Technical Report
Congestion Monitoring System Automated Data Collection Project
PSRC Project #96204003

INITIAL DESIGN DOCUMENT FOR THE TRAFFIC DATABASE FOR THE PUGET SOUND REGION

by

Mark Hallenbeck
Director

Washington State Transportation Center (TRAC)
University of Washington, JD-10
University District Building
1107 NE 45th Street, Suite 535
Seattle, Washington 98105-4631

Prepared for

Puget Sound Regional Council

February 1998

TABLE OF CONTENTS

LIST OF FIGURES

TRAFFIC DATABASE FOR THE PUGET SOUND REGION

INTRODUCTION

A number of PSRC analytical tasks require traffic data to describe the use and performance of regional roadways. This type of data is also often requested from PSRC by private companies and member agencies. Although measurements of the use of Puget Sound roadways are routinely collected by a variety of agencies (primarily cities, counties, and WSDOT) within the region, these data are not always readily accessible to the PSRC. However, the agencies that collect the data store much of this information electronically.

This report documents the design of a central database whose objective is to provide PSRC (and others) with easy access to the traffic data already collected in the metropolitan region. Electronic copies of collected traffic statistics will be obtained from member jurisdictions, reformatted, and geocoded within PSRC's geographic information system (GIS). These data will then be available to PSRC staff and member agencies as needed.

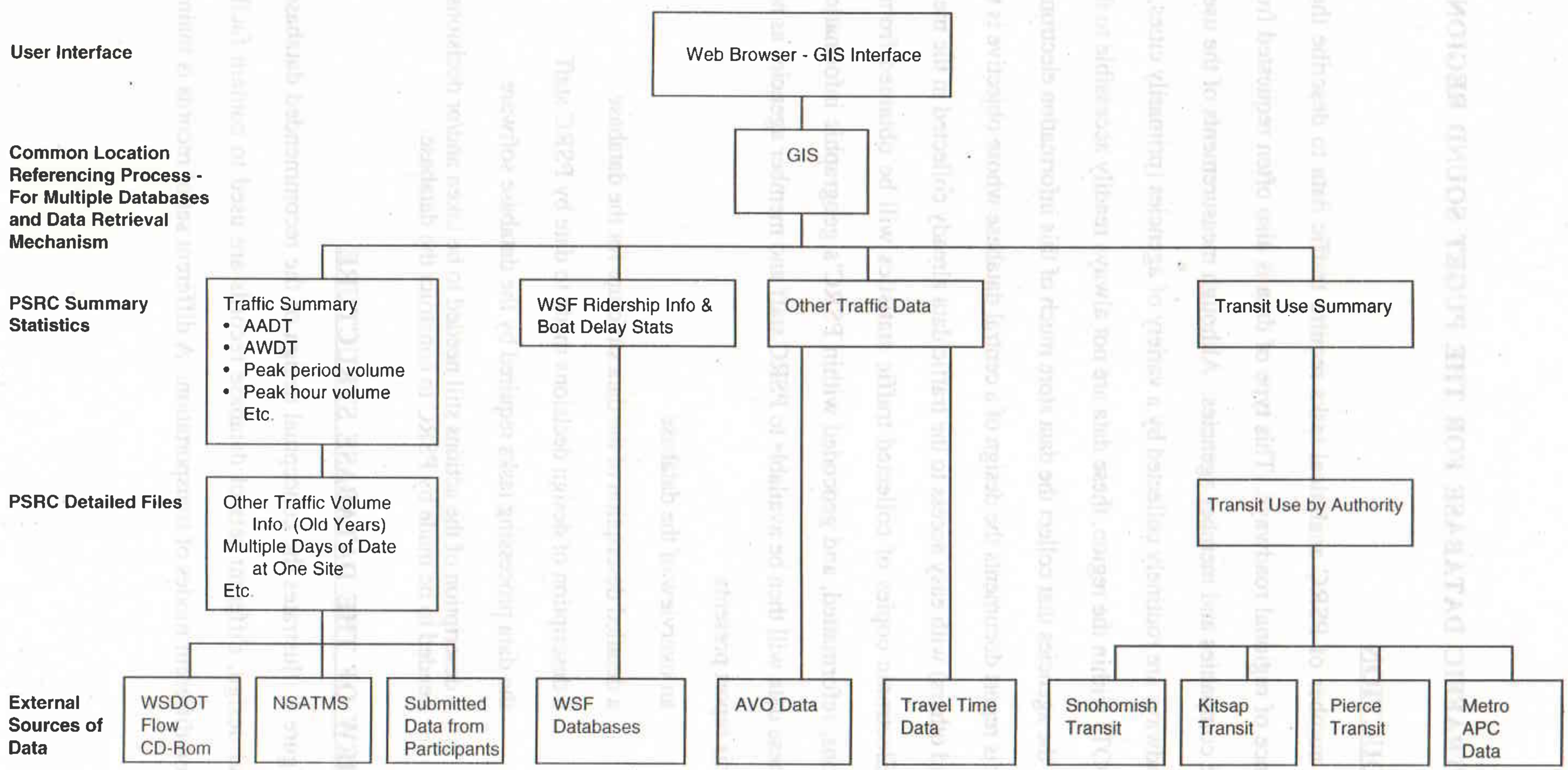
This report presents

- an overview of the database
- a detailed description of the data structure for the database
- a description of design decisions made to date by PSRC staff
- the data processing tasks required by the database software
- a description of the actions still needed to be taken and/or decisions still needed to be made by PSRC to construct the database.

OVERVIEW OF THE DATABASE STRUCTURE

Figure 1 illustrates the conceptual design of the recommended database. In this conceptual design, different sets of database records are used to contain facility usage statistics on different modes of transportation. A different set of records is maintained for

Figure 1. Conceptual Design of PSRC Traffic Databse



each transportation mode.[1] These different database records are linked via the PSRC Geographic Information System (GIS) system.

This database design will allow the system to be expanded to meet new requirements by accommodating the addition of new database records as needed and by providing the GIS links that will allow those new data collection points to be associated with the existing database records. In most cases, the GIS linkage will be based on the location (the roadway segment) for which the performance measure was taken.

Each type of performance data will reside at one of several levels of aggregation, starting with the measurement device used to collect the data and the jurisdiction that performs that function. For example, the Washington State Department of Transportation (WSDOT) collects traffic volumes at 20-second intervals on the freeway system. These data can then be aggregated to represent 1-minute, 5-minute, 15-minute, hourly, daily, seasonal, and annual traffic flows.

Agencies that collect traffic data determine the lowest level of aggregation that will be stored. (For example, currently WSDOT only stores traffic data in 5-minute increments, even though it collects data in at 20-second increments.) PSRC will have access only to the levels of data that the participating agencies store and maintain. However, in most cases, PSRC does not need data at the most disaggregated levels for which data are available.

Maintaining data at these more disaggregated levels creates large, unnecessary database files that require additional resources to create, process, and maintain. Therefore, PSRC should only obtain and store only those levels of traffic (and other modal) data that are routinely required for PSRC analyses. If special projects require more disaggregated data, these additional data can be obtained from the original data collection agency on a special case basis.

---

1 These separate records can also be combined in some cases if combining them would make database design, construction, and maintenance easier.

Discussions with PSRC staff indicate that the vast majority of PSRC's data needs can be met with hourly and higher aggregations of data. However, even hourly data are more detailed than is needed for most PSRC analyses, which tend to require aggregated statistics such as peak hour, peak period, and average daily estimates.

Consequently, PSRC should maintain two levels of traffic data, Summary Statistics (the "upper level" of the database) and Submitted Estimates (the "lower level" of the database). The Summary Statistics section of the database will contain data values that PSRC staff most commonly use. The Submitted Estimates section will contain all of the data provided to PSRC by participating agencies.

This design will also help solve another problem faced by PSRC, that of what happens when more than one data value is available for a single location. Because PSRC will have many locations for which more than one traffic count is available, it will have to differentiate between these different measurements. In most cases, PSRC staff simply need a single traffic value. In other cases, they need to examine the variability in the traffic volumes measured at a given site. The "two tiered" approach to the traffic database will allow a managed approach to both of these requirements.

When a single, "official" value is needed for a given road segment, this value will be contained in the "Summary Statistics" level of the database. Thus, the "upper level" of the database will meet the majority of PSRC needs. However, when more detail is needed, PSRC staff will be able to examine the "lower level" records, to review other traffic counts taken for the same road section. This approach will keep users from having to sort through multiple data records to determine the "correct" traffic statistic to use (this task will already have been done as part of creating the "upper" level of the database), while also preventing two separate data users from viewing the same data but choosing different summary statistics.

Thus, creating a summary database level should both reduce the time PSRC staff require to obtain traffic estimates (because they do not have to sort through multiple data records) and help ensure PSRC's consistent use of traffic statistics. Finally, a minor advantage of this approach is that the upper level of the database will be physically smaller than the lower level. (It will contain fewer records.) The smaller size should reduce data access time and improve system performance for most data requests.

Figure 2 summarizes the types of traffic data that will be stored in these two levels of the PSRC database. Specific variables are discussed later in this report.

Note that PSRC is not advised to duplicate the large database systems that participating agencies (such as King County Metro and the Washington State Department of Transportation) maintain. Instead, only summary statistics should be obtained from these sources. When data at a greater level of detail is required, PSRC should make special data requests. A good example of this is the WSDOT CDR system, which stores and provides access to freeway data on CD-ROM. This system can be used whenever detailed traffic data are needed. Only summary data (AADT, AWDT, etc.) should be loaded into the PSRC database.

The remainder of this report deals only with the upper two levels of traffic volume and transit ridership data shown in Figure 1.

The traffic data required for the database will include the data that can be obtained from the respective PSRC members and the subsequent summary statistics produced from those raw values. The data recommended for inclusion in the "upper" (readily available to PSRC staff) level of the database include the following statistics:

- AADT (average annual daily traffic - total both directions)
- AWDT (average annual weekday traffic - total both directions)
- peak hour volume (directional)
- AM peak period volume (6 - 9 AM)
- PM peak period (3 - 7 PM)
- count location and date information
- truck percentage information.

Figure 2: Summary of Database Records

	Location References	Date / Time	Traffic/Transit Variables[1]	Other
Upper Level of the Traffic Database (Summary Statistics)	Road segment ID Street name Direction of travel indicators	Count year Count date Length of longest count	AADT, AWDT, ADT, Directional weekday traffic, Peak hour traffic (AM/PM), Peak period traffic (AM/PM), Truck percentages, Number of buses	Adjustment factors Multiple count indicator ID of user that makes manual adjustments
Lower Level of the Traffic Database (Submitted Statistics)	Road segment ID Street name Cross street Directional reference from cross street Direction of travel indicators	Count year Count date Count duration	AADT, AWDT, ADT, Directional weekday traffic, Peak hour traffic (AM/PM), Peak period traffic (AM/PM), Truck percentages, Hourly traffic volumes (by direction)	Source of Estimate Adjustment factors Axle count versus loop count indicator
Upper Level of the Transit Database (Summary Statistics)	Road segment ID Street name Direction of travel indicators	Count year Count date	Daily transit ridership Number of buses (daily, peak hour) Peak hour ridership (AM/PM) Peak period ridership (AM/PM)	
Lower Level of the Transit Database (Submitted Statistics)	Road segment ID Street name Cross street Directional reference from cross street Direction of travel indicators	Count year	Daily transit ridership Number of buses (daily, peak hour, peak period) Peak hour ridership (AM/PM) Peak period ridership (AM/PM)	Source of Estimate

[1] In the upper level of the traffic database will be values computed or selected by PSRC. (That is, PSRC may provide an AADT value for locations where a city only supplied an ADT value. In addition, PSRC will need to determine a single AADT value to report when more than one jurisdiction submits statistics for a given roadway segment. Only this one value will be present in the upper level of the database. However, all submitted estimates will be present in the lower level of the database. In the lower level of the database only those values submitted by participating agencies will be stored.

The following transit variables will also be stored in the top level of the database:

- peak hour ridership
- peak period ridership
- daily ridership.

Unlike traffic volumes, transit ridership values will represent average Fall weekday conditions. This is because the nature of the transit ridership counting program does not lend itself to the calculation of average annual days. (Two factors limit the calculation of average annual conditions. The first is incomplete sampling of buses on weekend days during the Fall and Spring schedule periods. The second is the fact that the summer schedule period is too short for an accurate sample to be taken with Metro's existing automated passenger counting (APC) system.) Because most PSRC policy work is concerned with weekday transit usage, the Fall weekday ridership estimates have been chosen to represent "normal" transit ridership. These data are best used with and compared to average annual weekday traffic (AWDT) values stored in the dataset.

Geographic Stratification

The GIS basis for the database determines both the overall size of the database and the level of detail that will be provided to users. At the same time, it limits the amount of data that participating cities and counties need to provide to PSRC.

PSRC has decided that the network will contain all road segments in the Metropolitan Transportation System (MTS). This includes all principal arterials in the four-county region, as well as all other roads of "significance for inter-regional travel."

Each of these roads will be divided by PSRC staff into specific road segments. Every segment will be assigned "anchor points" that uniquely define that segment. Traffic and transit volumes will be considered constant throughout these segments. This means that when data entry takes place, a process will have to arbitrate between two alternative count values (traffic volumes or transit counts) when more than one count exists for that road segment. A recommended procedure for this task is described in the Data Input subsection of the Data Processing Steps section of this report. In addition, the data input step must correlate the location referencing system the various participating agencies use for their traffic database to the road segment identifiers the PSRC uses. Although this task will be fairly difficult, by limiting the database to the MTS network, the size and scope of this effort is considerably smaller than it might otherwise have been. (Note that nothing in the design of this database will prevent the addition of more road segments if PSRC decides to expand the scope of the database.)

The exact nature of the translation between the traffic databases and the PSRC's GIS will depend on the referencing systems used by the jurisdictions that collect and store the traffic data. PSRC has hired a consultant to examine linear referencing systems in the Puget Sound region with the eventual goal of establishing a common base of geospatial transportation data that could be used by a range of jurisdictions. If a common referencing system is established and adopted by the participating jurisdictions, the process of placing the count locations into the PSRC GIS will become quite simple. However, this type of integration is not expected in the near future.

Number of Years Stored

The traffic database should store six years of traffic and transit information (including the current year). Data older than six years should be archived and removed from the database. The archive process is described under the heading "Increment the Year Function" in the Data Processing Steps section of this report.

Data older than three or four years become suspect in their ability to represent current traffic conditions, especially in fast growing metropolitan areas. However, five-year-old data do serve as a good comparison value that will allow PSRC to measure changes in traffic conditions over time. Holding data for six years will provide a two-year "window" (since many locations are not counted every year) from which trends of at least five years can be computed.

Vision of the Retrieval Process

Two basic data retrieval mechanisms are suggested for the database, a text-based query system and a map-based query system. Each summary level record in the traffic database will be associated with a specific roadway segment (see previous section). The roadway segment will be the primary mechanism for retrieving the data.

Map-based retrieval is expected to be the primary mechanism by which data will be extracted from the database. In this vision, the user would access a GIS map. The map would display the MTS street network, with the individual road segments color coded to indicate the presence of (or lack of) traffic volume and/or transit ridership data. (Note that the user should be able to define the type of data desired—AADT, AWDT, peak period volumes, etc.—that would then be used to define the color coding on the map.) The user would select the road segments for which traffic data were desired by highlighting them. A dialog box would then ask for the type of output to be provided (which specific variables, whether the output should be directed to the screen, a file, or the printer).

Text-based retrieval would be the alternative method for accessing database information. The current vision of the text-based retrieval system is that the user would specify a type of data to be obtained, and the database system would respond with a list of road segments (sorted by county, city within county, and then alphabetically within each city) for which that type of data was available. (Multiple lists might be available depending on data age limits specified by the user.) Users would then select the road segments for which they wanted data. A dialog box would then request the variables to be output (if other variables for those locations were also to be printed) and whether the output should be directed to the screen, a file, or the printer.

Caveat Files

It is important that all data retrieved from the database be accompanied by a disclaimer that adequately describes what the data represent. This is particularly important with the transit data because changes in route structure and the fact that Metro's transit ridership numbers represent unlinked[2] trips mean that false conclusions can easily be reached if the data are used improperly. For example, when transit routes in the northern end of Seattle switched to a hub and spoke route system, the number of unlinked trips increased. However, because more riders had to transfer between buses, the number of linked trips really did not change appreciably (at least early in the new route structure). Similar types of errors can occur when routes shift from one street to another, producing what looks like a significant change in transit ridership if the ridership on a single street is reviewed before and after that service change. (For example, say routes changed from University Way NE to 15th Ave NE. If only ridership on University Way NE was examined, transit ridership would appear to have declined significantly, while if only ridership on 15th Avenue NE were examined, ridership would appear to have increased. In reality, no significant change in ridership occurred.)

For traffic data, an important caveat to report will be that most of the peak hour and peak period estimates for arterials in the PSRC database are based on a limited number of short duration, weekday counts. They are not based on continuous counts, and therefore, they do not represent either average annual or peak annual conditions. (That is, they are only estimates of average peak hour conditions, not a true measure of the "average" condition or the "worst of the year" condition.) Factors are not currently computed that would allow these estimates to be adjusted to represent "average annual" peak conditions. However, these values can be used with the transit ridership estimates in the PSRC database to provide a reasonable estimate of peak period conditions.

---

2 For "unlinked" trips, each boarding is considered as a separate trip. Thus, a trip in which a passenger takes an initial bus and then transfers to a second bus to reach his final destination is not counted as one linked trip but as two unlinked trips. Converting unlinked trips to linked trips requires the application of factors computed from transit surveys that account for the number of transfers taking place.

Data Labeling

It is important to correctly label the data stored in the traffic database. For example, a 24-hour count taken at a location is an ADT, not an AADT or an AWDT. Similarly, as noted above, PSRC must avoid presenting measured peak period or peak hour volumes as design hour information. This requires not only careful labeling of the data (both the names given in menus used to select data and the descriptive documentation provided with the database), but care that, where possible, necessary adjustments are made to data obtained from different agencies, so that all data in the database represent similar traffic attributes. (For example, if PSRC receives simple 24-hour daily counts from one jurisdiction, seasonal and/or day-of-week adjustments should be applied to those values before AADT is estimated.)

To ensure that users understand the data they have requested and/or received, a glossary of terms must be available to all users. Ideally, the glossary should be available on-line, as an option that can be accessed from the disclaimer screen that is shown before data are output.

DATABASE DESIGN

This section of the report describes the database itself, including the variables stored in the database and the attributes of those variables. Two levels of the database are described. In the top level are the summary traffic and transit statistics. In the lower level are all of the traffic and transit volume measurements provided by the cities and counties. If a city and a county submit different traffic volumes for a specific road segment, both will be maintained in the lower level of the database, but only one value will be stored for that road segment in the top level of the database. If two different transit agencies submit volumes for a road segment, both will be stored in the lower level of the database, but an aggregated transit estimate will be stored in the upper level of the database.

The following description assumes that in the top level of the database, the transit and traffic volume data will be stored in separate files. The MTS segments and naming conventions will be the same for both sets of records. It would be possible to store these variables in a single record, but current analysis indicates that separating these variables into two different files/records will benefit overall system design, operation, and maintenance.

The six years of data for both transit and traffic records could be stored as either six separate sets of data files, or as one large data file for each type of record. The advantage of using six files is that it would make archiving a year of data easy. It would also allow more efficient maintenance of the database, allow the files to be sorted more quickly (once a year is complete, those records no longer need to be included in future sorting efforts), and allow them to be stored more efficiently on disk (smaller files do not need large spaces to be stored in contiguous sectors, thus further speeding up disk access times). The drawback to multiple files is that the data retrieval function would have to know that it might need to look at all six files (depending on the data request), rather than at a single file. This might result in a slower data retrieval process for multi-year data requests. Current thinking by PSRC staff is that having a single file for each year of data is preferable to having a single large file for each record type.

Each year of data will have the same variable fields. However, because some locations are counted in one year but not the next, not all files will have data for the same MTS segments. PSRC made a design decision to not attempt to estimate traffic or transit volumes for a specific road segment if no count was provided to PSRC for that segment for that year. (That is, old counts will not be factored and stored in the database as a means for estimating current facility use.)

Each data file should be sorted by roadway. (That is, all traffic volume data for a given year on NE 45th Street should be stored in consecutive records.) This will allow more efficient data processing, data retrieval and database maintenance.

Traffic Variables Stored in the Top Level of the Database

Each of the six files that contains the top level of the traffic portion of the PSRC database should contain the following data items. (Some variables will be blank for many records.)

Road segment ID	This can be either the anchor points (two separate variables) for the road segment, or the unique segment ID, depending on the needs and/or capabilities of the PSRC GIS as defined by the ongoing PSRC location referencing project.
Street ID (name)	The name of the street on which the count was collected. This would include the street directional identifier in its appropriate location, e.g., the "NE" in "NE 45th Street" would be included in the street ID field.
Count Year	(e.g., 1997. Assumed to be a four digit value to ease some year 2000 calculation problems)
AADT	Computed average annual daily traffic (i.e., both directions combined)
AWDT	Computed average annual weekday traffic (i.e., both directions combined)
ADT	Represents a count for a specific day (or combination of days), but does not account for (either or both) the day-of-week or seasonal adjustments needed to estimate average annual conditions. It should represent vehicles, not axles.
Date of count	(dd/mm) - probably expressed as two separate variables. This date describes the start date of the count listed in the ADT variable.
Length	The length of the count performed to compute the ADT value.
Multiple count designator	A two-part flag that shows that multiple data counts exist within the roadway segment. The first part of the flag should be a number that indicates the greatest number of days of data available at a single location (and from a single source) within that MTS section. The second portion of the flag should be an indicator of the number of physical locations within that section at which counts are available. This second value could be alphabetical, where A = data are available at one location, B = two locations, etc. (So an MTS section with a seven day count at one location and three other one-day counts at three different minor intersections within that MTS section would have a code of 7D.)
DADT1	Directional daily traffic in direction 1 (roughly half of ADT)

Direction 1	The compass direction associated with DADT1
DADT2	Directional daily traffic in direction 2 (roughly half of ADT)
Direction 2	The compass direction associated with DADT2
Peak hour traffic 1	The peak hour volume computed in direction 1 (uses the same direction 1 as noted above). Note that this is assumed to be an "average weekday" condition. It is not a design hour volume.
Hour of peak traffic 1	The start time (hh:mm) of the hour (four consecutive 15-minute periods) during which the peak hour of traffic occurs in direction 1
Peak hour traffic 2	The peak hour volume computed in direction 2 (uses the same direction 2 as noted above)
Hour of peak traffic 2	The start time (hh:mm) of the hour (four consecutive 15-minute periods) during which the peak hour of traffic occurs in direction 2
AM Peak Period 1	The directional traffic volume from 6:00 AM to 9:00 AM measured in direction 1 (uses the same direction 1 as noted above)
PM Peak Period 1	The directional traffic volume from 3:00 PM to 7:00 PM measured in direction 1 (uses the same direction 1 as noted above)
AM Peak Period 2	The directional traffic volume from 6:00 AM to 9:00 AM measured in direction 2 (uses the same direction 2 as noted above)
PM Peak Period 2	The directional traffic volume from 3:00 PM to 7:00 PM measured in direction 2 (uses the same direction 2 as noted above)
axle adjustment	A value (from a look up table) used to adjust an axle count-based traffic count. (Leave blank if vehicles rather than axles are reported by the submitting agency.)
seasonal adjustment	A value (from a look up table) used to adjust the ADT count to represent AADT
Adjuster	The initials (3 letters?) of the person entering AADT, AWDT, truck percentage, or other summary statistics. (It should be possible to hand enter summary statistics for the following variables: AWDT, AADT, percentage of trucks (and buses) by classification, and the directional identifiers.) If these variables exist, but the "adjuster" variable is blank, the data were supplied by the county/city submitting the data. *(Note, that it might be best if individual flags were set to indicate specific variables that were added and/or changed as*

	a result of this manual process. This would allow users to know which variables had been reviewed and either added or changed as a result of that review. However, it would require the addition of a fairly large number of additional variables to the database.)
Percent single unit trucks	The percentage of the daily traffic stream that is composed of single unit trucks
Percent combination trucks	The percentage of the daily traffic stream that is composed of combination (tractor-trailer style) trucks
Percent multi-trailer trucks	The percentage of the daily traffic stream that is composed of multi-trailer (a tractor pulling a semi-trailer and a second trailer) trucks
Number of buses	The total number of buses on this road section in a day (See the transit record for more detailed transit volume information.)
capacity	Hourly roadway capacity obtained from the submitting jurisdiction. (Should be the limiting capacity for the road segment.)

Transit Variables Stored in the Top Level of the Database

To determine what data to store in the transit database requires clarification of the analyses the PSRC intends to perform with those data. The problem is that analyses that focus exclusively on transit operations and performance will require different data than analyses that focus on combined transit/traffic operations and performance.

An illustrative example of these differences is the definition of the "peak hour" for which transit data are provided. Strictly speaking, the transit "peak hour" and "peak period" will be different than the traffic volume peak hour and peak period. (That is, the hour in which either the most transit coaches or the most transit passengers pass a given point will almost always be different than the peak hour during which the most cars pass that point.) Thus, when PSRC defines "peak hour" for transit data, they must define which of several "peak hours" they actually want. Furthermore, each request for "peak hour" transit ridership data requires considerable effort on the part of the transit authorities to provide those data to PSRC. Consequently, PSRC will want to limit itself to a single definition of "peak hour" except for special cases. The options include the following:

- a specified hour for each location (for example, always between 7:00 AM and 8:00AM)

- the peak hour identified for traffic volumes for that location

- the peak hour (four consecutive 15-minute periods or standard clock hour?) during which the most transit passengers pass that point

- the peak hour (four consecutive 15-minute periods or standard clock hour?) during which the most transit coaches in revenue service pass that point.

Each of these definitions will result in a different value stored in the database, and each will support a different set of analyses. In addition, some of these will be easy for the transit authorities to provide, while others will entail considerable work by both PSRC staff (to develop the request) and transit agency staff (to meet that request).

The first alternative would allow PSRC to compare transit volumes during what approximates the peak hour consistently across the metropolitan area. Unfortunately, this would not capture the "true"[3] peak volumes at most sites. However, transit authorities could meet this request with relative ease. In addition, this value would serve as a reasonable approximation of transit volumes during peak travel times. In addition, because peak transit volumes occur at different times of the day in different parts of the metropolitan area, the selection of the hour of the day for requested traffic will have an impact on the relative transit volumes reported in different parts of the area. (For example, in the morning, ridership peaks in suburban areas earlier than it does in the downtown core areas.)

The second alternative would allow the use of transit data accurately within the "peak hour" analyses associated with traffic volume measurements. This would be most useful from a congestion management standpoint. However, it would require PSRC to determine the actual peak hours at each desired location, send that list of locations and time periods to transit agencies, and then have transit agencies create specialized data queries at

---

[3] Where the truth is either the maximum number of people passing that point during any given hour of the day, or the number of transit riders passing that point during the hour when the peak traffic volume occurs.

each data location. While most useful to the CMS from a congestion monitoring perspective, this alternative would require the most work of any. In particular, it would require a much more complicated query process at the transit agencies.

The third alternative would be most useful for transit analyses. It would allow PSRC to analyze the maximum hourly transit usage at a point. However, because the timing of the transit peak usually does not coincide directly with the traffic volume peak, this might not be the type of analysis that PSRC needs to concentrate on for congestion monitoring purposes. (It is the type of analysis usually performed by the individual transit agencies.)

The last alternative would be most useful if an analysis of the "maximum impact" of transit on facility operations was needed. It would provide a measure of the greatest level of transit service during the day on that road segment, as well as a measure of the greatest transit impact on traffic flow.

The preliminary recommendation is for PSRC to select a given peak hour and request data for as many MTS segments during that hour as possible from the transit authorities. PSRC may wish to use slightly different peak hours in the CT and Pierce County requests than in the Metro request. This would allow the outlying areas to provide peak hours selected that more closely approximate the true peak hour transit ridership.

Similarly, peak period transit ridership can be based on the peak periods used for traffic volumes, or the peak periods used for transit fare collection. It is recommended that PSRC use the traffic volume peak periods (6:30 AM to 9:30 AM and 3:00 PM to 7:00 PM), since the congestion monitoring uses of this database are more in line with PSRC priorities than the transit operation and revenue analyses. (These are primarily the responsibility of the transit authorities.)

Given the above decisions, each of the six files that contain the upper level of the transit portion of the PSRC database should contain the following data items. (Some variables will be blank for many records.)

Road segment ID	This should be the same road segment ID used for the traffic file.
Street ID (name)	The name of the street on which the count was collected, including the street directional identifier in its appropriate location, e.g., the "NE" in "NE 45th Street" would be included in the street ID field.
Daily transit ridership	Total daily (Fall schedule revision) ridership passing this point in both directions (includes all transit authorities' data combined)
Number of buses	The number of buses on this road segment each day (includes all transit authorities combined)
AM Peak hour ridership 1	AM peak hour ridership (Fall schedule revision) passing this point in direction 1, where direction 1 is the same as defined in the traffic counting section.
AM Peak period ridership 1	AM peak period ridership (Fall schedule revision) passing this point in direction 1, where direction 1 is the same as defined in the traffic counting section. Peak period will be the same time period as that used for traffic volumes (6:30 - 9:30 AM)
PM Peak hour ridership 1	PM peak hour ridership (Fall schedule revision) passing this point in direction 1, where direction 1 is the same as defined in the traffic counting section.
PM Peak period ridership 1	PM peak period ridership (Fall schedule revision) passing this point in direction 1, where direction 1 is the same as defined in the traffic counting section. Peak period will be the same time period as that used for traffic volumes (3:30 - 7:30 PM)
Peak Buses Direction 1	The number of buses in the peak hour
Direction 1	The compass direction associated with Direction 1 (Same direction 1 as with the traffic segment for this link.)
Hour of peak traffic 1	The start time (hh:mm) of the hour during which the peak hour of traffic is provided in direction 1
AM Peak hour ridership 2	AM peak hour ridership (Fall schedule revision) passing this point in direction 2, where direction 2 is the same as defined in the traffic counting section.
AM Peak period ridership 2	AM peak period ridership (Fall schedule revision) passing this point in direction 2, where direction 2 is the same as defined in the traffic counting section.

PM Peak hour ridership 2	PM peak hour ridership (Fall schedule revision) passing this point in direction 2, where direction 2 is the same as defined in the traffic counting section.
PM Peak period ridership 2	PM peak period ridership (Fall schedule revision) passing this point in direction 2, where direction 2 is the same as defined in the traffic counting section.
Direction 2	The compass direction associated with Direction (Same direction 2 as for the traffic segment for this link.)
Hour of peak traffic 2	The start time (hh:mm) of the hour during which the peak hour of traffic is provided in direction 2
Peak Buses Direction 2	The number of buses in the peak hour

Traffic Variables Stored in the Secondary Level of the Database

As noted above, the second level of the database will include all records submitted to PSRC. It will differ from the first level of the database in several respects. The most important aspect is that for any given MTS roadway segment there may be more than one traffic volume record. Multiple records can occur for several reasons:

- The PSRC road segment is so long that an agency has performed counts at more than one location within that segment (for example, counts performed at minor, signalized intersections as part of a signal timing study).
- More than one agency counted at some point within that roadway segment (for example, the WSDOT and a local jurisdiction).
- More than one count was taken at a specific location during the year (for example, on the freeway system where data is collected year round).
- A count that was performed lasted for more than one day, and all days of data were submitted to PSRC.

Creation of the second level of the database will allow all of these details to be maintained when they are submitted to PSRC but will keep PSRC staff from having to deal with them unless they need to see that level of detail.

The traffic data stored in the lower level of the database will include all those records transmitted by the cities and counties that apply to MTS roadways. The data will be converted to a common format from the various formats submitted by the cities and counties. A source identifier will be included in the record to indicate which agency submitted that particular record. Figure 3 lists the codes to be used for these purposes.

Each record will contain data for only one location, one day, and one direction of travel (or both directions combined). Thus, an intersection count taken during one day that included all four legs of the intersection in both directions would be stored as eight different records, since that count would translate into four different locations, each of which would have two directional counts. A count that included all four legs of the intersection but that included only the direction of traffic "entering" the intersection would result in four traffic records. Note that turning movement data would not be explicitly stored in the PSRC database. These data are considered to be too detailed for the PSRC application and will need to be obtained from the individual agencies on a case by case basis.

Lower level records should be stored as a series of flat files, with each file containing one year of data. Each yearly file should be sorted by location (MTS road segment), date, and source. The data to be stored in the lower level records are shown below.

Road segment ID	The same name used in the upper level of the database.
Street ID (name)	The name of the street on which the count was collected, including the street directional identifier in its appropriate location, e.g., the "NE" in "NE 45th Street" would be included in the street ID field.
Source of estimate	Code number supplied by the submitting agency (if any) that is used by that agency as a location identifier.
Cross Street location	The name of the street near which the count is located (obtained from the location information sent with the count by the city/county responsible for that count)
Extra location information	Other information provided by the city/county used in locating the count (usually something like "north of" the cross street)
Date of count	(dd/mm/yyyy)
Count duration	(number of days)

Figure 3
Codes Used For City/County Identification

County Codes

King	10
Kitsap	20
Pierce	30
Snohomish	40

City Codes

King County

Algona	11
Auburn	12
Beaux Arts	13
Bellevue	14
Black Diamond	15
Des Moines	16
Duvall	17
Enumclaw	18
Federal Way	19
Hunts Point	1a
Issaquah	1b
Kent	1c
Kirkland	1d
Lake Forest Park	1e
Medina	1f
Mercer Island	1g
Milton	1h
Newcastle	1i
Normandy Park	1j
North Bend	1k
Pacific	1l
Redmond	1m
Renton	1n
SeaTac	1o
Seattle	1p
Shoreline	1q
Skykomish	1r
Snoqualmie	1s
Tukwilla	1t
Woodinville	1u
Yarrow Point	1v

Kitsap County

Bainbridge	21
Bremerton	22
Silverdale	23
(others?)	

Pierce County

Bonney Lake	31
Carbonado	32
Dupont	33
Eatonville	34
Fife	35
Fircrest	36

Figure 3
Codes Used For City/County Identification
(Continued)

Milton	37
Orting	38
Puyallup	39
Roy	3a
Ruston	3b
South Prairie	3c
Steilacoom	3d
Sumner	3e
Tacoma	3f
Wilkeson	3g
Snohomish County	
Arlington	41
Bothell	42
Brier	43
Darrington	44
Edmonds	45
Everett	46
Gold Bar	47
Granite Falls	48
Index	49
Lake Stevens	4a
Lynnwood	4b
Marysville	4c
Mill Creek	4d
Monroe	4e
Mountlake Terrace	4f
Mukilteo	4g
Snohomish	4h
Stanwood	4i
Sultan	4j
Woodway	4k
Transit Agency Codes	
King County Metro Transit	51
Kitsap Trnasit	52
Pierce Transit	53
Snohomish Transit	54
Other Codes	
WSDOT	60
PSRC Collected Data	61
University of Washington	62
Other	63

Hourly Count value midnight - 1 AM

Hourly Count 1 AM - 2 AM

Hourly Count 2 AM - 3 AM

Hourly Count 3 AM - 4 AM

Hourly Count 4 AM - 5 AM

Hourly Count 5 AM - 6 AM

Hourly Count 6 AM - 7 AM

Hourly Count 7 AM - 8 AM

Hourly Count 8 AM - 9 AM

Hourly Count 9 AM - 10 AM

Hourly Count 10 AM - 11 AM

Hourly Count 11 AM - 12 noon

Hourly Count 12 noon - 1 PM

Hourly Count 1 PM 2 PM

Hourly Count 2 PM - 3 PM

Hourly Count 3 PM - 4 PM

Hourly Count 4 PM - 5 PM

Hourly Count 5 PM - 6 PM

Hourly Count 6 PM - 7 PM

Hourly Count 7 PM - 8 PM

Hourly Count 8 PM - 9 PM

Hourly Count 9 PM - 10 PM

Hourly Count 10 PM - 11 PM

Hourly Count 11 PM - midnight

ADT	The daily traffic volume measured at this location for this particular count.
axle count	Variable identifying whether this was an axle count

axle adjustment	A value (from a look-up table) used to adjust an axle count-based traffic count
seasonal adjustment	A value (from a look-up table) used to adjust this count to represent AADT
AADT	Computed average annual daily traffic
AWDT	Computed average annual weekday traffic
Direction	The compass direction associated with this count (may be both directions combined)
Peak hour traffic	The peak hour volume computed and submitted as a peak hour volume by the agency
AM Peak Period	The directional traffic volume from 6:00 AM to 9:00 AM
PM Peak Period	The directional traffic volume from 3:00 PM to 7:00 PM
Percent single unit trucks[4]	The percentage of the daily traffic stream that is composed of single unit trucks
Percent combination trucks	The percentage of the daily traffic stream that is composed of combination (tractor-trailer style) trucks
Percent multi-trailer trucks	The percentage of the daily traffic stream that is composed of multi-trailer (a tractor pulling a semi-trailer and a second trailer) trucks
Percent buses	The percentage of the daily traffic stream that is composed of buses

---

4 It is not clear how cities and counties will submit data on truck volumes. The volume databases reviewed for this project did not include truck volume information. Most cities and counties do not count vehicles by classification routinely; however, with increasing emphasis on intermodal transportation, the number of classification counts being performed is increasing. In addition, most vehicle classification equipment does not perform accurately in urban (stop-and-go, closely spaced) traffic conditions.

The records in this database will provide variables to hold summary truck travel information, but these variables will not be sufficient for submittal of true vehicle classification counts. (The lower level of the database would need four records for each daily count if four vehicle classes were counted.) However, if classification records are not submitted, it is unclear how truck volume (or percentage) information will be submitted. This paper assumes that the cities and counties will eventually provide summary truck percentage data as part of their normal traffic submittal. The only major problem with this assumption is that it does not allow for differentiating truck travel during the peak and off-peak periods, which will be important to PSRC.

These issues need to be resolved at some point in the future as PSRC's truck data needs become clearer. Revisions to the database at that time may be necessary.

Transit Variables Stored in the Lower Level of the Database

The lower level of the transit database will relate to the upper level in much the same manner as the two levels of the traffic database. Each record within the six files that contain the lower level of the transit portion of the PSRC database should contain data from a single agency for travel on a specific road segment in a single direction (or both directions combined). As with the traffic records, there could be multiple records for each road segment at this level of the transit database, because more than one transit authority could be operating on a given road segment and each could have submitted ridership and other statistics. However, unlike the traffic statistics, ridership estimates from different authorities should be additive for a given road segment, whereas the traffic estimates will be alternative estimates of the same traffic condition. That is, if Metro and CT both submitted transit ridership estimates for I-5 at NE 45th Street, the total transit ridership would be the SUM of these two values. (In the traffic database, the two estimates would be alternative estimates of the same traffic statistic.) In addition, because transit counts are often made over several days (some routes are counted on one day, other routes are counted on another day), there is no specific date field for the transit count. Thus, the database has no method for handling (as it does not expect to handle) more than one record per MTS roadway segment from a single transit authority.

The one major problem that is not resolved is how the database user (or PSRC) will know when one or more transit authority data points should be, but are not, present for a given road segment in the database. For example, on I-5, south of the King/Snohomish county line, the database may have both a Metro transit and RTA ridership count. Thus an upper level summary statistic can exist. However, without also having a CT ridership estimate for that road segment, the computed transit ridership estimate would underestimate the total transit ridership on that road segment. This issue needs to be further explored by PSRC.

The transit variables to be stored in the lower level of the database are as follows:

Road segment ID	This should be the same road segment ID used for the traffic file.
Street ID (name)	The name of the street on which the count was collected, including the street directional identifier in its appropriate location, e.g., the "NE" in "NE 45th Street" would be included in the street ID field.
Count Year	(e.g., 1997. Assumed to be a four digit value to ease some year 2000 calculation problems)
Daily transit ridership	Total daily (Fall schedule revision) ridership passing this point in the given direction from this transit authority
Authority[5]	The name of the transit authority
Direction	The compass direction associated with this record. (Note that these compass directions need to be coded in the same manner as those of the traffic counts. The directional coding also needs to be able to indicate when the count includes both directions combined.)
Number of buses	The number of buses on this road segment each day
Peak Buses	The number of buses in the peak hour at this location.
AM Peak Period Buses	The number of buses in the peak period
AM Peak hour ridership	AM peak hour transit ridership (Fall schedule revision) passing this point in this direction
AM Peak period ridership	AM peak period transit ridership (Fall schedule revision) passing this point in this direction.
PM Peak hour ridership	PM peak hour transit ridership (Fall schedule revision) passing this point in this direction
PM Peak period ridership	PM peak period transit ridership (Fall schedule revision) passing this point in this direction.

Database Programming Language

The database application can be written in a number of alternative languages. The basic alternatives include

- a conventional programming language (e.g., C++ or Visual Basic)

5 Note that the database is designed to incorporate bus ridership estimates. However, if the light rail line or monorail is coded into the MTS network, there is no reason why ridership on these two modes can not be included in the transit statistics portion of the PSRC database.

- a standard database application (e.g., Microsoft Access)
- a spreadsheet database (e.g., Microsoft Excel).

It may also be best to use some combination of these techniques. For example, the initial process of converting city/county data files into a common format might be written in C++, with the remaining database functions written in Access. Each of these options has strengths and weaknesses.

C++ would require the greatest level of programming but would provide the most flexibility in software design. It would also be the most effective at automatically converting the various file formats submitted by the cities and counties into a consistent output format. However, C++ is the most complex of the programming options, and would require the most technically sophisticated staff to build and maintain. C++ is also the worst choice among the programming alternatives in terms of obtaining add-hoc reports from the database. Special data requests that could not be fulfilled by the PSRC GIS system could only be obtained by loading the entire database into some other database system or by writing new C++ code.

Use of Visual Basic (or some other programming language) might provide many of the same advantages as C++ but within a much easier programming environment. Visual Basic also has a number of features that make it particularly useful for working with Excel and MS Access files. The downside to Visual Basic is that it is neither as powerful nor as easy to optimize as C++.

MS Access (or some similar database program) has the advantage of being readily compatible with the PSRC GIS, as well as being designed to perform routine database functions quickly and easily. It is also easier to program than C++, and PSRC should already have trained technical support staff capable of undertaking required system upgrades or modifications in the future. Access could also be used fairly easily to create ad-hoc reports to help PSRC meet various analytical requirements. The primary limitation with using a program like Access is that most database programs are not well suited for the "fuzzy logic" operations needed to decode multiple county and city record formats and coding schemes in order to produce data in the consistent format the PSRC GIS system needs. This preliminary (but vital) task is a key component of the database system.

Excel is the simplest language of the three major alternatives. It could also feed data to the PSRC GIS efficiently. It would be the easiest for the staff to use and understand. However, it is primarily a spreadsheet, and its database functions are limited. Like the Access software, converting the initial county and city submittals into a consistent format within Excel would be difficult and might require use of extensions to the Excel program itself (i.e., a separate C++ program that ran before the data were inserted into Excel). In addition, Excel is likely to be the slowest system when searching for data as part of an ad hoc query. Finally, while very simple ad hoc queries would be easy for a novice user to undertake within Excel, more complex queries might be quite difficult to accomplish because of the spreadsheet orientation of the database.

Preliminary discussions with the PSRC staff have indicated that the preferred alternative for the database system is the use of a combination of Visual Basic and Microsoft Access. The main database would be stored in Access, with a Visual Basic preprocessor that converted the files from their native format to an Access readable file and included much of the location referencing conversions. Visual Basic may also be used to prepare some of the output reports.

Geocoding

As mentioned earlier, in the short term, the traffic and transit counts will be geocoded by matching county and city location referencing systems to the MTS network segments defined by PSRC staff. The city and counties all use street names and directional identifiers to locate counts. (For example, a count is taken on 1st Avenue, north of the intersection with Pine.) In a few cases, where counts are taken repeatedly at the same location each year, count locations are also given identifying numbers. (Note that this latter case is true for all data obtained from the WSDOT FLOW system.)

Where the count ID system is used, PSRC staff can create (by hand) a simple look-up table that maps specific count locations to individual MTS road segments. In all other cases, a new system will be needed to use the text-based location references provided by the cities and counties and translate them to the appropriate MTS roadway segment.

This geocoding task will require several steps, including the following:

- defining the MTS road segments
- reading and reformatting the city and county submittals into a consistent format so that individual text segments represent specific parts of the location reference
- running a program that takes these location references and matches them against PSRC's roadway segments
- outputting a list of locations that could not be matched to PSRC's MTS locations
- performing a manual review of the unmatched records to determine whether those sites were not located on the MTS (in which case they are simply ignored), or the location information supplied by the city or county could not be decoded by the database software
- manually adding the appropriate roadway segment IDs to the unmatched records that are located on MTS roadway segments
- storing the MTS location codes within the new database records.

MTS roadway segments should be defined by PSRC GIS staff on the basis of the overall agency needs for the GIS. The traffic inputs to the roadway segmentation process are simply to

- make the roadway segments as homogeneous as possible with regard to traffic and transit volumes
- break roadways into new segments whenever MTS roadways intersect
- break roadways into new segments at each major intersection (even if that intersection does not involve another MTS roadway)
- create roadway segment boundaries at locations where traffic congestion is likely to be a cause for concern (for example, at railroad crossings or other potential bottlenecks).

For each roadway segment there will be (at most) one traffic and one transit record in the upper level of the database. However, because traffic volumes change continuously along most arterials, it is probable that in many cases cities and counties will count (at least in some cases) in more than one location within a PSRC-defined roadway segment. When this occurs, both will be geocoded as having the same location. This is a limitation of the database. However, the multiple count designator in the upper level of the database will allow a user to know which roadway segments are based on more than one location and/or more than one day of data.

Because all traffic volume records on MTS roadway segments will be stored in the lower level of the database, users will be able to review and select other volume estimates for that roadway segment if they so desire. In most cases, this should not be a problem. Once the database has been created, it may be useful to perform an ad hoc study with the data in the database to determine how often these multi-count roadway segments occur and what those multiple counts tell PSRC about traffic variability and the MTS network. As a result of such a study, adjusting the MTS network to improve the database may be appropriate. (For example, counts may show that PSRC needs to split a roadway segment into two or more segments because significant volume changes occur within the originally defined segment.) This review may also indicate problems with one or more of the participating jurisdiction's traffic counting procedures.

The geocoding step itself will have to entail a look-up process that first identifies the street on which a count was taken and then determines within which road segment the cross street and extra location identifier place the count. The GIS system may have this function built in. If not, the basic requirements for performing this step are as follows:

- creating a table that relates each MTS roadway segment (and its two anchor points and/or segment ID) to a given street name
- building a table of pseudonyms for MTS street names (e.g., 1st Avenue, 1st Ave, First Avenue, etc.) to identify all of the possible names for the MTS roadway segments
- building a cross reference table for each MTS roadway segment that identifies all cross streets within that segment (each cross street will also need a table of pseudonyms)
- for cross streets at the end points of the segment, indicate which direction ("north of") from that end point is within the road segment in question.

The logic required to geocode the city/county location will be to initially identify the street on which the count was located. This will reduce the number of potential MTS roadway segments to a small number. Then, within this group of roadway segments, the cross street location provided in the submitted record will be checked against cross street names for a possible match. If two MTS segments match (i.e., the cross street is the boundary point between two adjacent MTS roadway segments), then the information on the orientation of the count, relative to that intersection point, will be used to select between MTS segments.

Four possible outcomes can occur with this process. These outcomes, and the actions the PSRC should take when presented with each, are as follows:

- The road segment is correctly identified and is coded accordingly (no additional PSRC action is needed)
- the road on which the count was taken can not be identified
  - if human review shows that the count is not on the MTS, that volume information may be discarded.
  - if human review shows that the count should have been placed on an MTS roadway segment, the PSRC will need to update the roadway segment naming tables so that this count can be correctly identified the next time it is submitted.
- the road can be identified as an MTS roadway, but the cross street information submitted does not allow the geocoding process to identify an MTS roadway segment
  - it will be necessary to update the cross street table to include the identification keys used for that record (this may entail something other than cross street names). The revised table should use the same reference points used by the submitting jurisdiction.
- the road can be identified as an MTS roadway, but the cross street information submitted combined with the look-up tables identifies more than one MTS roadway segment
  - it may be necessary to refine the process used to differentiate a count location that uses a cross street that happens to be the end point of an MTS roadway segment.
  - it may be necessary to refine the geocoding process to include additional information in the location identifier (such as the agency that submitted the count) to differentiate between similarly named streets in different counties.

From a programming perspective, if the geocoding process is not successful, the important points are that the geocoding process indicates what types of matches have been completed, what remaining information has not been successfully matched, what the original record that was being geocoded looks like, and from what organization the data were submitted. These results can probably be printed to an output file, where they can be dealt with by PSRC staff.

In the longer term, the geocoding of the database may be simplified by the development of either a common regional linear referencing system or the creation of links that correlate the referencing systems used by the various jurisdictions to the PSRC GIS.

Auxiliary Data In Related Data Files

Most of the traffic volume data supplied by the cities and counties will be submitted in the form of daily traffic volumes. In some cases these estimates will be based on axle counts. In other cases, they will be based on loop activations, and in still others, they will be based on axle-sensor based vehicle classifiers.

Axle correction factors should be used whenever simple volume counts are taken with axle sensor-based traffic volume counters. However, in many cases, these factors may be essentially equal to two, the factor often built into the counter. Where truck traffic is very light (common in many urban settings), and most of that truck traffic consists of delivery vehicles, the average number of axles per vehicle may be very close to two, and errors in the axle sensor response means that the ratio of axles per vehicle is essentially equal to 2. In these cases, it is not necessary to create a separate axle correction factor. However, some measured proof of this relationship should be provided by a city or county wishing to use this adjustment factor.

A good source of this proof would be any vehicle classification work done to examine freight and goods movement in the PSRC region. (That is, if truck routes are developed in the metropolitan region, those routes should probably not use a simple axle correction factor of 2.0. Instead, they should have more route-specific numbers based on a special data collection effort.)[6]

Seasonal and day-of-week adjustment factors are also necessary to create AADT and AWDT values from short duration counts. According to WSDOT data, seasonal changes in traffic on state routes in the PSRC region range from 7 to 15 percent of average annual conditions. Luckily, such differences are fairly small, given the size of day-to-day variations in traffic volume, and such differences are not significant for most PSRC analyses. However, these differences are important when some analyses are performed (such as measuring annual traffic growth when counts are taken at different times of the year).

In addition, many urban streets have consistently higher weekday traffic volumes than weekend volumes. For many urban roads, weekday/weekend differences are much more significant than seasonal differences. When average daily conditions are needed (as opposed to average weekday conditions), it is important to factor in day-of-week adjustments as well.

Adjustment factors that can correct these biases come from permanent counters that record volume changes during the year. WSDOT operates 15 permanent counters within the PSRC boundaries (in addition to the FLOW system) that are used to compute seasonal adjustment factors for state routes. Kitsap County has indicated that it is in the process of deploying several permanent counters themselves. It is likely to create its own seasonal adjustment factors with data from these counters.

Both of these sources (as well as perhaps others) can be used to develop seasonal factors for use in computing AADT when needed. The PSRC should create a table of seasonal adjustment factors that can be referenced by database users as needed. These

---

6 Initially, PSRC should ask specifically for "vehicle" counts from each jurisdiction and accept that the jurisdiction counts vehicles appropriately. However, follow-up work should be done to ensure that axle corrections (and seasonal corrections) are performed correctly by each agency. If not, these differences will lead to discrepancies in the PSRC database.

factors can simply be obtained from WSDOT and from other PSRC members on a once a year basis. (The WSDOT factors are printed each year in the Annual Traffic Report.)

The difficulty with seasonal adjustment factors is that these factors vary depending on the amount and type of seasonally affected travel on a given road. For example, the traffic on SR-16 across the Tacoma Narrows bridge is affected by two major types of traffic, urban traffic traveling between Gig Harbor and Tacoma and recreational travel heading to and from the Kitsap peninsula. Because of this recreational component, the WSDOT traffic counter on the Tacoma Narrows bridge may not be a good estimator of seasonal traffic on I-705 in Tacoma, even though it is geographically close to I-705.

Given such differences, it is recommended that the PSRC review these factors and document which factors it recommends to use for traffic counts on specific types of roads. This will require an investigation of the location of available permanent traffic counters and of the differences in the seasonal patterns measured at those locations.

A variety of other adjustment factors could be developed and maintained in the PSRC database. However, these factors are not recommended for inclusion as part of the initial database construction. The most important adjustment would be adjustments to account for equipment error. Both axle sensors and loop sensor-based equipment have problems counting vehicles accurately in urban areas because of close vehicle spacing and the acceleration/deceleration of vehicles as they cross sensors.

Air tube axle sensors traditionally undercount traffic when stretched across more than one lane of traffic because the sensor response is not fast enough to record nearly simultaneous hits. Similarly, loop sensors located at the stop bar used for both traffic signal actuation and vehicle counting often undercount vehicles because the close vehicle spacing at the loop prevents differentiation of vehicles by the loop detector.

A special study by PSRC (or the cities and counties) that would help determine the accuracy of their counting devices would allow creation of any necessary adjustment factors, if PSRC requires the extra accuracy such an adjustment would allow.

The final set of adjustment factors that might be useful for PSRC would be the creation of a table that allowed estimation of design hour volumes for roads, given an average daily weekday volume. This might be useful in determining the latent demand for a road, given a specific daily volume. Such an estimate would not be a realistic measure of expected 30th highest hourly volume because capacity constraints might limit the actual volumes possible, but these estimates might be useful to PSRC staff and others using the database.

DATA PROCESSING STEPS

This section describes the various data processing steps needed to obtain, enter, process, and report the traffic volume and transit data described earlier in this report.

Data Input

To successfully perform data input, the system will need to have an initial data input screen that requests the city or county and the year for which data will be entered. This information will then be used to select the appropriate translation program for converting the incoming data files to the required PSRC format.

Once the city/county information has been entered and processed, the operator will have to select the files to be processed. This may be best performed with a simple file look-up process in which a mouse is used to select individual files from a directory or drive. (This procedure might be easier than requiring the user to remember and type input file names.) This procedure might also include an option in which users could indicate the file format for each input file. (For example, King County might supply data in either an ASCII or EXCEL format. Having such an option would allow the input processing program to select the appropriate input format.) For simplicity, only one year should be entered at a time.

Once the input processor has established that the files are in the appropriate format, it will check to see whether the database is ready to accept data for that particular year. If the database is not ready to accept data for that year (data are already present for six earlier years and no data are present for the year in question), then the program will indicate that the "increment the year" function should be run. Depending on the program design, one of three actions will take place.

- The program will terminate input processing.
- The program will temporarily suspend input processing and perform the "increment the year" function, and when that function is complete, return to processing the new input data.
- The program will complete the input processing function and either store the new data in a temporary file (until the increment function has been performed) or store the data as a seventh year of data in the database.

Once the database is ready to receive new records, the "reformatting" process will begin. In this process, the input records will be decoded on the basis of the expected input format for that city/county, and the results will be written to the lower level of the database.

As part of the decoding/recoding process, the input data have to be checked to ensure that the files being read match the formats expected, and that the data included in the files are reasonable (e.g., the dates of the traffic counts are for the correct year). That is, some type of quality assurance function must be run to help ensure the integrity of the data. Data records that are found to be "unreasonable" should be written to a separate file along with a description of why those data are considered unreasonable. These data must then be reviewed manually to determine whether they should be entered into the database. (Note, if the entire file, or even a large portion of the file, failed the quality control process, each record in that file would not have to be written to the "discard" file. Instead a different error message should be written to the user. This output message could be something like, "Input records are not in the expected format. Please check the record format, and resubmit the file.")

For the data records that pass the quality control process, the next step in the input process will be to geocode each of the traffic counts in the file. This may entail multiple geocoding efforts for each record in the input file. (For example, in the Tacoma database as many as four separate traffic counts, representing the four legs of an intersection, are incorporated in a single database record.) The geocoding process will entail determining the street on which each count is located and using the cross street indicators (and the location of count relative to that cross street) to determine the particular GIS road segment on which the count lies. Finally, the geocoding effort will require that any counts that can not be successfully geocoded must be written to an "exceptions" file, which must be reviewed manually. This file should be separate from the one used to hold records that fail the quality assurance tests.

Once a traffic count has been successfully geocoded, it can be written to the lower level of the database.

Upon completion of the data processing, the input program should provide the user with the option of either ending the program operation or entering a "manual" mode that will allow the user to learn what records have been either rejected and/or failed to be geocoded. The user should be able to physically edit the records that can not be geocoded and should be able to view the records (in some consistent format) that did not pass the quality control checks.

For records that can not be automatically geocoded, the manual process should allow for hand entry of the correct roadway segment ID information. This process should also allow rejection (and removal) of the submitted traffic count record, if the record turns out to have been made on a roadway that is not included in the PSRC database. (For example, the count is on a local, residential street within a city boundary.) Users should also be able to leave a record in the "non-geocoded" file on a semi-permanent basis, if they cannot supply a correct road segment ID. This would allow users to add the correct geocoding to this record at a later date, once the location of the count was determined.

If an initially rejected record can be successfully geocoded by hand, the information needed to geocode that site will have to be added to the geocoding tables to allow records entered at a later date to be geocoded automatically. Thus, the "hand update" must make two separate changes, 1) to the record about to be entered into the database, and 2) to the GIS look-up table that cross references expected incoming location information with a specific GIS roadway segment ID.

Users should also be able to view those records that fail the quality assurance test. This viewing effort will allow users to determine the cause of rejection. (Note that the reason for the rejection should be provided along with a formatted record.) For example, if the date is not appropriate, users need to know what information was in the expected date field to understand whether the file format was incorrect, or whether the wrong information was provided by the submitting city and/or county.

Creation of the Upper Level Traffic Volume Records

Once lower level records have been entered into the database, users will be able to create upper level (summary) records. Upper level records will summarize the lower level records available for a given MTS road segment. That is, if only a single record is available for a given road segment and direction, that record will serve as the basis for the road segment's summary volume. However, if more than one traffic count has been taken at that location, either because multiple days of data have been submitted or because counts have been taken at more than one location within a given MTS road segment, the PSRC database will have to determine how to either choose between those counts and/or average the data together to provide a single summary value for that road section. This process is called "arbitration," although it uses summarization, averaging, and other combination techniques as well as pure arbitration between alternative counts.

In some cases (for example, the City of Seattle), the counts submitted to PSRC will have already been through an arbitration process. (Seattle collects week-long counts and submits data for that week, including average weekday and average weekend estimates.) In other cases, the PSRC will need to perform the data arbitration process.

There are two methods for performing the arbitration function, manual and semi-automated. In the manual process, the database would bring up all counts available for a given year and display them (or print them) for the user. Along with these records, the database would provide a blank upper level record with the MTS road segment ID and count year filled in. Users would then be expected to fill in the appropriate summary traffic statistics with the available data.

The advantage of this approach is that it would allow users final control over how to treat the different pieces of traffic information that are available. For example, they could treat a 7-day, City of Seattle count as more trustworthy than a 3-day King County count. At the next site, however, the 3-day King County site might be more trustworthy than a 1-day City of Seattle count. Personal computation of summary statistics would also allow the application of adjustment factors (seasonal and axle correction factors) when those adjustments were warranted.

This technique would also eliminate the need to "out guess" the data submitters, so that a single set of data processing rules did not have to be developed to allow the computation of annual summary statistics from any combination of available data. It would also allow users to supply, or choose not to supply, estimates for any or all of the summary statistics maintained in the upper level of the database. (For example, users might choose to not supply an AADT value because seasonal adjustments for the available counts were not available.)

The advantage of this technique is that it would greatly simplify the programming processes needed to create a single traffic volume estimate from an unknown (and unknowable) number of inputs. It would also allow human input into the selection between two different counts for the same road section. This would allow users to telephone the submitting agency to learn more about specific counts and/or apply their knowledge of a specific road segment to select between counts to best represent the road segment in question.

The disadvantage of this technique is that it would require the operator to be knowledgeable about traffic data (or at least to be able to gather that information). It assumes that knowledgeable operators would be able to make the correct decisions about how to treat different data values to come up with the "best" summary statistic. However, someone that was not familiar with either traffic counting issues or (in some cases) the geography and traffic volume variations of specific road sections could incorrectly select between alternative data summarization alternatives.

Consequently, the second alternative is to create an automated process that would compute an upper level of the database from the records in the lower level of the database. This has the advantage of removing human error from the process and would ensure that all upper level values were computed with a similar set of rules. The difficulty with this technique would be in creating a process that could accurately account for all of the possible variations in data availability in the lower level of the database.

The most likely baseline scenario would be to have the upper level of the database accept the count data from the longest count available for that year and that site. If the data for a multi-day count were submitted as a single record (as in Seattle), the database would need to account for the actual count duration (as opposed to the number of records) when making that determination.

In addition to selecting between alternative counts, the automated process would have to be able to differentiate between one-direction and two-direction counts. For example, if directional information was included in the upper level database, then the combined direction ADT would have to equal the sum of the two directions. If there was a "better" combined direction count, the automated process would have to decide between the value of having that "better" count posted in the upper level database (but not having directional information in that level of the database) and the value of having the directional information but a less reliable combined value.

Dealing with directional counts also means that the automated program that computes the upper level record would have to be able to sum directional level 2 records to produce a single upper level record. It would also require that the automated program have some built-in quality assurance controls. For example, what would happen when four directional counts were present, but they showed four different roadway directions (i.e., one pair said the road went east/west and the other pair of records indicated the road went north/south)? This kind of discrepancy often occurs in traffic records, not because the record is "in error" but because most roads wind and do not follow strict compass directions. Consequently, field staff often enter inaccurate compass directions in the counters in the field.

One solution to the debate between automated and human computation of the upper level traffic records is for the process to be semi-automated. That is, an automated process would compute the upper level records. However, whenever more than one record (other than simple directional records) was from one location, the program would compute only a "recommended" upper level record. These "recommended" records would then be reviewed by an operator, who would also be provided with all of the lower level records available for that site, to accept or revise that estimate before it was stored in the upper level of the database.

This technique would require more staff time than a purely automated process, but it would allow the use of human oversight to determine the flaws in the automated process (which could then be refined over time). It would also allow the addition of knowledge not included in the database itself (for example, the computation of AADT values with seasonal factors) into the process. Finally, by having human intervention (at least at first), it might be possible to create an entire second level of quality assurance to ensure that data used by PSRC and/or released to the public were as accurate as possible.

Creation of the Upper Level Transit Data

The creation of the upper level transit records will be a less complicated procedure than the creation of the upper level traffic records. This is because there should not be any situation of "competing" transit records. That is, all transit records present in the lower level of the database should be added to create the upper level record. This should be a fairly simple, automated task.

The one difficulty with transit volume records is knowing (or letting the user know) that more than one transit agency should have ridership included in a given volume estimate and that volume is (or is not) present. For example, if CT submitted a transit ridership estimate for a segment of I-5 in King County, but Metro did not submit a ridership estimate for that same segment (Metro's count location might be farther south on the next roadway segment), an automated process, such as described above, would create a valid upper level record for that roadway segment. However, because no Metro data were included in that estimate, that ridership estimate would badly under-report the actual transit ridership for that roadway segment.

It is not clear how to resolve this issue. One alternative might be to create a look-up table that indicated which roads in the MTS had transit service from more than one transit authority. However, it is not clear at this time whether this would be a cost-effective approach to this task. More work is needed in this area as the database system is constructed.

Data Adjustments

For the initial version of the database, there should be a manual data adjustment process. Basically, this should consist of a screen entry procedure that allows authorized users to call up a specific traffic or transit record, determine specific adjustment factors (by accessing the auxiliary tables that are stored in relation to the database), calculate, and enter appropriate summary statistics. At the completion of this task, users should enter their initials and the date of the data adjustment. (The date might be added automatically.) The database would then be updated with these new summary statistics, along with the date of the change and the name of the individual who made the changes. Changes should not be allowed unless the user enters identification information.

Authorized users should be able to add and/or change the following data points in the upper level of the database:

- road segment ID
- direction of counts (Direction 1 or Direction 2)
- AADT
- AWDT
- roadway capacity
- percentage of single unit trucks
- percentage of combination trucks
- percentage of multi-trailer trucks
- number of buses
- daily transit ridership
- peak hour transit ridership
- peak period transit ridership

The remaining variables should only be changed by the addition of new data to the lower level of the database and the automated database update procedure. (Note that it should be possible to remove entire records from the lower level of the database if the city/county notifies the PSRC that a previously submitted record was not valid. In such a case, removal of the "bad" record, and addition of a new, "good" record, should replace the existing "bad" upper level record.)

Missing Data

Not all road segments will have data associated with them. That is, some road segments will not be counted during a given year. Similarly, some road segments will not have transit ridership counts. When data are requested for these road segments, the primary database system response should be to provide a message that states, "Data are not available for the requested road segment." It would then be useful to provide the user with a color coded map that would illustrate the nearest road segments for which data are available.

Depending on the nature of the data request, it may also be necessary to qualify the system response with dated information. For example, a response to a request for 1996 traffic on a particular road section might be, "Data for 1996 are not available, but a count was performed in 1994." Users would then have the ability to select (and adjust as they felt necessary) the older count data, use data from a different location, or pay for new data collection at the requested location. (Note that such a data collection request would be external to the PSRC database system.)

One currently foreseen instance may make it appropriate for the database system to interpolate between traffic counts: the production of a region wide traffic volume "flow" map. This map is a (often color coded) bandwidth plot of traffic volumes on major road segments. To produce this plot requires that every street segment to be plotted has a traffic volume associated with it. However, because the map is simply a bandwidth plot, the precision of the traffic volumes does not have to be high. (It is not possible to read a bandwidth precisely, and therefore the volume estimate upon which the width is based does not have to be very precise.) Consequently, the errors associated with interpolating between traffic counts are not significant.

When bandwidth plots are printed, volume numbers are customarily printed along with the volume bands. No interpolated volume numbers should be printed. Only actual volume counts should be printed in numeric form on these maps, and those numbers that are printed should be rounded up to no less than 100 vehicles (assuming AWDTs are being printed).

An algorithm for interpolating between available traffic volumes is presented below.

Interpolating Between Traffic Counts

Because the MTS network will be working with segmented roadways, it is assumed that traffic volumes within each roadway segment will be constant. Therefore, changes in roadway volume will occur only at segment ends (which are normally intersections). This means that it will not be necessary to compute smooth transitional volumes within a segment, even when given knowledge about the number of intersections within a segment. Consequently, the simplest procedure for computing the traffic volume on a segment with no volume data will be to average the volumes associated with the roadway segments on either end of the segment without traffic.[7]

Where more than one consecutive segment does not have traffic information, the difference between the two available traffic counts should be divided by the number of segments without traffic data plus one, with the result added to the lower volume segment's traffic volume to obtain the next segment's volume estimate. The incremental volume is then added once again to the newly computed roadway segment's volume to obtain the next roadway segment volume. This procedure is followed until all segments have assigned traffic volumes. This produces a smoothed transition (by segment) from the lower traffic volume to the higher traffic volume. It is a reasonable approach for producing a traffic volume map. It is not necessarily an appropriate approach for estimating traffic on these road segments for other, more exacting analyses.

This process is illustrated in Figure 4 below. In this example, the daily volume for segment A is 1000 vehicles per day, while the daily volume for segment D is 2000 vehicles per day. The incremental difference added to the lower volume (or subtracted from the higher volume) to compute estimated volumes for segments B and C is 333 [(2000-1000)/3 = 333]

---

7 It may be possible to perform this task with better accuracy using a more sophisticated approach, but given the scope of this project, and the difficulties in getting the database up and operating in the first place, investigation of better ways to perform this process can be left to the enhancement phase of the database development.

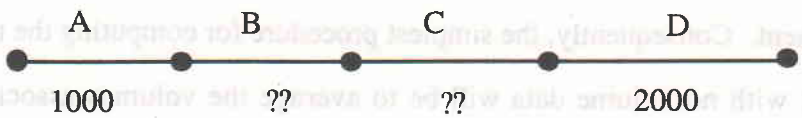


Figure 4
Interpolation of Missing Roadway Segment Volumes

Thus segment B is given an estimated volume of 1333. Segment C has an estimated volume of 1666.

If no count was performed on segment D and the road ended at this point, then the value for segment A would be used for all four segments shown in this figure.

Increment the Year Function

The year increment function is fairly simple. Once per year, the files that corresponds to the oldest traffic counts should be archived to tape (or optical disk) and removed from the database. New blank records should then be created for each roadway segment in the database,[8] and the database file should be saved.

This process could be performed at any time before entry of data for a seventh year into the database.[9]

---

8 Note that this assumes that all roadway segments in the database start with blank records. These records are then filled in when new traffic count information is uploaded from the lower level of the database, as data are submitted by the cities and counties. The alternative is to not have a record for a given road segment until data are present in the lower level. At this time, it is not clear which option is better. Data that might be "preloaded" into blank files include the directions that a road can operate (E/W/N/S), the year of the count, and any other static information (capacity?) that PSRC wishes to maintain with the upper level records.

9 If the database follows the convention that no record exists in the upper level until a traffic count exists to fill that record, this process can be run at any time. The database will simply contain more than six years of data. This will cause the database to grow in size more than intended, but it should not significantly affect the entry of data into the database.

System Output

It is envisioned that the system will be able to output data to

- computer screen
- printer
- electronic media (disk).

This will allow users to either obtain a single number (usually returned to their screen for review or for hand entry to some analytical process), a formal paper report, or an electronic file containing many traffic estimates that can be fed into some other analytical process.

The output format (single value, formatted table, or delimited file) should also be selectable by the user, although the program should develop a default for each of the available output types. (For example, the default output style for the electronic media output would be a delimited file.) The default output style for the paper report would be a tabular format.

All output must contain the appropriate disclaimer statements.

Long-Term Database Upkeep

The database has to include a maintenance process that will allow changes to the underlying MTS street network while maintaining backward compatibility with the traffic data already stored within the system. The most common requirement will be to either add roadway segments to the database, delete them from it, or change the roadway segments.

Two primary tasks will have to take place when a change occurs in the MTS network;

- a revision of existing database records
- alteration of the geocoding process used to identify specific GIS location references (segment ID or anchor points) based on the location referencing system used by the various cities and counties.

Revision of the database records will probably be the easiest of these changes.

Addition of a roadway segment would require the creation of blank road segments in the traffic database (if blanks segments are used) and the creation of new table entries for the geocoding process.

Deletion of a roadway segment could be accomplished by simply writing a program that would search on the appropriate roadway segment ID and delete any traffic volumes that existed for those records. However, to prevent new records from being entered into that old MTS roadway segment, the segment references would have to be removed from the geocoding database used to locate counts within the PSRC GIS.

Changing the existing MTS network (for example, splitting an existing roadway segment into two or more segments, or combining two segments into a one large segment) would require both the deletion and addition steps mentioned above, as well as a "reprocessing" step. To change the database, the first step would be to revise the geocoding process. That is, to remove the table references that identified the old segment and replace them with new table references that identified the new segments. The upper level of the database would then print out (for later use and as a safety precaution) all traffic volume records that applied to the roadway segments being changed. In addition, the transit information for these segments would be written to a separate file. Once these records had been printed, they could be deleted from the database.

The revision process would read the lower level database records that corresponded to the segments being changed. Using the revised geocoding table, these "old" counts would be relocated to the revised MTS network roadway segments. These revised records should then replace the original records and be used to compute new upper level database records. (This process could be done just as if they were newly added lower level records.) However, the procedure would need to be performed for each of the six years of data stored in the database.