Research Report
Research Project T9903, Task 29
Traffic Data Acquisition and Distribution (TDAD)

TRAFFIC DATA ACQUISITION AND DISTRIBUTION (TDAD)

by

D.J. Dailey, D. Meyers, L. Pond, and K. Guiberson
ITS Research Program
College of Engineering, Box 352500
University of Washington
Seattle, Washington 98195-2500

Washington State Transportation Center (TRAC)
University of Washington, Box 354802
University District Building
1107 NE 45th Street, Suite 535
Seattle, Washington 98105-4631

Washington State Department of Transportation
Technical Monitor
Pete Briglia

Prepared for

Washington State Transportation Commission
Department of Transportation
and in cooperation with
U.S. Department of Transportation
Federal Highway Administration

May 2002

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
WA-RD 484.1		

4. TITLE AND SUBTITLE	5. REPORT DATE
Traffic Data Acquisition and Distribution (TDAD)	May 2002
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
D.J. Dailey, D. Meyers, L. Pond, and K. Guiberson	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
Washington State Transportation Center (TRAC) University of Washington, Box 354802 University District Building; 1107 NE 45th Street, Suite 535 Seattle, Washington 98105-4631	11. CONTRACT OR GRANT NO. Agreement T9903, Task 29
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Research Office Washington State Department of Transportation Transportation Building, MS 47370 Olympia, Washington 98504-7370 Gary Ray, Project Manager, 360-705-7975	Final research report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

This study was conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. ABSTRACT

The wide variety of remote sensors used in Intelligent Transportation Systems (ITS) applications (loops, probe vehicles, radar, cameras, etc.) has created a need for general methods by which data can be shared among agencies and users who own disparate computer systems.

In this report, we present a methodology that demonstrates that it is possible to create, encode, and decode a self-describing data stream using the following:

1. existing data description language standards
2 parsers to enforce language compliance
3. a simple content language that flows out of the data description language
4. architecture neutral encoders and decodes based on ASN.1.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Travel time, data mine, JDBC, data user services, ADUS	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22616

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
None	None		

Disclaimer

Table of Contents

List of Tables

List of Figures

CHAPTER 1. INTRODUCTION

The Traffic Data Acquisition and Distribution (TDAD) project was completed in several stages. First, an analysis of possible users was undertaken. In this analysis, the possible uses of the data were identified, and the personnel who might take advantage of a data repository were interviewed. Second, an overall design for what has come to be called "Archived Data User Services" in the ITS National Architecture was developed. This design included an initial database schema design. This schema formed the basis for the self-describing data (SDD) stream that feeds the TDAD data repository. Third, a set of software components to implement the archive were developed itteratively. Database layouts were considered and a compact, efficient format was selected. Finally, a set of cooperating components were created to allow user interaction with the data repository. These stages are described in the following chapters.

Several projects/systems bear resemblance to the TDAD project. They include the ATIS (Advanced Traffic Information Service) Corporation, which provides real-time traffic information to Tokyo, Japan [Ito, 1995]; the Trip Reduction Information Management System (TRiMS) in Menlo Park, California, which provides transit route schedules and fares, bicycle routes, ride-matching services, and personal vehicle routing services (including ride and park); the Gary-Chicago-Milwaukee Corridor project, which provides real-time traffic data for Chicago and nearby regions [Dillinburg]; and TRAVLINK, which is being developed by the State of Minnesota. It provides both real-time traffic and transit information [Wright, 1993]. Other similar systems are described in [Kobayashi, 1995; Sobolewski, 1993; Wallace, 1993; Turnbull, 1993; Wallace, 1993].

The TDAD project is more versatile than any of these projects because it provides data for any time and location, and for any length of time interval. The goal of the TDAD project was to supply traffic data for a wide area, such as King County in Seattle, Washington, over long periods of time to enable various traffic agencies to perform research activities. The

literature survey we conducted revealed very few systems that store traffic transit data for long periods of time. These systems do not make the historic data available because it is very difficult to retrieve data from tape or other mass storage in an easy and flexible way [Ashbrook, 1983]. Turnbull [Turnbull, 1993] and Casey et al [Casey, 1993] point out that stored transit information can be used to enhance long-term route and schedule planning but give no specific examples of systems that actually implement it. Farber and Paley [Farber, 1993] used traffic data files (containing data recorded from a pair of traffic sensors called loop detectors) provided by the New Mexico Federal Highway Administration to develop a model for studying a collision warning system, but they do not mention any user interface through which they could access the traffic data files. Ashbrook [Turnbull, 1993] points out that the traffic data collected by Washington State DOT is aggregated at the five-minute level and saved on tape and is not easily accessible to an external user. We observed that in all these systems that store data for a long time, an external user does not have free and easy access to the stored data. TDAD distinguishes itself from all the existing systems by providing the stored data over long periods in an easily accessible and flexible form to either traffic researchers or interested users. This feature is an important contribution of TDAD.

CHAPTER 2. USER ANALYSIS FOR DESIGN

Traffic congestion is a serious concern for residents of the Puget Sound region (consisting of King, Kitsap, Pierce, and Snohomish counties). Operations personnel, researchers, and planners are working to alleviate this congestion, and they need traffic data in order to examine or respond to congestion problems. A variety of traffic data are currently being collected for real-time traffic analysis to enable traffic operations groups to keep the freeway and arterial road systems operating smoothly. However, if this traffic data could be archived and readily available, they would be extremely useful for historical purposes, such as planning and research.

The Federal Highway Administration (FHWA) would like to see a model demonstrating the use of new information technology to gather, store, and make available historical traffic data. It is interested in increasing the efficiency and decreasing the costs of traffic data collection and storage, and, therefore, the FHWA has funded the Traffic Data Acquisition and Distribution (TDAD) project. TDAD will access available traffic data in the Puget Sound Region and store them in a database for historical, research, and planning purposes. TDAD will provide for interagency and multi-jurisdictional sharing of data without interrupting existing operations or degrading the quality of data. The database will be flexible enough to add sources of traffic data as they become available. In addition, a central repository of traffic data will eliminate the duplication of data gathering, a situation which is prevalent today.

The first task in the TDAD project was to interview the potential users and providers of archived traffic data. The organizations that were interviewed were the Metro Planning Division, Puget Sound Regional Council (PSRC), Washington State Department of Transportation (WSDOT) Northwest Region Planning Office, WSDOT Planning Office, WSDOT Transportation Data Office, and WSDOT Office of Urban Mobility. In addition, researchers from Transportation Northwest (TRANSNOW), University of Washington Civil Engineering

Department, University of Washington Electrical Engineering Department, and Washington State Transportation Center (TRAC) were interviewed. (See Appendix A for a list of people attending the interviews and Appendix B for the notes from the interviews.) Each of these organizations and researchers uses or provides historical data in some aspect. While many agencies and individuals in Washington State use and provide historical traffic data, the organizations listed above play an important role in the traffic management of Washington State and the Puget Sound Region.

PSRC is the metropolitan planning organization (MPO) and the regional transportation planning organization (RTPO) for the Puget Sound region. These titles mean that all projects affecting regional transportation must first gain PSRC's approval before funding is received. The WSDOT Planning Office is responsible for the Congestion Management System (CMS) for the whole state. The WSDOT Transportation Data Office is in charge of meeting the federal data requirements for Washington. The Office of Urban Mobility examines new traffic projects to determine whether they meet their stated goals for city and county. The Metro Planning Division analyzes data procured through the automatic vehicle location system used on Metro buses. WSDOT's Northwest Region Planning Office gathers a variety of specific traffic data upon request for authorized agencies. Lastly, the researchers from the University of Washington, TRANSNOW, and TRAC explore technologies that may directly affect transportation in Washington State and, particularly, the Puget Sound region. Many of these groups are potential data providers as well as potential data users. This report will focus on these agencies' roles as data users.

Some patterns in the usage of historical traffic data emerged out of our interviews. These organizations use traffic data in any of five ways: (1) reporting, (2) long-term planning, (3) project planning, (4) performance monitoring, or (5) research. Organizations may use traffic data for all of these purposes or focus on some subset. The following section explores each of the users of historical traffic data, the goals of each type of use, and the specific traffic data needed to reach that goal. A summary of the traffic data currently used

by the agencies that were interviewed can be found in Table 1 at the end of this chapter. Table 2, also at the end of this chapter, summarizes the additional traffic data that the organizations indicated would be useful for their purposes.

2.1 REPORTING

A traffic report is a compilation of traffic data presented in a meaningful format. There are as many reports possible as there are kinds of traffic data. For example, subjects of traffic reports can be the geometrics of the roadway, the history of traffic patterns, accident reports, or even state route listings. However, when traffic data are reported for any federal purpose, it is important that the data meet the guidelines set forth in the federal Traffic Monitoring Guide (TMG), AASHTO Guidelines for Traffic, the Highway Performance Monitoring System (HPMS) Manual, and the new requirements for a Traffic Monitoring System for Highways (TMS/H). These guidelines offer a standard for collecting and reducing traffic data to ensure that data from many disparate sites can be recorded, catalogued, and reported consistently. The WSDOT Transportation Data Office is the only agency of those interviewed that uses traffic data for reporting.

The WSDOT Transportation Data Office produces many different reports. In fact, one of this office's responsibilities is to comply with the Federal Highway Administration's (FHWA) reporting requirements. For example, the Transportation Data Office sends the FHWA the Highway Performance Monitoring System (HPMS) report that is required of all states. The HPMS report must include the state's annual average daily traffic (AADT) count, average vehicle miles traveled (AVMT), vehicle classifications, and truck weights.

Another report that the Data Office produces, the Annual Traffic Report, is used by a wide variety of individuals and agencies in Washington State. The Annual Traffic Report is a list of the annual average daily traffic counts at each of WSDOT's automatic data collection locations. Included in this report is the percentage of trucks in the traffic.

The Ramp and Roadway Report is also widely used in the Puget Sound region. While this report is actually produced by the Northwest Region Traffic Systems Management

Center (TSMC), it is a compilation of data collected by the WSDOT Transportation Data Office and TSMC. This report includes volume data on the on- and off-ramps that are collected by the WSDOT Transportation Data Office and volume data on the freeways that are collected by TSMC.

To produce these and other reports, the WSDOT Transportation Data Office maintains and relies on traffic data in the Transportation, Information, and Planning Support system (TRIPS). The TRIPS system is WSDOT's database for storing reduced traffic data that meet federal requirements. In fact, the TRIPS system was partially derived directly from the Highway Performance Monitoring System (HPMS) guidelines. The database includes accident data, data from automatic data dollectors (ADC), truck classifications, roadway geometrics, and data from special requests. (See Appendix C for a schematic of the TRIPS system, provided to the TDAD research team courtesy of the WSDOT Transportation Data Office.)

Some of the traffic data that are stored in the TRIPS database are obtained from WSDOT's 150 permanent collection sites. Automatic data collectors (ADC) at these sites gather traffic volume data, and a few sites record vehicle classification data. The data from traffic volume counts and truck classifications can be reduced to AADT volume, average weekday (AWD) volume, average weekend day (AWED) volume, peak hour split percent, peak hour traffic percent, peak hour truck percent, and traffic volume truck percent for reporting purposes. (See Appendix D for a glossary of terms.)

The Transportation Data Office expressed an interest in receiving more of the same type of traffic data from other sources, as well as including traffic data that are not currently in the TRIPS database. If other agencies are gathering the same data that some of the WSDOT collection sites are gathering, then the Data Office could check, augment, or even remove the WSDOT collection sites. It could save money on the time and cost needed to maintain the equipment without diminishing its supply of data. In particular, the Data Office would like more vehicle classification data. As for different types of data, the Transportation Data

Office hopes to include speed data and vehicle occupancy data in its TRIPS database. In addition, it would like to have access to hourly and 15-minute volume data.

The Transportation Data Office would also like to establish an electronic connection between its TRIPS system and TSMC's freeway data. Currently, the Data Office receives only a hard copy of freeway data. Having these data in an electronic format would greatly help the Data Office reduce and compare TSMC's data with its own.

2.2 LONG-TERM PLANNING

When making plans for a long term or range of time, traffic systems are viewed at a broad level. In this application, traffic data are used to project traffic needs of the future. The Puget Sound Regional Council, the WSDOT Planning Office, and the Office of Urban Mobility are the three groups of those interviewed that use traffic data for long-term planning.

2.2.1 Puget Sound Regional Council

As mentioned previously, PSRC is the designated MPO and the RTPO for the Puget Sound region. It is required by the federal Intermodal Surface Transportation Efficiency Act (ISTEA) to adopt a growth plan and transportation strategy for the region. This plan must outline how the region will conform with other federal mandates such as the Clean Air Act Amendments. In addition, as the MPO, the Council is responsible for the Congestion Management System for the region. This, too, is included in the long-term plan, which it has named Vision 2020. The transportation strategy presented in Vision 2020 is intended to be a guide for other organizations working on traffic projects in the region.

To successfully plan a long-term strategy, PSRC relies on forecasts of the population, economic trends, and traffic demands for the next decade. For traffic demands, PSRC uses a travel demand model that forecasts the volumes of traffic that will be flowing across screenlines in the base years of the decades (1990, 2000, 2010). PSRC defines a screenline as a slice across a corridor of travel. The planners use actual traffic data to validate the

forecasts. If the computer model has forecast volumes that are within four to five percent of the actual traffic counts, PSRC considers the model to be producing an accurate forecast.

Currently, PSRC uses 24-hour volume counts, A.M. peak period counts, and vehicle miles traveled (VMT) data to validate its traffic forecasts. It receives volume counts from the individual cities in the region in the form of maps with volume counts written beside the roadways. PSRC also receives the Annual Traffic Report and the Ramp and Roadway Report in a hard copy format. From the Annual Traffic Report, PSRC takes the AADT volumes, and from the Ramp and Roadway Report, it obtains volume counts and peak-hour volume counts on the freeways and ramps. (See Section 2.1 for a description of these reports.) In addition to the Annual Traffic Report, the WSDOT Transportation Data Office sends PSRC the VMT as reported in the Highway Performance Monitoring System (HPMS).

PSRC would also like to check the validity of the model's forecasts by comparing interzonal model speeds with directly measured speeds. An hourly average speed and the distribution of speed within that hour would meet its requirements. In addition, PSRC would like to validate the forecasts with vehicle classification and average vehicle occupancy (AVO) data. Neither type of data is available for the whole region.

2.2.2 WSDOT Planning Office

The WSDOT Planning Office also uses traffic data for long-term planning. The Planning Office is responsible for the Congestion Management System (CMS) for the whole state. In each of the major urban areas (Puget Sound, Spokane, and Vancouver), the Planning Office works with the MPO on the planning and implementation of the CMS for those cities.

The Planning Office measures congestion by determining the efficiency of the freeways and road systems. It finds efficiency by calculating the volume of cars over the capacity of the roads. It charts the efficiency over time and extrapolates out efficiency figures for a decade or two to help form a plan to reduce traffic congestion. The Planning Office receives the volume counts from the information published by the WSDOT Transportation Data Office in the Annual Traffic Report. It also receives forecasted efficiency data from PSRC.

However, the Planning Office needs other types of traffic data that are not currently available. Foremost, this office needs measured travel times for both freeway and arterial systems. This type of traffic data could be obtained, for example, by using a fleet of probe vehicles with automatic vehicle identification (AVI) tags. Like PSRC, the Planning Office also wants vehicle classification and vehicle occupancy data. Lastly, it would like to see figures on hourly and peak hourly traffic volumes rather than just average volumes for the day.

2.2.3 WSDOT Office of Urban Mobility

The WSDOT Office of Urban Mobility also uses traffic data for long-term planning. Specifically, the Office of Urban Mobility refers to such planning as long-range planning. As indicated in its Mission and Activities statement, it is "responsible for long-range, multi-modal transportation system planning in the Puget Sound region." To accomplish this, the Office of Urban Mobility participates in Growth Management Act activities with a focus on integrating the local plans from the four counties in the region with WSDOT long-range plans. This agency also performs the corridor studies that affect either regional issues or the transportation network as a whole. In addition, the Office of Urban Mobility is involved in both the State Systems Plan and the Regional Transit Project, which is a plan for a regional high occupancy vehicle (HOV) system.

To find the traffic data that are needed for long-range planning, the Office of Urban Mobility first turns to a hard copy of the Annual Traffic Report that it receives from the WSDOT Transportation Data Office. The Office of Urban Mobility uses the AADT volumes listed in this report and determines what other types of traffic data it needs. Some examples of these additional data are travel times, speed, vehicle occupancy, turning data, specific car counts, transit use, vehicle classification, pedestrian counts, and bicycle counts. The Office of Urban Mobility also receives projected volume data from PSRC.

After examining the traffic data that it currently has, the Office of Urban Mobility often turns to the WSDOT Northwest Region and Olympic Region planning offices to supplement

those data. The regional planning offices conduct manual turning movement, vehicle occupancy, and vehicle classification counts upon request. In addition, they gather volume counts at specific locations on arterial systems. To obtain the rest of the traffic data, the Office of Urban Mobility may hire consultants to further supplement the information.

2.3 PROJECT PLANNING

In project planning, the traffic data used are specific to the area that a proposed project will affect. The data can help in the project development process, or they can be used to check that a project will meet its goals. Since projects cannot rely on averages of traffic data over the region, and the type of data needed depends on each project, the specific data must often be manually gathered. There are many traffic projects, and subsequently, there are many different users of traffic data for project planning. Of those interviewed, the WSDOT Office of Urban Mobility uses traffic data for project planning.

For every long-range plan or activity that the Office of Urban Mobility is involved with, it takes a specific segment of the highway network that is affected by the plan and examines this section on a subarea level. Therefore, corridor studies, the HOV Regional Transit Project, and the integration of Growth Management Act plans are all segmented and studied at the subarea level. In addition, the Office of Urban Mobility's Mission and Activities Statement indicates that it "facilitate[s] a comprehensive review of major development proposals," especially when the projects cross jurisdictional boundaries.

To accomplish project level planning, the Office of Urban Mobility uses the same traffic data that it uses for long-term planning. It uses these data to determine whether the project will accomplish its goals not only with the demand of today's traffic but also with the expected traffic demand of tomorrow. Again, the traffic data it requires includes AADT, travel times, speed, vehicle occupancy, turning data, specific car counts, transit use, vehicle classification, pedestrian counts, bicycle counts, and forecasted travel data. (See Section 2.2.3 for a complete description).

2.4 PERFORMANCE MONITORING

Archived traffic data can also be used for performance monitoring, a relatively new application of traffic data. Transportation professionals have realized that it is not enough to just plan traffic projects and programs. Current traffic conditions must be constantly monitored to ensure that the plan is working. Performance monitoring can be used to track which traffic programs and projects are succeeding, which are failing, and to what extent a program has worked. The information from performance monitoring becomes an excellent source of feedback for both long-term and project planners.

Much of the push to establish performance monitoring programs has been from the government. The Washington State Growth Management Act of 1990 and the Clean Air Act both have provisions requiring performance monitoring. The federal Intermodal Surface Transportation Efficiency Act (ISTEA), which requires the state to have a Congestion Management System, also requires the state to closely monitor the effectiveness of congestion reduction strategies. PSRC and the WSDOT Office of Urban Mobility are both required to initiate performance monitoring programs.

2.4.1 Puget Sound Regional Council

The Puget Sound Regional Council must monitor the performance of its long-term Vision 2020 plan. Within this are the steps to meet the Clean Air Act Amendment, the plan to meet the Commute Trip Reduction Act, and the Congestion Management System (CMS). All of these programs within the regional plan must be monitored to ensure that they are being implemented properly.

Having just undertaken this project, PSRC is finding that the usual traffic data it uses for long-term planning are perhaps insufficient to implement performance monitoring. The information it receives in the Annual Traffic Report and the Ramp and Roadway Report is often not specific or timely enough. PSRC would like more data on measured speed and travel time. It wants vehicle classification and average vehicle occupancy data. It would also

like the AVL data that are currently being collected on Metro buses for Metro's Operations Department. This information would provide timely traffic conditions in the region and possibly even actual travel times on freeways.

2.4.2 WSDOT Office of Urban Mobility

The Office of Urban Mobility uses traffic data to monitor performance on the Congestion Management System (CMS) network. By monitoring performance, the Office of Urban Mobility can identify and focus on areas of high congestion. In keeping with the policies in PSRC's long-range regional plan, Vision 2020, and countywide plans, the Office of Urban Mobility recommends a performance monitoring system for the region that centers around the movement of persons and goods over vehicles.

The Office of Urban Mobility currently uses volume counts and incident data to monitor performance of the CMS network. Like PSRC, the Office of Urban Mobility has discovered that some of the traffic data that are most effective for performance monitoring are not currently available. Travel time data would be particularly useful in determining the movement of automobiles, trucks, carpools, etc. The Office of Urban Mobility would also like to have vehicle occupancy and transit use data available across the entire CMS network.

2.5 RESEARCH

Finally, archived traffic data are used for research purposes. The goals of research are to develop both implementable transportation services for the traveling public and more effective and efficient methods of traffic control and monitoring for transportation managers. Researchers from Transportation Northwest (TRANSNOW), the University of Washington, and the Washington State Transportation Center (TRAC) were interviewed.

Researchers engage in a variety of research activities related to transportation and, consequently, use a variety of traffic data. However, most researchers use volume counts to aid in their research efforts. Volume counts can be obtained from the Ramp and Roadway Report, from the Annual Traffic Report, or directly from the Traffic Systems Management

Center (TSMC). (See Section 2.1 for more detail about the reports.) From TSMC, researchers request volume counts in time increments of 20 seconds, 1 minute, 5 minutes, or 15 minutes. Generally, researchers like to have access to short-time increments of traffic data. Data can always be aggregated, but aggregated data cannot be expanded back to the original data. With raw data, researchers can extract the data that fit their particular information needs.

Researchers also obtain speed, vehicle classification, and vehicle occupancy data from several different sources. However, the amount of speed, vehicle classification, and vehicle occupancy data collected by the current sources is limited. Most researchers would like to have more of this type of data over a broader region. Researchers would also like to receive travel time data.

2.6 CONCLUSIONS

From interviews with various users of traffic data, five distinct usages were found. The traffic data used for each of those purposes are summarized in Table 1. All the organizations interviewed indicated that they obtain traffic data from the Annual Traffic Report, and the majority also use the Ramp and Roadway Report. The remainder of traffic data are acquired from a variety of specialized sources. These specific data are most commonly procured from tube count collections, the WSDOT Northwest Region Planning Office, and the Puget Sound Regional Council. The Annual Traffic Report and the Ramp and Roadway Report are both produced from automated systems of data collection, meaning that the gathered data are automatically recorded in an electronic medium. As a result, that information should be easily incorporated into the TDAD database system. On the other hand, the traffic data from the specialized sources that were mentioned above are not collected with an automated system. Therefore, it would be much more difficult to include this information in an automatic archiving system.

Table 1: Data and Sources

Type of Data Use	User	Data Used	Source
Long-Term Planning	PSRC	AADT Volume HPMS VMT 24 Hr. & Peak Volume Counts 24 Hr. Volume Counts	Annual Traffic Report WSDOT Data Office Ramp & Roadway Report City & County Tube Collections
	WSDOT Planning Office	Volume Counts Forecasted Efficiency Data	Annual Traffic Report PSRC
Performance Monitoring	PSRC	AADT Volume 24 Hr. & Peak Volume Counts 24 Hr. Volume Counts	Annual Traffic Report Ramp & Roadway Report City & County Tube Collections
Long-Range Planning & Project Planning	WSDOT Transportation Data Office	AADT Volumes Projected Volume Data Turning Movements Vehicle Occupancy Vehicle Classification Specific Volume Counts Travel Time & Speed Transit Use Pedestrian & Bicylce Counts	Annual Traffic Report PSRC NW Region Planning Office NW Region Planning Office NW Region Planning Office NW Region Planning Office Consultants Consultants Consultants
Performance Monitoring	WSDOT Office of Urban Mobility	Volume Counts Incident Data	TRIPS system TRIPS system
Research	TRAC, TRANSNOW, & University of Washington Researchers	20 sec., 1 min., 5 min., 15 min. Volume Counts & Lane Occupancy Peak Volume Counts AADT Volumes Speed Vehicle Classification Vehicle Occupancy	TSMC Ramp & Roadway Report Annual Traffic Report ADCs, autoscope WSDOT Data Office, ADCS, autoscope TRAC

The additional data that these organizations feel they need to effectively perform their tasks are outlined in Table 2. All of the organizations interviewed would like to have access to speed, travel time, vehicle classification, and vehicle occupancy data. Two potential sources of these additional data, the WSDOT Northwest Regional Planning Office and Metro's Planning Division, were also interviewed by the TDAD research team.

The Northwest Region Planning Office undertakes short-term data collection projects on arterial routes by request from authorized WSDOT departments. It collects whatever traffic data are necessary to determine coverage counts, traffic flow, signal timing, or site impact studies, to name a few. These data usually take the form of vehicle occupancy, turning movements, vehicle classification, speed, average daily traffic counts, etc. All of the col-

lected data are organized in a standard spreadsheet. Supervisors in the Northwest Region Planning Office have already been working with their counterparts in other WSDOT regions and Puget Sound cities and counties to standardize collected data by placing them in a spreadsheet format.

Metro has an AVL system and an automatic passenger counter (APC) system for their bus fleet. Currently, any data collected by these systems are intended strictly for internal operations. However, since the AVL system tracks the buses by polling them for the vehicle location along a planned route, the system could provide useful data on the speed of traffic on the highways, particularly the HOV lanes. It is more difficult to determine speed data for traffic on arterial routes because the buses must stop to pick up passengers. The APC system, which counts and records the number of passengers entering and exiting the bus, is an excellent source of vehicle occupancy data.

As can be seen, the Northwest Region Planning Office and Metro collect a limited number of vehicle classification, vehicle occupancy, speed, and travel-time data. Similar to the Annual Traffic Report and the Ramp and Roadway Report, speed and travel-time data from Metro's AVL system are collected with an automated system. On the other hand, the Northwest Region Planning Office collects vehicle classification and vehicle occupancy data manually. As mentioned previously, non-automated sources of traffic data are more difficult to incorporate into an automatic archiving database.

In summary, this chapter focused on how historical traffic data can be used for reporting, long-term planning, project planning, performance monitoring, and research. Chapter 3 will translate this user analysis into a system design plan for archiving historical data, and Chapter 4 will discuss the communications necessary to acquire and deliver these data.

Table 2: Data Requested by User

User	Type of Data Use	Additional Data Requested
PSRC	Long-Term Planning Performance Monitoring	Speed Travel Time Vehicle Classificatio Vehicle Occupancy Hourly Volume Counts Speed Travel Time Vehicle Classification Vehicle Occupancy AVL Data
WSDOT Planning Office	Long-Term Planning	Travel Time Vehicle Classification Vehicle Occupancy Hourly Volume Counts
WSDOT Transportation Data Office	Reporting	Speed More Vehicle Classification Vehicle Occupancy Hourly & 15 min. Volumes Electronic Loop DAta Other traffic data collected under federal guidelines
WSDOT Office of Urban Mobility	Long-Range Planning & Project Planning Performance Monitoring	Speed Travel Time Vehicle Classification Vehicle Occupancy Truning Movement Travel Time Vehicle Occupancy Transit Use
TRAC, TRANSNNOW, & University of Washington Researchers	Research	Speed Travle Time Vehicle Classification Vehicle Occupancy Vehicle Miles Traveled Origin/Destination Data

CHAPTER 3. DATABASE SCHEMA

Currently, loop data from the Traffic Systems Management Center (TSMC) and automatic vehicle location (AVL) data from the Municipality of Metropolitan Seattle (Metro) are available for the TDAD database. These data provide information on freeway volume, freeway occupancy, and to some extent, freeway speed and freeway travel time. These data types are useful to the operations personnel who require them for real-time analysis, and they are also useful to researchers and planners who analyze archived traffic data. The researchers and planners who were interviewed at the start of this project indicated that they currently use freeway volume and freeway occupancy data and that they would like to have access to freeway travel time and freeway speed data (see Chapter 2. User Analysis).

To provide the data in a form that is accessible and usable by researchers and planners, the database schema is designed as two sets of tables. The first set of tables completely defines all the information for the loop data, and the second set of tables completely defines all the information for the automatic vehicle location data on Metro vehicles. This structure will enable data from new sources, such as the North Seattle Advanced Traffic Management System (NS ATMS), to be added as they become available.

The development of each of these sets of tables is described in the following two sections. Section 3.1 describes the schema for the loop data, and Section 3.2 describes the schema for the AVL data. In addition, Appendix E provides all of the tables in TDAD, giving the table name, the columns in the tables, the datatypes for each column, the default settings for the columns, and other special restrictions on the columns. For a complete written description of what data goes into the columns, see Appendix F.

3.1 LOOP SCHEMA

Before the way that the loop data are stored in the TDAD database is described, the way that the TSMC collects the data must be understood. TSMC collects data from the sensor

loops in the freeway every 20 seconds. These loops record the number of vehicles that pass over them (volume) and the percentage of time that it took those vehicles to pass over (occupancy). Some loops also record the bin counts (length of the vehicle) and the speed the vehicles were traveling. For archival purposes, TSMC combines all of the information from the loops into one data stream.

Data from one specific loop can be found by knowing how many loops are contributing data to this data stream and the placement of the specific loop's data relative to the other loops. The loop's placement relative to other loops is known as the loop's offset. For example, 500 loops sensors may be active and sending data to the TSMC. If data from a specific loop, call it X, are listed third, then all the data from X can be found by going to the third element in the data stream and then counting off 500 elements to the next bit of data from X. Counting off another 500 elements will bring up X's next bit of data, and so on.

TSMC continues storing the data stream in this manner until the status of the loops changes, which occurs either when one of the loops currently in the data stream becomes inoperable or when another loop that is not currently in the data stream becomes operable. In either case, the number of loops contributing data and the loops' offsets change. Therefore, TSMC starts a new data stream and includes a header that gives information about the new number of loops in the stream and the offset of each loop. This data stream and the accompanying header are broadcast across the Internet for storage in the TDAD database (see Chapter 4 for more information on the communication protocols).

Given that this large number of data was already structured, it made sense for the TDAD database to store the data stream on a separate and large storage medium, such as an optical jukebox. Then, only the pointers to the locations of the specific loops would need to be stored within the database itself. This design decision would enable users to easily access loop data without letting the loop data overwhelm and exceed the storage available in the database.

Information about the location of loop data in the data stream is stored in tables Loop_Measures, File_Scheme, and Loop_Time. A new row is added to File_Scheme every time TSMC starts a new data stream. This table stores the time that the new stream begins and the name of the file in the storage media where the data stream is being stored for TDAD. Also, this table stores the header information, which indicates how many loops are included in the data stream and how often data are collected from the loops.

The Loop_Time table provides the information for finding data from a unique loop in the storage medium. Every time that a new data stream begins, each loop's unique ID, the name of the file where the new data stream is stored, and the offset of each loop in the file are added to the table. This information, combined with the information in File_Scheme, enables a person to locate data from a particular loop.

Lastly, the information about what data the loop has collected is stored in Loop_Measures. Loop_Measures contains the unique loop identification and a description of what data the loop collects. A loop typically collects both volume and occupancy data, and some loops collect speed and vehicle length as well. Each type of data collected is specified by a unique row in the table. The information about the loop and the type of data collected is stored with information about the units of measurement (e.g., the number of vehicles for volume data), the storage size of the collected data, and the sequence of this data type among the other types of data that the loop collects. Also, each type of data collected has a status field associated with it, indicating whether the data are valid. The information about the size and the order in which the loop sends the status information is also stored in Loop_Measures.

In addition to the File_Scheme, Loop_Time, and Loop_Measures tables, TDAD contains four other tables that store additional information about the loops. These tables are Cabinets, Loops, Coordinates, and Loop_Locations. The Cabinets table stores general information about the TSMC's cabinets, including the freeway they are on, whether they are on a ramp, and a textual description of the locations. The Loops table stores information about which cabinet the loop belongs to, as well as information that is implicit in the loop's

ID. This additional information about the loops includes the type of road they are on, the direction of the road, the lane type, the lane number, and a textual description about each loop's location.

The Coordinates and Loop_Locations tables store information about each loop's geolocation. Since the geolocation can be found through a number of different coordinate systems, the Coordinates table stores the name of each different system and what each system measures. For example, the spherical coordinate system measures the latitude, longitude, and azimuth in degrees. The Loop_Locations table stores the name of the coordinate system used and the geolocation values of each unique loop. In addition, this table indicates which reference system was used for the coordinate system, as well as the reference points and the accuracy of the measurements.

3.2 AVL SCHEMA

As with the loop data, a description of how AVL data are stored in TDAD must be preceded by a description of how Metro collects the AVL data. Metro determines the location of the vehicles relative to their preassigned travel routes by polling its vehicles about once every minute for information concerning how far the vehicle has traveled since it was last polled and what "signpost" signals it detected. With this information, the Metro control center determines whether the vehicle is "on-route." Metro then calculates the vehicle's position relative to that route. To collect and derive this information from all the active vehicles in the fleet, Metro must poll approximately eleven vehicles every second.

Thus, Metro has a new and complete data packet containing the all the information about a vehicle and its location every minute. Given the atomic nature of the AVL data, it made sense for the TDAD database to store each AVL data packet intact in the database. Unlike the loop data, the AVL data are not collected in a structured schema, making it impossible to store all of the data on a different storage medium and keep only the pointers to the data in the TDAD database.

Therefore, the AVL_Data table stores all the pertinent information from Metro's AVL data packets. The table contains the vehicle ID, the route it is traveling, the block number (which indicates the set of routes that the vehicle is scheduled to travel), the distance into the route that it has traveled, the last signpost signal detected, the distance it has traveled since the last signpost signal was detected, the deviation from the schedule, its performance status, and the name of the route pattern that it is following. All of this information is available from the Metro AVL data packet. This table also stores some additional information that can obtained from the University of Washington Intelligent Transportation System (ITS) Research Team's AVL server. This server takes the information given in the Metro AVL packet and calculates the current geolocation of the vehicle in latitude and longitude. From this server, the TDAD database also obtains the unique trip number that specifies what trip the vehicle is traveling, a number that directly maps into scheduling information. (The trip number was actually developed for the Seattle Wide-Area Information for Travelers (SWIFT) project.) With the information in this table, researchers and planners can access time and geolocation data from a particular bus to determine speed and travel-time data.

The next four tables are Route, Destination, Schedule, and Intersection, and they store information about the schedules for Metro vehicles. First, the Route table contains information about the route: the route number, the service day, whether it is an express route, the direction of the route (inbound or outbound), and the part (north, south, east, or west). Because the same route number can describe several different trips - e.g., Route 2 heading toward downtown Seattle going north or Route 2 heading away from downtown Seattle going south - the route number given in the AVL data packet does not give enough information about which part of the route the vehicle is traveling. Instead, the specific part of the route on which the vehicle is traveling can be found either with the trip number or with the block number and the time that is stored in the AVL_Data table. In addition, this table contains fields that indicate the dates during which the route information is valid, since Metro may change the specific information about a route over the years.

The next table, Destination, contains information about where a vehicle that is traveling on a specific route with a given destination and part is traveling to. For example, when a bus is traveling north on Route 2 headed outbound (going away from downtown Seattle), its destination is West Queen Anne Hill. Next, the Schedule table contains information about all the scheduled stops (or time points, as they are referred to in this table) and the time the vehicle should arrive at the stop for each trip. Lastly, the Intersection table further defines the "time points," or the scheduled stops, that are given in the Schedule table. Each time point is defined in terms of the textual description of the intersection where it is located, e.g., 45th and University Way, and is also defined by the latitude and longitude of the stop's geolocation.

The last AVL table is called Patterns. This table contains the name of all the route patterns and a pointer to where the pattern is stored. Because graphical pattern files consume so much disk space and because pattern files are static and easily identified by their names, the TDAD database is able to store only the pointers to the patterns and can store the actual patterns on a separate storage medium.

3.3 CONCLUSION

With all the tables for the loop data and Metro AVL data that have just been described, TDAD furnishes researchers and planners with an archived version of the data that are currently available to operations personnel. With these data, researchers and planners can find volume and occupancy data on the freeways for any point in time. In addition, TDAD gives them access to some speed and travel time data. The next chapter discusses the communications protocols for acquiring and delivering the archived data.

CHAPTER 4. TRAFFIC DATA ACQUISITION AND DISPLAY (TDAD) CONSTRUCTION

4.1 INTRODUCTION

The Washington State Department of Transportation has installed thousands of roadside inductance-loop sensors to support its highway monitoring and control operations in the Puget Sound area. A focus of the Intelligent Transportation Systems Research Group at the University of Washington is to make these data available to students and the public for statistical and quantitative analysis.

The group provides software tools that can present the real-time state of this system. Previous traffic investigations were carried out by using these tools to collect a representative sample of the data over a period of hours, days, or weeks before importing the data into analysis programs such as MATLAB. This was a tedious task for one of two reasons. First, if all of the data were collected, the resulting output file would quickly reach an ungainly size, the collection process would require frequent attention, and the data would need laborious processing to make them acceptable to the intended environment. If, on the other hand, a filter was applied to create a collection of only the necessary data (selecting only a few representative sensors, for example), then a program had to be written to implement this filtering process for each new investigation. If a different filter was found to be necessary, requiring changes to this program, then the prior investment of collection time was wasted.

Database technology presented an attractive alternative: what if the time-varying state of the sensors could be stored in a database? This would leverage all the expertise in storage and query facilities provided by a modern database management system (DBMS). Rather than making their own collections over time, researchers could access an authoritative history of the sensors since the inception of the database. This idea motivated the TDAD project. This chapter will discuss the design of TDAD, a system for storing and querying vast quantities of time- and location-dependent data.

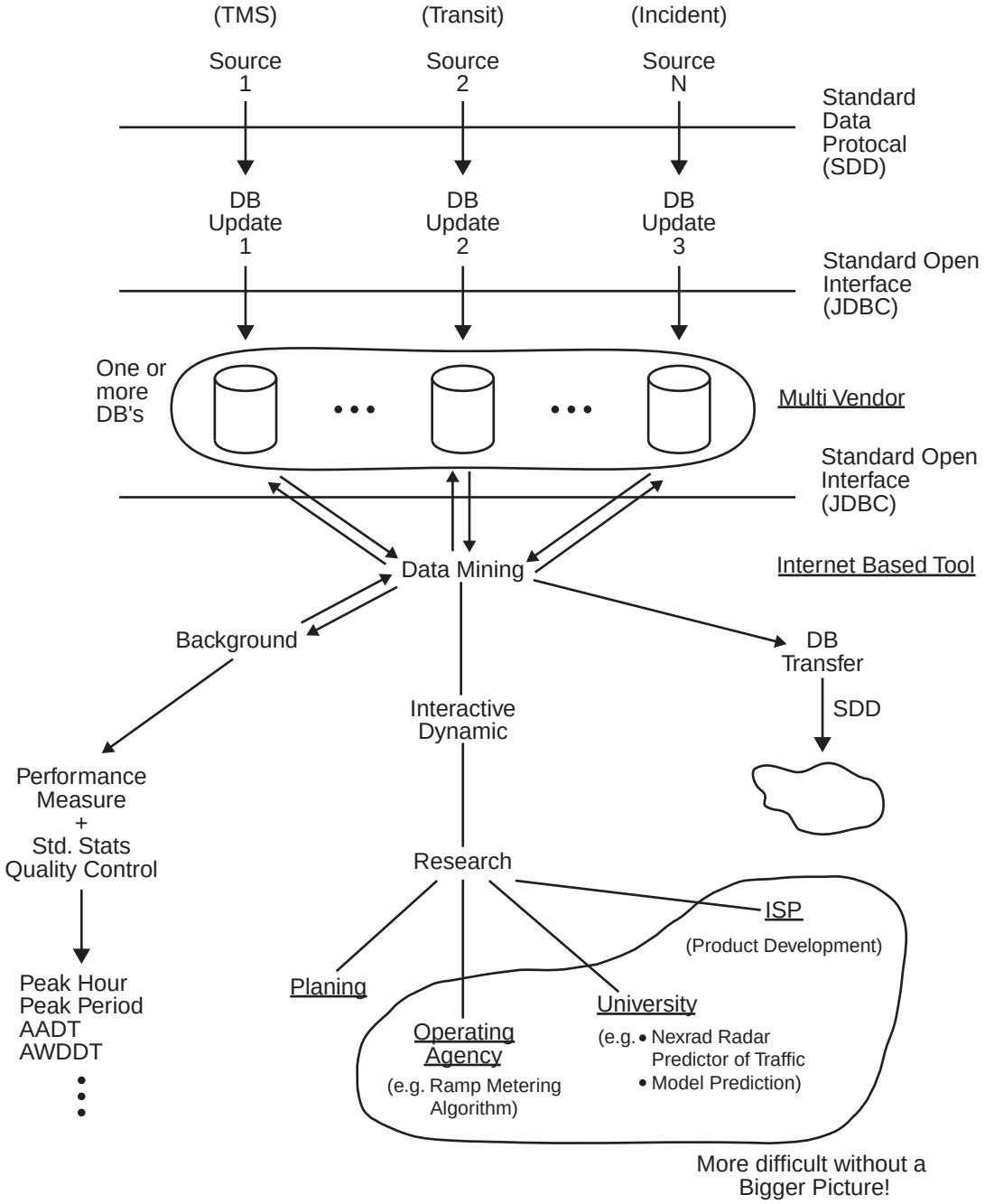


Figure 1: TDAD Query Interface ITS Data Archive/Use Model

4.2 ARCHIVED DATA USER SERVICES

In building a data mine for the WSDOT roadway data, we first describe the overall paradigm into which such a data mine will fit. Figure 1 shows the overall model for collecting data from a variety of sources and making those data available for different types of

activities. Several principals are important in this model: (1) Each of the data sources is interfaced to a standard data protocol; in our case, this is self describing data. (2) The data transfer protocol is used with a standard open interface to update one or more databases; in our case, this open interface to multiple vendors is the Java Database Connectivity (JDBC) that is available from all of the major database vendors. (3) Data mining is made available over the Internet; in our case, this is done using HTML and Java applets and makes access for a variety of already defined uses straightforward, as well as providing a clearly defined method for new uses to access the data mine. (4) There are several identifiable forms of performing the data mining: (a) a background activity that produces a standard set of performance measures and standard data sets that are widely required and used; in our case, this is a background process that produces HTML pages of the standard measures, (b) an interactive data interface that supports a variety of research activities; in our case, this is a Java applet accessed over the Internet, and (c) a large scale database transfer when it is desirable to transfer the entire database to a remote system for some large-scale processing. The TDAD data mine supports all the above activities and paradigms.

4.3 DATABASE DESIGN

4.3.1 An incoming flood of data

The output of WSDOT's Traffic Management System (TMS) is truly prodigious. It contains a dictionary component, which describes the name, position, and type of each of approximately 5,000 sensors. This dictionary is necessary to understand the sensor block component, which contains current measurements taken from each of the 5,000 sensors and is generated afresh every 20 seconds. The dictionary may change at any time to reflect the effects of road construction or installation of new sensors, though in practice several weeks may pass without observing a change.

Maintaining the distinction between dictionary and sensor block allowed the TMS developers to minimize their representation of the real-time state of the sensors. This was a

useful design goal for various reasons. Within the TMS, the previous hour's worth of sensor blocks are always in memory, so they must fit within a constrained resource. In the sensor block, the status of each sensor, containing several measured values, is squeezed into either 3 or 6 bytes, depending on sensor type. When all 5,000 sensors report this way, the total amount of storage required is around 22 kilobytes. The drawback to this representation is that the sensor block is not immediately legible to a human; a program must use the correct version of the dictionary to transform the sensor block into a textual representation.

The fundamental task of TDAD is to record both the dictionary (as it occasionally changes) and the information inside the sensor blocks (which are generated by the TMS every 20 seconds) in a database. Designing the database tables that would store this information was the necessary first step.

4.3.2 Most straightforward approach was unfeasible

An initial design of the TDAD database focused on ease and expressiveness of queries. Someone familiar with Structured Query Language (SQL) would be able to use the SQL tool of their choice to issue commands such as "Show me all the places where the average speed exceeded 65 miles per hour at 5 p.m. on Tuesday." This meant exposing the information within the sensor blocks as columns in a database table so that SQL statements could reference them. In this design, there was a table for each of the three sensor types, and the arrival of a new sensor block triggered the addition of a new row to one of these tables for each of the sensors in the block.

Although various measures were taken to economize, the growth rate of the sensor tables proved to be unmanageable. The fact remained that 15,000 new rows were added to the sensor tables every minute. In testing, a 520 megabyte partition was consumed in 13.3 hours, implying a storage consumption of 666 kilobytes per minute. As intended, complex SQL queries executed very quickly on this database. The sensor tables were fully indexed, so the performance of queries would not suffer if the tables continued to grow. However, this

design appeared to push the limits of what was possible with the DBMS and certainly consumed available storage far too quickly.

4.3.3 Storing sensor blocks directly

An obvious means of optimizing for storage size in the TDAD database was to continue using the highly compressed sensor block format invented for the TMS. A single database table would store the sensor blocks; every 20 seconds, a new row would be added that contained the sensor block, the current time, and the name of the most recently received dictionary. In practice, overhead within the DBMS was found to be negligible for this operation, so that the storage consumption was essentially the same as the sum of the sizes of the sensor blocks, 66 kilobytes per minute.

Given the factor-of-ten decrease in storage requirements, there was really no choice but to prefer this design. Unfortunately, this mandated a less straightforward query process, as the important sensor values remained buried inside the sensor blocks, unavailable to SQL. A special-purpose program would have to be written to perform queries.

At 66 kilobytes per minute, this database would grow by 93 megabytes each day if it were operating round the clock. To store a year of data, TDAD would require around 34 gigabytes of storage capacity. This magnitude of storage could be easily provided by affordable hard disks. There was no need to immediately procure all of it; since the DBMS offered a facility to split up a single table across multiple files, additional disks could be added as necessary. Using additional DBMS administrative programs, a backup strategy was devised to periodically export newly added sensor blocks to a tape drive, along with the complete history of the dictionaries.

4.4 FILLING THE DATABASE

Once the DBMS had been installed, the TDAD tables created, and their storage requirements carefully configured, the flood of traffic data could be channeled to this reservoir. A program was necessary to monitor the data produced by the TMS and update the TDAD

tables with the information in incoming dictionaries and sensor blocks. This program was directly responsible for the quality of the database, because any time it stopped running or made a mistake, a gap in the archive's coverage would occur, reducing its value.

The program is named TDAD Receiver, and its reliable network connection to the TMS is courtesy of the ITS Information Backbone. The Self Describing Data protocol defines the format of the dictionary and the sensor blocks, as well as the order in which they arrive. Taking advantage of this infrastructure, the process of writing TDAD Receiver consisted simply of responding to the arrival of a dictionary and responding to the arrival of a sensor block.

4.4.1 Storing a new dictionary

The dictionary contains infrequently changing information describing the sensors, as well as the essential mapping of each sensor name to the location inside a sensor block containing its values. TDAD Receiver must append new information to five database tables upon the arrival of a new dictionary. To differentiate the information from successive dictionaries, the tables contain a "serial number" column, which has a unique value for each dictionary. Consistent with the rules of a Self Describing Data transfer, the current dictionary is sent to TDAD Receiver each time it starts running, so it always checks that the dictionary's serial number has not already been processed and placed in the database.

If any errors arise during the processing of a new dictionary, the entire operation is "rolled back;" that is, the database is put back into the state it was in before the new dictionary was received. If the operation failed because of a lost connection to the DBMS, TDAD Receiver attempts to process the dictionary again immediately after securing a new connection to the DBMS.

4.4.2 Storing a new sensor block

As previously discussed, each new sensor block is placed directly into a table, along with its time stamp and its dictionary's serial number. No attempt is made to recover from an error in this operation; if the database is unavailable for some reason, the sensor block is lost.

4.4.3 Quality assurance

Because TDAD Receiver must keep doing its job forever, no matter what happens to the TMS or the DBMS, it contains sophisticated code to handle errors returned by the DBMS, behaving correctly if the meaning of the error can be anticipated, and otherwise notifying the software developer immediately.

An analysis of the TDAD database reveals gaps of varying lengths; these total 10 days out of 133 total days of coverage to date. What happened to those missing sensor blocks? As with most software, a few problems with TDAD Receiver have been discovered and fixed since it was first put into use, causing some initial downtime. A power outage once caused the loss of a day's worth of data. The DBMS once was unavailable for a day because of configuration problems. The most vexing cause of gaps, because it is beyond TDAD's control, is that the TMS periodically goes offline, producing no new sensor blocks to store, although time stubbornly marches on.

4.5 QUERYING THE DATABASE

With the DBMS and TDAD Receiver in operation, attention turned to creating a streamlined process for retrieving specific information from the growing collection of sensor blocks. This process would take the form of a program called TDAD Expand. Its specifications were derived from the design of the database, from the nature of the TMS, and also from anticipation of the types of requests that might be made of the system.

4.5.1 Query input and output

Remember that a sensor block is a snapshot of the state of the TMS at an instant in time, and that the sensor block table in the DBMS is a collection of blocks along with the date and time each was received. Therefore, the primary task of TDAD Expand must be to select a set of sensor blocks from the table and use its knowledge of their format to transform them into useful data. For the sake of simplicity, it was decided that a single run of the program would operate on a set of sensor blocks defined by a start time and an end time. Any cyclical

temporal request, such as “every weekday at 5:00 p.m. for the past three months,” could be built up from multiple invocations of TDAD Expand. There is no direct support for queries in which the time itself is in question; however, it is hard to imagine a useful query of the form “At what time does X occur?” that is not bounded by a time span of interest.

Supporting geographical limits on the output of TDAD Expand was desirable for many anticipated uses of the system, such as graphing traffic flow versus time at a certain spot along a highway. From a practical standpoint, TDAD Expand needs precise geographical limits in order to suppress unneeded output. The TMS dictionary specifies the location of approximately 239 roadside locations, the “cabinets” in which the sensors are housed. Each sensor’s name begins with the name of its cabinet. TDAD Expand produces output for individually specified sensors or for the sensors in a specified list of cabinets.

In order to interface well with the widest variety of analysis programs, the output from TDAD Expand consists of tabular ASCII data. The first line of output names the columns, and each remaining line provides the values reported from a single sensor at a single time. The sensors are grouped into three types, each with a different set of reported values. The “loop” sensors measure vehicle occupancy in a particular lane. The “station” sensors provide an average occupancy of all the lanes going in a particular direction. The “speed trap” sensors estimate the current speed in each lane. Since the tabular output looks different for each type of sensor, TDAD Expand requires the user to choose one of the three types.

4.5.2 Changing dictionaries

One further wrinkle in the design of TDAD Expand is that, in order to decode each sensor block, the dictionary associated with that block must also be retrieved from the DBMS. The program has to handle the case that a dictionary change (resulting in a new serial number on subsequent sensor blocks) occurs within the time span it is examining. The following flow of control implements all of TDAD Expand’s interactions with the DBMS in a natural, efficient manner:

For all serial numbers SN in the time span
    For all sensors X with serial number SN
        If X is in the requested list of sensors, add its name and sensor block offset to the list SL
    For all sensor blocks SB in the time span with the serial number SN
        For all sensors and offsets (S,O) in the list SL
            Decode the 3- or 6-byte sensor report at offset O in SB and print it out

4.6 A USER INTERFACE FOR QUERIES

To make the TDAD system accessible to the general public, a better interface than TDAD Expand was necessary. The reason is that to run TDAD Expand, a file must first be prepared containing its input parameters, and to do this requires knowledge of the DBMS representation of timestamps, as well as access to the cryptic information within the TMS dictionary about the location and name of the sensors. Another reason not to distribute TDAD Expand is that it connects directly to the DBMS and therefore could be used maliciously.

To provide a better public interface to TDAD, a map was made of the local highways, and a marker for each cabinet location was overlaid on the map. The resulting image was turned into an HTML “image map” for use within a Web page. The result is that the user sees all the cabinets and can select one of them (see Figure 2). Individual sensors from within each cabinet can then be added to the list of sensors to query. Additional form elements on the Web page simplfy selection of the sensor type, the start time, and the end time. Once the user has finished building a query, the Web server receives the query and invokes TDAD Expand on behalf of the user. As TDAD Expand performs its task, the Web server periodically sends a message back to the user’s browser giving notice of the percentage of the query that has been completed (see Figure 3). When the query has been completed, the Web server provides the user with the opportunity to download a file containing the results (see Figure 4).

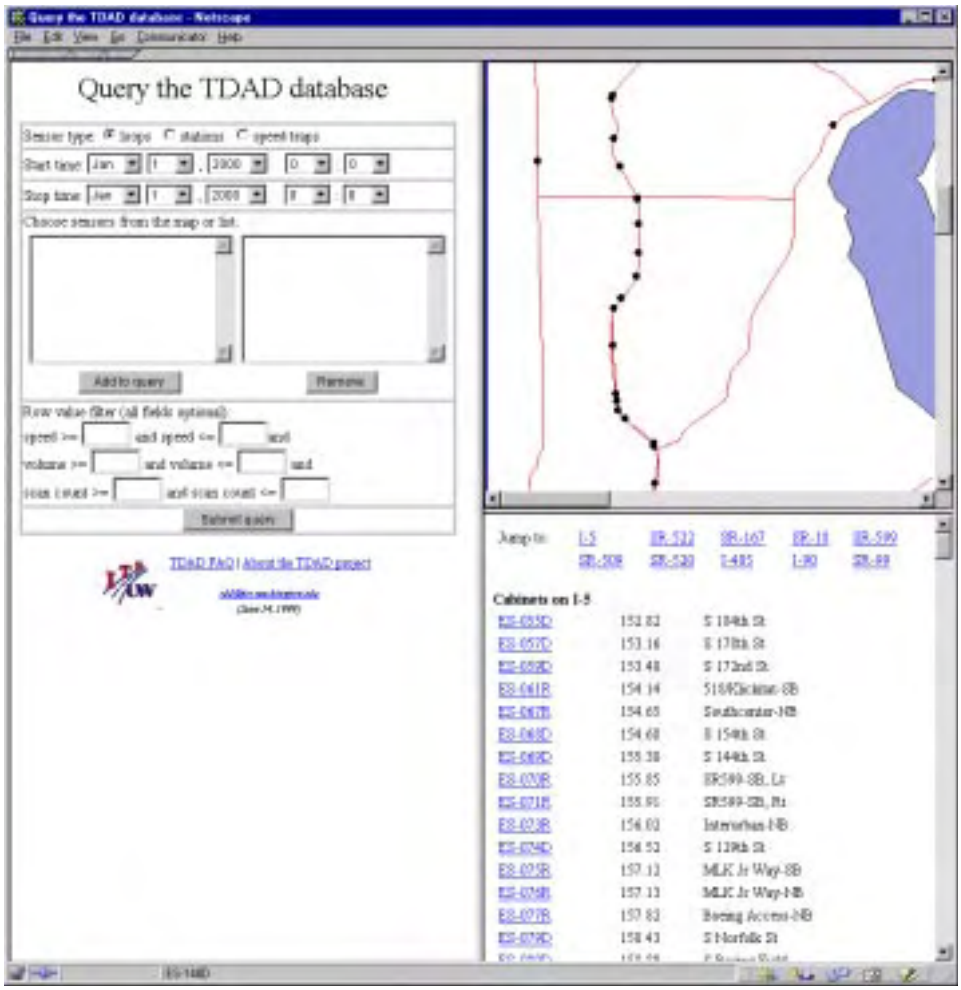


Figure 2: TDAD Query interface

Figure 3: Query results screen

949356346656.txt - Notepad

File Edit Search Help

```
SENSOR_ID,DATA_TIME,VOLUME,SCAN_COUNT,FLAG,LANE_COUNT,INCIDENT_DETECT
ES-055D:_MN___1,20000101000031000,0,0,0,1,0
ES-055D:_MN___2,20000101000031000,2,81,0,1,0
ES-055D:_MN___3,20000101000031000,1,14,0,1,0
ES-055D:_MN___4,20000101000031000,0,0,0,1,0
ES-055D:_MN___1,20000101000051000,1,11,0,1,0
ES-055D:_MN___2,20000101000051000,1,21,0,1,0
ES-055D:_MN___3,20000101000051000,0,0,0,1,0
ES-055D:_MN___4,20000101000051000,0,0,0,1,0
ES-055D:_MN___1,20000101000111000,0,0,0,1,0
ES-055D:_MN___2,20000101000111000,0,0,0,1,0
ES-055D:_MN___3,20000101000111000,0,0,0,1,0
ES-055D:_MN___4,20000101000111000,0,0,0,1,0
ES-055D:_MN___1,20000101000132000,1,12,0,1,0
ES-055D:_MN___2,20000101000132000,3,63,0,1,0
ES-055D:_MN___3,20000101000132000,0,0,0,1,0
ES-055D:_MN___4,20000101000132000,1,17,0,1,0
ES-055D:_MN___1,20000101000151000,1,12,0,1,0
ES-055D:_MN___2,20000101000151000,1,19,0,1,0
ES-055D:_MN___3,20000101000151000,1,13,0,1,0
ES-055D:_MN___4,20000101000151000,0,0,0,1,0
ES-055D:_MN___1,20000101000211000,0,0,0,1,0
ES-055D:_MN___2,20000101000211000,1,19,0,1,0
ES-055D:_MN___3,20000101000211000,1,13,0,1,0
ES-055D:_MN___4,20000101000211000,1,13,0,1,0
ES-055D:_MN___1,20000101000231000,0,4,0,1,0
ES-055D:_MN___2,20000101000231000,3,46,0,1,0
ES-055D:_MN___3,20000101000231000,1,13,0,1,0
ES-055D:_MN___4,20000101000231000,1,13,0,1,0
ES-055D:_MN___1,20000101000251000,3,32,0,1,0
ES-055D:_MN___2,20000101000251000,4,69,0,1,0
ES-055D:_MN___3,20000101000251000,2,19,0,1,0
ES-055D:_MN___4,20000101000251000,1,12,0,1,0
ES-055D:_MN___1,20000101000311000,0,0,0,1,0
ES-055D:_MN___2,20000101000311000,1,64,0,1,0
ES-055D:_MN___3,20000101000311000,1,13,0,1,0
ES-055D:_MN___4,20000101000311000,1,12,0,1,0
ES-055D:_MN___1,20000101000331000,2,60,0,1,0
ES-055D:_MN___2,20000101000331000,2,53,0,1,0
ES-055D:_MN___3,20000101000331000,1,19,0,1,0
ES-055D:_MN___4,20000101000331000,0,0,0,1,0
ES-055D:_MN___1,20000101000351000,0,0,0,1,0
ES-055D:_MN___2,20000101000351000,0,0,0,1,0
```

Figure 4: Data resulting from TDAD Query

CHAPTER 5. CONCLUSIONS

This report both reviews the historical usage of traffic data and describes the creation of a data mine for the roadway data from WSDOT Northwest Region.

The TDAD data mine has been used for a variety of purposes over its 18-month operational period. The records for TDAD use are based on the domain name of the user who has made a request and, subsequently, actually downloaded a file of loop data. The domains of the users include the following:

accessone.com
andrew.eng.uci.edu
atl1.da.uu.net
benchmark.com
ce.washington.edu
consultec-llc.com
coventry.ac.uk
cpsrta.org
DaimlerChrysler.COM
deainc.com
ee.washington.edu
fhwa.dot.gov
forthnet.gr
frisco.ch2m.com
fsmodem.washington.edu
gen.cadvision.com
gtei.net
korea.ac.kr
mcis.washington.edu
metapath.com
microsoft.com
mitretek.org
netcom.com
northgrum.com
ntu.edu.sg
odetics.com
olympia.wa.da.uu.net
ornl.gov rsandh.com
oscsystems.com
rsandh.com
saturn.bbn.com
sclfw.guidant.com
sea.lightrealm.net
sfba.home.com
trac.washington.edu
trapsoft.com
umd.edu
unimelb.edu.au
uoregon.edu
Virginia.EDU
wsdot.wa.gov
wolfenet.com

Many other domains, as well as Internet addresses, do not resolve to domain names. The list above is a sampling of the users and is not inclusive but rather demonstartive of the broad range of agencies, universities, and companies that have found this data mine with little or no publicity.

A diverse mix of users including agencies, local university researchers, and national/ international research institutions have found this type of data mine useful. It is, however,

noteworthy that these users had to seek out this infomation as there is no good, existing clearinghouse for such data. Therefore, we conclude that this kind of data mine is indeed useful but would be far more useful to more agency personnel if it was more widely advertized, perhaps on agency pages.

REFERENCES

1) Ashbrook IV, T.V. "Freeway surveillance data base design and development," Master's thesis, University of Washington, 1983.

2) Casey, R.F., L.N. Labell, and J.C. Schwenk. "Evaluation plan for AVL implementations," *Proceeding of the IVHS AMERICA* , 1993.

3) Dillinburg, E. "http://www.ai.eecs.uic.edu/GCM/GCMDescription.html"

4) Farber, E. and M. Paley. "Using freeway traffic data to estimate the effectiveness of rear-end collision countermeasures," *Proceeding of the IVHS AMERICA* , 1993.

5) Ito, K., K. Fujita, and E. Keitoku. "Advanced traffic information service," *Proceedings of the Second World Congress on Intelligent Transport Systems*, Vol. II, 1995.

6) Kobayashi, T., T. Watanabe, and T. Iwata. "Development and management of interurban expressway vehicle information & communication system (VICS)," *Proceedings of the Second World Congress on Intelligent Transport Systems*, Vol. II, 1995.

7) Sobolewski, M. and J.L. Wright. "Rural applications of IVHS in Minnesota," *Proceeding of the IVHS AMERICA* , 1993.

8) Turnbull, K.F. "The use of information generated from transit AVL systems," *Proceeding of the IVHS AMERICA* , 1993.

9) Wallace, C.E. and A.K. Kilpatrick. "IVHS applications for rural highways and small towns," *Proceeding of the IVHS AMERICA* , 1993.

10) Wallace, C.E. and A.K. Kilpatrick. "IVHS implications for transportation demand management," *Proceeding of the IVHS AMERICA* , 1993.

11) Wright, J.L. "TRAVLINK- can information influence choice," *Proceeding of the IVHS AMERICA* , 1993.

APPENDIX A. PARTICIPANTS IN INTERVIEWS

ENERGY OFFICE:
Brian Lagerberg - Program Evaluation Transportation Specialist

METRO - MUNICIPALITY OF METROPOLITAN SEATTLE:
Jon Flug - Transit Information Planner
Tom Friedman - Senior Planner

PUGET SOUND REGIONAL COUNCIL:
Larry Blain - Senior Planner, Travel Demand Modeling
Brad Brooks - Network Administration
Jay Clark - Geographic Information System Analyst
Judy Leslie - Senior Planner

TRANSPORTATION NORTHWEST:
Nancy Nihan - Director and Professor in Civil Engineering

UNIVERSITY OF WASHINGTON, CIVIL ENGINEERING DEPARTMENT:
Fred Mannering - Professor

UNIVERSITY OF WASHINGTON, ELECTRICAL ENGINEERING DEPARTMENT:
Dierdre Meldrum - Assistant Professor
Cyndi Taylor - Research Assistant

WSDOT ADVANCED TECHNOLOGY BRANCH
Morgan Balogh - Urban Traffic Engineer

WSDOT NORTHWEST REGION PLANNING OFFICE:
Pani Saleh - Traffic Planning Supervisor

WSDOT PLANNING OFFICE:
Charlie Howard - Manager of Transportation Planning
Toby Rickman - Manager of Regional Planning

WSDOT RESEARCH OFFICE:
Art Lemke - Assistant Research Director

WSDOT TRANSPORTATION DATA OFFICE:
Debbie Daugherty - Assistant Travel Data Supervisor
Irene Hertwig - Travel Data Supervisor

WSDOT OFFICE OF URBAN MOBILITY:

Miguel Gavino - Planning/Program Coordinator and Supervisor
Thaier Hassan - Transportation System Planner

WASHINGTON STATE TRANSPORTATION CENTER:

Mark Hallenbeck - Director

APPENDIX B. INTERVIEW NOTES

Puget Sound Regional Council
July 28, 1994
Participants:

Judy Leslie - Senior Planner
Jay Clark - Geographic Information System Analyst
Brad Brooks - Network Administration
Larry Blain - Senior Planner, Travel Demand Modeling

How do you use traffic data?

PSRC uses traffic data to validate the outcome of its Travel Demand Model. This model forecasts changes in population and employment and resultant trip time and travel demand for the Puget Sound Region over the next decade. A full run of this model is done every two to three years, and it takes six to nine months with three to four people working on it to complete the whole forecast. Figure 5 is a diagram of PSRC's modeling process.

The model starts with a zone by zone trip generation. From household travel surveys and the Puget Sound Transportation Panel, PSRC planners determine how many trips each zone produces and attracts and classifies these trips according to their purposes (for instance, Home Based Work). The number of productions and attractions is determined by the number of businesses, number of residents, etc. The production of a trip is similar to origin, and the attraction of a trip is similar to destination. However, there are some differences. For instance, the production of a Home Based Work (HBW) trip is at the home regardless of whether the trip starts at the home (to go to work) or starts at work (to go to the home).

The trip generation tables are then used to make trip distribution matrixes of production and attraction. In this step, the production trips are linked with attractions, and the attractions are linked with production trips. To determine the distribution of productions and attractions across the zones, PSRC uses a "gravity" principle. The law of gravity states that the force of attraction between two objects is proportional to the mass of the two objects divided by the square of the distance between the objects. Similarly, if a zone has a lot of attractions, it can

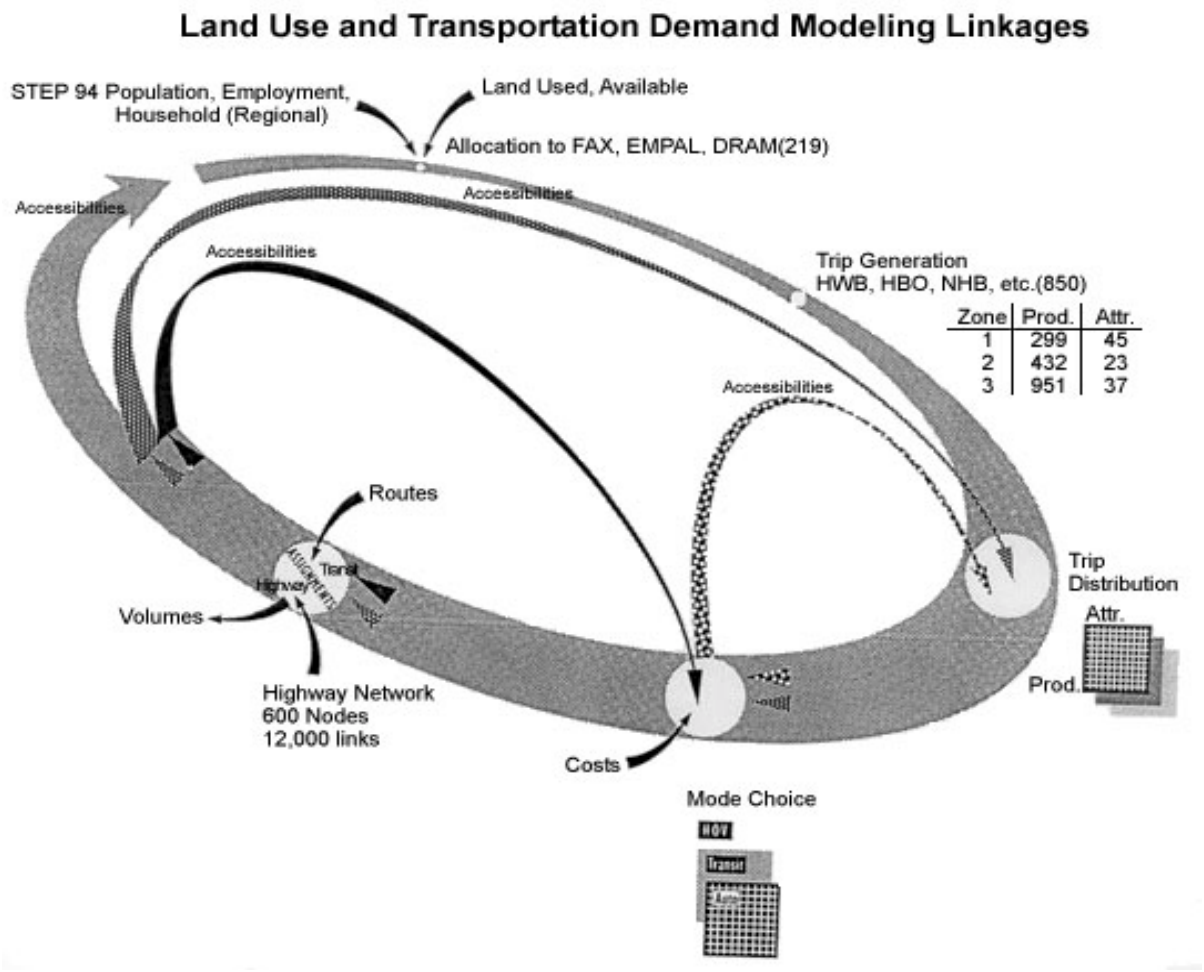


Figure 5: PSRC's modeling process

be thought of as having a large mass. Therefore, this zone will have a greater "force" in attracting trips. Yet, the amount of attraction to this zone drops off as the distance and travel time to the zone increases. Because of the impedance that traffic congestion creates for a trip, travel time is a better indicator of attraction to a zone than distance. With this principle, PSRC's computer links attraction and production trips by assigning a distribution of trip types and trip lengths throughout the zones.

After determining the number of trips and the production and attraction of the trips, the next step in the model is the determination of the mode of transportation chosen for each trip. PSRC calculates a matrix for each type of mode (HOV, transit, auto, etc.). Mode information

is obtained from census and transit surveys, which indicate the percentages of people who choose each type of mode.

The fourth step in the travel demand model is the assignment of traffic volumes that flow across the region's screenlines; PSRC defines a screenline as a slice across a corridor. It has provided a map for this report, Figure 6, of the Puget Sound region with the screenlines drawn in. Assignments for the network are only given for arterials, highways, and freeways. In addition, interzonal travel times are calculated at this step. The volumes and travel times are fed back into other steps of the model, and the feedback helps the computer model to adjust and fine tune trip distributions and mode choice.

The traffic volume assignments generated by the model are compared with actual traffic counts. If the computer model has forecast volumes that are within 4 to 5 percent of the actual traffic counts, PSRC considers the model to be producing an accurate forecast. If the actual counts do not validate the model's results, PSRC's planners will hand adjust some of the model's assignments. Since the assignments are fed back to the other steps in the model, the adjustments will change the forecasts to make them more accurate.

PSRC uses its Travel Demand Model to forecast for the base years of the decades (1990, 2000, 2010, etc.) and to develop a long-range plan. It develops or modifies its long-range plan every two to three years. To determine the forecasts for years other than the base years, PSRC planners may interpolate trip table information to get an approximate forecast.

Some of PSRC's other responsibilities include allocating federal funds for transportation projects. Since PSRC has been given the metropolitan planning organization (MPO) and regional transportation planning organization (RTPO) status for the region, including King, Kitsap, Pierce, and Snohomish counties, any project that affects the operation of traffic must first gain its approval. Each project must pass the following three levels of approval: 1) be included in the long-range plan, Vision 2020, 2) be included in the Trip Improvement Program, and 3) have an Environmental Impact Statement (EIS) for the immediate vicinity. Of

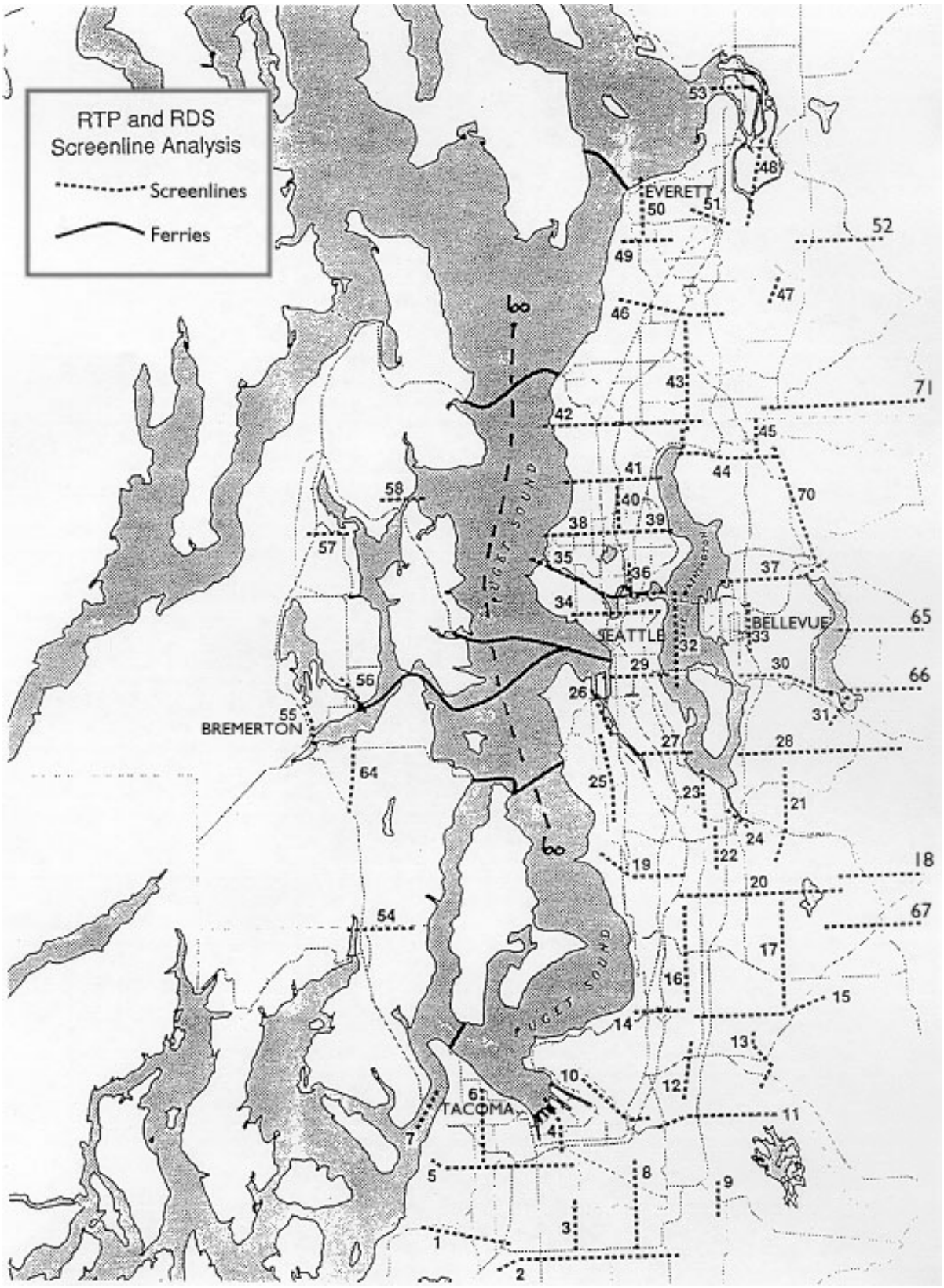


Figure 6: Map of Puget Sound Region with screenlines

course, to be part of the long-range plans, each project must meet the requirements of the Clean Air Act Amendments.

PSRC is starting a new project that will also use traffic data. Its new task is performance monitoring. For this job, it must track real annual data to determine the outcome that projects have on traffic in the region. The performance monitoring team must be able to answer global questions such as, “How has average vehicle occupancy been affected?” and “Is the long-range plan helping the region conform to the Commute Trip Reduction Act?”.

This organization also performs a 3-hour peak analysis. This is an analysis of the worst three hours of traffic, regardless of what time they occur.

What traffic data do you currently receive?

PSRC currently receives hard copy maps from the cities and counties in the region that perform volume counts beside the roads. It also receives the WSDOT Annual Traffic Report and Northwest Region’s Ramp and Roadway Report. Specifically, PSRC uses the average daily traffic, average weekday traffic, average weekend traffic, and truck percentage figures in these reports.

The Regional Council receives GIS maps from a variety of sources. The maps do not always match each other very well, and consistency becomes a prime concern.

What traffic data do you ideally want?

PSRC wants counts of traffic as it flows across corridors of travel. From this, the data need to be manipulated into average daily traffic, average weekday traffic, and average weekend traffic counts. (average weekday traffic is measured only on Tuesdays, Wednesdays, and Thursdays.) PSRC also needs speed and travel time data. It still uses the Bureau of Public Roads formula to determine speed and travel time from volume and lane occupancy. However, it does not feel that this algorithm is very accurate. The senior planner would really like measured speed instead of calculated speed. Specifically, the senior planner needs the hourly average speed and the distribution of speeds within that hour. PSRC also requires more vehicle occupancy data.

PSRC examined only vehicle trips in the past, but it is expanding into non-motorized trips and movement of goods. For the latter, PSRC needs to know the volume and classification of trucks traveling within the region.

One of the most important aspects of the data is that they must be consistent. Data gathered across a screenline on a weekday one year should be accessible at the same time and place the next year and the year after that.

The performance monitoring group within PSRC needs some additional information. People on this project would like access to real-time bus information.

Judy Leslie received data similar to what the PSRC planners require when she worked at Metro. Tom Friedman and Jon Flug in the Research and Market Group at Metro have been collecting consistent traffic data for ten years. PSRC does not currently get any data from this source. However, the format and content of data have met Judy Leslie's needs in the past and are therefore good examples to study.

In what electronic format would you like to see data?

The network administrator said an SQL database would best meet PSRC's needs. It wants to have control over how it uses the traffic data by forming its own inquiries to the database. However, it would like the data manipulation to be done at the database instead of bringing the large volume of raw data to a PSRC computer site.

PSRC is not currently on the Internet. However, work is being done right now to hook it up. The Council should be on the Internet by next summer.

WSDOT Office of Urban Mobility
July 29, 1994
Participants:

Miguel Gavino - Office of Urban Mobility
Thaier Hassan - Office of Urban Mobility
Morgan Balogh - WSDOT Advanced Technology Branch

How do you use traffic data?

The Office of Urban Mobility uses traffic data for corridor studies of new traffic projects. This office examines the plans on a more specific level than the Puget Sound Regional Council does. PSRC examines projects in the framework of how they affect the environment and how they fit with the regional plans. Urban Mobility examines projects at the intersection and interchange level to determine if a project will meet the specific goals for the city or county.

Similar to PSRC, the Office of Urban Mobility is becoming involved in performance monitoring.

What traffic data do you currently receive?

Urban Mobility does not work with raw traffic data. It receives processed data from the WSDOT Data Office and PSRC. From WSDOT, it receives the Annual Traffic Report and hard copy maps with traffic volumes written beside the roads. From PSRC, Urban Mobility receives current and projected volume data. It then determines what data are not provided that are needed for the particular project. To supplement what it has received, Urban Mobility hires consultants to manually collect the necessary data. Some examples of these extra data are travel times, car counts for particular areas, occupancy, measure of transit use, speed, commercial vehicle counts, pedestrian counts, and bicycle counts. Occasionally, it needs to have incident information. It feels this particular information is readily available through the Department of Transportation.

Urban Mobility also gets electronic area maps from WSDOT and PSRC. There are a few problems with matching the different maps.

Participants mentioned that they receive turning movement counts from the Traffic Studies Section in WSDOT Northwest Region. The people to contact at Northwest Region are Pani Saleh and Cary Berger.

What traffic data do you ideally want?

The Office of Urban Mobility wants information similar to what PSRC has asked for. Examples of this are average weekday and average weekend counts. It would also like travel time, speed, average vehicle occupancy, ferry information, maps, commercial vehicle counts, and turning movements.

This office wants accurate data. Currently, some of the volume counts do not balance. The data are high in some places and low in others.

In what electronic format would you like to see data?

Urban Mobility is operating on a PC platform and is getting access to the Internet. It will e-mail us to let us know when it is connected. However, it is concerned that its ethernet connection is already being overloaded.

Project Management

The Office of Urban Mobility is more than just a potential user of the TDAD database. It proposed TDAD to the FHWA and is understandably concerned with the management of the project. Miguel Gavino expressed concern that the project deliver what was promised. The promise was an ITS service that accesses operations data, archives them, manipulates them, and gives them to the appropriate planners and researchers. He would like to see the summary report of the interviews at the end of August. He would also like to have another meeting in a month or two in order to detail what realistically is going to be included on the database.

Washington State Department of Transportation, Olympia
August 8, 1994
Participants:

Debbie Daugherty - Data Office
Irene Hertwig - Data Office
Charlie Howard - Planning Office Manager
Brian Lagerberg - Department of Energy
Art Lemke - Research Office
Toby Rickman - Planning Office

TRANSPORTATION DATA OFFICE

How do you use traffic data?

The WSDOT Data Office collects traffic data and puts them on its Transportation Information and Planning Support (TRIPS) system. It collects traffic data from its own collection sites as well as other external sources. It reduces these data into meaningful numbers, such as annual average daily traffic (AADT) volumes, average weekday traffic (AWT) volumes, average vehicle miles, and truck percentages of volume. These figures become the source of data for agencies such as the Federal Highway Administration and the Puget Sound Regional Council. From traffic data on TRIPS, the Data Office compiles the Annual Traffic Report, which includes the AADT and truck percentages. The purpose of collecting and reporting on traffic data is to maintain a balanced traffic file to aid in "the planning, design, construction, and maintenance of the approximately 7,000 miles of roads in Washington State" (1992 Annual Traffic Report, Introduction Page).

However, the Data Office is quick to point out that any design project should obtain specific traffic data for the areas that would be affected. The Data Office participates in some of these special project counts.

All traffic data that are input into the TRIPS system meet the Highway Performance Monitoring System (HPMS), AASHTO, Traffic Monitoring Guide (TMG), and Traffic Monitoring Systems for Highways (TMS/H) guidelines.

What traffic data do you currently receive?

The Data Office has 150 permanent data collection sites. Sensors at these sites count the number of vehicles passing them. The office also has 30 classification sites that record vehicle classification information. It closely follows the federal Traffic Monitoring Guide in the implementation of all data collection.

The Data Office also receives data information from a variety of other sources. It gets a hard copy of freeway loop data from WSDOT Traffic Systems Management Center (TSMC).

The Data Office feels that more traffic data are always better, as long as it is able to manage all of them and the data have been collected in a manner consistent with the federal guidelines.

What traffic data do you ideally want?

The WSDOT Data Office would like hourly and 15-minute volume data. It would like to have speed monitoring information and vehicle classification data. Average vehicle occupancy and traffic counts in single HOV lanes are also desired. The Data Office would like to receive enough traffic counts from other sources to reduce its own number of data collection sites. This would help it cut down on maintenance costs and reduce its budget.

Ideally, the Data Office would like all incoming data to be extremely accurate. Currently, it is concerned with the consistency of the loop data. Representatives feel that they have to do a lot of work to weed out the bad data. In addition, it is important that data from a collection site or loop be collected for a whole year. However, they haven't seen any loop data after the most recent corrections to the system, and some of their concerns may have already been addressed.

In what electronic format would you like to see data?

The whole Washington State Department of Transportation will be connected to the Internet within the next two years. The Data Office uses the mainframe to hold its main database. The information it feeds into the database is data that have already been checked and validated. Its software is the TRIP System. Jan Myhr is the MIS TRIPS coordinator and database expert.

PLANNING OFFICE

How do you use traffic data?

The Planning Office uses traffic data to determine the efficiency of the freeways and roads in Washington State. Efficiency is measured as volume over capacity. The Planning

Office is not concerned with extremely accurate data since it is tracking the change in the traffic over time, often a 20-year span. Congestion Monitoring Systems (CMS) are also part of the Planning Office's responsibilities. The metropolitan planning organization (MPO) in each of the three urban areas (Puget Sound region, Spokane, and Vancouver) is responsible for the CMS in its city. The Planning Office is in charge of the CMS for the rest of the state.

What traffic data do you currently receive?

The Planning Office receives traffic volumes from the WSDOT Data Office. It receives efficiency data from the Puget Sound Regional Council's planning model. If the TDAD database meets the requirements for PSRC, the personnel in the Planning Office feel that it will therefore meet all their needs.

What traffic data do you ideally want?

The Planning Office needs hourly traffic volumes, peak hourly traffic volumes and average vehicle occupancy. It would like to have vehicle classification data as well. The latter data have always been a missing item.

Most importantly, it really wants travel times on both freeways and arterials in urban areas. Currently, freeway loops and an AVL system for Metro buses can provide some travel times. However, these data are only found on the freeways. The Planning Office believes that a fleet of probe vehicles with AVI tags is the only way to obtain travel-time data on both freeways and arterial systems.

In what electronic format would you like to see data?

The Planning Office, along with the rest of the Washington State Department of Transportation, will be connected to the Internet within two years.

DEPARTMENT OF ENERGY

The Department of Energy uses traffic data to aid in evaluations of programs relevant to the Department. One such program under evaluation is the implementation of the mandates

in the Commute Trip Reduction Act (CTR). The Department of Energy has to examine how this act has changed traffic patterns, but it does not need to examine traffic data at very specific levels. For example, it does not need to know the number of cars that are no longer making commute trips.

The Department of Energy receives traffic data from Mark Morris at WSDOT Northwest Region. Traffic data are downloaded from Mark's computer onto a tape. The Department of Energy also uses Traffic Reporter as an on-line way to establish occupancy. It is currently connected to the Internet.

Traffic Planning
WSDOT Northwest Region
August 15, 1994
Participant

Pani Saleh - Traffic Planning Supervisor

The Traffic Planning Group in Northwest Region was established two and a half years ago for the purpose of centralizing data requests. This group collects counts of vehicle classification, vehicle occupancy, average daily traffic (ADT), turning movements, and speed. It collects these data primarily on state routes and arterial streets. Highways and ramps fall under the jurisdiction of the Traffic Systems Management Center (TSMC) for safety reasons. Counts are done by request from any of several agencies. When requests come in for specific data, it is important for the Planning Group to know why these data are requested. It can determine whether there is a need for more depth and authorize a full scope project.

To collect either ADT counts or speed studies, personnel lay tubes across the roads in the desired locations. These tubes usually remain in a particular spot Monday through Friday for a weekday count and Monday through Monday for a count including weekends. Data collected by these tubes can be processed to give either the ADT or speed.

Traffic Planning receives a quarterly signal priority list from Traffic Operations. Traffic Planning needs to then manually collect turning movement data at these sites during the a.m., p.m., and mid-day peaks. It uses these data to determine whether the site meets warrants 1, 2, and 6. These warrants are the District's standards for installing signals.

Traffic Planning is trying to be more conscientious about collecting classification data. It manually collects these data and groups vehicles into heavy, medium, and light categories. Other agencies that collect classification data use tube counts. This method determines the number of axles on a vehicle and groups it into one of 11 categories.

The last type of data collection that Traffic Planning does is vehicle occupancy. This is done manually and not very often.

All of the data that this group collects are organized in spreadsheet format. "NWCOUNTS" is an automated traffic counts database available on the WSDOT Northwest Region Novell Server that provides updated traffic count information for all state owned facilities within the region. This database contains both diagram files and counts files, which are sorted by state route (SR), milepost (MP), and count date (1991 to present) and include mechanical counts (ADT), turning movement counts (TM), occupancy counts (OC), truck classification counts, and speed information.

This system will enhance the efficiency of all future traffic count requests for design and analysis purposes. It will eliminate unnecessary delays in receiving existing counts data and present the system user with an on-line request form should the data in question not be available in the existing database. Planning hopes to expand this database in the future to include level of service (LOS) information at all intersection locations.

The Traffic Planning Division is connected to the Internet. Pani Saleh would be very interested in having the TSMC loop data available on a database over the Internet. She receives many requests for data that actually must come from TSMC. Often she calls up TSMC and gets the data for the requester herself.

Metro
September 7, 1994
Participants:

Tom Friedman - Senior Planner
Jon Flug - Transit Information Planner

Metro collects data on the number of passengers riding the bus by manual observations or with automatic passenger counters. A technology known as “fare boxes” will also help collect data, but it is not yet perfected. In addition, Metro has a radio-based automatic vehicle location (AVL) system that helps it determine schedule adherence for all of its buses.

Automatic passenger counters (APC) collect the majority of reliable bus data for Metro. Twelve percent of the bus fleet has an APC at any given time. The Planning Department suggests which bus to put an APC on. Of course, the final control of the placement of the APC is in the dispatchers’ hands. Still, the end result is that 95 percent of the weekday trips have been counted at least five times.

At each stop, the APC counts how many passengers get on and off the bus, how many people stayed on the bus, and the time that the bus arrived at the stop. These very detailed data are stored in a bus stop file. APCs also have a time point interval (TPI) file that is a summary of the data in the bus stop file. The TPI file has the aggregated data between two time points, including maximum load during the time period and schedule deviation. At the end of the week, the data are manually downloaded from the APCs at Metro headquarters.

The APC system has been in effect since 1980. However, it took a while for the data to stabilize, and reliable data were not consistently available until around 1987. The data are processed and stored on an IBM mainframe in King County. Currently, records are stored on the mainframe from as far back as 1990. SQL inquiries to the mainframe are done in RAMOS. However, only a few people in the Metro Planning Department know how to access the information in this manner. There is a FOCUS system on Metro’s network that enables the majority of users to access common summarizations of the data. Any other data needs are directed to either Tom Friedman or Jon Flug. They can both access the mainframe and accommodate the special requests.

The AVL system is a relatively new system for Metro. Metro first implemented it in the spring of 1992 and is still working on minimizing errors. One cause of error is incorrect odometer calibration. This is scheduled to be fixed, but the maintenance department has many more pressing duties to complete first. There are 80,000 time-point events each day, and Metro is only picking up AVL data for about 40,000 events. For AVL (and APC) to work, three things must happen: 1) The map of the bus route must be correct, 2) the odometer must be correctly calibrated, and 3) the bus drivers must do what is expected of them. Personnel working with the AVL system are still attempting to make all the components work.

The other side of gathering data is answering the requests for the data. Requests in the past have been anything from broad aggregates of the data to extremely fine details. The questions are usually one of two types: (1) What is going on? and (2) How does what is going on compare to what used to go on?

Data have been used only for purposes internal to Metro. In fact, 90 percent of the data users are in the Planning Division. Outside agencies may request bus data, but they are only retrieved by someone in the system. Metro is afraid that a person uninitiated to the Metro system will be able to retrieve data but will actually obtain only meaningless numbers.

Metro's data are applicable for performance monitoring, project planning, long-term planning, and reporting data. The time point interval files from the APCs are given to Section 15 for federal reporting. The Puget Sound Regional Council uses data from Metro's origin/destination survey to calibrate the decade model. Projects use the Metro data to determine how many people riding Metro travel over the section of road that will be affected by the project. The levels of bus ridership are useful for performance monitoring.

RESEARCHERS

Fred Mannering - Professor,
University of Washington, Department of Civil Engineering
January 19, 1995

How do you use traffic data?

Fred Mannering is a professor in the Civil Engineering Department at the University of Washington, and he uses traffic data in some of his research projects. Mannering correlates traffic data, such as volume and speed, with accident rates. He also uses traffic data to check or validate the forecasted values that he obtains from his computer-simulated model of the traffic network.

What traffic data do you currently receive?

He receives peak hourly volumes from the Ramp and Roadway report that is published by the Traffic Systems Management Center (TSMC). In addition, he receives hourly volumes, speed, and vehicle classification data from three different data collectors on rural I-90. He also obtains traffic data from an autoscope. An autoscope videotapes traffic as it flows by the camera location and automatically calculates the volume counts by lane and speed of the vehicles.

What traffic data do you ideally want?

Fred Mannering would like to have 15-minute and hourly volume counts for full 24-hour time periods. He would also like to have 1-minute volume counts during the peak traffic volumes. He feels that it is best to receive the smallest increment of data available because the data can always be aggregated, but once aggregated, the data cannot be decomposed into constituent time periods. Mannering would also like speed and vehicle classification data from more sites around the Puget Sound region. He would like to have access to vehicle occupancy data throughout the region. Lastly, he would like to have all these different types of traffic data available for arterial streets.

Mark Hallenbeck - Director
Washington State Transportation Center (TRAC)
January 20, 1995

How do you use traffic data?

One of TRAC's goals is to meet the traffic needs of the Washington State Department of Transportation (WSDOT). To this end, Mark Hallenbeck uses traffic data to validate the outcomes of computer models of the traffic network. He also uses traffic data to conduct research analyses and to aid in signal design. With relevant traffic data, he can model how well a facility performs before and after modifications.

What traffic data do you currently receive?

For his research, Mark Hallenbeck obtains 15-minute volume data and peak period volume data on the freeways from the Traffic Systems Management Center (TSMC). With the volume and lane occupancy data he receives from TSMC, he can calculate speed. Hallenbeck also uses HOV vehicle occupancy data that are collected by TRAC. From the WSDOT Transportation Data Office's Annual Report, he receives the annual average daily traffic volumes (AADT). He also uses average weekday traffic volumes (AWDT) and vehicle classification data that are available from the Data Office.

The vehicle classification data from the WSDOT Transportation Data Office are collected in two different ways. One classification scheme divides vehicles into 4 categories - light, medium, etc. This method of classification relies on the total vehicle length. A second classification scheme places vehicles in one of 13 categories. These categories have been designated by the Federal Highway Administration (FHWA), and the correct classification is determined by the spacing of the vehicle's axles. In most locations, the WSDOT Transportation Data Office classifies vehicles into the FHWA 13 different categories. However, in urban areas, axle spacing is difficult to determine because of the typical congestion and volume of vehicles. As a result, vehicle classification data in cities is mostly of the 4-category classification type. When Hallenbeck has both types of vehicle classification data, he usually collapses the 13 categories down to the four categories by following a standard procedure approved by Washington State for this type of data compression.

The amount of traffic data that can be obtained from these sources is not always sufficient for Hallenbeck's purposes. Sometimes he has to have 15-minute volume and vehicle classification data collected from additional sites. He must also manually gather travel-time data on an as-needed basis.

What traffic data do you ideally want?

Mark Hallenbeck would like the traffic data that he uses - volume counts, travel time, vehicle classification data, speed, HOV vehicle occupancy - to be readily available from a large number of sites across the Puget Sound region and throughout the state. He would like to have access to traffic data before they have been aggregated. With raw data, he can extract the data that will fit his particular information needs, even as those needs change from one analysis to the next. Specifically, he would like to have access to peak-hour volume counts in 5-minute increments or smaller and daily volume counts in 5- to 15-minute increments. He would like for traffic data to be available in an electronic format, preferably a flat ASCII file.

Dierdre Meldrum - Assistant Professor
Cyndi Taylor - Research Assistant
University of Washington, Department of Electrical Engineering
January 23, 1995

How do you use traffic data?

In conducting their research projects, Dierdre Meldrum and her research assistant, Cyndi Taylor, require current traffic data. Their research includes training neural nets to predict traffic data and determining the effects of ramp metering on traffic flow. They use traffic data to calibrate their simulation model of the traffic network.

What traffic data do you currently receive?

Meldrum and Taylor currently use volume and lane occupancy data from the Traffic Systems Management Center (TSMC). They use these data in 20-second, 1-minute, and 5-minute increments. With these data, they calculate speed. They get the data from TSMC in a

text file that they obtain either by file transfer or on disk. Then they must check the validity of the data and sort out data from inoperable or inaccurate loop detectors to leave only valid data.

What traffic data do you ideally want?

Meldrum and Taylor would like to have performance criteria, such as measured speed and travel time. They would also like to obtain vehicle miles traveled data. Even though they are mainly interested in freeway data, they would like to have access to volume and occupancy data for time increments of 20 seconds or a minute on the arterial streets leading to a freeway on-ramp.

Nancy Nihan - Director and Professor
Transportation Northwest (TRANSNOW)
University of Washington, Department of Civil Engineering
February 9, 1995

How do you use traffic data?

Nancy Nihan uses traffic data to analyze the level of service on the freeways. She also inserts traffic data into a statistical time series model to predict future travel trends. In addition, she uses freeway traffic data to aid in research on ramp metering methods, and she uses traffic data from arterial roads to aid in research on signal timing.

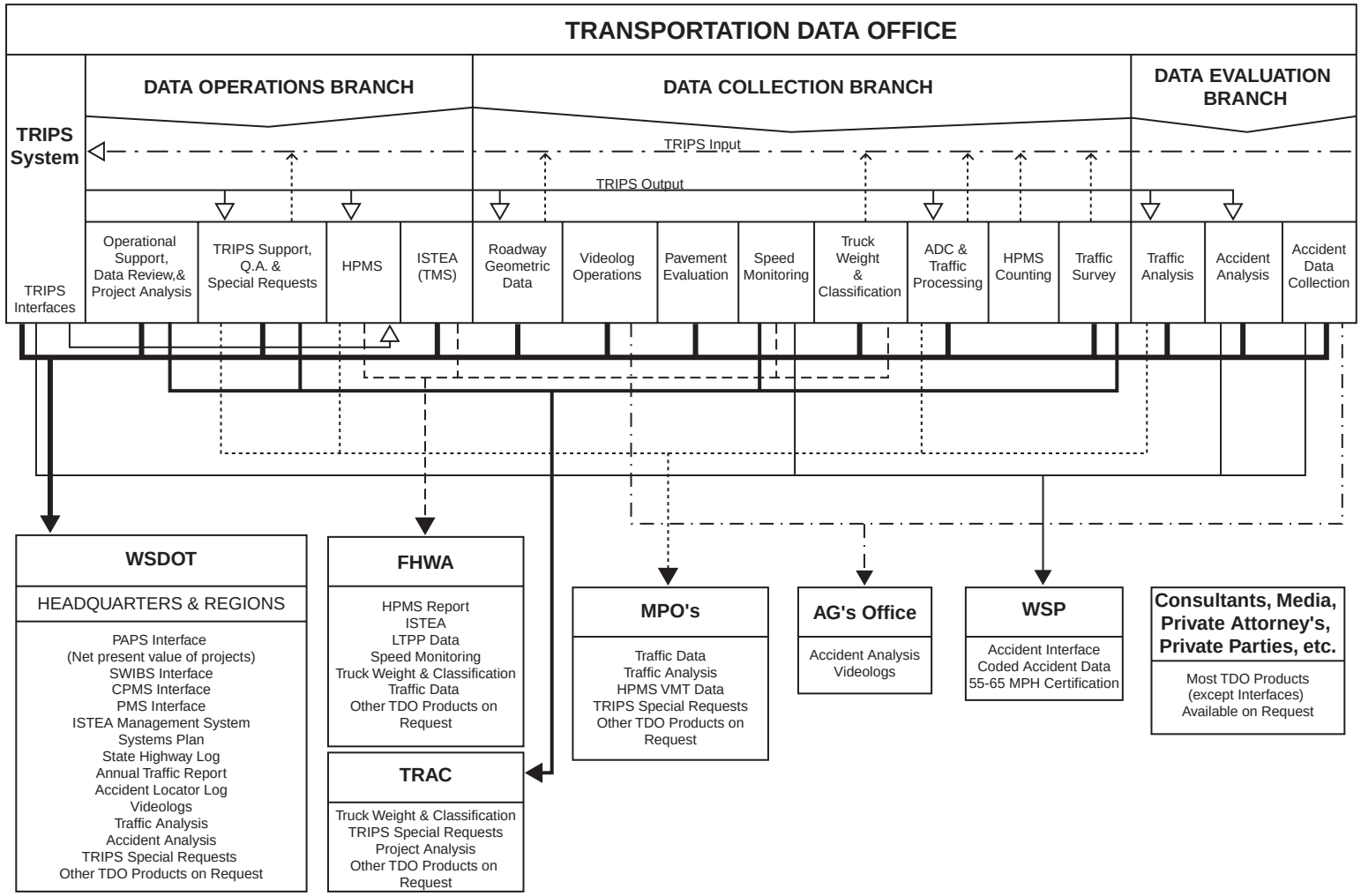
What traffic data do you currently receive?

Nihan uses volume and lane occupancy data from freeways in 20-second and 1-minute increments. She obtains these data directly from the WSDOT Traffic Systems Management Center (TSMC). She also gets traffic data from video cameras. The autoscope videotapes traffic and automatically records volume, lane occupancy, vehicle length, and speed. In addition to an autoscope, she is using a new video recorder prototype known as the Mobilizer. This new equipment should gather travel times as well as volume, lane occupancy, vehicle length, and speed.

What traffic data do you ideally want?

Nihan would like to receive travel-time data and origin/destination data on the freeways that are accurate and not expensive. She would also like to have access to vehicle classification, vehicle length, and measured speed data.

APPENDIX C. TRIPS SCHEMATIC



APPENDIX D. GLOSSARY OF TERMS AND ACRONYMS

AADT: Annual Average Daily Traffic Volume - a traffic count of the number of vehicles that pass a collection site in both directions for a specified period of time. The result is multiplied by the appropriate seasonal and axle correction factors.

APC: Automatic Passenger Counter - counts the number of passengers using the transit system

AVI: Automatic Vehicle Identification - the identification of a specific vehicle

AVL: Automatic Vehicle Location - tracking or locating specific vehicles

AVMT: Average Vehicle Miles Traveled - the average number of miles traveled by vehicles

AVO: Average Vehicle Occupancy - the average number of people per vehicle

AWD: Average Weekday Volume - the traffic volume count on a weekday averaged over 24 hour increments occurring from Monday noon through Friday noon

AWED: Average Weekend Day Volume - the traffic volume count on a weekend day averaged over 24 hour increments occurring from Friday noon until Monday noon

CMS: Congestion Management System

FHWA: Federal Highway Administration

HOV: High Occupancy Vehicle

HPMS: Highway Performance Monitoring System

ISTEA: Intermodal Surface Transportation Efficiency Act

MPO: Metropolitan Planning Organization

Peak Hour: maximum hourly traffic volume during the day

Peak Hour Split Percent: percent that peak hour traffic in one direction is of the total traffic volume

Peak Hour Traffic Percent: percent that peak hour traffic is of the Average Daily Traffic volume

Peak Hour Truck Percent: percentage of trucks in the traffic during the peak hour

PSRC: Puget Sound Regional Council

RTPO: Regional Transportation Planning Organization

TDAD: Traffic Data Acquisition and Distribution

TMG: Traffic Monitoring Guide

TMS/H: Traffic Monitoring System for Highways

Traffic Volume Truck Percent: percentage of trucks in the total traffic volume

TRIPS: Transportation, Information, and Planning Support

TSMC: Traffic Systems Monitoring Center

VMT: Vehicle Miles Traveled - the combined number of miles traveled by all vehicles

WSDOT: Washington State Department of Transportation

APPENDIX E. DATABASE SCHEMAS

Experimental Design

```
CREATE TABLE TIME_STAMPS (
TS_INDEX INTEGER,
        TS CHAR(18),
        PRIMARY KEY (TS_INDEX)
)
CREATE TABLE SENSOR_NAMES (
SN_INDEX INTEGER,
        SN CHAR(15),
        PRIMARY KEY (SN_INDEX)
)
CREATE TABLE LOOPS (
        TS_INDEX INTEGER,
        SN_INDEX SMALLINT,
        VOLUME SMALLINT,
        SCAN_COUNT SMALLINT,
        FLAGS SMALLINT,
        PRIMARY KEY (TS_INDEX, SN_INDEX),
        FOREIGN KEY (TS_INDEX) REFERENCES TIME_STAMPS,
        FOREIGN KEY (SN_INDEX) REFERENCES SENSOR_NAMES
)
CREATE TABLE STATIONS (
        TS_INDEX INTEGER,
        SN_INDEX SMALLINT,
        VOLUME SMALLINT,
        SCAN_COUNT SMALLINT,
        FLAGS SMALLINT,
        PRIMARY KEY (TS_INDEX, SN_INDEX),
        FOREIGN KEY (TS_INDEX) REFERENCES TIME_STAMPS,
        FOREIGN KEY (SN_INDEX) REFERENCES SENSOR_NAMES
)
CREATE TABLE SPEED_TRAPS (
        TS_INDEX INTEGER,
        SN_INDEX SMALLINT,
        SPEED DEC(4,1),
        LENGTH DEC(4,1),
        FLAGS INTEGER,
        PRIMARY KEY (TS_INDEX, SN_INDEX),
        FOREIGN KEY (TS_INDEX) REFERENCES TIME_STAMPS,
        FOREIGN KEY (SN_INDEX) REFERENCES SENSOR_NAMES
)
```

Final design

```
CREATE TABLE DICTIONARY (
LOOP_ID CHAR(16) NOT NULL,
cabinet_id CHAR(7) NOT NULL,
serial_num CHAR(17) NOT NULL,
metered CHAR(4) NOT NULL,
road_type CHAR(30) NOT NULL,
direction CHAR(12) NOT NULL,
lane_type CHAR(20) NOT NULL,
lane_num SMALLINT,
SENSOR_TYPE CHAR(12) NOT NULL,
DATA_OFFSET INTEGER NOT NULL,
PRIMARY KEY (loop_id,cabinet_id,serial_num),
FOREIGN KEY (cabinet_id,serial_num) REFERENCES CABINET)
)
CREATE TABLE SENSOR_BLOCK (
serial_num CHAR(17) NOT NULL,
time_stamp CHAR(17) NOT NULL,
data LONG RAW NOT NULL,
PRIMARY KEY (serial_num,time_stamp)
)
// this corresponds to the SDD Contents table named CABINETS
CREATE TABLE TDAD_CABINET (
    cabinet_id CHAR(7) NOT NULL,
    serial_num CHAR(17) NOT NULL,
    freeway CHAR(10) NOT NULL,
    description CHAR(255),
    RAMP SMALLINT NOT NULL,
    PRIMARY KEY (cabinet_id,serial_num))

// this corresponds to the SDD Contents table named LOOPS
CREATE TABLE TDAD_SENSOR (
    LOOP_ID CHAR(16) NOT NULL,
    cabinet_id CHAR(7) NOT NULL,
    serial_num CHAR(17) NOT NULL,
    metered CHAR(4) NOT NULL,
    road_type CHAR(30) NOT NULL,
    direction CHAR(12) NOT NULL,
    lane_type CHAR(20) NOT NULL,
    lane_num SMALLINT,
    SENSOR_TYPE CHAR(12) NOT NULL,
    DATA_OFFSET INTEGER NOT NULL,
    PRIMARY KEY (loop_id,cabinet_id,serial_num),
    FOREIGN KEY (cabinet_id,serial_num) REFERENCES TDAD_CABINET)

// this corresponds to the SDD Contents table named COORDINATES
```

```
CREATE TABLE TDAD_COORDINATE (
    coord_type CHAR(40) NOT NULL,
    name1 CHAR(40) NOT NULL,
    name2 CHAR(40) NOT NULL,
    name3 CHAR(40) NOT NULL,
    unit1 CHAR(40) NOT NULL,
    unit2 CHAR(40) NOT NULL,
    unit3 CHAR(40) NOT NULL,
    PRIMARY KEY (coord_type))

// this corresponds to the SDD Contents table named MEASURES
CREATE TABLE TDAD_MEASURE (
    coord_type CHAR(40) NOT NULL,
    authority CHAR(30) NOT NULL,
    ref_system CHAR(128) NOT NULL,
    ref_pnt1 DEC(11,8),
    ref_pnt2 DEC(11,8),
    ref_pnt3 DEC(11,8),
    accuracy1 DEC(11,8),
    accuracy2 DEC(11,8),
    accuracy3 DEC(11,8),
    PRIMARY KEY (coord_type,authority),
    FOREIGN KEY (coord_type) REFERENCES TDAD_COORDINATE)

// this corresponds to the SDD Contents table named CABINET_LOCATION
CREATE TABLE TDAD_LOCATION (
    cabinet_id CHAR(7) NOT NULL,
    serial_num CHAR(17) NOT NULL,
    coord_type CHAR(40) NOT NULL,
    authority CHAR(30) NOT NULL,
    value1 DEC(11,8),
    value2 DEC(11,8),
    value3 DEC(11,8),
    PRIMARY KEY (cabinet_id,serial_num,coord_type,authority),
    FOREIGN KEY (cabinet_id,serial_num) REFERENCES TDAD_CABINET,
    FOREIGN KEY (coord_type,authority) REFERENCES TDAD_MEASURE)

// a row is added to this table when an extractor frame
// (with a new serial number) is received
CREATE TABLE TDAD_EXTRACTOR (
    serial_num CHAR(17) NOT NULL,
    data LONG RAW NOT NULL,
    PRIMARY KEY (serial_num))

// a row is added to this table when a new data frame is received.
```

```
CREATE TABLE TDAD_BLOB (
    serial_num CHAR(17) NOT NULL,
    time_stamp CHAR(17) NOT NULL,
    data LONG RAW NOT NULL,
    PRIMARY KEY (serial_num,time_stamp))
```

APPENDIX F. SOFTWARE IN USE

- The DBMS is Oracle 8.0, running on Windows NT Server 4.0.
- TDAD Expand and TDAD Receiver are implemented in Java, and both use the Oracle JDBC “thin” driver to connect to the DBMS.
- TDAD’s web server is Microsoft IIS 4.0, and the form handling is done with a Perl CGI script.
- ESRI ArcView 3.1 was used to create the map from King County GIS data.
- The Dynamic HTML interface was tested with Microsoft Internet Explorer 4.0 and Netscape Communicator 4.5.