. Multiple spatial referencing systems inevitably are used (census tracts, traffic zones, rasterized pixels/grid cells), both as the primary level of spatial precision for a given variable (e.g., census data provided at the census tract level) and as multiple descriptors of the same piece of data (e.g., a trip origin might be coded with both a traffic zone number and a census tract number). Multiple aggregations of a given base referencing system ("planning district" aggregations of traffic zones, municipalities which are aggregations of both traffic zones and census tracts, etc.) generally exist, as well as mappings from one zone/referencing system to another (e.g., census tracts to traffic zones and *vice versa*). Also note that, most fundamentally, all locations can always in principle be referenced by an x-y geocode. What differs from one location representation to another is the spatial precision of this geocode, which generally can range from a zone centroid (e.g., the typical assumption in four-step travel models is that all trips begin and end at traffic zone centroids) to fine-grain grid cells (e.g., 100x100m grid cells) to actual street addresses.

Without getting into an explicit object design herein, this discussion makes clear that several "spatial objects" exist within the urban system and are required to characterize the system. These include zone systems (defined by centroids and boundary files and possessing additional attributes such as area), mappings between zone systems, and an x-y coordinate system. Geographic information systems (GIS) generally are used to store spatial data and to provide mapping and spatial analysis capabilities. Non-GIS data files generally will also include in each data record for spatial referencing purposes one or more zone designations, as well as, possibly, an x-y geocode to facilitate spatial cross-referencing of data and the interface with spatial mapping and analysis software.

Finally, to illustrate the application of the object-oriented data model to a practical urban transportation context, Appendix A presents a mapping of current City of Edmonton datasets into the model.

4. DATA COLLECTION METHODS

As discussed in considerable detail in Part II of this report, *Literature Review*, a wide variety of data collection methods exist to address the data needs discussed in the previous section.

Table 4.1 provides a representative mapping of survey- and non-survey-based methods to the collection of different types of data. As indicated by the columns of this table, data collection methods can be loosely grouped into three broad categories: household-based surveys; choice-based and other types of surveys; and other types of data colelction, which depend upon a variety of technologies, both standard and emerging.

Household-based surveys are separated out from other types of surveys in recognition of the major role that they play as a standard tool for collecting travel behaviour and other transportation-related data. As shown in Figure 4.1, 57% of the Canadian agencies responding to the survey conducted by the project team concerning data collection methods in Canada either conduct or commission household travel surveys, while another 18% either plan to use such surveys or would like to, resources permitting; i.e., 75% of the responding agencies either use or would like to use household travel survey data.[59] Of the agencies reporting that they conduct or commission household surveyrs, 94% have undertaken at least one survey over the past 10 years.

Figure 4.2 summarizes the sampling frames and interview methods currently used in Canadian household travel surveys. As indicated in this figure, telephone-based mehtods dominate Canadian survey methods, with 55% using telephone listings for their sampling frame (and other 4% using random-digit dialing) and 57% using telephones as their interview medium. Web-based methods are also seen to be emerging as the next most popular form of interview method (24%)

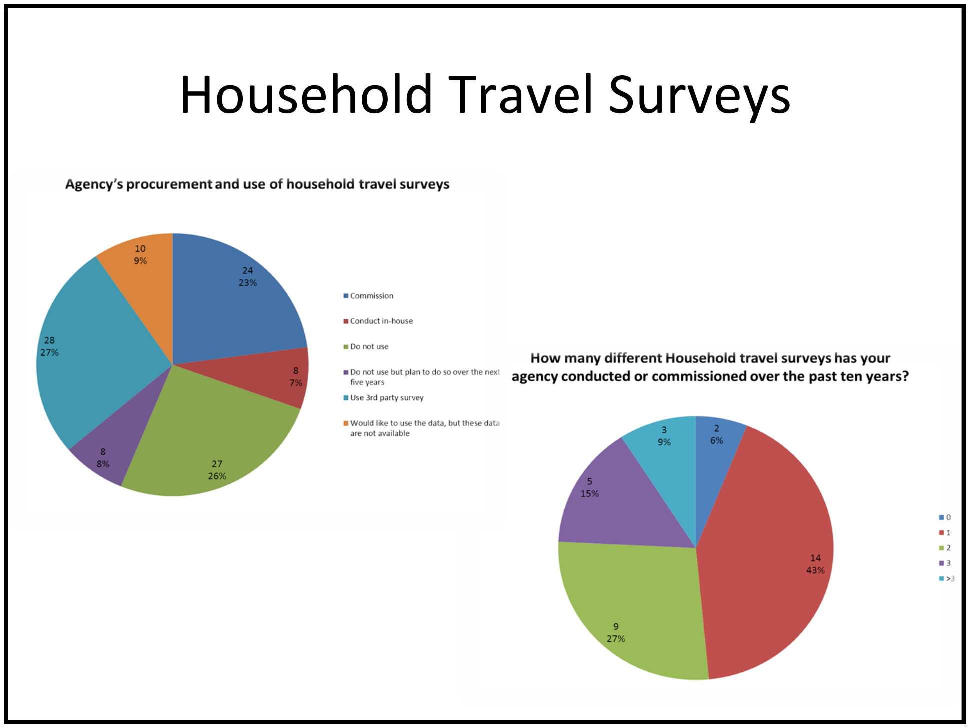
---

[59] For full documentation of this survey, see Part IV in this report, *Survey of Data Sources for Urban Transportation Applications*.

Table 4.1: Data Type/Method Taxonomy

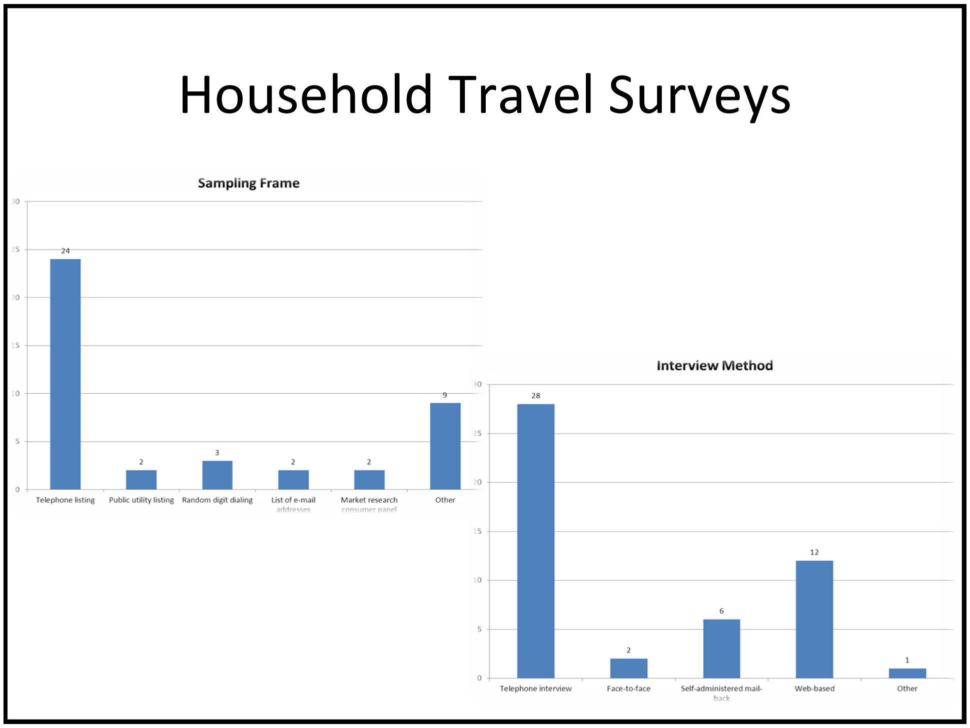
Data Type	**Household-Based Surveys**				**Choice-Based Sample Surveys**		**Standard Technology-Based Methods**			
	Face to face	Telephone	Mail-back	Web- based	Road-side	On-board	GPS	Roadside detectors	Smart-card	Mobile phone
Household activity / trip-making behaviour	X	X	X	X	X	X	X		X	X
Count data					X	X		X	X	
Transportation system characteristics							X	X	X	X
Transportation cost and service levels	X	X	X	X			X	X	X	X
Land use characteristics										
Population socio-economic information	X	X	X	X	X	X				
Attitudes / opinions / stated choices	X	X	X	X	X	X				
System impacts (e.g. emissions)								X		

	Emerging Technology Based Methods							
Data Type	Remote sensing (satellite / aerial)	Web “apps”	Social network software	Smart phone	Acceler-ometers	Personal health sensors	Environ-mental sensors	Blue-tooth
Household activity / trip-making behaviour		X	X	X	X			
Transportation system characteristics	X							X
Transportation cost and service levels	X			X				X
Land use characteristics	X	X					X	
Population socio-economic information		X	X	X		X		
Attitudes / opinions / stated choices		X	X	X				
System impacts (e.g. emissions)					X	X	X	



Household Travel Surveys

Figure 4.1: Canadian Usage of Household Travel Surveys



Household Travel Surveys

Figure 4.2: Household Travel Survey Sampling Frames and Interview Methods

Many other types of surveys are undertaken to gather travel-related data in addition to household surveys. These are too numerous to exhaustively list here, but common types include: transit on-board surveys of transit riders, roadside intercept surveys, place of employment based surveys of workers' commuting behaviour; and intercept surveys at "special trip attractors" (shopping malls, airports, etc.), among many others.

In addition to surveys, many types of counts and inventories are undertaken to gather data concerning travel behaviour, transportation system characteristics and behaviour, land use / urban form attributes, etc. Table 4.2 summarizes the range of counts and inventories currently undertaken by responding agencies in the project survey.

Table 4.2: Counts and Inventories Used by Canadian Transportation Agencies

Counts and Inventories

	Commission	Conduct in-house	Do not use	Do not use, but plan to do so over next 5 years	Use 3rd party survey
Road traffic counts, automated	16	31	36	0	10
Road traffic counts, manual	17	32	31	0	13
Transit ridership (boarding) counts, automated	3	37	39	7	7
Transit ridership (boarding) counts, manual	6	53	24	4	6
Screenline/cordon counts	14	12	55	1	11
Truck/freight/commercial vehicle counts	10	15	58	3	7
Parking counts	5	32	45	1	10
Bicycle counts	4	24	51	8	6
Pedestrian counts	6	26	50	4	7
Other counts	2	14	69	2	6
Roadway inventories (number of lanes, lane widths, etc.)	5	40	34	1	13
Transit service inventories (route, frequencies, etc.)	1	57	24	1	10
Parking inventories (number of spaces, parking charges, etc.)	4	33	40	1	15
Bicycle inventories (bikeways, etc.)	2	32	46	3	10
Sidewalk inventories (sidewalk widths, etc.)	2	42	36	2	11
Other inventories	1	11	78	0	3

An alternative approach to categorizing data collection methods is in terms of the nature of the interaction between the data collector and the subject of the data collection. That is, data collection methods can be categorized as follows:

- *Active surveys*, in which the respondent agrees to participate in the survey and voluntarily provides information through this participation (e.g., conventional household surveys).
- *Passive data stream collection*, in which data are collected without the participants' explicit knowledge, often for primary purposes other than transportation planning (e.g., credit/debit card transactions, mobile phone usage records, smart card transactions).

- *Hybrid active/passive methods*, in which passive data collection is combined with some level of participant consent and participation (e.g., GPS tracking with a prompted recall survey).
- *"Proactive" data provision*, in which information is voluntarily provided for a variety of reasons without explicit prompting by the data collector, but which can then be mined for useful transportation related purposes (e.g., social network postings).
- *Administrative datasets*, often collected for primary purposes other than transportation planning. These also include transit schedules, real-time vehicle location feeds, and other data collected for operational purposes.

As presented in Figure 3.4 in the previous section, urban passenger transportation data can be divided into information on system behaviour (which further subdivides into travel demand and network performance and impacts) and system characteristics (which further subdivides into activity system and transportation supply attributes). Table 4.3 provides an extensive (although possibly not totally comprehensive) listing of data collection methods for both system behaviour and system characteristics, categorized by the typology presented immediately above.

Table 4.3: Data Collection Methods by Type & Application

System behaviour	Surveys (Active)	Household travel survey
		Place of employment travel survey
		Roadside OD
		Transit onboard or transit stop survey
		Truck/freight/ commercial vehicle survey
		Bicycle
		Pedestrian
		Parking usage
		Special traffic generator or other location specific
		Stated Response
		Attitudinal
		Census
		Other StatCan surveys (e.g. General Social Survey - Time Use Survey, Travel Survey of the Residents of Canada)
		Other federal or provincial surveys
	Passive / active hybrid	GPS with prompted recall
		GPS with active recall
		Canadian Vehicle Use Survey
		Accelerometers and other personal sensors
	Passive stream data collection	Counts and other roadside detectors (road and transit)
		Smart card (transit, parking, car sharing, bike sharing transactions)
		Credit card and debit transaction cards
		Mobile phone usage records
		GoogleMaps (frequent updates, e.g. congestion)
		Web “apps”
		Remote sensing (satellite/aerial)
		Environmental sensors
	“Proactive” data provision	Crowdsourcing
System characteristics	Administrative datasets	Vehicle registry
		Property assessment files
		Driver license records
		Inventories
		Remote sensing (satellite/aerial)
		GoogleMaps (base maps)
		Google Transit Feed Specification (GTFS) files
	“Proactive” data provision	OpenStreetMap

5. MEETING DATA NEEDS: ISSUES, TRENDS & OPTIONS

5.1 INTRODUCTION

The literature review undertaken in this study provides clear indications of the main trends and issues that urban data collection faces. First, there is agreement that in the transportation field surveys are a primary means of gathering the required data on travel behaviours and that they constitute an essential data collection tool. It is the project team's view that they will remain central in the toolbox of transportation planning for the upcoming years. Still, the ways in which surveys are conducted and the methods used to conduct them are changing and will continue to change to adapt to new issues and benefit from emerging opportunities.

The most critical issues faced by Canadian household travel surveys, which are by far the most important sources of information on people movements, are:

- Decreasing comprehensiveness of the usual sampling frames (typically: directory of residential telephone land-lines) due notably to the increasing number of people who solely use cell phones.[60] This trend is expected to continue over the years.
- Increasing difficulties to contact and recruit respondents and consequent declining response rates.
- Difficulty of retrospective and proxy responding to correctly report home-based discretionary trips and non-home-based trips that are expected to continue growing in importance in the future.

Hence, various efforts have been made recently to address some of these challenges:

- Implementation of dual-frame sampling to overcome issues faced when relying on landline-based samples.
- Use of a multi-methods approach to conduct the survey (Web, GPS).
- Passive data collection methods are increasingly being investigated as cost-effective alternatives to traditional active collection methods.

Of course, with the diversification of survey methods and sources of information, the community is facing the issue of relevantly integrating/fusing the data to make them usable for planning and modelling purposes, including the challenge of maintaining compatibility of datasets over time as collection methods change.

Other trends have been observed, for instance that stated preference/response methods are increasingly being used in a variety of transportation planning contexts.

The review has also confirmed that new technologies are providing interesting opportunities for the travel survey community:

---

[60] As the Canadian population continues to age, it may also be the case that an increasing percentage of people may live in household settings without land-lines (e.g., group homes) as well.

- Combining multiple survey instrument will help reach a mode complete and representative sample of individuals and households.
- Web-based surveys are being used to complement (sometimes replace) traditional survey and it is compatible with the increasing proportion of people having access to the internet.
- GPS (on-board and portable) and other portable devices are increasingly used in travel survey context, allowing to gather spatial-temporal details of trips and decreasing respondent burden.

While household travel surveys will remain central to the gathering of data on the movement of people, they probably will change in terms of method and content. It may also be the case that more “person-based”, as opposed to “household-based” approaches will be increasingly adopted, at least for some applications, to better match technological options with data needs. Care must be taken, however, to not lose understanding of the basic household context within which person travel occurs, including the interdependencies among household members (notably the need to serve children’ and other dependents’ travel needs).

Some current data gaps will also need to be filled, for instance, the lack of data on long-distance travel (no national passenger travel survey is available in Canada as it is with the NHTS in the US for instance), trips by non-residents or week-end / occasional trips. It may be opportune to think of what data collection method is the best suited to fill those various gaps.

In order to better understand future data collection needs the next section briefly explores likely future trends with respect to society and travel behaviour, technology and planning needs. Section 5.3 then attempts to summarize these trends and their implications for urban passenger transportation data collection.

5.2 A LOOK INTO THE FUTURE

5.2.1 Societal Trends – Travel Behaviour Trends

Transportation data collection methods must adapt to deal with changing trends in urban travel behaviour. These include:

- Less regularity in daily travel. Travel surveys may not be the best method for collecting information about the variability of travel.
- Loss of regularity in transportation decisions due to decline in the importance of the work trip in daily travel.
- People may use expensive private modes more sparingly as the cost of travel increases.
- Increasing numbers of perturbations that disrupt travel patterns (weather, climate change, peaks in unemployment, road construction delays): do we have enough / the right data to manage these situations?
- Population aging, and planning services and facilities to meet needs of this demographic (e.g., increasing health related, non-work travel).
- Immigrant populations will continue to grow. Their travel behaviour is in some ways unique. Can we fuse data from census to help explore this?
- Privacy and how comfortable people are with their private information being disseminated.
- Households are getting smaller on average.

- People will have more ability to act spontaneously. How do we predict people's unpredictability? How do people simplify their decision-making in the context of increasingly complicated systems and more information?
- Which aspects of travel behaviour will it always be necessary to ask?
- Weekend travel is increasing important, with large volumes of travel that are very different in purpose, spatial-temporal distribution and modal usage than weekday travel.
- Teleworking continues to increase and evolve.
- eCommerce / eShopping
- New transportation modes and/or modes of activity?
- Vulnerable populations (the elderly, the poor, etc.) are growing in numbers and complexity of spatial distribution.
- Evolution from a manufacturing-based to a service/knowledge-based economy.

5.2.2 Trends in Technology

Technological advances are changing both travel behaviour and our means for observing this behaviour. Trends include:

- Nature of interaction between transportation and ICT – availability of information will change the way people make decisions about travel (route choice, other choices).
- Availability of trip planners will increase.
- Paradigm shift – continuous reaction to continuous streams of demand data; control-based sampling/modelling/planning may evolve to less-controlled sampling methods; everyone will have real-time information (e.g., on network conditions) and uses it to make decisions.
- The web will be more present. The web is just the beginning: what comes next?
- New datasets are increasingly available (e.g., smart cards) that indicate inadequacies in household travel surveys. Should we use external datasets to calibrate household travel survey-based models, or should we be stitching together datasets to develop a model?
- Our data expansion (weighting) methods may not always be valid spatially and don't capture time variability. Matching available demographic data suitable for sample expansion with the survey data can be challenging (e.g., if survey year and Census year do not match), and is becoming more so as with increasing use of continuous surveys. An explicit multi-instrument approach may help with this, but the ability to weight data obtained via various new technologies needs to be investigated.
- We may be able to use smart card data and GPS data and other sources (using fusion techniques) to expand the scope of household travel surveys.
- GPS provides a radically different view of travel behaviour because it can provide information over long time periods. It provides new insights into the spatial organization of travel over longer periods of time. It also provides information about trip-makers' responses to interventions in the system.
- Impacts of RFID technology on data collection?
- Considerable potential exists to use downloadable apps to collect information from respondents. As with other web-based methods, sample representativeness is a concern with such approaches.

- Information Technology provides opportunities to run “controlled” experiments so that we can learn more about how people adapt (e.g., user pricing).

5.2.3 Trends in Planning / Policy Issues / Needs / Funding of Data Collection

The urban transportation planning environment is also changing in a variety of ways that affect data collection needs and options. Trends/issues include:

- Can we merge data collection with proactive suggestions (e.g., from trip planners) on travel decisions?
- Carbon pricing? If end user pricing of carbon is introduced, what would it mean for travel patterns? Survey methodologists may be asked to find ways to collect the carbon taxes.
- Many organizations are collecting huge amounts of information. We may find ways to make more/better use of third-party databases (although a better understanding of the strengths, weaknesses and costs of these databases is generally required).
- Individuals versus households – do we need household samples? Proxy responses in current surveys are certainly problematic.
- What is the future of transportation models (continuing disaggregation? probabilistic? simplified? more holistic?).
- Managing an evolutionary process. Addresses continuity and compatibility. Overlapping instruments as a general principle. Software that allows us to make valid inferences.
- Sustainability concerns, fuel dependency, and other emerging policy issues will generate new questions to be addressed by models and supporting data. Road pricing and other user-pay policies are increasingly important to be addressed.

5.3 IMPLICATIONS FOR DATA COLLECTION

The preceding section has sketched a picture of a very dynamic and uncertain future for Canadian urban regions and their transportation data collection needs. In general, travel behaviour will continue to become more complex as urban regions continue to grow in spatial scale and population, social and economic systems continue to become more heterogeneous, and technology (ICT and other) continues to evolve and pervasively influence our behaviour. Urban transportation planning will be increasingly challenged to deal with new demands for transportation systems, with these demands being greater, more varied and substantively different than in the past. In addition, these growing and changing demands will need to be addressed within continuing challenging fiscal environments.

These trends, as has been discussed in Part II and elsewhere, are also providing both challenges and opportunities with respect to data collection. Conventional household survey methods are becoming increasingly less viable in the face of changing patterns in telephone ownership/usage, as well as in the lifestyles of the people being surveyed. At the same time, technological innovation (smart cards, internet surveys, personal ICT apps, etc.) provide promising opportunities for collecting travel-related data in new and potentially cost-effective ways.

As is typically the case, new technology is not, however, without its own challenges and costs. Many of the biggest challenges with respect to most new data collection technologies/methods are the “classic” ones that pertain to any data collection method, such as sample representativeness, (non)-respondent bias and privacy/confidentiality. In addition, increasing use of passive data streams, etc. will require increasing exploitation of sophisticated data fusion methods in order to bring together diverse datasets to provide the comprehensive picture of travel and trip-makers required for transportation analysis and modelling purposes.

Although not the focus of this study, the implications for travel demand modelling of these trends in behaviour and data collection should also be noted. Increasingly complex behaviour and increasingly complex policy analysis questions require increasingly detailed and sophisticated modelling methods to capture these behaviours and address these policy questions. The state of the art in travel demand modelling today is, in fact, quite sophisticated, and is arguably well-placed to address these challenges. But it is often limited in practical application due to lack of appropriate data, as well as lack of technical and financial resources within transportation agencies to support such sophisticated modelling (Miller and Hatzopoulou, 2008). Implicit in much of this project's work is the understanding that improved data collection methods are needed to support the development of models and other decision-support tools that are essential to effective decision-making in today's and tomorrow's complex world. In particular, the data needs of models should be explicitly considered in the development of an agency's data collection program to ensure that these needs can, indeed, be adequately and efficiently addressed.

Also, as data collection methods and associated types of data (passive data streams, etc.) change, it is important that models change as well so that they can best exploit these new datasets. In particular, the emerging availability of massive (but possibly relatively anonymous) passive data streams represents a possible "new paradigm" with respect to how we, first, observe the transportation system, and, second, subsequently model it. This issue lies well beyond the scope of the present project, but it is one that model developers should seriously consider.

6. DATA MANAGEMENT

6.1 INTRODUCTION

While the focus of this project is on data collection methods, data management is a critical component of data collection: it is little use to collect data if they are not properly maintained. Similarly, a critical element of data management is the security of the collected data and the maintenance of appropriate privacy and confidentiality protocols. Comprehensive treatment of these issues of data management and security would require whole reports unto themselves and is well beyond the scope of this study. The following two sub-sections do, however, provide high-level guidance with respect to key aspects of both these issues.

6.2 DATA MANAGEMENT

The over-riding point to be made concerning data management is that data must, indeed, be managed. Any organization must have an explicit, systematic data management plan that is consistent, comprehensive, on-going and well-documented. All too often data are “lost” in terms of their usability over time due to failure to pay proper attention to their maintenance within a working data management system. Just as data collection requires resources and expertise if useful data are to be collected, so too must resources and expertise must be invested in maintaining these data over time. A well-maintained and documented database not only provides the factual basis for planning analysis and modelling, it also provides a concrete basis for “institutional memory” as individuals come and go within an agency.

Key elements of any data management system include:

- *Storage*. Clearly data must be maintained in a location, format and medium that: persist over time; protects both the security and integrity of the data over time; and provides convenient, efficient access to the data to authorized users. This can be challenging in the face of changing storage media technologies, data management software, computing hardware, etc., as well as evolution of organizations and their functions over time.
- *Documentation*. Undocumented data are worthless. Users must understand: how the data were collected, definitions of variables, formatting of data fields, how to access the data, etc. *Metadata* (i.e., data/information about data) are as critical as the data themselves in terms data usability. These metadata also require maintenance over time as datasets grow, are edited/cleaned, etc. Every new piece of data must be documented at the time of its collection (or when it is modified) in a standardized manner of it is to be usable. As part of its metadata, each agency should ideally have a “data map” of its data holdings.
- *Completeness (contextual data)*. Too often surveys (or other data collection efforts) are undertaken without parallel effort being put into collecting by other means contextual data that are essential for interpreting the survey data. In particular, travel demand models cannot be constructed from household survey data without a large amount of contextual data (transportation networks and their associated travel times; travel cost information (fuel prices, fares, etc.); parking supply and cost information; land use data; and socio-economic data (unemployment rates, interest rates, immigration rates, etc.). Agencies often neglect to collect

and/or maintain over time these contextual data, thereby significantly reducing the usefulness of the survey data. This issue will be even more challenging as the use of passive data streams becomes more common, since it may be very difficult to collect the contextual data required to properly interpret and use such data streams.

- *Access*. What tools/interfaces are available for accessing the data? Documentation of these tools/interfaces and training of users in their use are required. What are the protocols for data access (who may access the data, for what purposes)? While "open data" initiatives are increasingly common in many agencies, many barriers to data sharing exist, including: privacy concerns (discussed in the next sub-section) and organizational cultures that may work against data sharing.
- *Dissemination*. People need to be aware that data exist before they can use them: how is information concerning data to be posted/distributed/advertised. How can data be provided to authorized users? Agencies often have limited resources to respond to requests for data that may prevent widespread dissemination, even in cases where data sharing is nominally encouraged.
- *Quality control*. Perhaps over and above all other issues, maintaining the quality of the data is paramount. As discussed in the literature review, the entire data collection process needs to be designed around maximizing the quality of the data being collected. This concern for quality must be maintained throughout the data management process as well.

Changes in data collection methods also pose challenges to maintaining the usability of major datasets over the long term. As new data collection methods are introduced, retro-compatibility with older data can be a challenging problem. In addition, a particularly important challenge with transportation datasets is maintaining "zonal consistency" across diverse data sets and over time. Traffic zone systems, even within a given agency, often change over time (as does Census geography from one Census year to another); municipal boundaries, etc. change over time as well, etc. Different agencies often maintain zone systems which may be suitable for their internal purposes but these may not be compatible with other agencies' systems. Geocoding of primary data is an obvious solution to solve zone compatibility issues, but not all datasets are geocoded and not all disaggregate, geocoded datasets are sharable across agencies. Use of raster (grid-cell) data, as well as maintenance of well-documented mappings among diverse zone systems are also means for managing zonal consistency.

In addition to managing and sharing data within a single agency, sharing of datasets across agencies should be strongly encouraged, but, of course, faces the same sort of issues and barriers as within-agency sharing. Nevertheless, examples of successful inter-agency data-sharing arrangements exist world-wide, often involving collaboration with universities who often undertake the data management/hosting tasks. These include:

- The Data Management Group housed at the University of Toronto, which collects and maintains travel survey data for transportation agencies within the Greater Toronto-Hamilton Area (GTHA) (Data Management Group, 2011; http://www.dmg.utoronto.ca/).
- The "Metropolitan Travel Survey Archive", maintained at the University of Minnesota, archives travel surveys from around the US (http://www.surveyarchive.org/).

- The Montreal Origin-Destination travel survey technical committee, headed by AMT, maintains continuous collaboration with researchers from Polytechnique Montreal (MADITUC, Chaire Mobilite) that provide interview software and methodological assistance for the collection, validation, analysis, visualisation and dissemination of data. The technical committee manages data storage and diffusion. All researchers showing interest in travel survey analysis can get access to the travel survey microdata file for research purposes.
- Although not explicitly transportation-related, the University of Essex maintains the “UK Data Archive”, the Unite Kingdom’s largest archive of social and economic data (http://www.essex.ac.uk/depts/ukda.aspx).
- A survey dataset archive at ETH Zurich (http://tda.ethz.ch/).
- ONE-ITS is a collaboration of universities, government agencies and the ITS industry to develop a web-based environment for the sharing of data, models, etc. relating to transportation operations, modelling and planning (http://one-its-webapp1.transport.utoronto.ca:11080/web/guest).
- The Centre for Data and Analysis in Transportation (housed at Laval University) is a clearinghouse, depository, and analysis centre for transportation data from Canada (http://www.cdat.ecn.ulaval.ca/en/index.php?pid=764).

Open data initiatives among a growing number of municipal and other governments is an extremely encouraging trend which holds obvious promise for data-sharing and for maximizing data usage. The rate with which government datasets are currently being made readily available to the public in Canada varies considerably from one urban area to another. Some municipalities are very actively embracing the open data concept, while others are moving much more slowly, presumably due to a variety of concerns (lack of resources to support data requests, concern about how the data will be used, etc.). Nevertheless, the trend towards open data is one that seems to be gathering momentum and is one that should prove to be of considerable promise to transportation agencies, given the wide variety of data they require.

One final issue worth noting is that many excellent datasets are maintained by agencies over time, but they do not archive these data as new data becomes available. That is, while a “current snapshot” of the data being collected is maintained and continuously updated over time, historical data are not saved. The result is that the opportunity to collect and maintain longitudinal datasets that might be extremely valuable in supporting a wide variety of trend analyses and modelling is lost. Typical examples include motor vehicle registry data, assessment data, roadway inventories, Google Maps, etc.

6.3 PRIVACY/SECURITY

Maintaining the privacy of individuals and the confidentiality of their information has always been a primary concern of transportation data collectors. Canadian transportation agencies, however, are facing increasingly strict (and conservative) regulations concerning data collection methods and data dissemination and use protocols through FIPPA/FOIPP type legislation.[61] Over and above meeting the

---

[61] University researchers also face strict internal survey protocols imposed by their Research Ethics Committees.

letter of such laws, it is critical that transportation agencies collecting data display a pro-active sensitivity to privacy/confidentiality concerns so as to maintain both the perception and the fact that the data being collected are being protected and used appropriately.

There are four key elements with respect to privacy in the transportation data collection process:

- Gathering private information (e.g., phone numbers).
- Analysts' access to confidential data (both respondents' contact information and their actual survey responses).
- Being able to identify individuals in the analysis of survey data.
- Archiving data in secure and confidential manner.

As emphasis in data collection increasingly focusses on disaggregate data and increasingly detailed personal datasets become available via passive data streams (and other means), concerns about privacy/confidentiality obviously increase. This problem, however, is not unique to transportation. The health sector, in particular, has been successfully dealing with these issues for some time. The key to this success is the concept of "privacy by design" (Cavoukian, 2010) in which protocols are built into both data collection and management processes at the time of their design so that confidentiality, privacy and data security are ensured while simultaneously maintaining data access/usability. In particular, these issues need to be addressed at every point in the data collection process:

- Survey design.
- Accessing respondents and gaining their consent to provide information.
- Analysis of data.
- Storage/maintenance of data.

Elements of a successful privacy policy include:

- Simple, clear protocols for data storage and access, strictly adhered to.
- Anonymization of data records.
- Taking care in how records are linked and analyzed/post-processed so as to avoid identification of individuals.

At the same time that governments are increasingly concerned about intruding into citizens' lives, huge databases are being assembled by private sector firms (banks, credit card companies, telephone and internet service providers, etc.) containing very detailed, very personal information about individuals. Ethical issues concerning accessing such information for transportation planning purposes (assuming that such information can, indeed, be accessed) need to addressed and guidelines/protocols need to be established concerning the use (or not) of such data.

Similarly, the voluntary disclosure by millions of people of personal information on social media sites such as Facebook and Twitter raises intriguing new possibilities for data mining of people's opinions, attributes and behaviours that might be of use in transportation planning applications. Since such disclosures are totally voluntary and public, presumably use of such data for planning purposes is feasible, providing that individuals are not identified in the assembled databases and subsequent analysis. Many technical questions concerning sample representativeness, the potential to provide "apps" for people to voluntarily report on themselves, and how such data might actually be put to good use within planning analyses exist and will likely be the subject of survey methodological research for some time to come (Miller and Cottril, 2012). The extent to which such voluntary disclosures might also include biased responses of various sorts ("making oneself look good", etc.) also must be considered.

7. THE CORE-SATELLITE DESIGN PARADIGM

7.1 INTRODUCTION

As discussed in more detail in the literature review, there is an increasing trend in travel surveys to use "multi-instrument" surveys in which two or more surveys (or combinations of a survey and other data collection methods such as GPS traces) to capture different elements of the behaviours being studied. This approach reflects both the complexity of these behaviours (which may not be able to be captured within a single instrument) and the need to keep response burdens within a single survey within reasonable levels.

A particularly attractive and generalizable multi-instrument survey design is the *core-satellite design* proposed by Goulias, *et al*. (2011) (see Figure 7.1). This paradigm is defined in detail in the next section. Sections 7.3 and 7.4 then present two prototypical "use cases" to illustrate how the core-satellite design paradigm can be applied in practice, as well as how it interfaces with the object-oriented data model discussed in Section 3.5

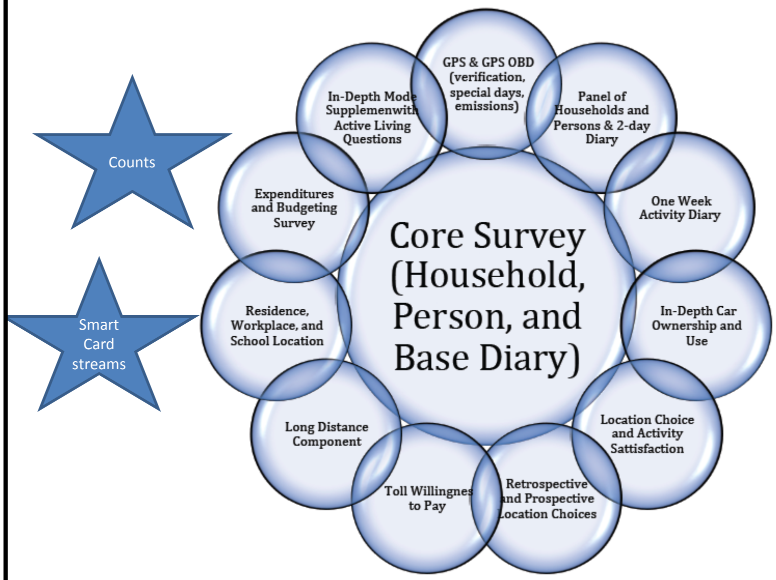
7.2 PARADIGM DEFINITION

As illustrated in Figure 7.1, the core-satellite approach involves the following components:

- A *core survey*, which is a large-sample survey which gathers primary information concerning the respondents and their key behaviours.
- Any number of *satellite surveys*, which are smaller-sample, more focussed surveys (or other data collection methods) designed to gather more detailed information about specific behaviours of interest.
- Additional, independent, *complementary* surveys/datasets that might be used to augment the core-satellite database, but may not be directly linkable to the core-satellite data.[62]

---

[62] These independent datasets have been added by the study team to Goulias, *et al.*'s original paradigm, which only include the core and satellites.



Adapted from: Goulias, et al., (2011)

Figure 7.1: The Core-Satellite Multi-Instrument Survey Design

Characteristics of the core survey include:

- It includes key data that are fundamental to the agency's primary policy/planning needs (their core business).
- It includes attributes of the respondents that permit core data to be linked to common variables in the satellite surveys so that core and satellite data can be jointly used. This linkage is direct if the satellite survey is a sub-sample of the core survey. If the satellite respondents are not a sub-sample of the core respondents, then these common variables must permit fusing of core and satellite data, as discussed in Part III*: Data Integration/Fusion Methods*.
- Its sample size is large enough to make statistical inferences concerning variables of interest.
- It is expandable to make statements about the full population.
- It is consistently applied over a large geographical region (e.g., the urban region of interest and, ideally, its immediate hinterland).
- It is stable (but not necessarily static) over time and is applied relatively frequently (or continuously) so as to provide consistent time-series data.
- It is relatively short, so as to minimize response burden and to permit large sample sizes to be cost-effectively collected.

Satellite surveys provide the opportunity to enrich the core dataset by filling gaps and adding detail to the core that would not be feasible and/or cost-effective to collect as part of the core. They can be used to gather data for special models that can be linked to core behaviours or to augment the core sample for small sub-populations of special interest. They must be statistically linkable to the core survey, either by being sub-samples of the core or via common attributes of the respondents in the two datasets. Types of satellite surveys include:

- Extra questions and/or instrumentation of a subset of the core sample.
- An additional survey (separately administered subsequently to the core survey) of a subset of the core survey.
- Increased (stratified) sampling of specialty populations (the elderly, transit riders, etc.).
- Passive data sources (e.g., smartcard data), where a reliable means of linking to the core survey (through common variables) exists.
- Surveys conducted on different samples but with common data items for linking/fusion.[63]

Satellites may be either "one-time" surveys or on-going in nature. They may or may not cover the full geographic area of the core. They can be much more flexible in terms of choice of method and can be a means for experimenting with new methods. They may well involve greater response burden, more detailed methods, etc. Examples of possible satellite surveys might include:

- Car type/usage.
- Bicycle ownership and usage.
- Walking behaviour.
- Multi-day/week trip or activity diary.
- Residential mobility and dwelling type choice.
- Heath, fitness, physical activity, active transportation
- Route choice (chosen routes, choice sets considered; by mode
- Parking location/usage/cost

The third category in the paradigm: independent, complementary surveys/datasets consists of any other data, collected by any means or obtained from any other source which is part of the agency's overall database. Such data generally will not be fusible with the core-satellite database, except in special cases in which common variables and sufficient commonality in sampling frame, time of collection, geography, etc. permit some form of integration.[7] This category is added to the Goulias, *et al*. paradigm for the sake of completeness and to emphasize the need to think comprehensively about one's overall data collection, management and analysis program. Complementary surveys and datasets might include:

- Special generator surveys (airports, hospitals, etc.).
- Visitor/non-resident surveys.
- Taxi usage records.
- Smart card records.
- Transit onboard/ridership surveys.[64]

---

[63] It must be noted that significant challenges in data fusion can exist if the satellites are conducted at significantly different points in time and/or upon non-core samples.

- Cordon/screenline counts.
- User satisfaction surveys.
- Census Place of Residence – Place of Work (POR-POW) questions, including the usual mode to work question.

The core-satellite paradigm is an extremely flexible and generalizable approach to meeting different agency needs. It is defined around content (i.e., what content is core, what can be collected via a satellite process) rather than method. Different agencies will use different methods for both their cores and their satellites, depending on their data needs, resources, etc. These methods can and should evolve over time (with satellite methods generally evolving more quickly and more often than the core), while recognizing the need to maintain data compatibility over time for time-series analysis and consistency in modelling. One approach for maintaining such compatibility is to use both old and new methods during transition periods so as to be able to test in a controlled way for the impacts of the changes in methods on survey results.

The definition of what is "core", "satellite" and "complementary" will vary from one application to another. For example, as illustrated in the examples presented in the following sections, a transit onboard survey might be a satellite (or even complementary) survey for a multi-modal transportation planning agency tasked with developing a regional travel demand model forecasting system (since it provides supplemental information to the household travel survey which provides the planning agency's core dataset), while it is the core for a transit service planning group whose focus is on understanding (and possibly modelling) short-term route ridership (with a household travel interview survey providing useful satellite or complementary data to enrich the analysis). Similarly, cordon counts and/or roadside OD surveys may be core to a highway department's operational planning activities but will typically be complementary or satellite to the regional planning department's core household travel survey.

Thus, core-satellite paradigm (ideally combined with the object-oriented data model discussed in Section 3.5) allows each agency to think through the structure of its own data collection and management methods in a comprehensive, consistent manner to best meet its own analysis, modelling and planning needs. Ideally, if all (or at least many) agencies within a given urban region are working within this same framework, this will facilitate consistent, effective collaboration in data collection and sharing, with each agency undertaking components of the overall data collection task that it is best suited to do and has the most direct interest in, while at the same time providing a basis for sharing data among agencies (my core becomes your satellite, etc.).

Further, the approach permits a systematic "enrichment" of the database over time by permitting the incremental addition of new satellites (and/or complementary data) over time as need, time and resources permit. That is, it permits agencies to define priorities within their overall desired data collection program and to build their data collection program in a structured, manageable way that recognizes these priorities and short-run constraints while working in a systematic way towards long-term goals. Satellites also potentially provide the flexibility to address "hot-button issues" in a more cost-effective, flexible and timely fashion that would otherwise be possible. In particular, the satellites

---

[64] For further discussion of transit onboard surveys see the next section.

provide the opportunity for both testing new methods and/or addressing immediate issues with a quicker response/turnaround time than the five-year cycle that is typical of many current Canadian travel survey programs. Finally, the core-satellite approach provides a flexible framework for the inevitable "stitching together" of datasets derived by various actors and various time for various purposes.

7.3 EXAMPLE APPLICATION 1: TRAVEL DEMAND MODELLING

All large and most medium-sized urban regions require a regional travel demand forecasting model for a variety of short-, medium- and long-range planning applications. This model needs to be multi-modal, deal with all trip purposes, cover the entire urban region, work at a spatially disaggregate level (at least at a traffic zone level), and deal with the time period(s) of planning interest.

Figure 7.2 sketches an example core-satellite data design for a typical travel demand modelling exercise. The core survey is a large-sample home-interview survey. This survey collects key data about households and persons across the urban region, as well as details for each trip made by each person surveyed. Also shown in this figure is the large number of ancillary data concerning the road and transit systems required for model development and calibration.

Figure 7.3 uses the object-oriented data model presented in Section 3.5 to illustrate the variables currently collected in a typical Canadian home interview survey (highlighted in red in this figure). As indicated in this figure, typical Canadian home interview surveys focus on gathering detailed person trip information and selected, key person and household attributes. By limiting the survey to these core data items large sample sizes can be cost-effectively gathered that provide statistically reliable population-level estimates at traffic zone levels of spatial precision. The trade-off with this approach is that limited or no information is collected about certain types of travel behaviour (vehicle usage, walk/bike trips, HOV usage, etc.) and/or sub-groups within the population of particular policy interest (e.g., the elderly). Expansion of the core survey to include a large number of additional questions and/or sample size is generally not a practical, cost-effective option.

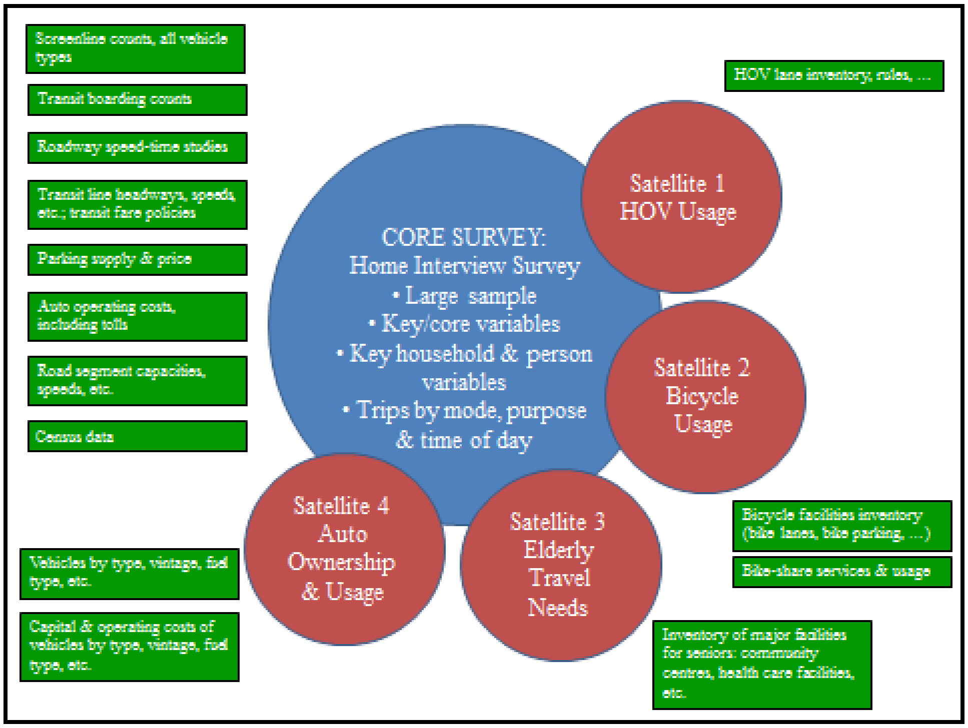


Figure 7.2: Travel Demand Modelling Data Design

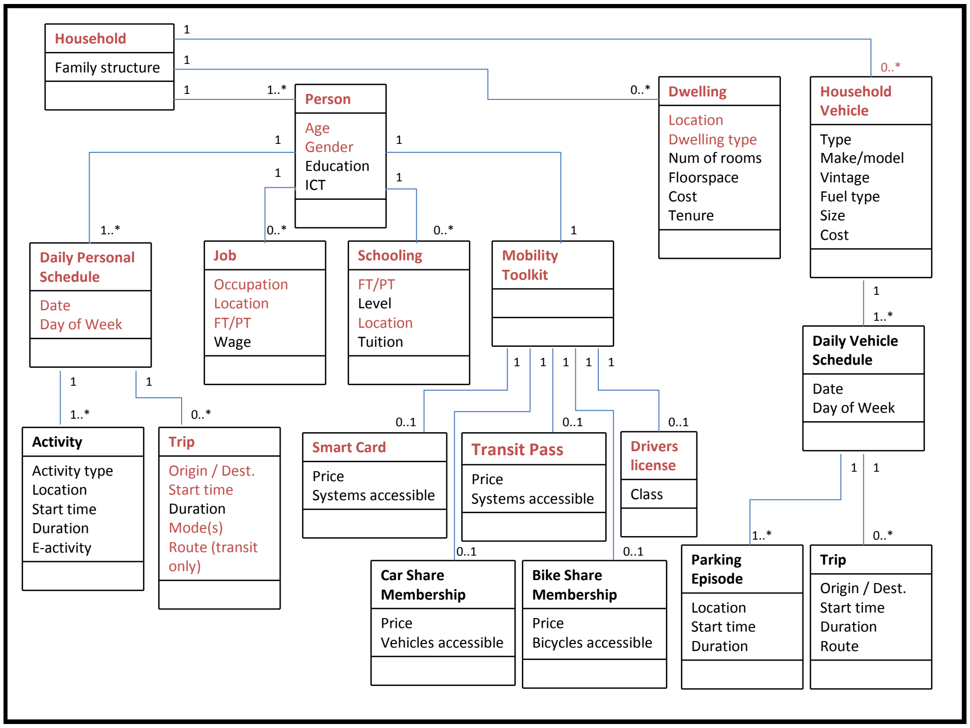


Figure 7.3: Data Collected in a Typical Core Home Interview Survey

Satellite surveys can address this problem. Smaller-sample, targeted, special-purpose surveys can be undertaken to address special modelling/analysis needs. These can either be add-on/follow-up surveys of a sub-sample of respondents to the core survey or stand-alone surveys of a fresh sample of respondents. Add-on surveys have the advantage that they can be directly linked to the core survey responses and so can be fully integrated into the core survey dataset, whereas stand-alone surveys must be fused to the core survey dataset through the use of common (overlapping) variables in the two surveys (as discussed in detail in Part III*, Data Fusion/Integration Methods*). On the other hand, stand-alone surveys can be advantageous in that they:

- Can be undertaken at a convenient point of time, independent of the core survey, in particular, after the core survey as need for the data arise.
- Permit specialized sampling frames, focused on the behaviour or sub-population of interest, to be used.
- Facilitate the use of relatively complex survey instruments.

In the hypothetical example shown in Figure 7.2, four satellite surveys are shown for the sake of illustration. These deal with HOV usage, bicycle usage, travel behaviour of the elderly and auto ownership and usage behaviour. Gathering detailed information for each of these within a general core survey adds considerably to the cost of the survey and the response burden on the respondents. Also, many questions are not relevant to all respondents in the core survey (e.g., non-HOV users, non-elderly, etc.). Well-designed satellite surveys can much more cost-effectively address specific behaviours/issues such as these relative to trying to cover them within an omnibus survey. A key in all cases, however, is to ensure that variables are collected in each satellite survey that all satellite data to be linked to core data in useful ways.

Although not shown in Figure 7.2, another type of satellite survey that can be usefully “spun off” from a core travel survey involves exploring selected behavioural or policy issues with a suitable sub-sample selected from the main sample. Examples of such issue-focussed surveys include (but certainly are not limited to):

- Stated preference/response surveys designed to explore specific elements of travel behaviour (e.g., Would people use a new transit mode if it was introduced? How do people trade time and money off in a route choice context?).
- Policy-issue specific questions (e.g., exploration of responses to a new fare policy).

7.4: EXAMPLE APPLICATION 2: TRANSIT SERVICE PLANNING

All transit agencies require data about the performance of their service on a route-by-route basis, both in terms of the service being provided (capacity, headways, travel times, reliability, etc.) and the ridership that this service attracts, in order to do short- to medium-term service planning. Figure 7.4 presents one possible core-satellite data collection design for transit service planning applications.

In this application, the core survey is a transit on-board survey that is applied on an on-going basis to transit lines within the system so that over a period of time every line is surveyed. Of particular importance is to survey a given line both before and significant service changes to the route in order to capture riders’ responses to the service change. As shown in Figure 7.5, transit on-board surveys generally capture very limited socio-economic (S-E) information concerning the trip-maker along with information concerning the current trip, but not any information concerning any other aspect of the respondent's travel behaviour.

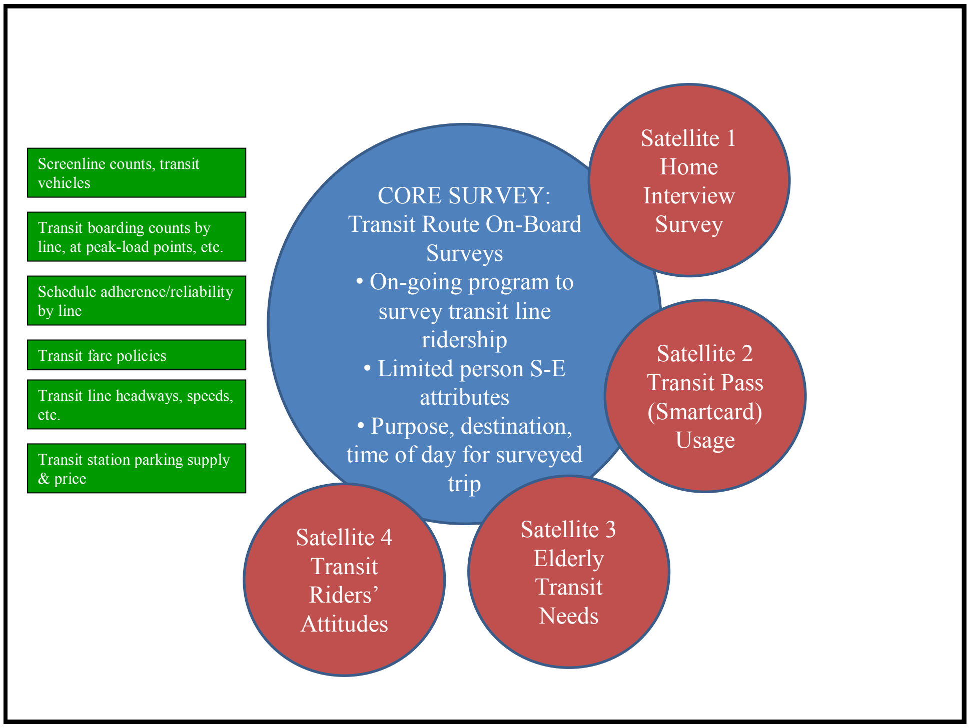


Figure 7.4: Transit Service Planning Data Design

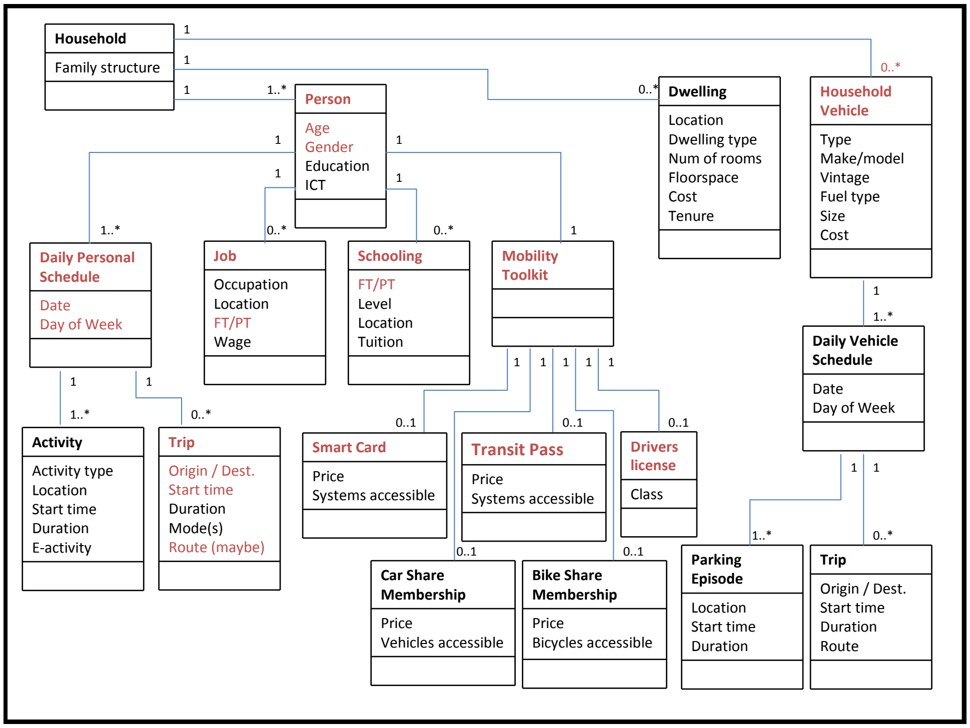


Figure 7.5: Data Collected in a Typical Transit On-Board Survey

The transit on-board survey is core to transit service planning since it provides key data about who is actually using what routes in the system, when, for what purposes. If systematically and consistently collected across the system and over time it permits the transit agency to track ridership in response to service changes, develop route-level ridership models and develop systematic procedures for assessing proposed route-level service changes over time.

What on-board surveys don't do, however, is inform the transit agency about the overall travel market; i.e., who is not currently using transit and why. Thus, a potentially very important satellite for transit service planning purposes is a multi-modal home interview survey that can provide detailed information concerning the overall travel market in the urban region and hence provide a basis for the development of mode choice models and other tools that can help the transit agency assess the potential for attracting new riders to the system. If transit route choice information is collected within the home interview survey (as is the case, for example, in both the Toronto and Montreal home interview survey programs), then this information can provide a powerful complement to the on-board survey data in terms of providing additional route-level ridership information for analysis and modelling purposes.

Comparing Figures 7.2 and 7.4, it is seen that the home interview survey has changed roles from being core for regional travel modelling purposes to being a satellite for service planning purposes. This illustrates that the context/application within which a data collection process is being designed is critical to that design. It also illustrates the way in which different agencies and their data needs can and should interact. Regardless of whether it is considered core or satellite, the home interview survey is extremely useful for a variety of transportation planning purposes and agencies, and it should be designed and administered with this multiplicity of uses and users explicitly in mind. The on-going

Montreal and Toronto home interview survey programs are excellent examples of this approach, with, in both cases, the transit agencies playing a very strong role in the design and administration of the surveys in recognition of the importance of the home interview survey data for both service and strategic transit planning.

Similarly, it should be noted that in Figure 7.2, transit boarding counts are shown as ancillary data for regional travel demand modelling purposes. It is possible, however, for transit on-board surveys to become a satellite to the core home interview survey, if sufficient socio-economic information about the surveyed transit riders is obtained to permit some level of fusion/integration between the on-board and home interview surveys. This is a particularly useful approach in urban regions in which transit ridership is a relatively small percentage of total, trips, in which case insufficient observations of transit riders may be gathered within the home interview survey to develop statistically significant models of transit ridership. In such cases, the on-board survey data can be treated as "choice-based survey data" in the development of mode choice (see Part III*: Data Fusion/Integration Methods*).

Finally with respect to Figure 7.4, the figure illustrates how other satellite surveys might be used to complement on-board survey data for a variety of purposes. Of particular emerging importance is the possible uses that might be made of smartcard data as they become increasing available.

8. A FRAMEWORK FOR URBAN TRANSPORTATION DATA COLLECTION

8.1 FRAMEWORK CRITERIA

A framework for urban passenger transportation data collection should:

- Establish the context for data collection.
- Clarify best practices.
- Provide a process to evolve the toolkit in a forward thinking way.
- Get people involved in data collection on the same page.
- Keep up with :
    - Changing personal activity patterns.
    - Changing data collection technology, methods.
    - How people interact with technology.
    - Emerging methods of travel (auto/bike sharing; new transit modes, etc.).
- Be applicable to different types of agencies
- Promote cost-effective data collection and management methods.
- Be responsive to modelling needs (usable data for modelling).
- Facilitate benchmarking/continuity over time (usable data for trend analysis).
- Be pragmatic, practical and implementable
- Allow for the collection of both fundamental, commonly used content and less frequently used information.
- Be adaptive to different spatial scale / city sizes
- Create an environment of innovation and experimentation (research needs are a critical part of the framework).

8.2 ELEMENTS OF A NATIONAL FRAMEWORK

Based on the discussion in this report a six-element program is recommended for a Canadian data collection framework. The six elements are:

- An institutional/political commitment to data collection.
- Commitment to high quality data collection and management methods.
- Adoption of a standard (object-oriented) model of urban transportation data.
- Adoption of a core-satellite design paradigm for survey design.
- Commitment to evolutionary change in data collection methods.
- Growing urban surveys into province-wide surveys.

Each of these six framework elements are discussed in turn in the following sub-sections.

8.2.1 Institutional/Political Commitment to Data Collection

As has been expressed elsewhere in this report, without information we cannot plan, and without planning we cannot address the critical decisions facing our urban areas with any reasonable hope of success. While governments at all levels face on-going budgetary challenges, the cost of collecting, maintaining and applying the data needed to inform investment decision-making (not to mention the myriad of other transportation policy decisions) represents a small fraction of the billions of dollars of investment required to maintain and expand our urban transportation systems. A sound national data collection framework begins with the recognition by governments at all levels (local, regional, provincial and federal) of the need to the support data collection programs that will provide effective, efficient urban transportation planning in this country.

8.2.2 Commitment to High Quality Data Collection & Management Methods

Clearly, if data are to be collected and maintained, this must be done using best-practice methods, and there must be a commitment to maintain and enhance this best practice over time. Given the wide range of data collection needs and methods, it is not possible to codify this best practice in a few simple statements. This project's literature review provides an overview of best-practice procedures as well as lists the many survey manuals which provide excellent guidance in the design and conduct of surveys and other data collection procedures.

Important elements in a high quality, best practice approach to data collection and management include:

- Every data collection exercise must:
    - Be rigorously designed to ensure that it is collecting the needed data in a valid and usable manner.
    - Adhere to strict statistical standards.
    - Impose quality control during data collection, coding and storage.
- A well-designed, on-going data management system to maintain and preserve the data over time must exist. This is as important as the original collection of the data, if these data are to be productively used over time.
- The process must address privacy concerns while keeping data accessible and usable.
- Detailed, up-to-date metadata concerning agency data must be maintained. All data must be properly documented.
- Data must be readily and easily accessible if they are to be useful.
- Agencies should proactively share data of common interest and design their data programs to maximize collective efforts, budgets and expertise. Agencies should also share their experiences with data collection and management so as to collectively improve the general state of practice. Working in silos and not sharing information and experience is wasteful of resources, reduces the overall utility of whatever data have been collected and limits the sum total of what can be accomplished within any given total expenditure on data.

- Similarly, best practice involves making best use of existing data wherever it exists and monitoring for new sources of relevant data that may become available over time.

Data collection and management processes must also recognize the on-going, evolutionary nature of urban transportation planning and decision-making. New issues and needs emerge over time, as do data collection methods. A well-designed data management system should flexibly adapt to incorporate new data as these become available, facilitate consistent time-series analyses and adjust to new information needs and analyses over time (FHWA, 2011).

8.2.3 A Standard Model of Urban Transportation Data

As described in Section 3.5, a "standard model" of urban transportation data can be constructed using object-oriented design concepts. While it is not essential that all agencies universally adopt this model, it does provide a comprehensive, systematic framework for defining data needs and describing the relationships among different types of data. It can provide a basis for organizing data management systems, as well as provide a starting point for model system development. It also encourages the sharing of a common "vocabulary" among practitioners, leading to better sharing of experiences and knowledge. It is readily extensible if more detailed representation of some elements of the system is required.

8.2.4 The Core-Satellite Design Paradigm

As discussed in detail in Section 7, the core-satellite design paradigm provides a very general, flexible approach to survey (and more generally, data collection) design. Given the complexity of modern data requirements, multi-instrument approaches are inevitably required, even in smaller urban areas. The core-satellite approach establishes a core, base set of data that addresses fundamental elements of the data required that is useful in its own right but which also is the foundation upon which more detailed, specialized satellite data can be integrated via the data integration/fusion methods discussed in Part III.

What is "core" and what is "satellite" depends on the agency's data needs. The methods used to collect core and satellite data can and will also vary from one application to another. Indeed, the satellite approach provides an excellent opportunity for experimentation with new methods within a systematically controlled environment. Thus, the core-satellite paradigm is a framework for data collection design that is adaptable to a wide variety of situations and contexts, rather than a specific method.

8.2.5 Commitment to Evolutionary Change in Data Collection Methods

As is clear from the discussion throughout this report, the *status quo* in Canadian urban passenger transportation data collection is not sustainable. Data collection methods must adapt to fundamental changes in travel behaviour, technology and planning needs. In particular, household travel survey methods that have been successfully employed for decades are being increasingly challenged by changes in how people use telephones (e.g., fewer land lines and much more pervasive use of cell phones), as well as the increasing analytical and modelling demands being placed on survey data to support complex planning analysis and decision-making needs.

At the same time, technological options for data collection are emerging and evolving rapidly. While many of these new technologies are very promising, they are at various levels of maturity (with associated cost and risk implications), and, like any tool, they have their strengths and weaknesses. Thus, it is very difficult to “pick winners and losers” with respect to technology at this point in time. Also, of course, the wide variety of applications and contexts that exist in urban passenger transportation data collection makes narrow specification of a single approach, or even a small handful of approaches, impossible.

Thus, the overarching nature of this proposed framework is that it describes a **process**, not a static structure. Given that data needs/applications are constantly changing (and will continue to do so for the foreseeable future), what is needed is an adaptive, flexible approach to data collection. This approach must allow methods to change over time in response to both changing needs and methods for addressing these needs. But it also must respect the need for continuity and compatibility of our datasets over time. Given this “ying and yang” (the need for both change and continuity), what is required is a process of careful (but on-going) **evolutionary** change in our data collection methods and state of practice. That is, the framework needs to be a program for creating an ability/environment/process for innovation, for creating “channels into the future”

If this evolutionary change is to occur (and occur in a cost-effective way) it will require agencies to explicitly adopt an “R&D mentality” in their data collection program. They must:

- Critically assess their data needs and how these are changing, as well as how well their current data collection methods are meeting these current and emerging needs.
- Pro-actively and systematically experiment with new methods to update/improve their data collection methods in a controlled, evolutionary way, so that these improved methods are ready for operational use as they become needed.
- Recognize that every survey without an experimental component is an opportunity wasted. Planning for next time is part of the plan this time. That is, the only way in which new improved methods will be available tomorrow is if one experiments with them today.
- Instill within the agency an ability/willingness to innovate in an environment of risk.
- Share findings with and learn from the experiences of other agencies across the country and internationally.

Elements of a process for effective, evolutionary change include:

- Finding ways to reduce risks, encourage innovation and learn from experiments. This may well involve provision of subsidies to agencies to encourage risk-taking/experimentation.
- Undertake “side by side” experimentation in which a new method is tested in parallel with a standard method.
- Coordinated experimentation with shared cost between different urban areas (while maintaining local autonomy).
- Invite local municipalities and regions to suggest opportunities for collaboration.
- Shift data collection budgets somewhat to provide some money for experimentation.

- Perhaps undertaking common cores with different satellites (to testing different things in different contexts) across several cities.
- Experimentation with new technology is especially critical if technological promise is to be turned into operational implementation.
- Universities have a lot to offer in terms of R&D, as well as data management, fusion, etc. Everywhere that university-transportation agency collaborations exist they have paid big dividends, but this is an under-utilized resource in many urban regions.

8.2.6 Growing Urban Surveys into Provincial Surveys (Towards a National Survey)

Unlike many countries, Canada does not have a national urban travel survey. This is not particularly surprising, given the constitutional constraints on federal involvement in urban affairs which have historically limited the role of the federal government in urban transportation.[65] Given this, a direct, “top-down” development of a national survey is not likely to be a feasible proposition, at least within the short- to medium-term. Nor is it clear that it is necessarily the best approach to improving data collection standards and practice across the country.

Given that all major urban regions in the country routinely undertake major household surveys, and given that, at least in the largest regions, these already encompass a significant proportion of provincial populations, a much more promising approach to “growing” data collection capabilities and improving/standardizing best practice would be to expand the current urban surveys to province-wide levels. This approach would have huge benefits to each province in terms of:

- Providing uniform travel data across the province.
- Provide smaller- and medium-sized cities and towns with vastly improved travel data.
- Eliminate “boundary” effects in the analysis and modelling of urban areas with constantly expanding development boundaries.
- Provide a framework for gathering longer-distance intercity and rural travel behaviour as well as local/urban. For example, a satellite survey could be added to the local travel core survey to gather long-distance travel.

Nationally this approach would provide an “organic”, voluntary, one-step-at-a-time approach to evolving a national data collection program that would be driven “bottom up” by the provincial and municipal organizations, who are both closest to the problem and the greatest beneficiaries of access to improved data. It would promote and facilitate collaboration and the sharing of data and experience among provinces and their constituent urban areas across the nation. And it would encourage experimentation through the spreading of risk and, possibly, pooling of funds.

---

[65] Note, however, Federal funding of urban transportation infrastructure has increased dramatically in recent years through a variety of programs and initiatives.

8.3 ROLE OF NATIONAL ORGANIZATIONS

Both the federal government and national organizations such as TAC and CUTA have important roles to play in promoting the proposed framework. These may include:

- Advocating for and/or participating in a national data collection framework.
- Facilitating the sharing and dissemination of information.
- Promoting collaborations in joint data collection experiments among local and provincial agencies.
- Funding pilot studies.
- Subsidizing experiments testing new technologies/methods.
- Supporting the evolution of provincial travel surveys.

8.4 CONCLUDING REMARKS: A WAY FORWARD

As discussed at length throughout this report, efficient urban transportation systems are fundamental to Canada's economic and social well-being, and effective, efficient data collection and management programs are essential to the planning, design, implementation and operations of these transportation systems. Adequate urban transportation data collection and management is, therefore, an essential component of on-going evolution of Canadian urban transportation systems.

Canada has historically been a leader in urban travel behaviour survey methodology, both in terms of its large household surveys and with respect to the development and testing of new technologies, such as GPS tracking devices. We are, however, at a crossroad, in which our conventional survey methods are no longer adequate to the task at hand. We must adapt, and we must invest in the evolution of new, improved methods that are adequate to meet our changing needs.

The opportunity and the capability is before us to make this investment and to collectively put Canada on the map as a leader in the urban transportation data collection area through a systematic, coherent, aggressive approach to the problem of evolutionary change, as has been outlined in this report.

REFERENCES

Applied Research Associates, Inc. (2008). *Estimation of the Representative Annualized Capital and Maintenance Costs of Roads by Functional Class: Revised Final Report*. Prepared for Transport Canada. August. http://publications.gc.ca/collections/collection_2009/tc/T22-147-2008E.pdf

Cavoukian, A. (2010). *Privacy by Design: Achieving the Gold Standard in Data Protection for the Smart Grid*. Technical Report, Toronto: Information and Privacy Commissioner of Ontario, June.

Canadian Urban Transit Association (2012). *Summary of Canadian Transit Statistics 2011 Operating Data*, Toronto: Canadian Urban Transit Association.

City of Toronto (2012). *Report of the Expert Advisory Panel Regarding Transit on Sheppard Avenue East*, Toronto: Office of the City Manager, City of Toronto, March 15. (http://www.toronto.ca/legdocs/mmis/2012/cc/bgrd/backgroundfile-45908.pdf)

Data Management Group (2010). *Transportation Tomorrow Survey, 2006. Design and Conduct of the Survey*. Toronto: Joint Program in Transportation, University of Toronto. http://www.dmg.utoronto.ca/pdf/tts/2006/conduct2006.pdf

Data Management Group (2011). *Data Management Group, Annual Report 2010*. Toronto: Joint Program in Transportation, University of Toronto. http://www.dmg.utoronto.ca/pdf/reports/dmgannualreports/an_rpt2010.pdf

FHWA (2011). *Guide on the Consistent Application of Traffic Analysis Tools and Methods*, Report FHWA-HRT-11-064, Washington, DC: Federal Highway Administration, US Department of Transportation. http://www.fhwa.dot.gov/publications/research/operations/11064/11064.pdf

Goulias, K., R. Pendyala and C. Bhat (2011). Total Design Data Needs for the New Generation Large Scale Activity Microsimulation Models, forthcoming, *Proceedings of the 9th International Conference on Survey Methods in Transport, Scoping the Future While Staying on Track*, Termas de Puyehue, Chile.

Manheim, M.L. (1978). *Fundamentals of Transportation Systems Analysis Volume 1: Basic Concepts*, Cambridge, MA: MIT Press

Metrolinx (2008a). *The Big Move*. http://www.metrolinx.com/en/regionalplanning/bigmove/big_move.aspx

Metrolinx (2008b). *The Big Move Backgrounder: Modelling.* http://www.metrolinx.com/thebigmove/Docs/big_move/RTP_Backgrounder_Modelling.pdf

Miller, E.J. (2001). *The Greater Toronto Area Travel Demand Modelling System, Version 2.0, Volume I: Model Overview*, Toronto: Joint Program in Transportation, University of Toronto, January (http://www.jpint.utoronto.ca/PDF/doc86.html).

Miller. E.J. and C. Cottrill (2012). Workshop B5 -- Multi-Method Data Collection to Support Integrated Regional Models, forthcoming, *Proceedings of the 9th International Conference on Travel Survey Methods, Scoping the Future While Staying on Track*, Termas de Puyehue, Chile.

Miller, E.J. and M. Hatzopoulou (2008). *Organizational Structures for Regional Travel Demand Modelling*, final project report to Metrolinx, Toronto: Cities Centre, University of Toronto, November.

Meyer, M.D. and E.J. Miller (2001). *Urban Transportation Planning*, Second Edition, New York: McGraw-Hill.

MTQ (2009). *Évaluation des coûts de la congestion routière dans la région de Montréal pour les conditions de référence de 2003*. http://www.mtq.gouv.qc.ca/portal/page/portal/Librairie/Publications/fr/regions/montreal/etude_eval_couts_congestion_mtl.pdf

Roorda, M. and A. Shalaby (2008). *Transportation Data Collection in the Greater Golden Horseshoe: A Framework and Priorities for Improvement*. Final report for the Data Management Group, Toronto: Urban Transportation Research and Advancement Centre, University of Toronto.

Toronto Transit Infrastructure Ltd. (2012). Toronto Transit Back on Track: Sheppard Subway Development and Financing Study. Interim Report, January. http://www.toronto.ca/legdocs/mmis/2012/ex/bgrd/backgroundfile-44984.pdf

Transport Canada (2006). *The Cost of Urban Congestion in Canada*. Transport Canada Environmental Affairs. March. http://www.gatewaycouncil.ca/downloads2/Cost_of_Congestion_TC.pdf

Transport Canada (2010). *Transportation in Canada 2010: An Overview*. http://www.tc.gc.ca/media/documents/policy/overview2010.pdf

APPENDIX A –
OO MODEL CASE STUDY – CITY OF EDMONTON

The City of Edmonton was selected as a case study to demonstrate the use of the OO model to map existing datasets and identify categories of information that could be considered gaps.

Figures A1 to A6 map available datasets for the City of Edmonton onto the object oriented model. Each data class name and data attributes for which at least some data are available from databases found for the City of Edmonton are shown in coloured text. Symbols indicate the data source, as follows:

+ Edmonton Household Travel Survey (2005)

++ Edmonton Region Commodity Flow Study (2001)

+++ Parking Inventory (conducted in house by City of Edmonton)

† Orange Text: City of Edmonton EMME Model (Brownlee et al. 2003)

††: GeoEdmonton Spatial Data

http://www.edmonton.ca/city_government/initiatives_innovation/geoedmonton.aspx

& City of Edmonton Zoning Bylaw

http://webdocs.edmonton.ca/zoningbylaw/Zoning_Maps.htm

Google Gas Stations Edmonton

City of Edmonton Traffic Camera System http://www.edmontontrafficcam.com/

‡ Edmonton Transit System data

http://etstripplanner.edmonton.ca/RouteSchedule.aspx

http://www.edmonton.ca/transportation/ets/system-maps.aspx

http://etstripplanner.edmonton.ca/BusStopSchedule.aspx

http://www.edmonton.ca/transportation/ets/transit-centres.aspx

‡‡ Edmonton Transit System Data for Developers http://www.edmonton.ca/transportation/ets/ets-data-for-developers.aspx

Δ Bikemap.net bicycle routes (limited to recreational routes) and City of Edmonton Bike Map

http://www.bikemap.net/regional/Canada/Alberta/Edmonton

http://www.edmonton.ca/transportation/bike_paths_Sept_2011_Front.pdf

ΔΔ City of Edmonton Community Walking Maps: http://www.edmonton.ca/community-walking-maps.aspx

Any data classes, or data attributes within those classes, that remain in black font are those where the City of Edmonton appears to have no available data (at least from the databases that we have investigated). This is not necessarily problematic because not all data classes and attributes may be needed or even relevant in the City of Edmonton. For example, the City of Edmonton does not currently support a bike share program, so it is not possible to collect such information anyway. Table A1 summarizes data gaps for the City of Edmonton identified with the object oriented model.

Clearly this is only the starting point of analysis that could be done using this type of approach. It would be feasible to identify the quality of data in each class (e.g. when the data were last updated, completeness of coverage over geography and time, etc.).

Table A1: Data Gaps in the City of Edmonton

Super class	Data classes missing	Attributes missing
Household	• Smart Card Access • Car share memberships • Bike share memberships	• Transit pass details • Electronic activities • Drivers license class • Trip route choice information • Detailed information about the vehicles owned by a household • Detailed information about dwellings • Job wage and school tuition information
Business Establishments	• Opportunities at establishments and opening hours • Employee schedules, activities, trips	• Attributes of jobs and employees (except num of employees) • Attributes of buildings • Some attributes of commercial vehicles • Commercial vehicle trip routes
Land Use	• Buildings, (including residential buildings and non-residential space) • Other facilities	• Parking prices and type of parking
Road System	• Car share stations • Car share vehicles • Control System	• Fueling station capacity and fuel type • Road segment grade • Intersection phasing and turning restrictions
Transit System		• Line segment maximum and average speeds
Bicycle System	• Bike share bicycles • Bike share stations	
Walk System		

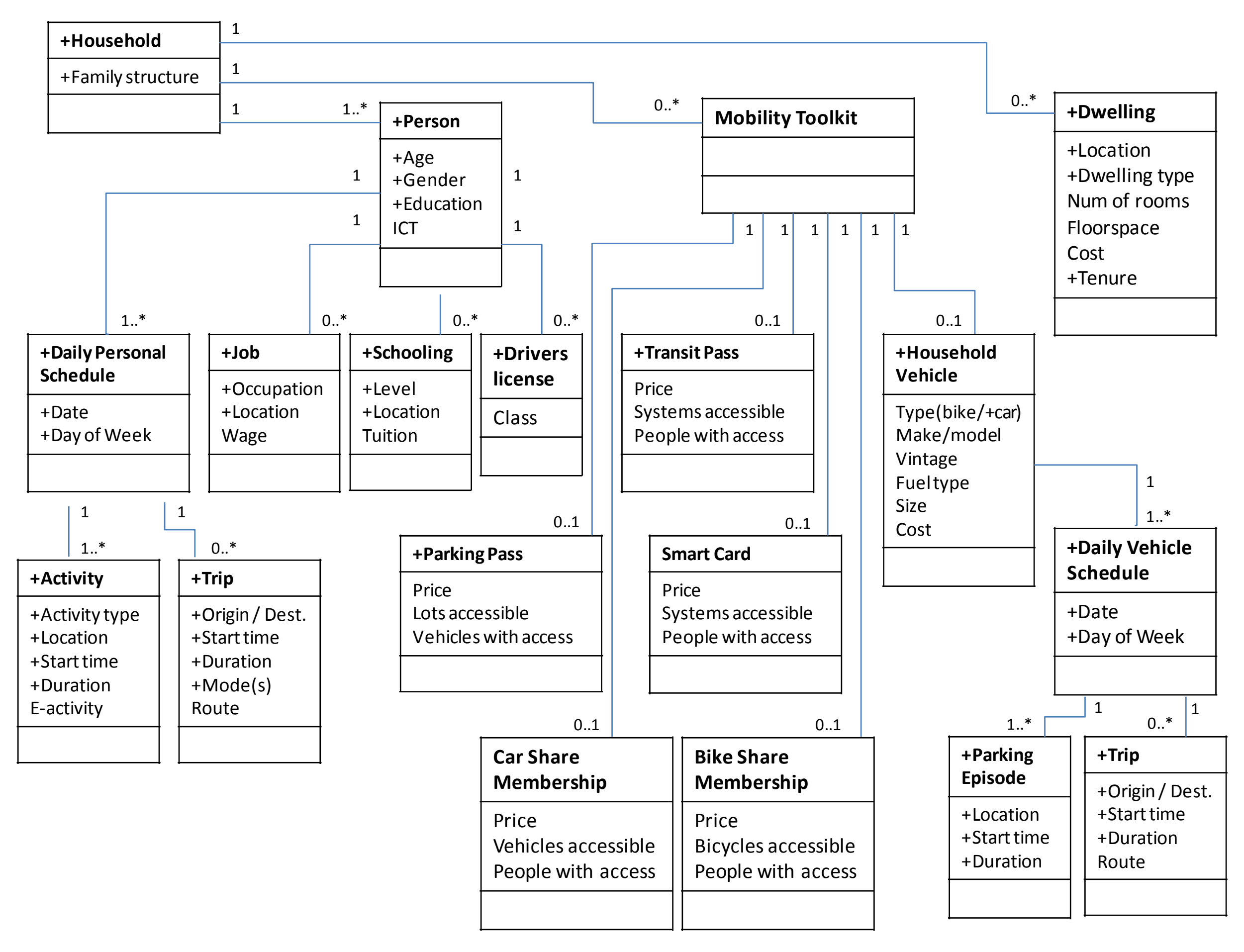


Figure A1: Object Oriented Model - Household - City of Edmonton

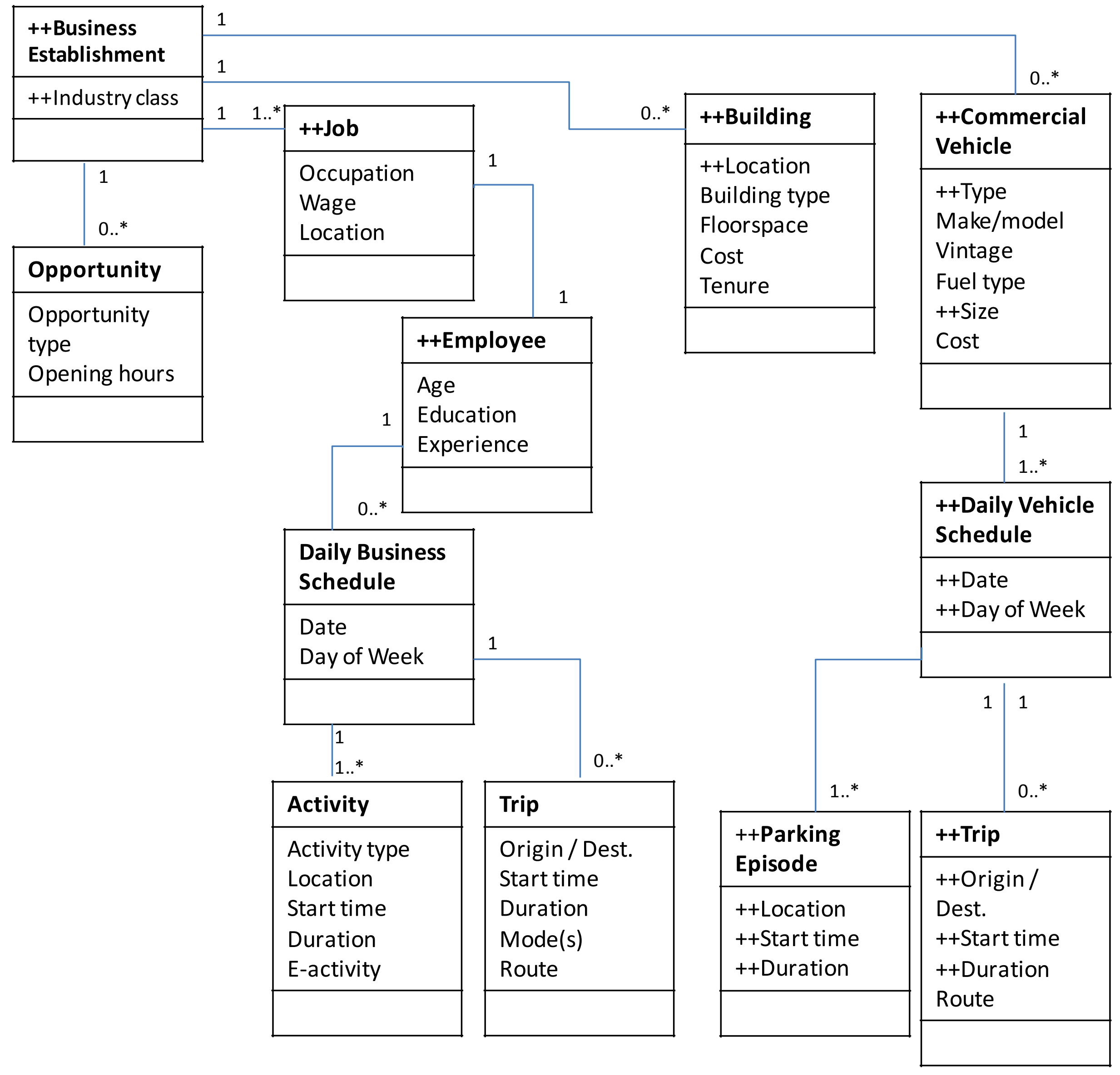


Figure A2: Object Oriented Model – Business Establishment – City of Edmonton

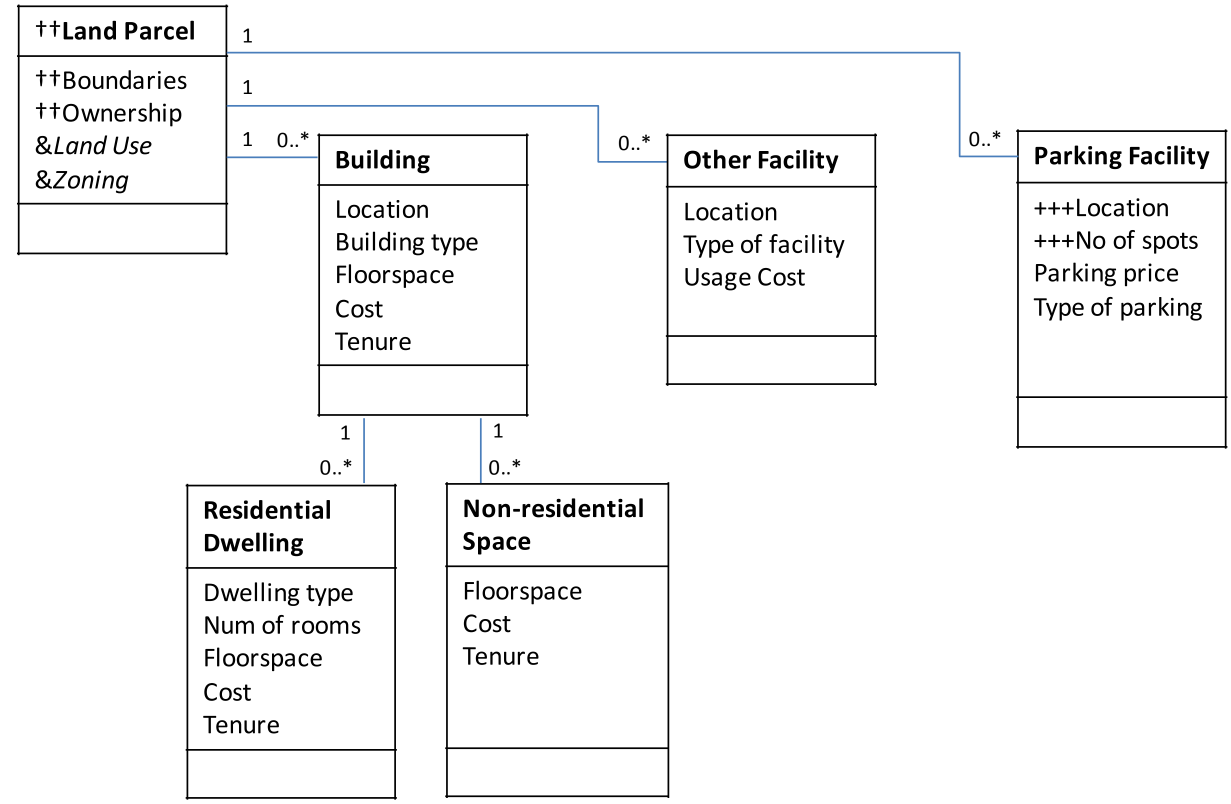


Figure A3: Object Oriented Model – Land Parcel - City of Edmonton

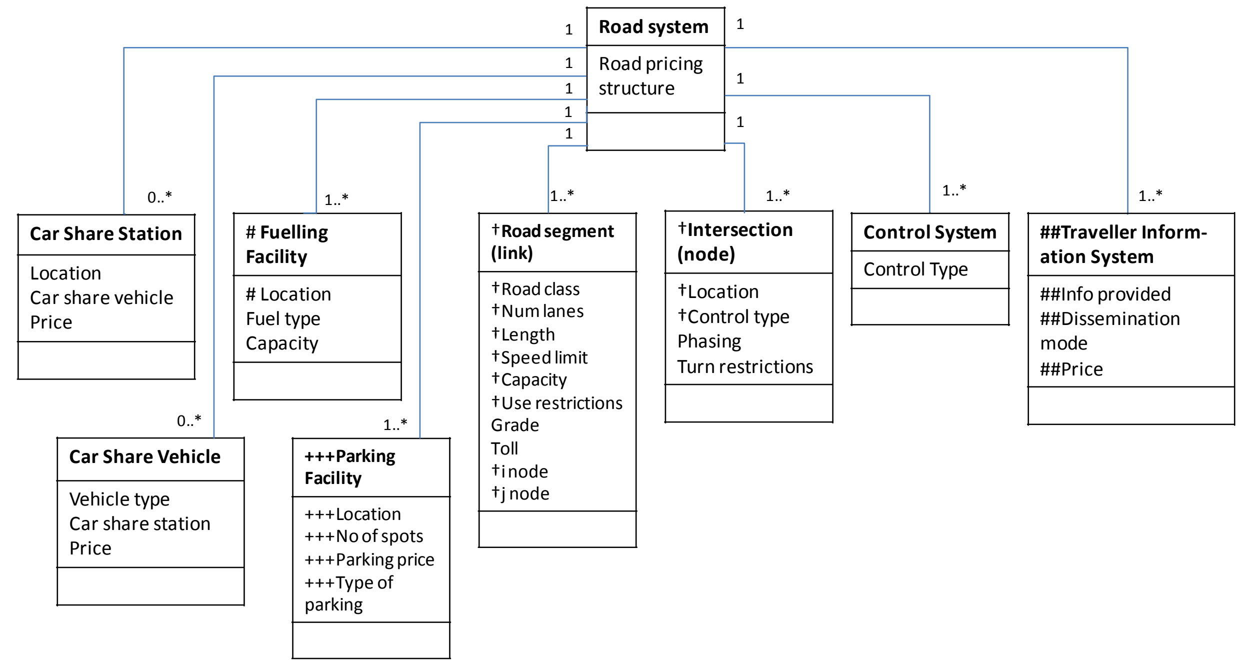


Figure A4: Object Oriented Model – Road System – City of Edmonton

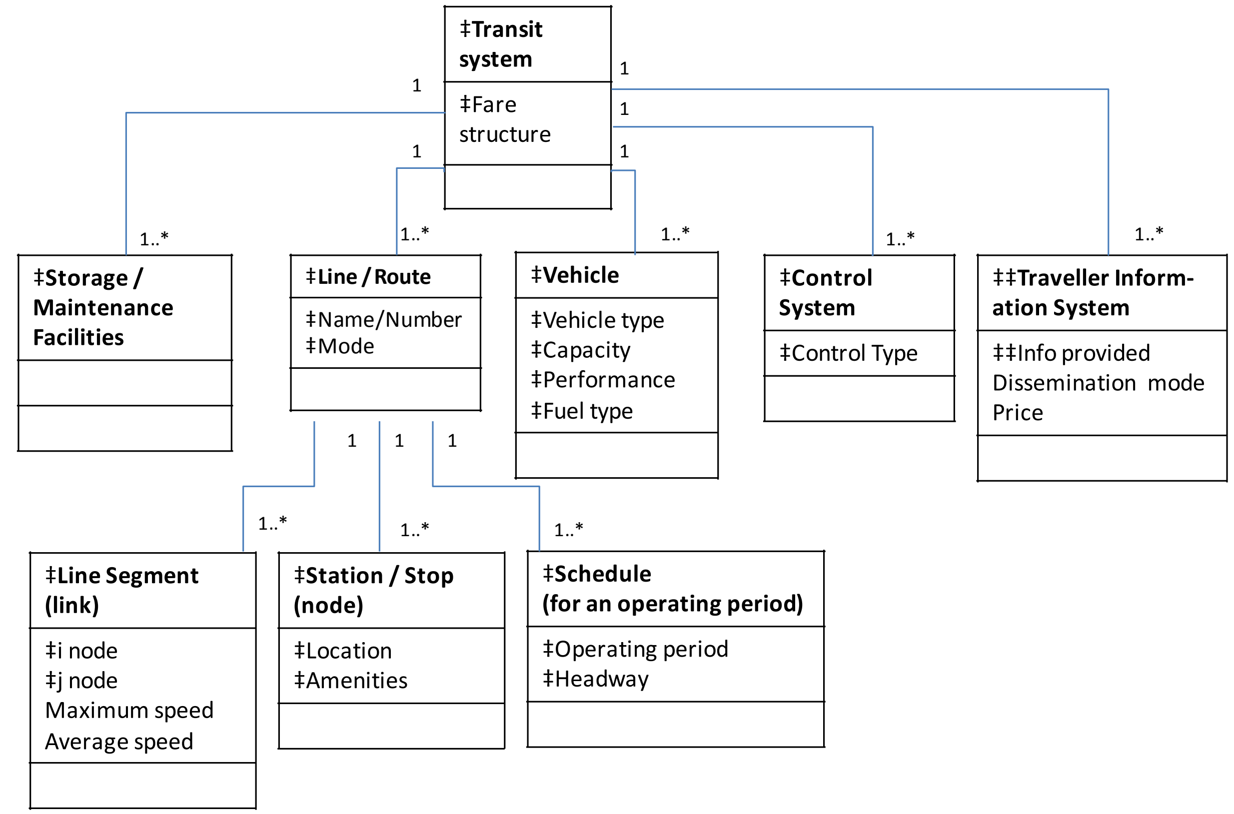


Figure A5: Object Oriented Model – Transit System – City of Edmonton

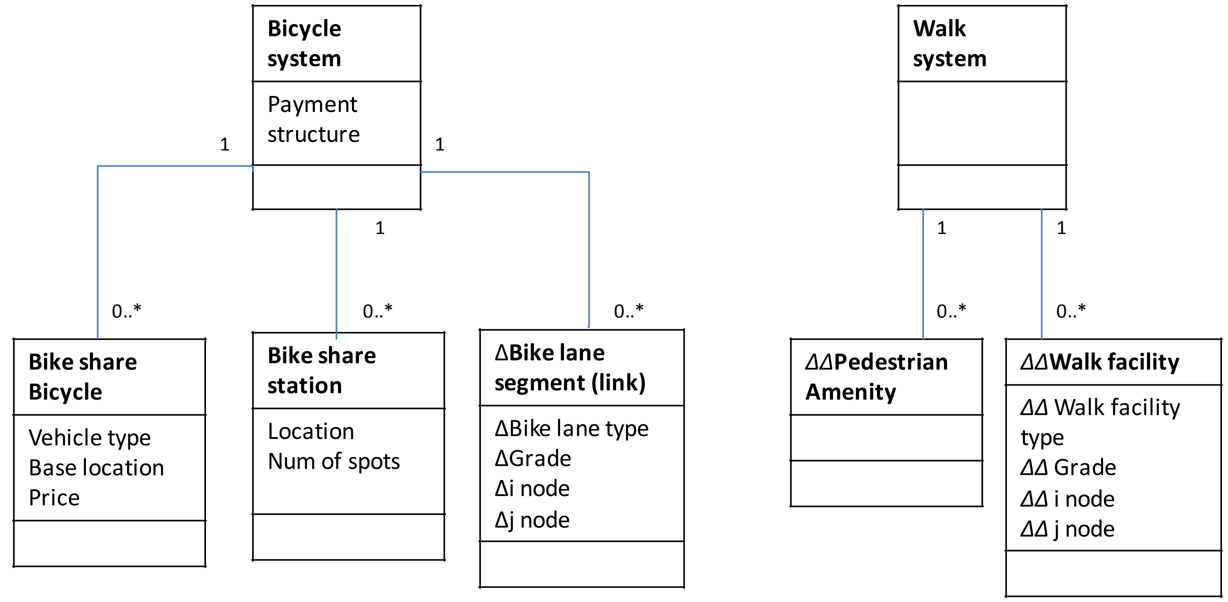


Figure A6: Object Oriented Model – Bicycle and Walk System – City of Edmonton