

Technical Memo

To: SHRP2 - C10 Tri-Agency Project Implementation Files
From: Bhargava Sana
Date: June 17, 2015
Subject: Transit Demand Data Standards - Working Draft

Introduction

The main purpose of this document is to propose a standard format for transit demand data as in input to Fast-Trips (Python version of [FAST-TrIPs](#)) for dynamic transit assignment. The raw output data from the existing activity-based models (ABM) needs to be restructured and reformatted before being sent as input to Fast-Trips. The feedback of Fast-Trips to the ABM will be dealt with in Task 6 - Software Implementation and Task 8 - Implementation & Testing.

Two related tasks helped inform the needs of the demand data standards. The first is *Task 4 - Transit Rider Behavior*, which involves estimating route/path choice models to be used by Fast-Trips during assignment. This means that all the potential route choice model explanatory variables pertaining to travelers (primarily demographic attributes) need to be obtained from the ABM and passed on to Fast-Trips along with transit trip information. The second task that will inform data standards is *Task 8 - Implementation & Testing*. The proposed data standard needs to take into account the outputs from both SFCTA and PSRC AB models.

This document first investigates and evaluates various approaches that have already been undertaken and then proposes a format and contents for transit demand. Appendix A contains a description of the proposed file format.

Background

This section reviews various approaches to demand file formats and then discusses the type of information that should be included in the demand file based on relevant findings from the Task 4 literature review on route choice.

File Format Approaches

This section describes three existing approaches for standardized demand being fed into dynamic network models: (1) origin-destination (OD) matrices; (2) activity file; and (3) trip list. The current version of Fast-Trips uses the trip list approach.

Origin-Destination Matrices

The Transportation Authority used OD matrices when it recently implemented a dynamic traffic assignment (DTA) model using INRO's Dynameq. Dynameq input was formatted as OD matrices in 30-minute time slice by vehicle class (auto, truck, etc.). Details about this project can be found [on the project website](#). The benefit of this approach is that trips with similar origin-destination-time-period combination can be efficiently combined and routed together. The primary drawback is that each additional market segment creates a 2X size increase in the number of matrices that need to be created and stored, which becomes inefficient at a very low number of market-segments. Additionally, 30-minute time windows may be too lenient. These can be narrowed, but at a storage and efficiency cost.

Activity File

Another approach, used in TRANSIMS, is the activity file, which consists of a chronologically ordered list of activity locations of all travelers. This was used in the SHRP2-C10A ABM-DTA integration project in Jacksonville. Further details about the effort can be found [in the final report](#).

Figure 1: TRANSIMS Activity File Example

HHOLD	PERSON	ACTIVITY	PURPOSE	PRIORITY	START	END	DURATION	MODE	VEHICLE	LOCATION	PASSENGER
1	1	111110	0	9	1	44520	44519	1	0	5937	0
1	1	11111	4	9	45480	57360	11880	2	1	13688	0
1	1	11121	0	9	58500	97140	38640	2	1	5937	0

The advantage of the disaggregate file is that person and activity-specific information is readily available, as are the durations of the activities that are associated with the travel. Activity-level variables such as 'priority' could be particularly important if travel becomes exceedingly difficult, as the traveller may decide to drop the activity in real life. There are two primary disadvantages: not all travel models readily produce activity- rather than trip-data; and it requires the software itself to piece together the activities into travel components. Additionally, it is not as legible a format without prior understanding (i.e. is the mode and vehicle associated with the trip before or after that activity; what about the start and end times?).

Trip List

The final documented approach is the trip list, which is what the current version of Fast-Trips uses, as specified in the [user's manual](#). The trip list file, `ft_input_demand.dat`, contains record for every passenger trip to be routed during the transit assignment.

Table 1: Existing FAST-TrIPs Demand File Format `ft_input_demand.dat`

Field Name	Details
passengerId	ID that uniquely identifies a passenger
OrigTAZ	Origin zone (TAZ)
DestTAZ	Destination zone (TAZ)
mode	Transit mode (e.g. 3 - walk to transit)
timePeriod	Demand time period (e.g. AM, PM, OP)
direction	Binary value indicating the passenger's direction. Outbound from home: Direction = 1 Inbound to home: Direction = 2
PDT/PAT	Preferred arrival/departure time in minutes after midnight. If direction==2: Preferred departure time (PDT) If direction==1: Preferred arrival time (PAT)

Demand file information

This section discusses the type of information that should be included in the demand file based on relevant findings from the Task 4 [literature review of transit route choice models](#) and discussions regarding how Fast-Trips would be implemented in various jurisdictions.

In addition to basic trip attributes that constrain the route choice problem to a specific origin-destination-time window, the team discussed a variety of approaches for dealing with park and ride lots. Specifically, the proposed specification should be flexible to whether the park-and-ride lot was constrained to a single lot in the ABM, or left to be determined by the network model.

At a minimum, the team should test explanatory variables that define the market segments already used within respective ABMs. These include income, household autos, and household workers. The latter two are combined to form an auto sufficiency calculation relating the number of autos to the number of workers. The literature review documented a range of potential explanatory variables and found that age, gender, and income were particularly significant in route choice model specifications. The following is a list of explanatory variables that warrant further exploration in Fast-Trips route choice estimation, although many implementations will be ultimately limited to the market segments that were already set up in the ABMs.

Explanatory variables:

- Value of Time
- Purpose
- Auto ownership
- Income
- Gender
- Age
- Children in household by young (can't take care of self), middle (can't drive) and older (can drive)
- Elders in household (65+ years)
- Household size
- Worker status (FT, PT, unemployed)
- Disability status

While real-time Information was identified in the Literature Review as an important variable for determining route choice, given that smartphone penetration is at about 75% in the US and that there are host of apps providing transit real-time information, this is no longer being considered as an explanatory variable.

Proposed Transit Demand Format

The following section discusses the proposed file format and file contents of the proposed Fast-Trips transit network demand. A full listing of the proposed format can be found in Appendix A.

File Format

List versus Matrix

Using transit demand in OD matrix form would be more efficient for assigning person trips between same OD pairs that belonged to the same market segment. However, considering that Fast-Trips works at the individual person level and the level of disaggregated demographic data available from ABM, **we propose that transit demand from ABM be provided as input to Fast-Trips in the form of a trip list** rather than OD matrix form. This format is in many ways consistent with the existing input demand format for FAST-TrIPs and would also enable us to pass along as many traveler demographic attributes as needed.

File Type

While the current file format is tab-delimited, **we propose a CSV-based demand format** that is more consistent with the GTFS structure of the transit network format. There shall be a header in the first row with case-sensitive variables names.

Separate or Joined Trip, Person, Household Data

Since Fast-Trips will be using information at the trip, person, and potentially household-level, there is potential to reduce repeated information by having separate files for trip, person, and household level data that could be joined within Fast-Trips by using Person and Household IDs. **We propose that we store all the person- and household-level attributes separate files to avoid redundant data.** The person and household files are optional in the event that the Fast-Trips specification does not use any data from that level, such as for visitor trips or other auxiliary demand.

File Contents

Ideally, Fast-Trips should be flexible enough to accommodate different route choice model specifications and consequently also accommodate different input data items. Input transit demand should serve two purposes:

1. Provide trip details: provide important characteristics of the trips so that Fast-Trips can assign them to the transit network
2. Provide route choice model explanatory variables: these include socio-demographic and other traveler characteristics that might influence route choice

Trip details are fairly straightforward and can be enhanced starting from the input demand file used by the existing version of FAST-TRIPs. Clarifications and proposed changes are as follows.

Mode can be specified as a tour or trip mode, so long as there is a valid specification for the modal hierarchy of networks modes in the Fast-Trips control file. Mode as defined here will specify what parameters are used for the path-building and valid network modes (including access and egress modes). Modes that are not valid for the purposes of transit assignment (e.g., “drive-alone”) can be discarded/ignored.

Purpose is an activity-level variable rather than trip- or tour-level. However, many models assign a single purpose to an entire tour (i.e. SF-CHAMP). The purposes of activity at both the origin- and destination- end may have an impact on the route choice decision-making. The proposed purpose field can include any segmentation of purpose that is deemed appropriate for the purposes of segmenting the route choice model. So long as there are corresponding parameters specified in the controls file, Fast-Trips can accommodate both a more general purpose of “other” and a very detailed “work-to-social”. The decision and logic behind what purposes are chosen can be specified in the demand creation control file.

Time of day for both trip departure and arrival times are available at the minute level from Soundcast. However, SF-CHAMP only assigns broad time periods. The demand conversion tool will have to assign each trip a time-of-day value to the minute level. In

the existing version (FAST-Trips) the time rigidity was decided by the field `direction`, which specified which “tour-half” that particular trip was on. Outbound direction meant that arrival times of the trips were more important and vice-versa. The proposed format keeps both arrival and departure times from ABM and has an additional column called `time_target` to let Fast-Trips know which time to give more importance to during path choice. The intent with the explicit specification of `time_target` is to allow the ABM to make the decision about whether the departure or arrival time is more important and to give that information to Fast-Trips rather than having the rule within Fast-Trips itself.

Park and Ride (PNR) lots are used during drive-to-transit trips. Some ABMs have a robust choice process that directly outputs the lot that they say should be used as part of the drive to transit path (e.g., Soundcast), while others (e.g., Travel Model One) allow the network model to choose the PNR lot. SF-CHAMP takes a middle-of-the-road approach by narrowing down the PNR lots to a few likely candidates before allowing the network model to make the final choice. The Fast-Trips demand specification must be able to accommodate all of these approaches. The proposed demand specification has an optional field, `pnr_ids`, that contains a comma-delimited list of PNR stations that will be passed to Fast-Trips to constrain the PNR station choice set. CHAMP currently could output two to three PNRs for each drive-transit trip whereas the specific PNR is already chosen by ABM in SoundCast.

Several socio-demographic attributes of travelers are available from the ABM to be used as explanatory variables in the route choice model in Fast-Trips.

Value of Time as specified in dollars per hour can be used in assignment and is already available in both Soundcast and SF-CHAMP at the trip level.

Disability Status will be used to determine if the path needs to be wheelchair or walker accessible. Passengers using wheelchairs and walkers will also have differing effects on dwell time compared to passengers at large. While passengers utilizing walkers will need to have access to seats, passengers utilizing wheelchairs need to be able to access one of the wheelchair slots in a bus. Disability status can be simulated from the variable called “DIS” (Disability recode) in ACS PUMS. Other disability status related variables in PUMS can help classify passengers into wheelchair and walker access statuses.

Special Cases

Auxiliary Demand such as visitor trips in SF-CHAMP do not have the type of disaggregate variables available from the microsimulation-based ABM. However auxiliary demand still needs to be assigned within Fast-Trips and skimmed for feedback into the trip-based visitor model. Most of the data available in the trip list file can be derived from or inferred from the auxiliary trip tables. Since the data in the

person- and household- files is optional, this can be left blank so long as the route choice model specification for that trip purpose (i.e. visitor) does not use these variables.

Appendix A - Demand File Format

Files

Filename	Description	Requirements
trip_list.txt	All trip characteristics necessary to assign travel to the network and trip-level explanatory variables.	Required
person.txt	Person-level explanatory variables which can be joined to trips by <code>person_id</code> .	Optional - based on specification
household.txt	Household-level explanatory variables which can be joined to trips by <code>hh_id</code> .	Optional - based on specification

Format

While the current file format is tab-delimited, **we propose a CSV-based demand format** that is more consistent with the GTFS structure of the transit network format. There shall be a header in the first row with case-sensitive variables names.

File Contents

trip_list.txt

One record for every trip. If mode is not in list of valid transit modes set in the Fast-Trips parameters, it will be skipped. All fields are required unless specified as optional.

Field Name	Details	SFCTA	PSRC
person_id	ID that uniquely identifies the passenger. Use 0 to identify trips that do not have a disaggregate person-record associated with them.	Available	Available
o_taz	Origin TAZ (int)	Available	Available
d_taz	Destination TAZ (int)	Available	Available
mode	Mode (string) Entry must match a valid specification in Fast-Trips for route choice parameters and modal hierarchy. Invalid entries for transit (i.e.	Available	Available

	<p>“drive_alone”) will be ignored. If applicable, entries should encapsulate the access and egress modes undertaken.</p> <p>Example entries:</p> <ul style="list-style-type: none"> ● walk-local_bus-walk ● PNR-commuter_rail-walk <p>Example main modes:</p> <ul style="list-style-type: none"> ● local_bus ● premium_bus ● light_rail ● commuter_rail ● street_car ● ferry <p>Example access or egress modes:</p> <ul style="list-style-type: none"> ● walk ● bike_own ● bike_share ● PNR ● KNR 		
purpose	<p>Trip purpose (string) Can include any segmentation of purpose that is deemed appropriate for the purposes of segmenting the route choice model so long as there are corresponding parameters specified in the controls file, Fast-Trips can accommodate both a more general purpose of “other” and a very detailed “work-to-social”.</p> <p>Example entries:</p> <ul style="list-style-type: none"> ● work ● school ● personal_business ● shopping ● meal ● social ● work-based ● other ● visitor 	Available	Available
departure_time	Departure time in HHMMSS from midnight.	5 time periods available. Each	Available

		trip needs to be assigned exact time.	
arrival_time	Arrival time in HHMMSS from midnight.	5 time periods available. Each trip needs to be assigned exact time.	Available
time_target	<p>Arrival/Departure preference indicator (string).</p> <p>Valid entries:</p> <ul style="list-style-type: none"> ● arrival = arrtime is more important (for assignment) ● departure = deptime is more important 	We could come up with some logic to determine this	We could come up with some logic to determine this
vot	Trip value of time (\$/hr) (float) Continuous value.	Available	Available
pnr_ids	<p>(Optional) A comma-delimited list of stations within brackets (list of int) Example: [1219, 3354, 9485] An empty list or empty field implies any accessible PNR can be used.</p>	Available (2-3 stations)	Available (DaySim considers all PNRs and chooses one)

person.txt

Person-based explanatory variables which is optional based on the details of the implementation. All fields are optional depending on the implementation.

Field Name	Details	SFCTA	PSRC
person_id	Person ID used to link to trip_id	Available	Available
hh_id	Household ID used to link household information. Use 0 to identify people that do not have a disaggregate household-record associated with them.	Available	Available
age	Continuous value (years) (int)	Available	Available
gender	(string) valid entries: <ul style="list-style-type: none">● male● female	Available	Available
worker_status	Worker status (string) valid entries: <ul style="list-style-type: none">● unemployed● full-time● part-time	Available	Available
work_athome	Works at home? (boolean) valid entries: <ul style="list-style-type: none">● False = No● True = Yes	Not available	Available
multiple_jobs	Works 2+ jobs? (boolean) valid entries: <ul style="list-style-type: none">● False = No● True = Yes	Available	Not available
transit_pass	Transit pass ownership (boolean) valid entries: <ul style="list-style-type: none">● False = no transit pass● True = transit pass owner	Not available (Come up with a simple model to predict this)	Available
disability	Disability status (string) valid entries: <ul style="list-style-type: none">● none● wheelchair● walker	Available in PUMS (needs to be pulled into synthetic population)	Available in PUMS (needs to be pulled into synthetic population)

household.txt

Household-based explanatory variables which is optional based on the details of the implementation. All fields are optional depending on the implementation.

Field Name	Details	SFCTA	PSRC
hh_id	Unique ID of household used to link to person and trip	Available	Available
hh_vehicles	# of HH vehicles in working order (int)	Available	Available
hh_income	Annual household income (int)	Available	Available
hh_size	HH size (int)	Available	Available
hh_workers	# of HH workers (int)	Available	Available
hh_presch	# of HH young children (0-4 years age) (int)	Available	Available
hh_grdsch	# of HH grade-school children (5-15 years age) (int)	Available	Available
hh_hghsch	# of HH high-school children (16-17 years age) (int)	Available	Available
hh_elders	# of HH elderly persons (65+ years age) (int)	Available	Available