The Technical Scope of the Auckland Public Transport Intercept Surveys (as defined in the tender documents)

SECTION C – SCOPE OF SERVICES

1. Specification

1.1 Scope

This document provides a specification of the intended Auckland Public Transport Intercept Survey for 2006 (APTIS2006). The aim of APTIS2006 is to collect information on the travel by public transport of residents of the region, for use in the new Auckland Transport Model (ATM2).

The scope of work for the Survey Consultant **for each of the three surveys** would encompass:

- questionnaires: the content, questions and coding frame would be specified by the client; agencies would comment on these specifications and agree the final version and would design the layout and print the questionnaires;
- design of the survey methodology **aimed at achieving the targeted response rate**, and sampling strategy; creation of sampling frame and selection of sample;
- liaison/consultation with operators and other relevant stakeholders;
- pilot survey;
- execution of the final survey;
- implementation of quality control to achieve the sample data quality targets;
- coding and editing the data, including geocoding of addresses;
- provision of a survey report.

The questionnaires for both the pilot survey and the main survey and the methodology employed to carry out the survey to obtain the targeted response rates shall be formally approved by the client before the surveys commence.

1.2 Background

This survey is being conducted as part of an overall project to recalibrate and develop the existing multi-modal transportation model maintained by the Auckland Regional Council, the Auckland Transport Model. The project consists of a review of the existing model, identification of survey data requirements associated with recalibration of the model, co-ordination of the surveys and recalibration of the model.

The specific objectives of this project are:

- to update the transportation model in a timely and cost effective manner;

- to ensure the satisfactory calibration of the model to 2006 Census of Population information;
- to prepare forecast models for future years;
- to ensure that the model development work receives approval from a Transfund New Zealand appointed Peer Reviewer
- to provide full documentation with respect to the modelling;
- to ensure training and familiarity of Council staff in the use of the re-calibrated model.

The specific objectives of APTIS2006 are

- to collect sufficient information on public transport trips by train, bus and ferry to enable the development of a reasonable OD matrix;
- to expand the sample data so as to represent the population of public transport passengers on a typical weekday.

1.3 Operators

Auckland Regional Council, with the assistance of Sinclair Knight Merz, will obtain agreement in principle to the surveys with the relevant train, bus and ferry operators (there are 6 bus, 4 ferry and 1 train operators in Auckland).

The Survey Consultant will need to remain in close liaison with these operators, consulting over the detailed design of the survey, its scope, method, timing, sample etc. It is for the Survey Consultant to ensure that the survey arrangements are approved/agreed by the operators. There is a formal liaison group sponsored by ARTA, which may form an appropriate medium for consultation.

Auckland Regional Council, with the assistance of Sinclair Knight Merz, will also make arrangements for access to operator information such as: routes and timetables. The ARTA website [www.maxx.co.nz] also has much detailed route and timetable information, and Council expects to arrange access to the associated databases prior to awarding the contract.

Processing of this information to assist with interviewer scheduling will be the Survey Consultant's responsibility.

1.4 Type of Survey

General objectives

This is a survey of weekday patronage on trains, public buses (refer to Section 2 in connection with school buses) and ferries, carrying passengers within the study area and between the study area and the rest of the Auckland region.

Passengers will be intercepted on their journey and asked to complete and return a short questionnaire.

A concurrent count of the sampled population is also required. Also required is the bus sample design data: for each route/corridor/time period, the number of buses intercepted and the total number of bus services. Using this and, it is presently

intended, operators' passenger count data (ETM data) the samples will be expanded by the client to represent a typical weekday's public transport trips.

In principle the surveys should cover the full operating period of the services, but in practice it is expected only to be practicable to sample during daytime/peak commuting hours (e.g. 06:30-18:30, hours to be agreed, but it is important that the commuting peaks are encompassed). Counts of passengers outside these hours will be needed on trains and ferries. It is presently intended that this be sourced from operators ETM data; should this not prove obtainable, then these counts may be included as a contract variation.

ETM data is also required for the purposes of generating bus and ferry passenger counts outside the survey period and, for ferry, counts in the non-surveyed direction. The availability of ETM data is not yet certain, so at this time it has been assumed for the purposes of awarding this contract that it is. If this data proves to be unavailable then other means for expansion and count data will need to be explored.

The survey consultant is required to produce a survey methodology designed to obtain the required information and to achieve the targeted response rates and have that methodology approved by the client. This includes the acceptance of questionnaires.

The required outcome from this survey is reliable data to use in the transportation and land use models previously mentioned.

Advice on Survey Method

A possible survey method could be as follows (although suggestions for improvements or more accurate or cost-effective alternatives are acceptable).

- Ferries. Passengers would be intercepted at the ferry terminal as they board each ferry, and handed a self-completion questionnaire and pencil/pen; boxes for returning the questionnaires would be placed at the exits of terminals. The additional value of stationing survey staff with the boxes to retrieve forms and/or of providing the option of returning the completed questionnaires by freepost should be considered. All passengers boarding the ferries would be counted (in 15min periods). It is expected that the survey would cover one full day at each boarding point in just one travel direction (probably towards the CBD) and the questionnaire would include information to enable the client to estimate the reverse patterns of travel.
- Rail. The network is small and the choices are on-train or on-platform interviewing, which are mainly issues of cost and practicality, either method is acceptable, with concurrent passenger counts. However our expectation is that the most convenient method would be similar to that proposed for the ferries, with passengers intercepted on the inbound platforms of all stations except Britomart. Further thought is required on the most effective methods of questionnaire return.

- Bus. Because of the wide scope of the network and the large number of bus services the survey is expected to be on-bus, passengers being intercepted (and counted) as they board the bus, and the questionnaires recovered on alighting. The possible option of postal return should be considered. Unlike the trains and ferries, it seems likely to be most convenient to interview in both directions of travel with a survey unit being dedicated to a particular bus route for a survey period (half day, day etc). In this case, there is some small risk of multiple interceptions and consequent potential response bias.

Potential response biases brought about by intercepting individual passengers more than once should be considered and mechanisms implemented to address this issue. It is for this reason that it is suggested that the ferry and train surveys should be in a single direction. This is unlikely to be practical for the bus surveys, but here because of the route sampling, the repeat interception risk seems likely to be low.

Specific Objectives for the Survey Outcomes

Target samples

The sample targets for train and ferry are based on what is needed for the modelling and also what is believed can be realistically achieved with one full day's passenger interception in one direction on each route. Additional days of interviews would inevitably risk extra respondent burden and consequent response bias as regular travellers experienced multiple interception.

Evidence of recent surveys suggests that reasonably high response rates of 60-80% can be achieved if questionnaires are collected during the survey (as distinct from leaving the respondent to return it by post). If all passengers are handed a questionnaire in the survey period, then it is assumed that 40-50% of the total passengers would return questionnaires, the difference being due to passengers not intercepted at times outside the survey period or because of their age (i.e. schoolchildren).

For all the intercept surveys, a specific systematic sampling rule must be used (i.e. select every nth passenger entering the platform or boarding the bus). This may be varied by time of day etc to allow for congested times.

Successful questionnaires

Successful questionnaires are returned questionnaires that meet the minimum information criteria. For each questionnaire:

- no missing data on Key Items of information (provisionally defined in a later section);
- no more than 2 missing responses for all other items (questions).

Across the entire sample of questionnaires, there shall be:

- no more than 5% of missing responses for each other item.

The Survey Consultant shall advise on the level of quality control checks they will use to ensure the veracity of data collected in the field by interviewers. There may

be situations where such constraints would undermine the representativeness of the data, in which case they may be relaxed by negotiation (see, for example, the discussion of short trip bias in the bus survey).

Ferries

Interviews:

- Approximately 10,000 passengers use the Auckland ferry services on a typical weekday (that is, 5,000 boardings in each direction). Interviewing in one direction, the achieved target successful questionnaire sample would be 2,000 successful questionnaires (40% of 5,000 passengers). **The minimum sample requirement is 1,000 questionnaires,** covering am and pm peaks and the interpeak.
- **The methodology should demonstrate that the tenderer can meet the target of 2,000 successful questionnaires with all assumptions used stated**.
- The Council may consider increasing the sample size to 3,000 successful questionnaires and seeks a price and methodology to accomplish this.
- In all time periods, it is assumed that most passengers will be intercepted and handed a questionnaire such that an even sampling rate by time of day and service is obtained. Some systematic reduction in the sampling rate during periods of peak loading would be acceptable if significant survey economies were achieved, and the overall interview target maintained. Normally it is expected that 100% of passengers would be intercepted but at congested periods 50% interception rates would be acceptable. The client however also wishes to maximise the sample achieved, and seeks the Survey Consultant's views on what could be achieved and how it would be done.

Counts:

- full 15 minute passenger counts shall be taken during the survey period of passengers boarding each ferry in the survey direction;
- if ETM data is not available then counts may be required outside the surveyed periods and in the non-surveyed direction of travel. This would be a variation to the contract.

Trains

Interviews:

- Approximately12,000 passengers use the Auckland rail network on a typical weekday (that is, 6,000 boardings in each direction). Interviewing in one direction, the achieved questionnaire sample would be 2,500 successful questionnaires. **The minimum sample requirement is 1,500 questionnaires**, covering am and pm peaks and the interpeak,
- **The methodology should demonstrate that the tenderer can meet the target of 2,500 successful questionnaires with all assumptions used stated**.

- The Council may consider increasing the sample size to 3,500 successful questionnaires and seeks a price and methodology to accomplish this.
- In all time periods it is assumed that most passengers will be intercepted such that an even sampling rate by time of day, service and station is obtained. Some systematic reduction in the sampling rate during periods of peak loading would be acceptable if significant survey economies were achieved, and the overall interview target maintained. Normally it is expected that 100% of passengers would be intercepted but at congested periods 50% interception rates would be acceptable. The client however also wishes to maximise the sample achieved, and seeks the Survey Consultant's views on what could be achieved and how it would be done.

Counts:

- full 15 minute passenger counts of passengers entering the station platform to board a train will be taken during the survey period in the survey direction; procedures for associating these counts with individual train services are required;
- it is expected that counts outside the survey periods and in the non-surveyed direction of travel would come from the boarding and alighting information currently held by ARTA.

Buses (Commuter Service only)

The survey and its design are considerably more complex for bus services.

Approximately 120,000 passengers use the Auckland bus services on a typical day (i.e. total passenger boardings). There are approximately 250 major bus services which with minor services could be combined into approximately 100 service corridors, in each of which all services broadly serve the same areas, distinguishing express from stopping services. The client will develop a draft specification of this grouping, which will form the basis of the survey, but which the Survey Consultant is required to comment on and modify if necessary, with the client's approval, to facilitate an efficient survey programme.

Within the survey time period on each weekday, the buses make about 6,000 trips (1-way) each day, corresponding to an average daily frequency of 12.5 trips in each direction.

Interviews:

- **The methodology should demonstrate that the tenderer can meet the target of 10,000 successful questionnaires from 500 service units with all assumptions used stated.**
- **The minimum sample requirement is 8,000 successful questionnaires**
- The Council may consider increasing the sample size to 15,000 or 20,000 successful questionnaires from 800 and 1,100 service units respectively (details below) and seeks a price and methodology to accomplish this.
- It is assumed that a sampling strategy which would typically allocate an interviewing team to a bus route and that team would usually accompany a bus

route from end-to-end and the journey return; thus each survey 'unit' would be the full length of a bus route in both directions (for scheduling reasons it may be appropriate to vary this in practice, but this does not affect the overall sampling concept).

- It is assumed that the interviewing team would hand out and collect the questionnaires on the bus.
- Three levels of sampling in each route corridor are to be considered as summarised in the table below. Overall, the 3 sampling levels would involve surveying 500, 800 and 1,100 service units respectively. Given the agreed sampling targets, the Survey Consultant is expected to devise individual service samples which are spread across the survey time periods and across the individual services with each grouping, insofar as this is feasible.
- These sampling rates are given as a guide only and, while remaining broadly consistent with this concept, the sampling may be designed to achieve survey efficiency.

Sample Option	**Number of service 'units' to be sampled in each route corridor**				
	AM Peak	**Interpeak**	**PM School Peak**	**PM peak**	**Total**
A	1	2	1	1	5
B	2	3	1	2	8
C	3	4	1	3	11

- Normally it is expected that 100% of the passengers on each service unit would be intercepted but during congested periods 50% interception rates or less would be acceptable. The client wishes to maximise the sample achieved, and seeks the Survey Consultant's views on what could be achieved and how it would be done.
- Assuming a 50% overall response rate, the achieved sample for the 3 sampling levels could be targeted at 10,000, 15,000 and 20,000 successful questionnaires respectively. In fact, with questionnaires issued and collected on the buses higher response rates are believed to be achievable. Therefore these should be taken as the target samples.
- Refinement of these statistics is expected to follow during detailed design and Council may consider varying the sampling options and/or apply different levels in different corridors, depending on passenger volumes and the service complexity within each corridor. In particular, the design of the school peak survey will need further consideration in regard to its practicality and the locations where it is most relevant.
- Suggestions as to variants to the survey strategy, which would maximise survey efficiency and the successful questionnaires while achieving reasonable representativeness by corridor and time period, would be welcomed.

Counts:

- full counts of passengers boarding will be taken at each bus stop, and the time should be recorded at each bus stop;
- if ETM data is not available then counts may be required outside the surveyed periods. This would be a variation to the contract.

Persons Intercepted on all 3 Modes

It is expected that all persons sampled other than very young children to be handed a questionnaire. Persons out of scope due to age or other reasons should be separately counted. Tenderers should indicate what constraints apply here and how children will be dealt with.

There is often ambiguity as to how sampling affects people travelling together (i.e. travel groups such as families or friends), which could lead to bias in sampling. Agencies should indicate how they would be handled in the survey.

Survey Designs

The detailed design of the survey, the scheduling of interviewers etc is the responsibility of the Survey Consultant. Detailed information on the number of passenger boardings at each rail station is available from the survey day held by ARTA. The operator ETM data is the source of boarding data for each bus service and ferry route; the availability of this has yet to be established.

Appendix A contains maps showing the frequency of services.

The tenderer will advise as to whether the questionnaires should also be printed in languages other than English.

Publicity

Appropriate survey advance advertising and publicity will be managed by the Survey Consultant in discussion with ARC.

1.5 Survey Study Area

The Survey Study Area will cover the whole of the Auckland Region, but the focus will be on the area of the new transport model, ART3. This area is expected to be less than the whole region with rural areas not included, though this has yet to be finally determined.

It will be at least the same as the current ART model area, that is, including the whole of Auckland, North Shore and Waitakere Cities, the urban areas of Manukau City, the whole of Papakura District, and parts of Rodney (in the northwest and the Hibiscus Coast). The boundaries of the current model are Hatfields Beach (north of Orewa) in the north, State Highway 17, Riverhead and Waimauku in the northwest, and Maraetai in the southeast and the Papakura boundary in the south.

Possible extensions to this for the new model include: Warkworth, Coatesville, Helensville, Pukekohe, and the Bombay area.

The survey is to include passenger transport services between the current ART area and rest of the region. Tenderers should indicate any specific design issues

regarding surveys design to obtain details of public transport trips crossing the study area boundary.

1.6 Survey Period

The preliminary planning phase for the APTIS2006 started in November 2005. The draft final design will be completed in January 2006. The Survey Consultant will be appointed in late March/early April 2006. The main survey window is the period from 1st May to 20th October 2006, excluding school holidays and university holidays and exam periods.

The concept of a Pilot Survey in April 2006, being followed by the main survey commencement in May 2006 should be considered, but not if such an early start would prejudice survey quality.

1.7 Survey Biases

In self-completion questionnaire surveys of this nature, there are concerns about sampling biases and response biases. Examples are: biases in selection of route, vehicle and passengers for interview, varying propensities of people of different ages and genders to complete the questionnaires, and the likelihood that passengers travelling short distances will be less likely to return the questionnaire. Advice is requested from the tenderers on how these biases may be managed, reduced and/or corrected.

Concerning short distance trips, there was evidence of a bias in this regard in bus surveys in Auckland in 2003. Practical means of identifying and correcting the problem are not immediately obvious, but might involve ideas like:

- asking questionnaires to be returned on alighting the bus even if they have not been fully completed, and recording the boarding and alighting bus stops (using some agreed coding frame);
- sequencing the questionnaires to obtain the most important data in the initial questions;
- minimising the length of the questionnaire.

Generally survey difficulties increase with crowding and congestion and Council would like to know how agencies would manage the process in such conditions.

1.8 Questionnaire Content and Design

The APTIS2006 will obtain data for each passenger on the characteristics of the journey currently being undertaken and of the passenger. The following table gives a provisional, illustrative list. Council recognise that particularly for the bus survey this may result in a questionnaire of too great a length, and some compromise will be needed. If this proves to be insuperable, Council may consider having two shorter questionnaires with a subset of these questions on each, and each to be distributed to half the sample.

The Survey Consultant will be provided with a complete draft questionnaire on the signing of contracts to undertake the survey. The Survey Consultant shall be

responsible for checking the wording of questions and response categories to ensure compatibility with common NZ usage of terms. It shall also be responsible for the final layout and format of the questionnaires. Final approval of questionnaire design rests with the client. After final approval is given by the client, the Survey Consultant shall be responsible for the printing of the questionnaires.

The design envisages a paper-based questionnaire, but suggestions for alternative technologies may be made.

Data Item	K: Key Data Item (K): possible key data item	Comments
Intercept details: - route number - location/boarding point - time	 K K	These identification details will need to be made as far as possible consistent with operators' codes. The location identifiers must relate to the concurrent counts. One method is to number each questionnaire and maintain a record of these number sequences handed out in each 15min period.
Travel details: - Activity at start of trip - Origin address - Mode of access to this vehicle - Activity at end of trip - Destination address - Mode of egress from this vehicle	 K K K K K K	In principle, for buses, the addresses will be placed at the end of the questionnaire if the boarding and alighting bus stops can be identified. In this case some shortcomings in a address responses would be accepted.
Boarding and alighting train stations Boarding and alighting ferry terminals Boarding and alighting bus stops	K K (K)	Boarding station may be precoded. Both terminals may be coded by Survey Consultant where there are no options If feasible, both would be coded by survey staff
Outward/Return trip: - Time of outward/return trip	 (K)	
Other data: - Ticket type – Vehicle availability - Sex - Age - Licence holding status	 - (K) - - -	

1.9 Concurrent Survey Counts

At all survey sampling points, a count of the sampled passenger flow will be obtained by hour/vehicle as appropriate. The questionnaires collected must be able to be related to the counts taken which, in the case of trains and ferries, may mean that count periods would be better related to train and ferry departures.

1.10 Pilot Survey

It is anticipated that a Pilot Surveys for each of the three modes will be conducted for the equivalent of one full day of the main survey to ensure that all survey operations have been tested under anticipated workload conditions. The pilot survey will test all aspects of the final survey, and will not be confined simply to a test of the questionnaire design. If significant changes are not made to the survey methodology

after conduct of the Pilot Survey, it may be possible to add the responses obtained in the Pilot Survey to the main survey data set.

1.11 Coding and Editing

Data will be entered into databases (preferably Microsoft Access).

All destination locations will be geo-coded to x-y coordinates. Given the vital importance of the accuracy of the geo-coded data to the subsequent construction of the travel models, the quality of geo-coding must be high. An example of an acceptable level of detail in geo-coding is provided in Attachment B, based on the geo-coding methods employed in the recent Melbourne Area Travel Survey (MATS).

Agencies are asked to provide details of their geocoding capabilities, in line with the methods outlined in Attachment B, including their ability to geo-code locations for which a complete address is not provided, but a landmark name is provided (e.g. the name of a shop, a school, a public building, a transport terminal, a restaurant/café and other similar locations).

As outlined in Attachment B, all geo-coded locations must have a supplementary code attached, indicating the type of geo-coding method used to obtain the x-y coordinate. This will serve as an indicator of the quality of geo-coding processes.

The Survey Consultant need only code all origin and destination locations to x-y coordinates. The client will be responsible for the conversion of the x-y coordinates to any required zoning system (such as meshblocks, traffic zones etc).

The data entry programs will contain sufficient range and logic checks to ensure the integrity of the data. The Survey Consultant will be required to submit a complete list of checks which are tailored to the final survey design for APTIS2006.

Missing data will be coded in a consistent manner (e.g. –1 for missing data, -2 for not applicable) to clearly distinguish missing data from fields with valid zeroes (e.g. number of cars in household). No blank fields should exist in the final databases. Instead, fields intentionally left without a data entry should have a distinguishing code (e.g. –3).

Address directories assembled for the household survey will be made available to the Survey Consultant undertaking this survey.

1.12 Data Management and Control

Full records should be kept of all administrative procedures employed in the survey to ensure that information on response rates and other aspects of survey quality can be calculated and reported on an ongoing basis.

The Survey Consultant will provide a weekly progress report to the client, summarising the procedural results obtained to-date and highlighting any problems encountered with field operations. Progress reports will be due by close of business on the Monday of each week during the survey period.

In addition, the Survey Consultant will provide weekly downloads to the client of raw data collected to date, to enable SKM to undertake their own audit checks on data quality. Data downloads will be due by close of business on the Monday of each week during the survey period.

1.13 Data for Expansion

The sample data from APTIS2006 will be expanded by SKM to match the counts of passengers on the sampled services.

1.14 Deliverables

The major deliverables from the project by the Survey Consultant are:

- Microsoft Access relational database files for the administration, questionnaire and count data;
- full documentation of the survey process, in the form of a survey report;
- coding frames for all data sets.

The data sets, coding frames and documentation should be contained on a single CD-ROM.

The final "clean" data set, including full documentation of the survey, all coding frames and the data files, will be presented to the Council by 1st December 2006.

1.15 List of Attachments

Refer to the CD separately couriered

2. School Buses

As a potential variant to the survey, Council may require the survey Consultant carry out a survey of schoolchildren using school buses.

There are just over 200 services contracted by ARC, plus an unknown number of additional services chartered directly by schools.

There is uncertainty as to the feasibility of this and the most efficient methodology and Council have considered three methods on which it seeks tenderer's views:

- handing out questionnaires to schoolchildren when they board or alight the buses at the schools, with a concurrent count of the bus users; the completed questionnaires would be returned by post or via the school:
  - the perceived advantage of this method is its interception efficiency;
  - it would be necessary only to obtain data from a sample of the schoolchildren, but it would be important to cover all buses insofar as practicable;
  - our concerns here are the difficulty of attracting children's interest as they arrive at school, their acceptance of formal sampling strategies and, in both cases, the potential for very low response rates;
- positioning an interviewer on the buses to collect the data by direct interview and counts; if a limited passenger sub-sample of interviews was conducted on

each bus service, then perhaps the interviewer time involved and their travel of the bus may be efficiently limited; this short interview might be feasible using hand held computers;

- arranging for travel to school questionnaires to be completed at school (in a previous Wellington survey an attempt was made to sample one class at each age in each school, which was reasonably successful, although not all schools/classes were prepared to participate).

In either of these methods, there may be issues concerning the differences in the use of these services to and from the schools. While it is not anticipated doubling the survey scope, where the survey applies in one direction only it may be useful to collect some supplementary data on the reverse direction.

All methods would require agreement by the schools, and in the process details of charter services would be collected.

The questionnaire/interview content would be:

- Intercept details (recorded by the interviewer): route number, location/school, time of trip;
- Travel/child details: home address, age/school class level.

Given the survey uncertainties, a target sample size would be negotiated, influenced by survey practicalities.

At this tenderers are asked to indicate their views on how best to undertake this survey, and to give a broad indication of whether the $50,000 allowed for as a Provisional Sum for this work is appropriate to meet the likely cost and sample size associated with their preferred approach. These views will only be considered as part of the overall methodology in assessing tenders.

Depending on the outcomes of the tenders received the ARC reserves the right to re-tender this portion of the survey.

Attachment A

Maps showing frequency of Services

These are included on the CD referred to in Cl 1.15

Attachment B Geocoding and the Use of Landmarks

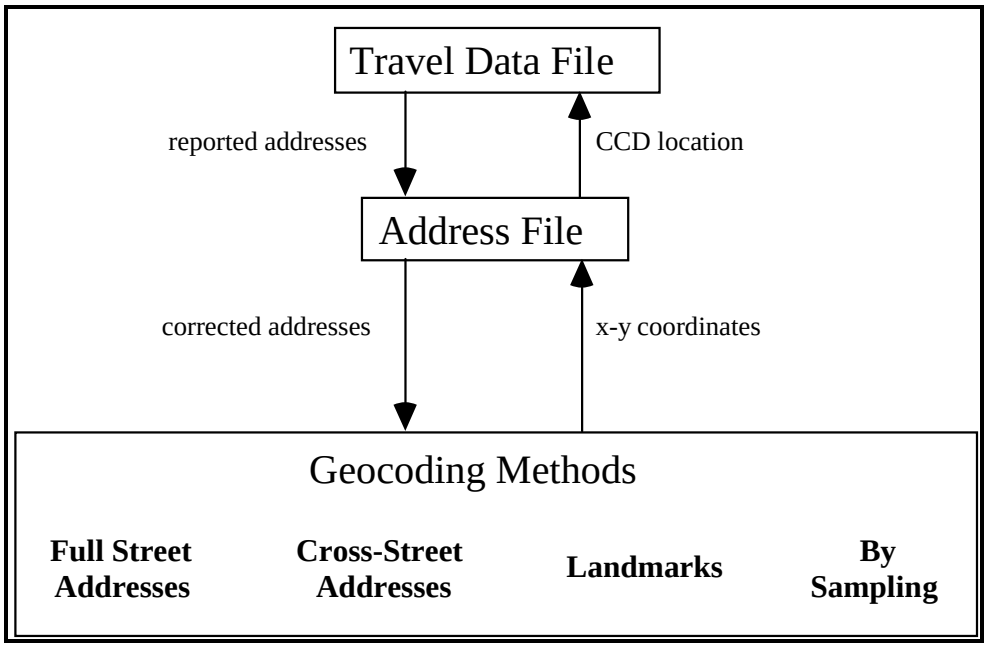
(**Note that the geocoding for this survey will need to be compatible with ARC's systems and those of it regional partners. These include the ESRI suite of GIS products including ArcGIS Desktop, ArcView, ArcIMS, ArcSDE. This means that the geocoding for this survey will need to be in a format allowing importation into the ARC GIS system. The ARC GIS system is an ESRI based GIS suite (ArcGIS 9.2), and x-y coordinates will need to be in New Zealand Map Grid (NZMG) projection).**

9.12 B.1 Introduction

In past travel surveys, destination locations have often been coded directly to rather aggregate traffic zones, at suburb level of detail, with the result that considerable information has been lost about the location of destinations. However, coding survey data to x-y coordinates and then to the level of the Census Collectors District (CCD) (meshblock in the NZ context) is extremely useful for the plotting of trip information, for more accurate calculation of distances between destinations, and allows greater flexibility for the design of more specific zoning systems (e.g. for the analysis of public transport corridors).

The most efficient method of coding locational information for large-scale travel surveys, is the direct geocoding of street names and suburbs (and other principal locations). This method involves the use of a computer program whereby the user enters an address, the computer checks through a list of such addresses (in much the same way as a coder would look through a street directory index), and then returns an x-y coordinate for that location. The computer program should be able to match coordinates with incorrect addresses (caused by misspelling of street, or the use of adjacent suburb name) and should take account of house numbers, especially on long streets.

A general procedure adopted for the geocoding of locations is shown below.



Geocoding may be performed at two points in the data coding process; either interactively as the data is first being coded, or in batch mode after all the data has been entered. While interactive geocoding is somewhat more complex, it has the major advantage that assigning geocodes to locations as the data is first entered enables a wider array of data editing to be performed interactively, thus allowing the data coder to detect and correct problems as they occur, rather than leaving them to compound throughout the remainder of the data coding process. The following description of the geocoding process applies equally to both methods, but is often worded in the context of batch mode geocoding. For interactive geocoding, simply assume that the “address file” contains only one address at a time.

Locational information is obtained from the travel data files in the form of reported addresses. These addresses may be from the sample frame database of residential addresses in the case of the household file addresses, or from the respondents, in the case of the stop file destination locations. These addresses are transferred to an address file which contains only the address and an identifier which enables the geocoded CCD location to be transferred back to the travel files at the end of the geocoding process. The locational information, especially from respondents, is of varying degrees of completeness and accuracy. Therefore, before attempting to geocode the address information, the addresses have to be corrected to put them in a format which is compatible with the GIS database of address coordinates. These corrected addresses are then geocoded by one of various methods of geocoding, as described below. The x-y coordinates of the addresses are then transferred back to the address file. By comparing these coordinates with the CCD (meshblock) boundary files, the CCD in which the address is located can be obtained, and this CCD number is then transferred back to the travel data files.

The geocoding procedure consists of a series of geocoding methods applied in a hierarchical fashion to obtain a likely geocode for an address. The accuracy of the geocode is dependent on the geocoding method used. Therefore, the more reliable methods are attempted first.

The degree of accuracy of the geocoding depends on two factors; the accuracy with which the respondent can supply the locational information, and the accuracy with which the GIS program (in this case, MapInfo©) can use that information to generate a set of coordinates. For example, a respondent might know that they went shopping at the Safeway supermarket in Ringwood. From their point of view, this is the most accurate description of their destination. However, whether MapInfo© can geocode this location correctly will depend on what information it has about the location of Safeway supermarkets. If all Safeway supermarkets are in a landmarks datafile, which gives the geocodes for named locations, then this should provide a very accurate geocode. However, if they are not in a landmarks file, then the very accurate locational information provided by the respondent will be of little use, unless an alternative method of locating Safeway supermarkets can be found. For example, it is possible to look up the Yellow Pages (either the paper version or the electronic version on CD-ROM database) and find that the Safeway supermarket in Ringwood is on the corner of Smith and Brown Roads. This information, in that form, is still not very useful since MapInfo© needs a street name and number to find a geocode. However, as will be described later, it is possible to write a special module which finds geocodes based on the specification of cross-streets. Therefore, the accurate locational information supplied by the respondent can eventually be converted into an accurate geocode. On the other hand, the information that MapInfo© is most accurate in working with (i.e. full street name, number and suburb) is often not easily supplied by the respondent. For example, very few people would know the street number of the Safeway supermarket in Ringwood, even if they knew what street it was on. If they provided only the street name, then Council would be forced to select a random position along the street within the suburb - providing a less accurate geocode than that provided by use of the shop name.

These geocoding principles have been employed in several recent travel surveys, including the Victorian Activity & Travel Survey (VATS), the South-East Queensland Travel Survey (SEQTS), the North-Eastern Suburbs Travel Survey (NESTS) in Melbourne, and the Melbourne Area Travel Survey (MATS). In the actual computer implementation of the geocoding methods, four program modules were developed for the MATS project. These are:

- geocoding using MapInfo© cadastre files;
- geocoding using a cross-street database;
- geocoding with a landmark file; and
- geocoding by sampling.

The next few sections will discuss how addresses are prepared to make them suitable for geocoding and then details of the four geocoding program modules mentioned above will be provided.

9.13 B.2 Preparation of the address data

A crucial factor in geocoding is the success of matching the address information (i.e. street name and suburb name) provided by the respondents to that used in the electronic reference maps. Slight differences in spellings result in a mismatch and consequently a geocoding failure.

Steps were made to minimise spelling mismatches in the MATS data (which used interactive geocoding at the time of data coding) by providing a pop-up dictionary of street names, suburb names and postcodes in the data entry program for the travel data. Using a pop-up list of street names and suburbs avoids the problems caused by different abbreviation of street types (e.g. Rd vs Road), and by reversals of suburb names (e.g. East Doncaster vs Doncaster East).

9.14 B.3 Geocoding full street addresses

The geocoding of full street addresses in MATS was performed using a database of address geocodes based on a cadastre file of all properties in the study area. A cadastre file contains the property boundaries of all properties, linked to an address and centroid coordinates. Cadastre files are available for all of Australia and New Zealand in MapInfo format from an variety of data retailers. The addresses and centroid coordinates were extracted from the cadastre file into an Access database, and functioned as a giant lookup table (with about 1.6 million records for all of Melbourne).

It is quite common that respondents give incorrect suburb information and so the address cannot initially be geocoded. This, however, is often circumvented by assuming that respondents are likely to give a suburb not far from the correct suburb. Respondents often upgrade their suburb to a nearby, more socially distinguished, suburb. By using this assumption, success in geocoding can be improved by re-attempting to geocode using an increasingly larger boundary file. Postcode boundaries are generally larger than suburb boundaries and so they are used in the geocoding process after the suburb boundary, by allowing the user to search within the postcode boundary of the nominated suburb, rather than using the suburb boundary. In this way, the street address was often found in a neighbouring suburb.

Once an x-y coordinate had been attached to an address, the CCD in which this coordinate was located was found by overlaying the boundaries of the CCDs on the geocoded coordinates. The region (CCD) in which the geocoded point was located was then transferred back to the travel data files as the most disaggregate description of the location of that address.

9.15 B.4 Geocoding cross-street addresses

For many types of destination, the best information that the respondent can provide in the nearest intersection. To enable geocoding using this information, a database of cross-street geocodes was constructed by finding the points where different streets intersected in the MapInfo database of streets. The data coder could then specify one of the street names (in a particular suburb) and the program would then show all intersections with that street in a pop-up list, from which the cross-street (and its geocode) could be selected.

9.16 B.5 Geocoding landmarks

It was allowable in the MATS survey for respondents to nominate a landmark as a destination address. Examples of landmarks include the name of a restaurant, a school, a bank, a government office, a shopping centre, a park, a beach, etc. To be effective as a valid address, a landmark has to be qualified to identify it uniquely from all others with a similar name. A bank, for example, needs to have the branch (usually a suburb) appended to its name.

The geocoding of landmarks requires the compilation of information from various sources such as Mapinfo features files, telephone books and street directories on a variety of landmarks, such as:

- schools, pre-schools and childcare centres
- universities and colleges
- shopping centres
- food outlets
- sporting centres
- places of interest
- parks, ovals and reserves
- caravan parks
- hospitals
- ambulance stations
- police stations
- fire stations
- churches
- bus and airline offices
- ferry terminals
- post offices
- public libraries
- council offices
- bays and beaches
- boat ramps
- theatres and cinemas
- hotels and motels
- commercial buildings
- racecourses
- golf courses
- bowling centres
- swimming pools

For the MATS survey, a landmark file with approximately 50,000 entries was assembled.

For each of these landmarks, an equivalent full street address or cross-street address was obtained from these sources, and then the geocoding methods for full street addresses and cross-street addresses (described earlier) were used to generate the geocodes.

Not all landmarks are easily identified by a street address. Finding an address for a landmark poses a problem in cases where one is not available and/or the area covered by the landmark is large (e.g. beaches and parks). For such large areas, the area centroid may represent a more appropriate definition of the location to be used as a geocode. Centroids of areas can be marked and geocoded in a MapInfo map, from which the geocode coordinates can then be extracted.

Using the landmarks for geocoding within MATS proceeded in a hierarchical four-level fashion using dropdown menus. For example, in order to find the geocode for a specific greengrocer, the coder would select Retail (at level 1), Food Store (at level 2), Greengrocer (at level 3), and then the specific greengrocer identified by name, street address and suburb (at level 4). Once a level 4 selection has been made, the geocode coordinates are automatically returned to the coding program.

9.17 B.6 Geocoding by sampling

Addresses provided by respondents are not always complete. Some respondents intentionally omit street numbers or just indicate their suburb or locality - possibly for privacy reasons. The approach that was used to geocode these cases in the MATS survey was to sample a point along the length of the street, if a street name was given, or to sample a point within a suburb, if a suburb was all that was available.

9.18 B.7 Recording the type of geocode

Because of the different accuracies of the various types of geocoding procedure, it is important to record which method of geocoding was used to find the coordinates for a location. This information helps later when editing the data by indicating the level of confidence that can be placed in the geocoded location. The types of geocoding employed in MATS (in decreasing level of precision) were:

- Survey Address
- Landmark
- Cadastre – suburb
- Cadastre – postcode
- Cross-street – suburb
- Cross-street – postcode
- Nearest Landmark
- Street – Suburb
- Street – Postcode
- Town – Suburb
- Town – Postcode
- Manual Geocoding
- Interstate
- Overseas