Automated Data Collection, Analysis, and Archival

Dr. Peter T. Martin, Associate Professor
Civil and Environmental Engineering

Peng Wu
Research Assistant

Department of Civil and Environmental Engineering
University of Utah Traffic Lab
122 South Central Campus Drive

November 2003

Disclaimer

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ADMS	Archived Data Management Subsystem
ADUS	Archived Data User Service
ATMS	Advanced Transportation Management System
DBMS	Database Management System
DOL	Delay On-line
EB	Eastbound
ER	Entity-Relationship model
FHWA	Federal Highway Administration
HPMS	Highway Performance Monitoring System
HOV	High Occupancy Vehicles
ITS	Intelligent Transportation System
MAG	Mountainland Association of Governments
MPO	Metropolitan Planning Organization
NB	Northbound
PeMS	Performance Measurement System
SB	Southbound
SDEV	Standard Deviation
SVG	Scalable Vector Graphics
SQL	Structured Query Language
TDAD	Traffic Data Acquisition and Distribution
TMC	Traffic Management Center
TMS	Traffic Monitoring Station
TOC	Traffic Operation Center
TTI	Travel Time Index
UDOT	Utah Department of Transportation
UTL	Utah Traffic Lab
VHT	Vehicle Hours Traveled
VMT	Vehicle Miles Traveled
WB	Westbound
WFRC	Wasatch Front Regional Council

1. INTRODUCTION

Intelligent Transportation System (ITS) applications and their sensors and detectors provide transportation system performance data [1]. This data is used for traffic signal control, traveler information services, and incident management.

As ITS operations expand in major urban areas, data collection is becoming more comprehensive. If this data is saved and made accessible it could be used for new purposes. For example, ITS-derived performance measures describing traffic throughput, delay, and variability could be applied in transportation planning, research, and other areas beyond present operational and monitoring uses. More traffic management operators are considering an automated system for archiving, managing, and disseminating data generated by traffic monitoring devices.

1.1 Problem Statement

The Utah Department of Transportation (UDOT) and other transportation agencies collaborated to deploy a comprehensive Advanced Transportation Management System (ATMS). Traffic Monitoring Stations (TMS) are a component of the ATMS. They are located on the interstate freeway system at approximately half-mile increments. Each TMS generally consists of a set of inductive loop detectors that cover each mainline and its on-ramps. Loop detector data is aggregated at 15-minute intervals and archived as ASCII-CSV files on a FTP server at a Transportation Operation Center (TOC). Data storage for one month requires 420 Megabytes (MB). Because large quantities of data exist, they are only temporarily stored and are overwritten monthly.

Presently, UDOT does not archive this information for long-term purposes. Fifiteen-minute-interval data currently is archived in an ASCII-CSV file. About 155,000 records are archived each day. This data temporarily is kept on-line and is overwritten monthly. The Utah Traffic Lab (UTL) began to download 15-minute TMS data in Feb 2002 and placed the data from previous months on back-up CDs. TOC and UDOT planning, traffic and safety, pavement management agencies require this information. Transportation professionals outside of UDOT, such as Wasatch Front Regional Council (WFRC), Mountainland Association of Governments (MAG), Salt Lake City/ County, and universities, also need this information.

1.2 Research Scope

This study explores multiple TMS data application purposes. It establishes guidelines for developing a TMS data archival, analysis, and retrieval system. The system makes data available to users in Utah. Its results should be evaluated and applied to make TMS data usable, to diversify TMS data uses, and to enable data dissemination. This research project discusses TMS data management system data analysis and archival processes.

This project consists of four objectives:

1. to conduct a review of nationwide data collection, analysis, and archival practices
2. to identify user requirements inside and outside of UDOT
3. to develop the TMS information system architecture
4. to provide recommendations for implementation in the system.

This paper consists of the following:

1. literature review – summary of standards of Archived Data Users Servers (ADUS), a component of National ITS architecture; review of existing ITS data archival practices
2. investigation of current procedure for TMS data collection, storage, and analysis; review of UDOT's current TMS data processing software, Martin Knopp's Delay On-line software
3. interview with the potential users of archived TMS data; identification of the types of data required and methods of data acquisition
4. development of a set of models to compute a range of key performance measures based on TMS data
5. definition of archiving system requirements
6. explanation of data quality control
7. recommendations for implementing TMS data archiving system; determination of institutions that should maintain data archiving routines

Figure 1.1 shows the process for this research study.

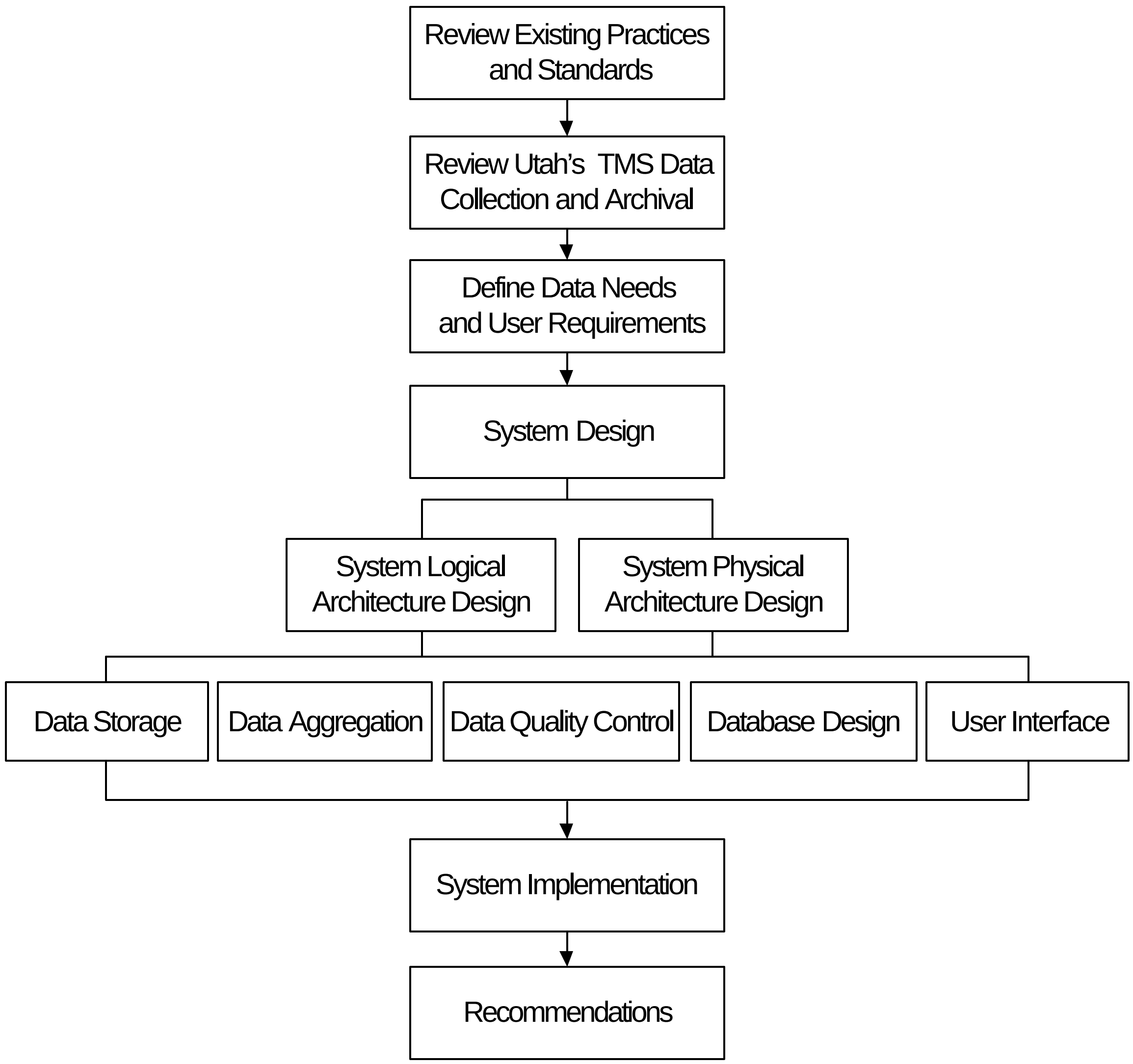


Figure 1.1 Research Process

2. LITERATURE REVIEW

This chapter reviews the advancement of ITS data archival and explains development of the Archived Data User Service (ADUS). Case studies summarize ITS data archiving practices and data applications.

2.1 Development of Archived Data User Service

ITS technologies generate massive amounts of operational data. This data presently is used in real time to effect traffic control strategies. ADUS meets the needs of ITS historical data archiving. It also provides a framework for the collecting, manipulating, retaining, and distributing data generated by ITS for use in other transportation activities.

In the mid-1990s, real time data from traffic and transit operations was archived and used for purposes beyond ITS control strategies. This led to the creation of ADUS. In December 1999, ADUS officially was incorporated into Version 3.0 of the National ITS Architecture. A five-year program (2000-2004) is being conducted to support the implementation of ADUS. ADUS provides the following functions [2]:

- collects, archives, manages, and distributes data from ITS sources
- provides proper formatting, quality control, and assigns necessary metadata
- performs data fusion, or the association and joining of data elements from numerous disparate sources
- prepares "data products" for input into federal, state, and local data reports
-

Figure 2.1 shows Archived Data Management Subsystem (ADMS), the physical entity in the National ITS Architecture that provides ADUS services.

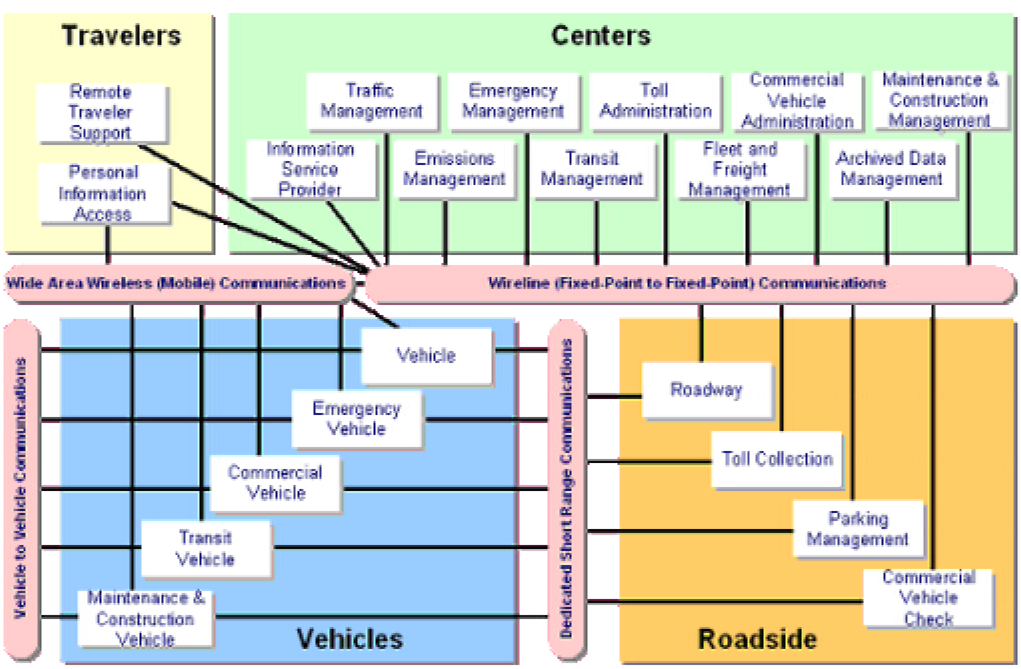


Figure 2.1 National ITS Architecture *(Source: http://itsarch.iteris.com/itsarch/)*

The following equipment packages carry out ADMS processes [1]:

- Traffic and Roadside Data Archival — collects and archives traffic, roadway, and environmental information for use in off-line planning, research, and analysis. This package enables importation of data from devices outside the ITS domain into an ITS data archive. For instance, traffic volumes from sensors maintained by transportation planners can be imported into an ITS data archive.
- ITS Data Repository — collects data and data catalogs from one or more data sources. And stores data in a focused repository for a particular set of ITS data users. This package is the basic data storage and management function that most relational databases provide.
- On-Line Data Analysis and Mining — provides advanced data analysis, summarization, and mining features that facilitate discovery of information, patterns, and correlations in large data sets. This package provides additional analysis functions that enable typical users to analyze large relational databases.
- Virtual Data Warehouse Services — provides capabilities to access "in place" data from geographically dispersed archives and coordinates information exchange with a local data warehouse. This package enables data sharing between agencies that maintain separate databases. For example, it could allow many users access to data maintained in a state highway patrol database and access to the traffic speed data are maintained in a DOT data archive.
- Government Reporting Systems Support — selects and formats data in an ITS archive to meet local, state, and federal government data reporting requirements. For example, this package could provide reports to FHWA's Highway Performance Monitoring System (HPMS) database.

ADUS establishes a general process for developing an ITS data archiving system and carrying out the functional requirements of the system. The main objective of ADUS is to store various data sources and maximize their applications. Due to limited data availability, initial efforts to implement ADUS usually start with archiving and mining traffic surveillance data. The following section reviews existing ITS data archiving practices.

2.2 Overview of Existing ITS Data Archiving Practices

Transportation agencies have realized that ITS data should not be limited to real time operational uses. Since the formation of ADUS many ITS data archiving activities have been conducted by different transportation agencies [5]. Table 2.1 describes these activities.

Table 2.1 Examples of Existing ITS Data Archiving Practices

Location	Archived Data
Phoenix, AZ	Loop detector data from freeways and arterials are archived; plans underway to archive all "relevant" data used by the Traffic Operations Center.
Los Angeles, CA	Developing an archival system for freeway loop detector data.
Orange County, CA	Developing an archival system for freeway loop detector data.

Location	Archived Data
Chicago, IL	Developing an archival system for arterial loop detector data.
Montgomery County, MD	Loop detector data from selected arterials are archived.
Detroit, MI	Loop detector data from freeways are archived.
Minneapolis-St.Paul, MN	Loop detector data from freeways are archived.
TRANSCOM, NY/NJ/CT	Travel times derived from AVI-equipped vehicles are archived.
Houston, TX	Travel times derived from AVI-equipped vehicles are archived.
San Antonio, TX	Loop detector data from freeways, travel times derived from AVI equipped vehicles, and incident management data are archived.
Seattle, WA	Loop detector data from freeways are archived
State of VA	Currently developing an archival system for ITS-generated data.

Table 2.1 shows roadway surveillance data. These are the primary data sources for current ITS data archiving system. The following archived data are applicable to various fields of transportation:

- *Freeway Performance Evaluation in Puget Sound Region, Wash.* Loop detector data monitors congestion patterns, including variability in speeds and travel times.
- *Evaluation of HOV Lanes in Houston, Texas.* Probe vehicle data compares travel times for HOV and non-HOV lanes.
- *Traffic Statistics in Chicago, Ill.* Loop detector data has been used to produce an "atlas" of traffic statistics.
- *Development of Travel Time Prediction Models, Texas.* Data from probe vehicles is being used to develop short-term travel time prediction models for traffic operations.
- *Ramp Metering Evaluation, Minneapolis-St. Paul, Minn.* Freeway and ramp loop detector data used to evaluate cycle lengths on ramps.
-

These applications indicate that archived ITS data can be applied in diverse areas. However, the data mainly is associated with highway performance analysis. The Federal Highway Administration's (FHWA) Office of Highway Policy Information conducted a nationwide investigation to determine ITS traffic detector potential in Performance Monitoring System (HPMS) reporting. It asked various states three questions about HPMS reporting [3]:

- Is the state traffic monitoring office aware of ITS detectors?
- Is the state using ITS detectors for HPMS reporting purposes?
- If the state is not yet using ITS detectors for HPMS, why not?
-

Answers to these questions were available for 43 states. Fourteen states answered positively to both questions and 16 states answered positively to the first question and negatively to the second. Survey percentages are shown in Figure 2.2.

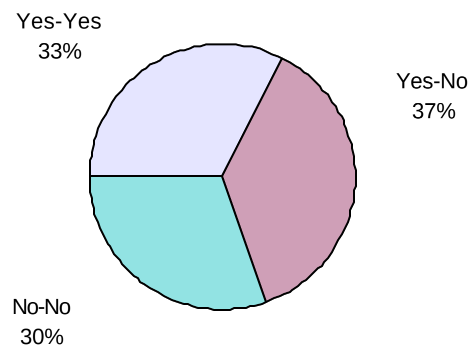


Figure 2.2 Investigation of ITS Data Uses in HPMS

(Source: Status of ITS Traffic Data for HPMS. The FHWA Office of Highway Policy information, August 2002.)

One-third of these states used ITS traffic detectors to supply HPMS traffic data. Several states noted that the number of ITS detectors available currently was limited, but was expected to increase in the future. Thirty percent of the surveyed states had no ITS traffic detectors at that time, though many were implementing the detectors. Several states reported that they were willing to use the detectors and expected to have them in the future. However, some states without detectors believed the detectors provided poor quality data. Others were still considering the benefits and drawbacks of ITS detector application. Overall, 70 percent of the states have ITS detectors available. Nearly one-half of these states currently are archiving some data for HPMS.

2.3 Case Studies

This section provides case study summaries of several well-known ITS data archiving and utilizing systems. It illustrates some basic processes of and principles for developing an ITS data archiving and analysis system. These systems have many applications.

2.3.1 PeMS in California

The California Department of Transportation (Caltrans) and researchers at the University of California at Berkeley (UCB) have worked together to create a freeway PeMS. In this PeMS, a front-end processor at the Traffic Management Center (TMC) receives data from freeway loops every 30 seconds. It formats this data and writes it into the TMC database. Data is sent over the Caltrans Wide Area Network from the TMCs to the PeMS computer, a data warehouse located at UCB. PeMS maintains a separate part of the database for each of Caltrans' 12 districts [5]. Through the software, users can query diverse information on traffic conditions using web browsers (http://transacct.eecs.berkeley.edu).

PeMS extracts traffic information from real time and historical data and presents it in a low-cost, easy-to-use format. This assists managers, traffic engineers, planners, freeway users, researchers, and transportation information service providers. Managers can use PeMS to determine how well the systems work. Engineers can use it to obtain detailed traffic analyses and to spot bottlenecks or malfunctioning equipment. PeMS also can be used by planners to evaluate management strategies, by travelers to find the quickest or shortest routes for their trips, and by researchers to validate theories and calibrate simulation models [4].

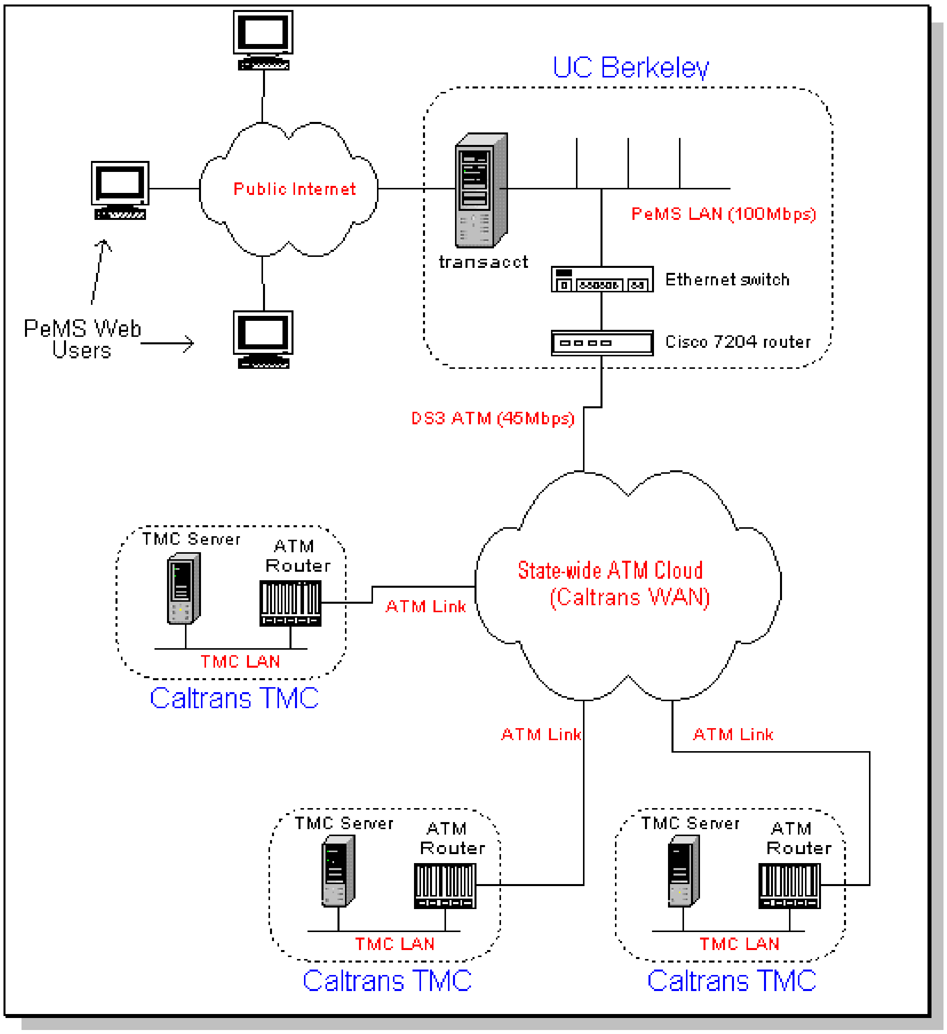
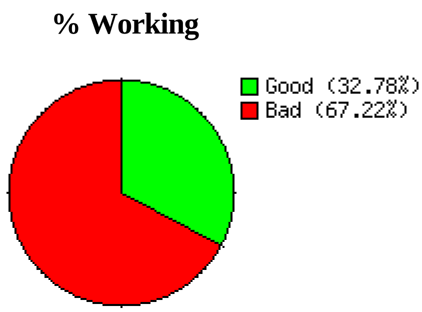


Figure 2.3 Architecture of PeMS

(Source: www.path.berkeley.edu/PATH/Research/presentations2001/alex_skabardonis/PeMS_Phase_II.ppt)

PeMS has different processes of data health checking and diagnosis than other ITS data archiving and processing systems. Data quality control is a major part of PeMS function. Figure 2.4 is an example of a PeMS data quality report. According to statistics, data are categorized into six groups based on their data error types. Only 32.78 percent of detectors worked well and provided reliable data. Aside from checking data quality, PeMS also developed sophisticated algorithms to compute values for bad and missing detectors. All web pages on the PeMS site use processed data rather than raw data.



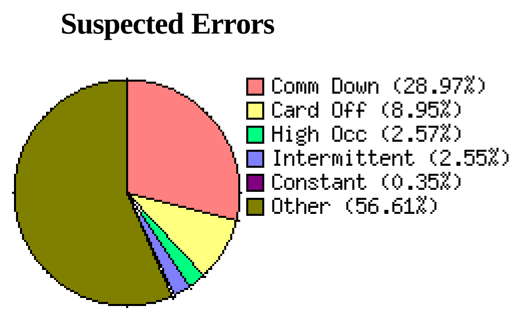


Figure 2.4 Example of PeMS Data Quality Check *(Source: http://pems.eecs.berkeley.edu)*

2.3.2 TDAD in Washington State

Researchers at the Washington State Transportation Center (TRAC) are working with the Washington State Department of Transportation (WSDOT) to demonstrate uses of archived ITS data and to extend its applications. They developed several analytical tools that now are used by TRAC and WSDOT.

TRAC and WSDOT collect freeway detector data every 20 seconds from field controllers. The data is then summarized by five-minute increments in the data archive. Before the detector data is loaded into a CD archive, a data quality control procedure is performed to document failed information. PC-based CD retrieval software retrieves and adds the five-minute loop data to local or network disk drives. CD Analyst software uses spreadsheet macros to analyze the loop data and to develop graphical presentations [5].

Presently, TRAC researchers are also developing a web-based ITS data mining application in the Traffic Data Acquisition and Distribution (TDAD) study. This prototype application has a map-based interface (http://www.its.washington.edu/tdad/). It is used to select loop detector locations of interest. At this time, the raw 20-second loop data populates the TDAD database. Approximately six months of data have been accumulated [6].

Archived ITS data in WSDOT is used for a variety of purposes. It tests and evaluates operational improvements such as ramp metering of HOV lanes, freeway performance monitoring, pavement design, and freight performance analysis.

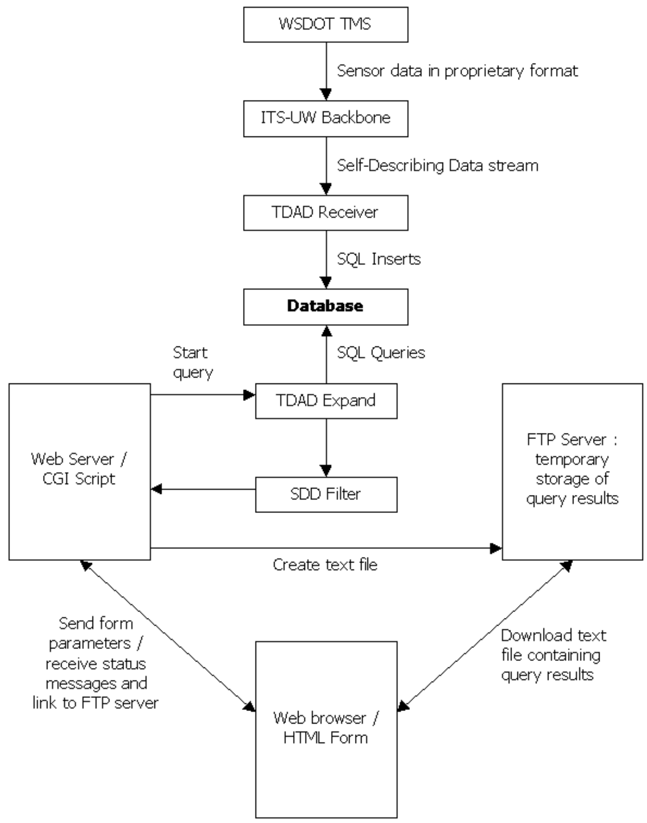


Figure 2.5 Traffic Data Acquisition and Distribution Framework
(Source: http://www.its.washington.edu/tdad/tdad.gif)

2.3.3 ITS Data Management in Texas

Texas uses ITS data archiving. The TransLink® ITS Research Program at Texas Transportation Institute has developed an on-line ITS data management system referred to as DataLink. This prototype DataLink system provides user-friendly access to 40+ Gigabytes (GB) of TransGuide loop detector data. The data has been stored since late 1997. A range of transportation system performance measures can be queried in a web browser interface. This eliminates the need for Structured Query Language (SQL) [5].

Furthermore, a FHWA performance-monitoring program was started in 2000 with support from the Texas Transportation Institute and Cambridge Systematics. It continuously archives freeway detector data from 10 cities. Houston and San Antonio are attendees to this program. The archived data is gathered in a variety of formats and is compiled into a standard five-minute, lane-by-lane format for further processing and analysis. This multi-city data archive primarily is meant to monitor data mobility and reliability at the city and national levels. In this program, raw data is saved off-line on CDs and summary data is kept on-line for analysis. Basic data quality control methods identify and remove suspect or erroneous data [7].\

2.4 Summary of Findings

According to nationwide research and practice review, many existing TMCs archive or plan to archive traffic surveillance data. Archived ITS data can be successfully applied to purposes beyond real time uses. Table 2.2 summarizes ITS data types and their archival and dissemination.

Table 2.2 Summary of Loop Detector Data Archiving Practices

Data Archiving Activi ties	**Data Source**		**Data Archiving**		
	Interval	Spacing	Aggregation	Storage	Distribution
Caltrans PeMS	30 s	0.5 mi	5 min	Current 40G online	Internet
Washington Seattle TDAD	20 s	0.5 mi	5 min	CD and online	Internet
San Antonio Transguide	20 s	0.5 mi	5 min	CD and online	Internet
Minneapolis TMC	20 s	0.33-2 mi	5 min	CD	
Long Island INFORM	60 s	0.5 mi	5 min	Tape, recent months online	Internet
Michigan ITS	20 s	0.5 mi	5 min	Tape, one week available online	
Phoenix AZTech	20 s	0.5 mi	5 min	Online and offline storage	
Salt Lake City TOC	20 s	0.5 mi	15 min	Overwritten each month	
Florida UCF	20 s	0.5 mi	15 min	Online	Internet

The following are common concerns and similarities in archiving activities.

Data Source

ITS generates a variety of data through its freeway management, incident management, and transit management components. However, initial ITS data archiving efforts were focused only on freeway detector data such as volume, speed and occupancy. This data usually is gathered for individual lanes by loop detector systems every 20 or 30 seconds. This data is a major source of current ITS infrastructure. It also provides basic information about transportation systems.

Data Aggregation and Storage

Data often is aggregated into different levels before it is archived. This reduces storage space and improves efficiency. Table 2.2 shows that data frequently is summarized by five-minute increments.

User Identification

Many TMC personnel are interested in making data accessible to secondary users, such as planners, traffic performance monitoring programs, and travelers.

Data Quality

Many ITS data archiving efforts emphasize the importance of data quality. Only accurate data is worth further analysis. Data quality control is performed in relatively mature archiving systems. This control includes checking data quality, identifying and correcting erroneous data, and providing missing data.

Easy Data Access

Stored ITS data are relatively inaccessible and unusable due to their size and format. Many TMCs want to decipher these data and distribute them via the Internet or CD rather. This process would make data more accessible than archiving data onto magnetic tape cartridges or off-line storage devices.

Institutional Issue

Transportation operations personnel usually own ITS data. Dedicated research institutions should manage archived data. The University of California at Berkeley, the University of Washington, and the University of Texas A&M maintain archival data and develop applications to distribute this data. Assignment of responsibility and adequate funding are important in managing data archives [8].

3. CURRENT TMS DATA COLLECTION AND ARCHIVING

ATMS is a computerized communications system that carries out traffic surveillance and management, incident detection, and information dissemination. UDOT has installed comprehensive ATMS. This section reviews the elements of ATMS and identifies available data produced by ATMS. It also discusses TMS data as a major ATMS data source. In addition, it discusses Delay On-Line, UDOT's current TMS data archiving and processing software.

3.1 ATMS Elements

UDOT's current ATMS includes the following elements [9]:

- A TOC
- TMSs on all freeways and some arterial roadways
- Interconnected traffic signals on many arterials, in addition to ramp metering
- Closed Circuit Television (CCTV) on all freeways and some arterials
- Variable Message Signs (VMS) on all freeways and a few arterials
- Highway Advisory Radio (HAR) stations at nearly a dozen locations
- Roadway-Weather Information System (RWIS) stations at a number of locations
-

Figure 3.1 diagrams the relationship among the above ATMS elements. The heart of ATMS is TOC. This center interconnects different elements of ATMS and communicates with other transportation communities. The diagram, TMS, CCTV, and RWIS provide data input for other elements. These data are processed in TOC. They are then broadcast via HAR, VMS, and the Internet. Major traffic information, such as volume, speed, and occupancy are collected by TMSs. Therefore, they are a basic component of the ATMS data source. The following sections discuss the specific data inputs, intervals, and formats for the TMS data currently available.

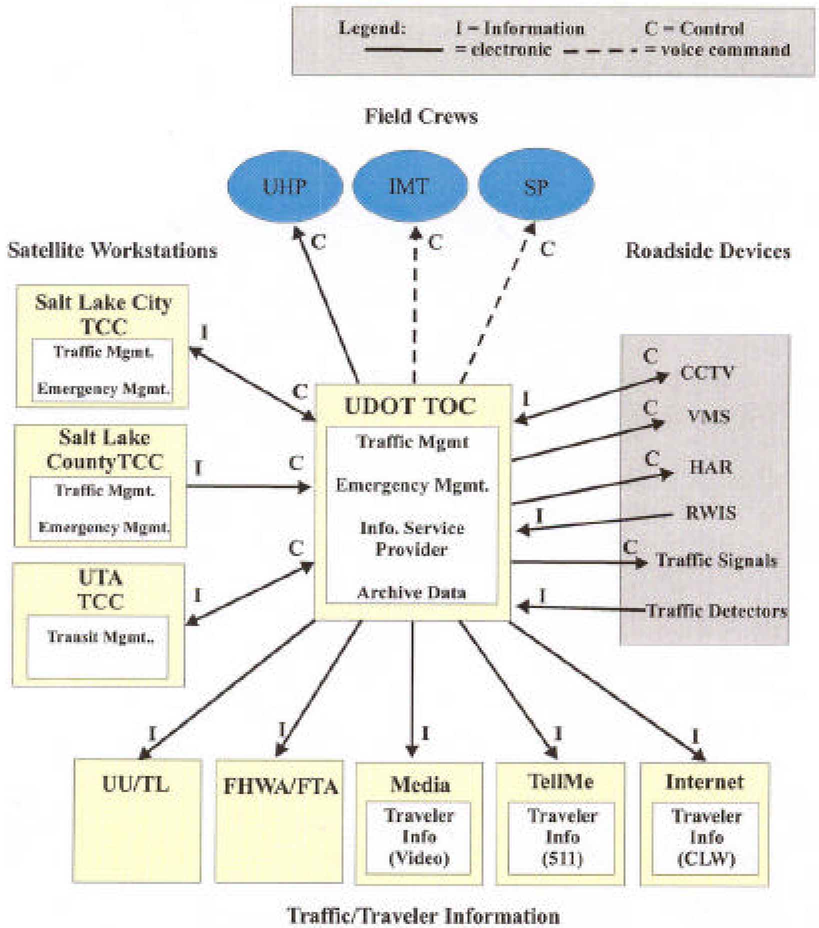


Figure 3.1 ATMS Architecture Diagram
(Source: Intelligent Transportation System at the 2002 Salt Lake City Olympic Games, Iteris, Inc., August 2002.)

3.2 TMS Data Collection

Currently, 641 TMSs connect to the TOC via communications links. This number will grow with TMS expansion. Almost all TMSs are located on the freeway system at approximately half-mile increments. A few arterials and state routes also house TMSs. Each TMS generally consists of a set of detectors, one detector in each mainline lane. It also has additional detectors on the on-ramps. Each mainline detector consists of double inductive loops to measure volume, speed, and occupancy. The detectors at the 23 metered on-ramps generally have several loops in each lane to detect calls, clearance, and queue backup. Detectors at non-metered on-ramps include fewer loops. Figure 3.2 illustrates the layout of a TMS. TMS data includes:

- speed, volume, and occupancy
- recording time
- TMS site ID
- detector ID
- lane ID at every TMS site
- milepost value of each TMS site
- location of the TMS site (including direction)

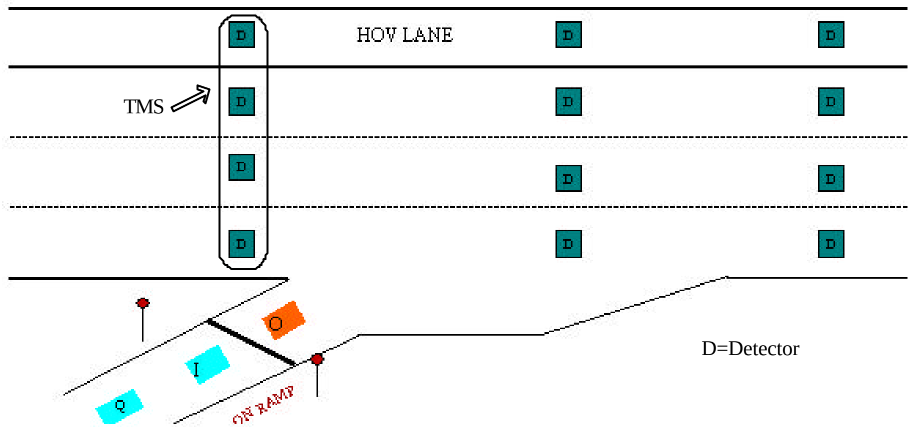


Figure 3.2 Layout of Traffic Monitor Station

3.3 TMS Data Storage and Processing

TMS data are collected in real time. The standard sampling period is 20 seconds. Through CommuterLink fiber-optic communications, average 20-second volumes, speeds, and occupancies continuously flow back to TOC. TOC uses the data to perform real time traffic management and traveler information functions.

Figure 3.3 illustrates 20-second and 15-minute TMS data stored in separate ways according to their uses. Tables 3.1 and 3.2 give the average volumes, speeds, and occupancies in 20-second and 15-minute levels. The malfunction records are coded as "–9".

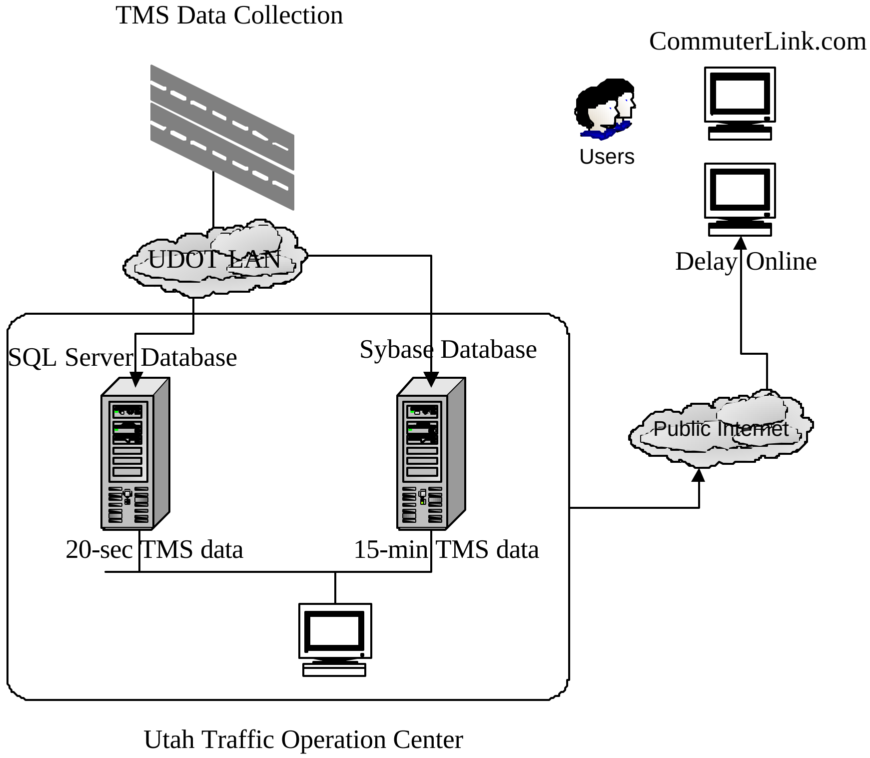


Figure 3.3 Current TMS Data Storage Process

Table 3.1 20-second TMS Data Sample

Station_id	Timestamp	Volume	Speed	Occupancy
1	2003-05-21 15:48:40.000	-9	-9	-9
2	2003-05-21 15:48:40.000	33	59	11
4	2003-05-21 15:48:40.000	23	66	6
6	2003-05-21 15:48:40.000	18	63	5
9	2003-05-21 15:48:40.000	16	67	4
11	2003-05-21 15:48:40.000	9	67	3
13	2003-05-21 15:48:40.000	14	71	3
14	2003-05-21 15:48:40.000	18	69	5
15	2003-05-21 15:48:40.000	19	71	4
16	2003-05-21 15:48:40.000	15	72	4
17	2003-05-21 15:48:40.000	10	60	4
20	2003-05-21 15:48:40.000	12	62	3

Table 3.2 15-minute TMS Data Sample

DetectorID	StationID	SampleStart	Dir&Lane	Mile_Post	Station_Length	Avg_Speed	Avg_Occupancy	TotalVolume	LocationText
1263	1	Feb 9 2003 12:00AM	12	0.5	0.3	0	0	0	Interstate 215 East Southbound @ Foo
1264	1	Feb 9 2003 12:00AM	13	0.5	0.3	0	0	0	Interstate 215 East Southbound @ Foo
1067	2	Feb 9 2003 12:00AM	10	1.3	0.6	-9	-9	-9	Interstate 215 East Southbound @ Inte
1068	2	Feb 9 2003 12:00AM	11	1.3	0.6	-9	-9	-9	Interstate 215 East Southbound @ Inte
1069	2	Feb 9 2003 12:00AM	12	1.3	0.6	-9	-9	-9	Interstate 215 East Southbound @ Inte
1058	4	Feb 9 2003 12:00AM	10	1.9	0.3	69	0	23	Interstate 215 East Southbound @ 330
1059.00	4.00	Feb 9 2003 12:00AM	11	1.9	0.3	65	1	78	Interstate 215 East Southbound @ 330

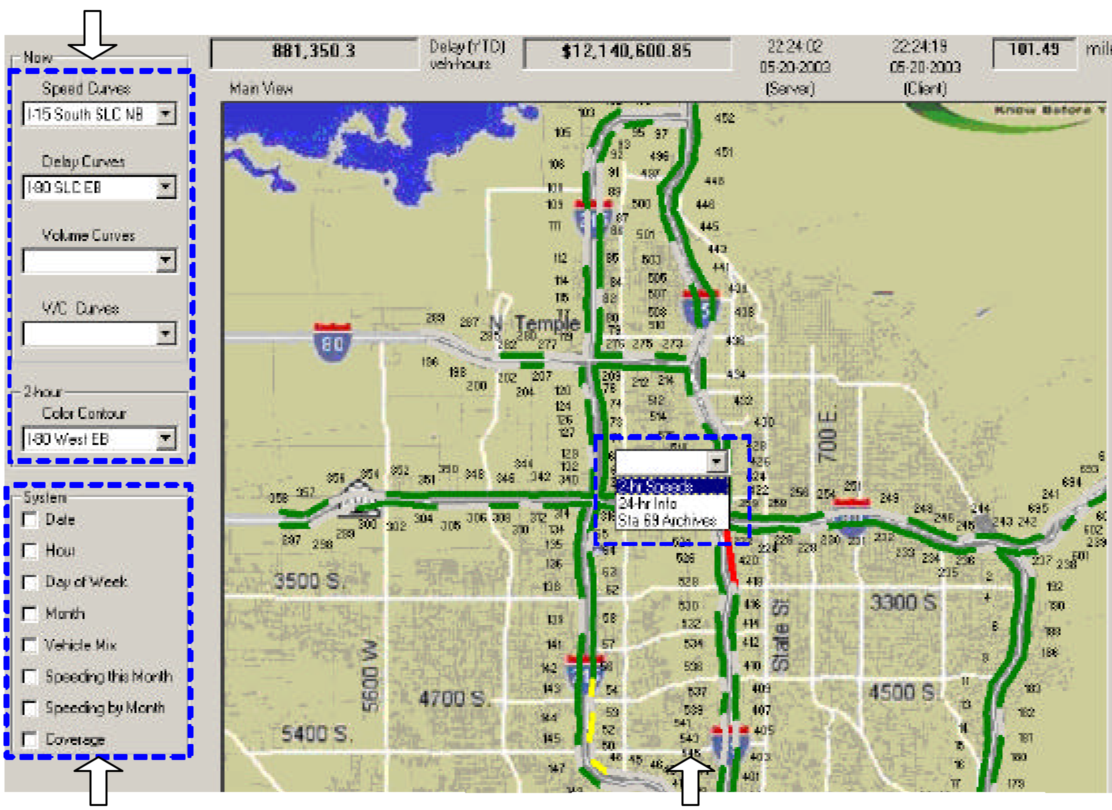
Twenty-second real time data is stored in a "buffer" in the TOC computer. The real time data is then fed to CommuterLink. Over the Internet, CommuterLink provides real time traffic speed information using different colors to display speeds in different ranges. The Website, http://commuterlink.utah.gov shows the functions of CommuterLink.

3.4 Delay On-Line

System Introduction

In 2001, Martin Knopp, a UDOT employee, released Delay On-Line (DOL) 1.1, a system that monitors total delay on detectorized freeway systems. This software obtains 20-second TMS data from TOC's data server, a collection and storage base for TMS data. Currently, the system provides real time display, and it aggregates and examines TMS data collected from Utah's freeways. Delay On-Line also provides a wide variety of performance measures focusing on traffic delay analysis. Any authorized user can access this system. However, before running the system, the user must to download and install the Citrix ICA Client software on their workstations. Figure 3.4 shows an easy-to-use graphical interface.

Corridor Wide Performance Measures



System Wide Performance Measrues

Site Specific Performance Measures

Figure 3.4 Delay On-Line User Interface *(Source: Delay On-Line software)*

Performance measures provided by DOL can be divided into three groups according to their coverage capacity: site-specific measures, corridor-wide measures, and system wide measures. Each group's measures can be categorized at a different temporal level. Table 3.3 provides a complete list of DOL performance measures organized by their spatial and temporal differences. Table 3.4 shows each Utah freeway broken down by corridors. The following sections describe performance measures along Utah's freeways. Twenty-second measures are described in real time performance measures. The others are discussed in historical performance measures.

Table 3.3 Delay On-line's Performance Measures

Space Level / Time Level	**Point**	**Corridor**	**Freeway Systems**
20 Seconds	Vol, Speed, V/C, Delay	Vol, Speed, V/C, Delay	
Hour	Vol, Speed, Occ, V/C, Delay, TTI		Delay
Day	ADT, Delay		Delay
Month	Speeding		Delay, Speeding
Year			Delay

Note: Volume (Vol), Occupancy (Occ), Travel Time Index (TTI).

Table 3.4 Corridors List

Freeway Name	**Corridor Name**
I-15	I-15 South SLC NB
	I-15 South SLC SB
	I-15 North SLC NB
	I-15 North SLC SB
	I-15 SLC NB
	I-15 SLC SB
I-215	I-215 SouthWest SB
	I-215 SouthWest NB
	I-215 NorthWest SB
	I-215 NorthWest NB
	I-215 East NB
	I-215 East SB
I-80	I-80 West EB
	I-80 West WB
	I-80 East EB
	I-80 East WB
	I-80 SLC EB
	I-80 SLC WB
StateRoad 201	SR 201 WB
	SR 201 EB

Real Time Performance Measures

Martin's software provides features similar to the UDOT CommuterLink system. Real time traffic speed information is visible on a map of the freeway system. The system updates this information every 20 seconds. Figure 3.4 shows the speed on each freeway segment. It is typically about one-half mile long and is shown as a color-coded band.

Based on 20-second data, connecting TMS data along a specific route makes it possible to monitor real time traffic conditions using volume, speed, V/C, and delay curve. For example, a V/C ratio curve along I-80 WB displays current congestion conditions at different locations. These could be supplied to traffic operators and others interested in monitoring traffic flow.

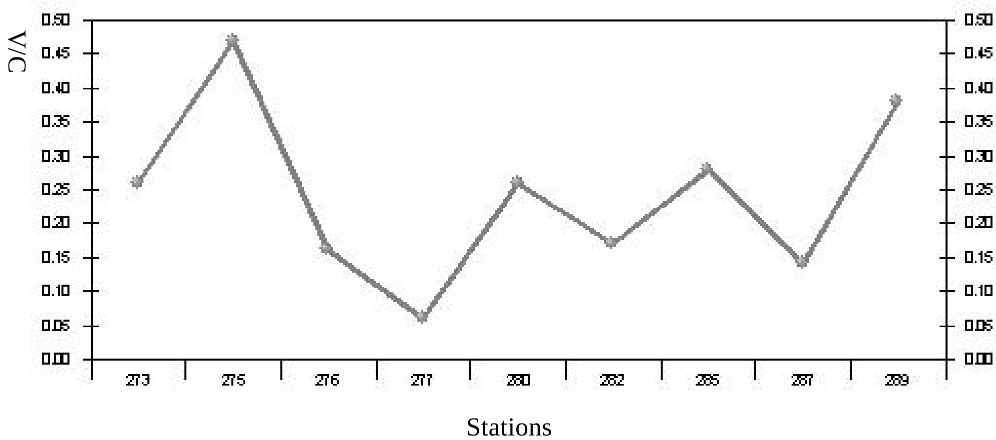


Figure 3.5 Volume/Capacity along I-80 WB *(Source: Delay On-Line software)*

Historical Performance Measures

This software aggregates 20-second raw data into one-hour data and archives it for historical and system-wide freeway performance analysis. Figure 3.6 show an interface for accessing historical data. DOL provides primary site-specific measures, such as speed, volume, occupancy, Travel Time Index (TTI), delay, and V/C for each specific day and station.

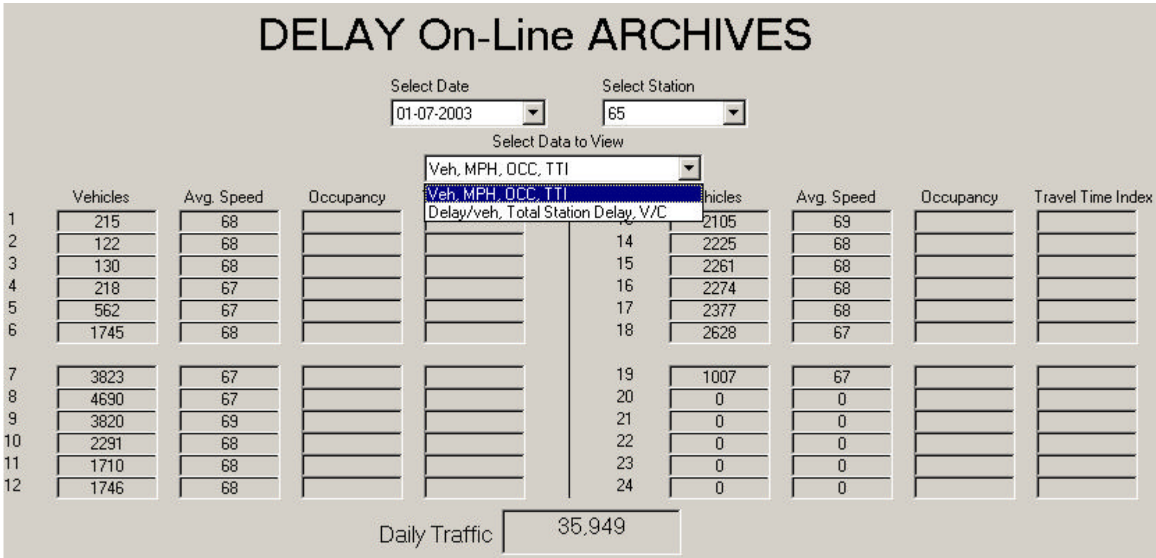


Figure 3.6 Site-Specific Performance Measures Interface *(Source: Delay On-Line software)*

Archived data is used to evaluate the freeway system's performance. System delay is computed by hour, day of week, and month. Figure 3.7 reveals that afternoon peak period traffic causes a significant delay every day. It also shows percentage of vehicle congestion.

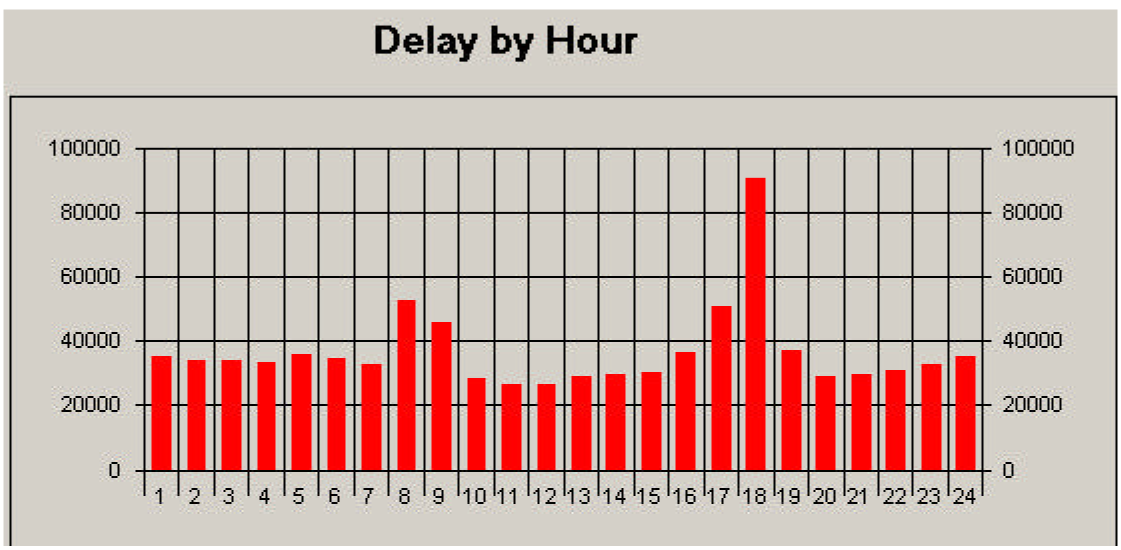


Figure 3.7 System Wide Delay Summaries by Hour *(Source: Delay On-Line software)*

Summary

DOL is particularly strong in delay analysis. However, developers currently are enhancing it. DOL does not perform sophisticated data quality control. All analyzed and displayed data is raw data, which can lead to possible bias of summarized performances.

System delay, speeding, and TTI analysis depend on accurate speed measurement. However, a significant number of current TMS detectors tend to provide higher than average traffic speeds. This can cause low estimates in system delay and high calculations in speeding percentages. Erroneous data includes missing data, unreported data, and data with "–9" values. If data "holes," are not filled, the accuracy of system measures also is questionable. Figure 3.8 illustrates a sample of speed data along I-80 Eastbound (EB) with a significant amount of missing and erroneous data.

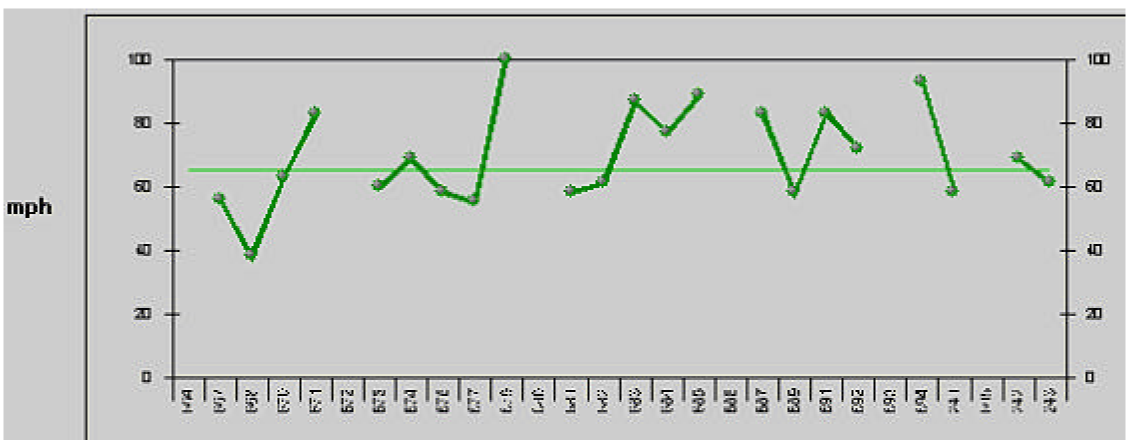


Figure 3.8 Real-time Speed along I-80 EB *(Source: Delay On-Line software)*

4. ARCHIVED TMS DATA NEEDS AND USER REQUIREMENTS

User requirements for ITS data influence the design of data archiving systems. The following section identifies user needs and requirements through the input of potential users. These users include operation personnel, planners, researchers, and traffic engineers.

4.1 Potential Users and Applications for Archived ITS Data

TMS archives a variety of ITS data. Table 4.1 lists the ITS data types used by different groups. Needs of different user groups may help determine appropriate designs for ITS archiving systems. This section describes several stakeholders listed in Table 4.1 and discusses the following:

- Typical data elements of interest
- Spatial/temporal aggregation levels of interest
- Additional data characteristics of importance
- Examples of common ITS applications

Table 4.1 Potential Users and Applications for Archived ITS Data

Stakeholder Group	Primary Transportation-Related Functions	Example Applications
Metropolitan Planning Organization (MPO) and state transportation planners	Identifying multimodal passenger transportation improvements (long and short-range); congestion management; air quality planning; develop and maintain forecasting and simulation models	• congestion monitoring • link speeds for Traffic Demand Forecasting and air quality models • AADT, K- and D-factor estimation • temporal traffic distributions • truck travel estimation by time of day • macroscopic traffic simulation • parking utilization and facility planning • HOV, paratransit, and multimodal demand estimation • congestion pricing policy
Traffic management operators	Day-to-day operations of deployed ITS (e.g., Traffic Management Centers, Incident Management Programs)	• pre-planned control strategies (ramp metering and signal timing) • highway capacity analysis • saturation flow rate determination • microscopic traffic simulation

Stakeholder Group	Primary Transportation-Related Functions	Example Applications
		-- historical -- short-term prediction of traffic conditions • dynamic traffic assignment • incident management • congestion pricing operations • evaluation and performance monitoring
Transit operators	Day-to-day transit operations: scheduling, route delineation, fare pricing, vehicle maintenance, transit management systems, and evaluation and planning	• capital planning and budgeting • corridor analysis planning • financial planning • maintenance planning • market research • operations/service planning • performance analysis planning • strategic/business planning
Air quality analysts	Regional air quality monitoring; transportation plan conformity with air quality standards and goals	• emission rate modeling • urban air shed modeling
MPO/state freight and intermodal planners	Planning for intermodal freight transfer and port facilities	• truck flow patterns (demand by origins and destinations) • HazMat and other commodity flow patterns
Safety planners and administrators	Identifying countermeasures for general safety problems or hotspots	• safety reviews of proposed projects • high crash location analysis • generalized safety relationships for vehicle and highway design • countermeasure effectiveness (specific geometric and vehicle strategies) • safety policy effectiveness
Maintenance personnel	Planning for the rehabilitation and replacement of pavements, bridges, and roadside appurtenances; scheduling of maintenance activities	• pavement design (loadings based on ESALs) • bridge design (loadings from the "bridge formula") • pavement and bridge performance models • construction and maintenance scheduling
Commercial vehicle enforcement personnel	Accident investigations; enforcement of commercial vehicle regulations	• HazMat response and enforcement • congestion management

Stakeholder Group	Primary Transportation-Related Functions	Example Applications
		• intermodal access • truck route designation and maintenance • truck safety mitigation • economic development
Emergency management services (local police, fire, and emergency medical)	Response to transportation incidents; accident investigations	• labor and patrol planning • route planning for emergency response • emergency response time planning • crash data collection
Transportation researchers	Development of forecasting and simulation models and other analytic methods; improvements in data collection practices	• car-following and traffic flow theory development • urban travel activity analysis
Private sector users	Provision of traffic condition data and route guidance (Information Service Providers); commercial trip planning to avoid congestion (carriers)	

Source: http://www.fhwa.dot.gov/ohim/its/tab1_1.pdf

4.2 Process of Establishing TMS Data Needs

Data needs are established through user input. This process requires coordination of different agencies. Users establish data needs in a series of one-on-one interviews via e-mail or telephone. However, these data needs may be beyond the capacity of the newly developed TMS database. More ITS data may become part of the database in response to existing needs. This process allows data providers to modify or adapt data collection practices. They also may sensor designs to meet data needs. However, if data needs are too exhaustive, the data provider may be incapable of providing all data elements. Table 4.2 lists the people who attended to the data needs survey. These potential users are from different agencies within and outside of UDOT, who have different purposes in traffic data collection.

Table 4.2 Data Needs Survey List

	Name	Organization	Phone No.	E-mail
Within UDOT	Stan Burns	Research	965-4190	sburns@utah.gov
	Sam Sherman	ITS	887-3744	ssherman@utah.gov
	Joe McBride	TOC	887-3716	joemcbride@utah.gov
	Gary Kuhl	Planning	964-4552	gkuhl@utah.gov
	Rob Clayton	Traffic/Safety	964-4521	robertclayton@utah.gov
	Robert Hull	Traffic/Safety	965-4259	rhull@utah.gov
	Tammy Kaeser	Traffic Statistics	965-4137	tkaeser@utah.gov
Outside UDOT	Chad Worthen	MAG	229-3811	cworthen@mountainland.org
	Wayne Bennion	WFRC	363-4230-115	wbennion@wfrc.org
	Kent Barnes	SL County	562-6422	kbarnes@co.slc.ut.us
	Kevyn Smeltzer	SL County	562-6490	ksmeltzer@co.slc.ut.us
	Jerry Blair	SLC	535-7103	
	Tom Stetich	SLC	538-6530	
	Richard Hodges	UTA	262-5626-2354	rhodges@uta.cog.ut.us

4.3 Summary of Data Needs Survey

Interviews with traffic data users reveal a variety of data applications. Table 4.3 summarizes these applications. The appendix lists each user group's specific data needs by data format and detail level.

Most organizations indicated that they obtain traffic data from the annual traffic report produced by UDOT. Traffic counting and monitoring stations like Automatic Traffic Recorder (ATR) are located throughout the state. These stations provide monthly average traffic counts for weekdays and weekends. This data is collected by the Transportation Monitoring Unit of UDTO. It is then developed and analyzed by UDOT's Program Development Division. This information is not included in an automatic archiving system. Therefore, these organizations must wait until the annual report is released. A more effective way to query data is needed as organizations want quick access to speed, volume, and occupancy data for different times and locations.

Some users found TMS data difficult to understand in its current format and size. Data users commonly worry about data quality. Some users mistrust the huge amount of data collected by TMS sites. Table 4.3 shows data needs and user concerns incorporated into a system design plan for an automated data management system.

Table 4.3 Data Needs

	User	Purpose	Data Needs
Within UDOT	Research	Research	AADT Speed/5min Traffic Volume/5min, peak hour
	ITS	Real Time Traffic Control / Management	AADT Incidents Speed/Hourly Travel Time Traffic Volume/Hourly Vehicle Classification
	TOC	Manage Commuter Line and Provide Instant Data on Road and Traffic Conditions; Congestion Management; Signal Timing.	Speed/15min Traffic Volume/Hourly, Real Time Turning Movement/Peak Times Vehicle Classification
	Planning	Long Range Planning HOV Analysis Capacity Analysis	Traffic Volume/Hourly Peak Hour Volume/Directional Split Ramp Volumes Vehicle Classification
	Traffic/Safety	Safety Studies Traffic Studies	AADT/AWDT Density/15min Speed Traffic Volume Vehicle Classification Turning Movement/15min
	Traffic Statistics	Traffic Statistic and Reporting	AADT Traffic Volume/15min Vehicle Classification/Length, Axle
	Maintenance	Road Maintenance	AADT Traffic Volume
Outside UDOT	MAG	Planning Signal Coordination Incident Analysis Congestion Analysis	Speed/Hourly Traffic Volume/Hourly Turning Movement Ramp Metering
	WFRC	Long Range Planning Validate Transportation Model	AADT/AWDT Speed/15min Vehicle Classification/Hourly Traffic Volume/Hourly Turning Movement
	SL County	Maintenance Signal Design	AADT Travel Time Turning Movement
	SLC	Maintenance Signal Design	AADT Travel Time Turning Movement

	User	Purpose	Data Needs
	UTA	Route performance Analysis Scheduling Evaluation and Planning	Speed/ 15min Incidents/Accidents Volume/Hourly, by lane Vehicle Classification
	Universities	Research	AADT Speed/5min Traffic Volume/5min Turning Movement/5min

5. DEVELOPMENT OF PERFORMANCE MEASURES

This chapter summarizes a variety of performance measures that can be generated from TMS data. A set of models is developed for the evaluation of transportation systems. It is based on site-specific measurements of volume, speed, and occupancy.

5.1 Available Performance Measures

Table 5.1 lists performance measures derived from archived TMS data. Performance measures are collected by point, link, corridor, and system. Points are locations of detector stations. They measure volume, speed, and occupancy for each detector and station. A link is section of lane or road holding detectors. Link-based measurements can be taken by lane or a segment. A segment is a collection of several lanes. Performance measures, such as VMT, VHT, travel time, and delay, show transportation system mobility at the corridor or system level. Just as measurements take place at different spatial levels, they also take place at different time levels, ranging from 20 seconds to one year.

Table 5.1 Performance Measures

Measure Level	Measure Type	Performance Measures
Point Measures	Detector Station	Speed Volume Occupancy
Link-Based Measures	By Lane Road Segment	Speed Volume Travel time Delay
Corridor or System Measures	By Lane Road	Travel Time Delay Vehicle Miles Traveled (VMT) Vehicle Hours Traveled (VHT)

5.2 Performance Measures Modeling

There are two ways to obtain performance measures. The first is to aggregate and summarize original data to calculate volume, speed, and occupancy measures. The second is to derive measures such as travel time, delay, VHT and VMT from existing data. Figure 5.1 shows the process of calculating performance measures. This section describes each measure and defines a set of algorithms to compute these measures.

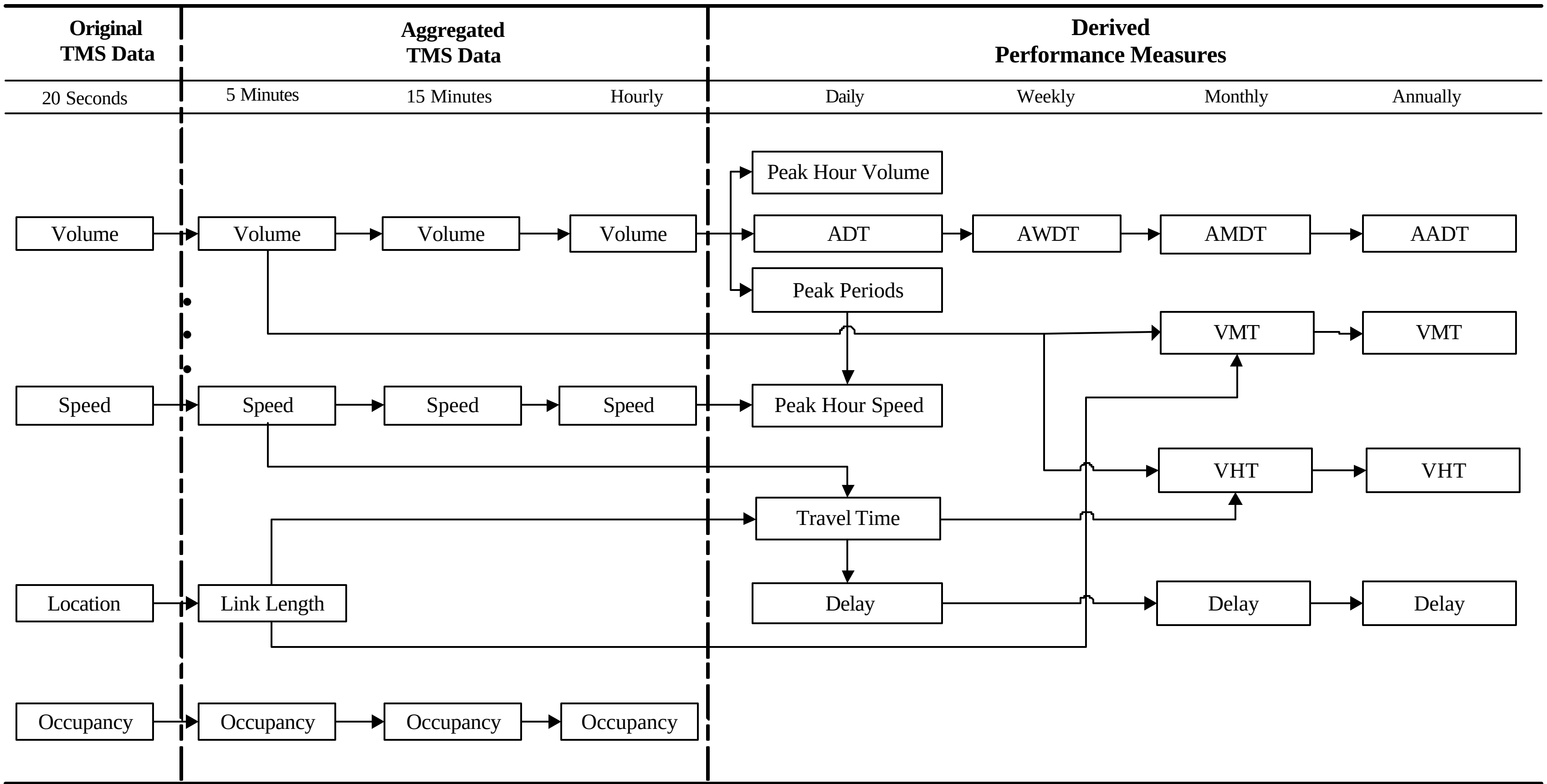


Figure 5.1 Data Flow of Performance Measures Calculation

TMS output is an average of 20-second volume, speed, and occupancy measures. Occupancy describes traffic density and often is used for short-term traffic pattern analysis. Volume and speed are the most commonly used traffic characteristics. They are measured at intervals varying from 20 seconds to one year. These intervals depend on anticipated use of the data. 20-second data can be aggregated to find data for different time intervals. Traffic volumes for Average Annual Daily Traffic (AADT), Average Daily Traffic (ADT), and Peak Hour Volume (PHV) are found by aggregating and averaging. However, certain transformations are needed to obtain average speeds.

TMS data is collected by lane. Therefore, weighting factors should be considered when determining the average speed at a particular point in all lanes. Weighting factors are based on the volume in each lane. The lane with a higher traffic volume is given a heavier weight. Equation 5.1 is used to calculate average speed at a specific site.

$$V^i = \sum_{m=1}^{n} (F^i_{Dm} V^i_{Dm}) / \sum_{m=1}^{n} F^i_{Dm} \qquad \text{Equation 5.1}$$

Where:

V^i = Weighted average speed at the *ith* TMS site for the specified period.

V^i_{Dm} = Average speed at the *mth* detector of the *ith* TMS site for the specified period.

F^i_{Dm} = Total volume at the *mth* detector of the *ith* TMS site for the specified period.

n = Number of detectors at the *ith* TMS site.

Similarly, when hourly or peak period speed is computed, the weighting process should adjust the result. Equation 5.2 shows that weight is the ratio of total volume in time of *t* to total volume in time of *T*.

$$V^i_T = \sum_{k=1}^{n} (F^i_{tk} V^i_{tk}) / \sum_{k=1}^{n} F^i_{tk} \qquad \text{Equation 5.2}$$

Where:

V^i_T = Weighted average speed at the *ith* TMS site for the specified period *T.*

V^i_{tk} = Average speed at the *ith* TMS site for the specified period *t.*

F^i_{tk} = Total volume at the *ith* TMS site for the specified period *t.*

n = The number of t intervals included in the T.

$$T = \sum_{n=1}^{n} t \; .$$

5.2.2 Travel Time

Travel time is the time it takes to travel along a particular segment of a corridor or road. It is computed with the average speed traveled on a segment and the distance between two points. Travel time and its variability measure service quality offered by the transportation system.

Travel time over a link can be computed precisely using real time speed when vehicles are traveling at a particular speed. An average five-minute speed typically is used to estimate travel time. Travel time $T_i(t)$ at time *t* over the *ith* segment is computed as follows:

$$T_i(t) = l_i / V_i(t)$$

Where:

$V_i(t) =$ average speed in a five-minute interval at the *ith* TMS at time *t* when vehicles travel over the *ith* segment.

$l_i =$ the length of the *ith* segment holding the *ith* TMS, which can be derived from the adjacent TMSs' locations marked by milepost value.

Assuming $x_1, x_2, ... x_n$ as locations of *n* TMSs on a directional roadway, l_i is calculated as follows:

$$l_i = (x_{i+1} - x_{i-1}) / 2 \qquad \text{Equation 5.3}$$

The length of the first and last segments are:

$$l_1 = (x_2 - x_1), \qquad l_n = (x_n - x_{n-1})$$

Equation 5.4 shows that travel times are aggregated over a set of links to find the total travel time *T* for an entire or specific section of a route.

$$T = \sum l_i / V_i(t) \qquad \text{Equation 5.4}$$

5.2.3 Delay

Delay is the difference between actual travel time and travel time assuming that vehicles are traveling at free-flow speed on the section being studied [10]. Delay $D_i(t)$ on a link during a five-minute period can be computed as follows.

$$D_i(t) = l_i \times F_i(t) \times [1/V_i(t) - 1/ f_i]$$

Where, $F_i(t)$ is the total volume at the *ith* TMS site for the specified period *t* and f_i is the free-flow speed at the *ith* segment. Equation 5.5 calculates travel time delays over a set of individual links. It represents the total delay on some routes or in a freeway system.

$$D = \sum l_i \times F_i(t) \times [1/V_i(t) - 1/ f_i] \qquad \text{Equation 5.5}$$

5.2.4 VHT and VMT

Vehicles Hours Traveled (VHT) and Vehicles Miles Traveled (VMT) describe the efficiency and productivity of a transportation system. VHT measures the total hours traveled by all vehicles on specific routes or links during a given time period. VMT is the number of vehicle-miles traversed in a given amount of time. Using average five-minute traffic facts, VMT and VHT for each link is calculated as follows:

$$VMT_i = l_i \times F_i , \quad VHT_i = l_i \times F_i / V_i$$

VMT and VHT can be added for any set of links (e.g., one corridor or a whole freeway system) and for any time interval.

5.2.5 Travel Time Index

TTI is the ratio of peak period travel time to free-flow travel time [11]. It represents the ease of getting to a destination. A TTI of 1.3 indicates that a 10-minute off-peak trip will take 13 minutes during peak traffic. A higher TTI value means the system is more congested. Equation 5.6 shows how TTI is calculated. TTI can measure either chosen corridors or a whole system, depending on the range of road links aggregated.

$$TTI = \sum l_i / V_i(t) \Big/ \sum l_i / f_i \qquad \text{Equation 5.6}$$

5.2.6 Variability and Reliability

Traffic demand variation, crashes, and other irregular events influence travel reliability. Variability and reliability usually are measured in terms of travel time. Reliable travel time increases predictability of a trip. Highly variable travel time indicates a low degree of reliability, or an unpredictable travel time.

ITS data is collected continuously and is comprehensive. Therefore, it provides consistent and accurate information regarding variability and reliability of traffic conditions. TMS data allows statistical analysis of traffic variability and reliability.

Standard Deviation (SDEV) is the difference between a given set of numbers. It can assess variability in travel time. Equation 5.7 shows how to determine variability based on historical travel data. A large number of trips were recorded to obtain the necessary inputs for Equation 5.7. This equation calculates variability of a specific route during a particular period. Higher standard deviation means higher variability. This results in a more unpredictable travel time.

$$s^2 = \sum (T_i - M)^2 \Big/ n-1 \qquad \text{Equation 5.7}$$

Where:

s = the estimate of travel time standard deviation.
T_i = the travel time of the *ith* travel crossing a specific route.
M = the mean travel time of a set of samples.
n = the number of sampling travels.

Level of confidence can quantify travel time reliability. The 95 percent confidence region contains 95 percent of the historical travel time data of a representative sample of the population. So, based on past trips, a traveler could obtain the suggested travel time with a 95 percent probability of actually arriving within the pre-planned time. For example, it takes an average of 13 minutes for a commuter to travel from Sandy to downtown Salt Lake City. However, due to congestion, this commuter must allow 20 minutes of travel time if he or she wants to arrive at his or her office on time 95 percent of the time. With increased traffic demand, the 20 minutes previously needed would likely only guarantee an 80 percent possibility of on-time arrival which results in lower travel time reliability.

Figure 5.3 shows the process for calculating traffic variability and reliability. TMS data provide basic inputs for this algorithm. Travel time can be derived from TMS speed data and location information. Users also must provide the expected number of trips, n, which should occur during the same traffic periods.

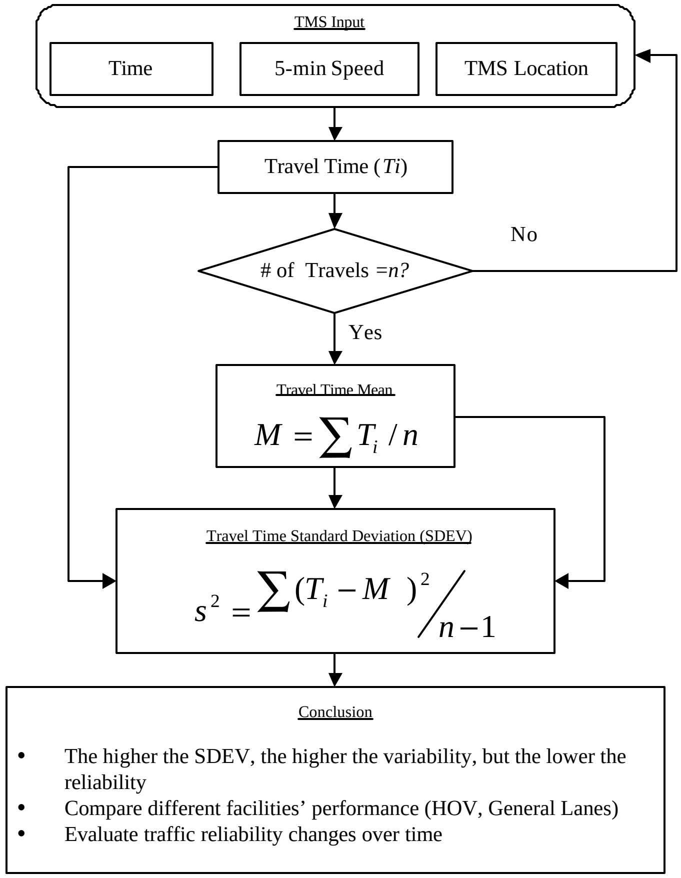


Figure 5.2 Traffic Variability and Reliability Algorithm

6. TMS DATA ARCHIVING SYSTEM ARCHITECTURE

This chapter describes the architecture of an automated TMS data archiving and performance measuring system. This architecture provides guidance for archiving, analyzing, and disseminating TMS data. Several capabilities are incorporated into the system's design:

- Capability to store and manage large amounts of data.
- Capability to aggregate data, derive performance measures at different spatial and temporal levels, and present them in tabular and graphical formats to users.
- A user-friendly point-and-click query interface to acquire measures of interest.

The following section considers the above features and documents the development of a TMS data archiving system used to archive, analyze, and present data from TMSs. This information can guide researchers to develop a prototype version of a TMS data management system.

6.1 Conceptual Architecture of TMS Data Archiving System

Figure 6.1 shows the six components of the proposed TMS system: data collection, data transmission interface, data aggregation and quality control, off-line data storage, database management system, and user interface.

TMSs provide the main source of archived data. They use double inductive loop detectors to measure traffic volume, speed, and occupancy across all lanes and on on-ramps and off-ramps at hundreds of freeway locations. The system receives real time data every 20 seconds from TMSs over UDOT's Local Area Network.

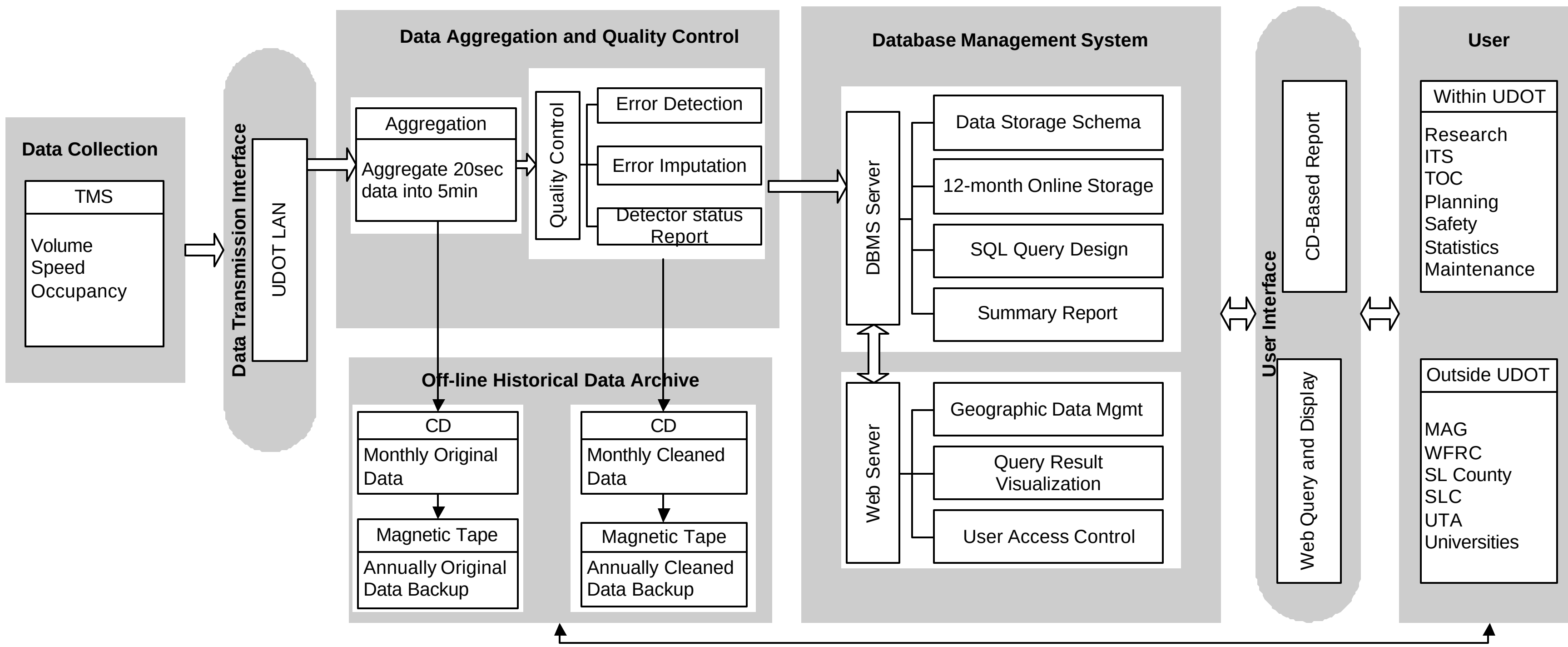


Figure 6.1 Logical Architecture of TMS Data Archiving System

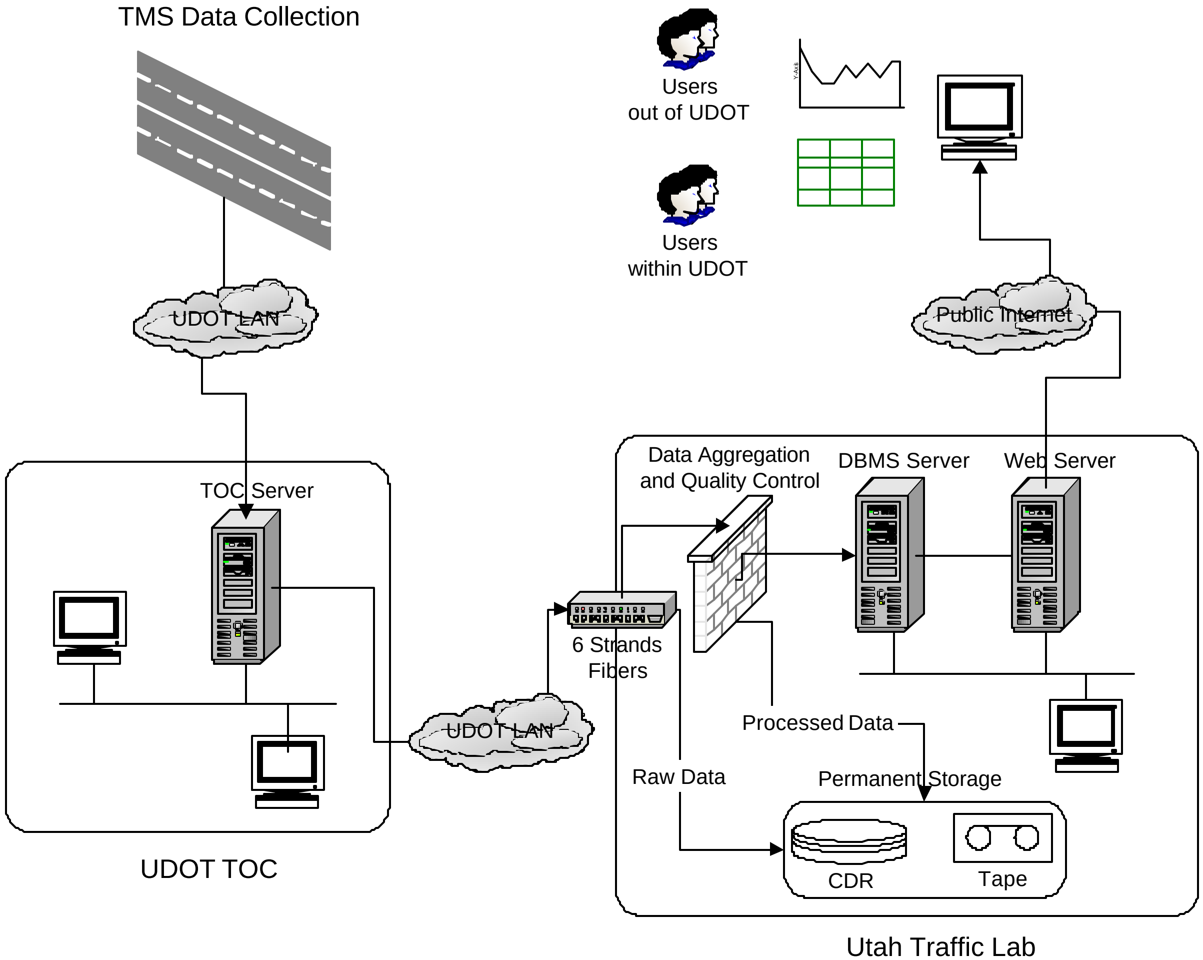


Figure 6.2 Physical Architecture of TMS Data Archiving System

The data processing block in Figure 6.1 is comprised of data aggregation and quality control. The system aggregates the received data into a level that can reduce data flow size without sacrificing user needs. To assist TMS inspection and maintenance, data quality control detects erroneous data. It also imputes missing data and produces a detector status report.

The system also provides on-line and off-line storage of raw and processed data. Original five-minute data must be permanently preserved off-line in the historical data archive. After the data quality control procedure, processed data also should be copied off-line. In addition, data processed in the last 12 months are kept in a database management system for on-line queries.

The archived data management system is a key subsystem for this PeMS. It implements all functions that provide data users with different types of data products. It includes a Database Management System (DBMS) server and a web server. Users can interact with DBMS by inputting queries on Web pages without acquiring DMBS knowledge.

Figure 6.2 demonstrates the structure of this system. The Utah Traffic Lab (UTL) is proposed to hold the TMS data archiving system because UTL has advanced communication infrastructures, is integrated with UDOT's network, and has strong expertise in database management. The functions of this system, such as data aggregation and quality control, data storage, database management, and data distribution are implemented by UTL.

6.2 Data Aggregation

The enormous volumes of TMS data require innovative storage and aggregation strategies. Currently at UDOT, TMS data coming from detectors at 20-second intervals is aggregated at 15-minute intervals and archived as ASCII-CSV files on a FTP server. However, users from different agencies require data at different intervals, ranging from five minutes to one hour. To meet data requirements for multiple applications we recommend that 20-second raw data be aggregated into five-minute intervals rather than the current 15-minute intervals. The archiving system should also be able to summarize five-minute data into various time periods, ranging from fifteen minutes to one hour to one day. This allows the database to serve a number of different purposes. These may include real time point analysis, corridor congestion monitoring, and long-term planning applications.

Data aggregated at five-minute intervals is reasonably detailed at 12 data intervals per hour. In addition, the size of five-minute data measurements is manageable in DBMS. Table 6.1 shows the required storage size for different aggregation levels. For example, each day of data measured at five-minute intervals requires approximately 33.6 megabytes (MB) of storage space. A month of the same aggregated data requires 1.0 gigabytes (GB). A full year of data requires 12.3 GB. The constant flow of large amounts of TMS data requires an efficient data storage strategy.

Table 6.1 Comparison of Data Size at Different Storage Levels and Times

Time \ Level	20 seconds	5 Minutes	15 Minutes	60 Minutes
Day	504 MB	33.6 MB	11.2 MB	2.8 MB
Month	15.1 GB	1.0 GB	336 MB	84 MB
Year	184.0 GB	12.3 GB	4.1 GB	1.02 GB

6.3 Data Storage

TMS data can be stored either as original 20-second data or as aggregated data. Table 6.1 shows that a full year of 20-second data is about 184.0 GB. It is not economical to save such a great amount of data. Because the previous analysis proved that five-minute data could provide a reasonable basis for future analysis, it is cost-effective to archive aggregated five-minute TMS data.

Figure 6.1 shows that data storage encompasses both on-line and off-line storage of raw and processed data.

Off-line Storage

The off-line Historical Data Archive is comprised of both raw data and processed data. Original data stored in the master archive should not be modified as a result of user-specified data requests or data manipulation. Data should only be manipulated (e.g., editing, formatting, aggregation, cleaning, or fusion of data) from copies of the master archives. Cleaned or otherwise transformed data can co-exist with the original data in the Historical Data Archive.

Table 6.2 describes the capacity, cost, and durability features of CDR and magnetic tape. Both are recommended as cost-effective data storage devices for the Historical Data Archive. Table 6.1 shows the size of aggregated five-minute data. Magnetic tape backs up yearly data.

Table 6.2 Storage Media Features

Media Type	Capacity (GB)	Unit Cost ($/GB)	Applications
CDR	0.65	0.39	Data and application distribution, PC desktop storage
Magnetic Tape	70	1.78	Permanent data archival or storage for PC desktop or enterprise system

On-line Storage

Most archived data are distributed on-line through web queries. On-line data storage provides flexibility for data analysis. Three types of on-line data are designed for storage in DBMS: current 20-second raw data, the last 12 months of five-minute data, and hourly data from previous years. Twenty-second current-day raw data provides a real picture of data received from TMSs. It can be compared with cleaned data to verify data integrity. Five-minute data can satisfy different data needs. Current DBMS restricts the capacity to store an unlimited amount of data. Therefore, long-term historical data is only stored in a 60-minute level. This reduces the amount of data stored on-line. Users usually require detailed data for the most recent time periods. Therefore, this storage strategy balances data needs and DBMS limitations. Furthermore, storage of long-term hourly data also can improve the efficiency of data analysis. It can be processed from 60-minute data instead of five-minute data, thus reducing data size.

On-line data storage also consists of metadata. This data provides analysts with an indication of collection and sampling conditions, variability, quality control procedures, edits, and transformations. These features should promote careful use of data and help analysts understand the nature of the data. Any data edited after the data quality control procedure must be flagged.

6.4 Database Management System

Choosing DBMS

The total size of on-line data stored in a DBMS is beyond 13.8 GB. Therefore, it is unlikely that current, medium-sized desktop database applications, such as Microsoft Access, can effectively store and manage data.

Many large businesses and corporations use "enterprise-class" relational databases on computer workstations to manage data requiring more than five GB. These large databases are commonly referred to as "data warehouses" or "data marts." A data warehouse is a "separate data store in which the data is stored in a format suitable for business intelligence and decision support systems, in which these systems do not interfere with the performance requirements of operational systems" [20]. A data mart is a scaled-down version of a data warehouse and typically contains between five and 15 GB of data. A data warehouse typically contains more than 30 GB of data. Several software developers (e.g., SQL Server, Oracle, Sybase, Informix, etc.) are currently marketing products for data mart and data warehouse management. A database product supporting data warehouse functions is recommended during the development of this PeMS.

Conceptual Schema of Database

The conceptual schema of any database must be established using a high-level data model. In the past, the Entity-Relationship (ER) model was used for most relational database designs. This model provides detailed descriptions of data entity types, attributes, relationships, and constraints. The ER model eliminates data redundancy by creating more tables. During the normalization process the database schema is grouped by attributes that satisfy certain functional properties.

In a data archiving system, the ER model plays a limited role. Data warehousing technology provides a new data archiving schema. In a data warehouse, data update is rare and many query variations are expected. Multi-dimensional models arrange data in fact and dimension tables. Database normalization is less restricted in the multi-dimensional mode and provides fast access to data. Variations of multi-dimensional schemas describe more complex data structures and support different data granularities.

Figure 6.3 tracks time, date, and location dimensions for TMS data. Volume, speed, and occupancy data are stored in the traffic fact table. The fact data is expected to grow as new TMS data is received. Each record in the fact table is linked to each of the date, time, and location dimensions. This allows cross-subject queries and analysis.

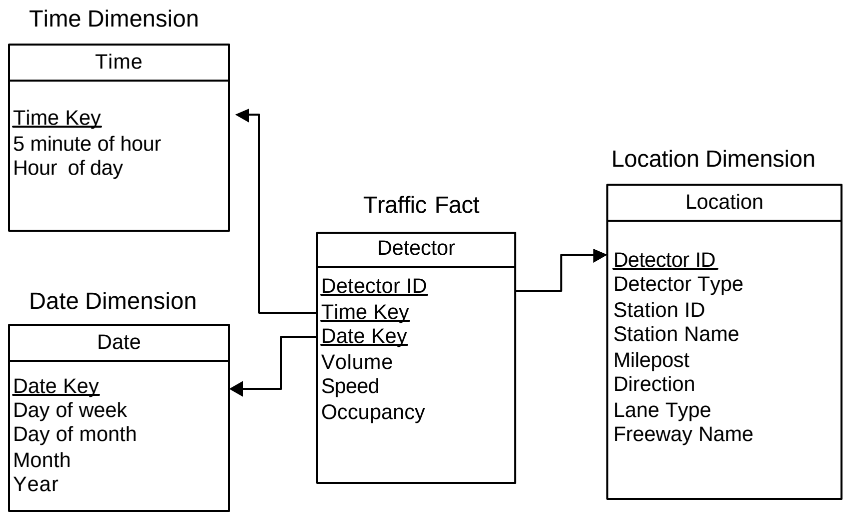


Figure 6.3 Multi-dimensional Implementation of TMS data

Time and date dimensions, in conjunction with the traffic fact table, can be used to analyze traffic on different temporal and spatial scales. This method allows such queries as average speed on the I-15 corridor on Friday and traffic volumes on the HOV lanes during different traffic peak periods. Similarly, the location dimension is associated with analysis of traffic at varying spatial scales. Possible queries include average volume at a certain TMS site or at a traffic bottleneck on I-15 where TMS sites frequently show the lowest speeds.

6.5 User Interface

The PeMS user interface enables novice database users to perform simple and complex data queries. The interface is fairly simple and user-friendly. Web browsers act as interfaces for TMS data archiving systems. Web browsers are increasingly being used to perform complex tasks over the Internet. Therefore, even novice computer users generally have some experience with them.

Figure 6.4 shows a prototype of visually archived TMS data. All functions in this prototype can be implemented with available web-based software.

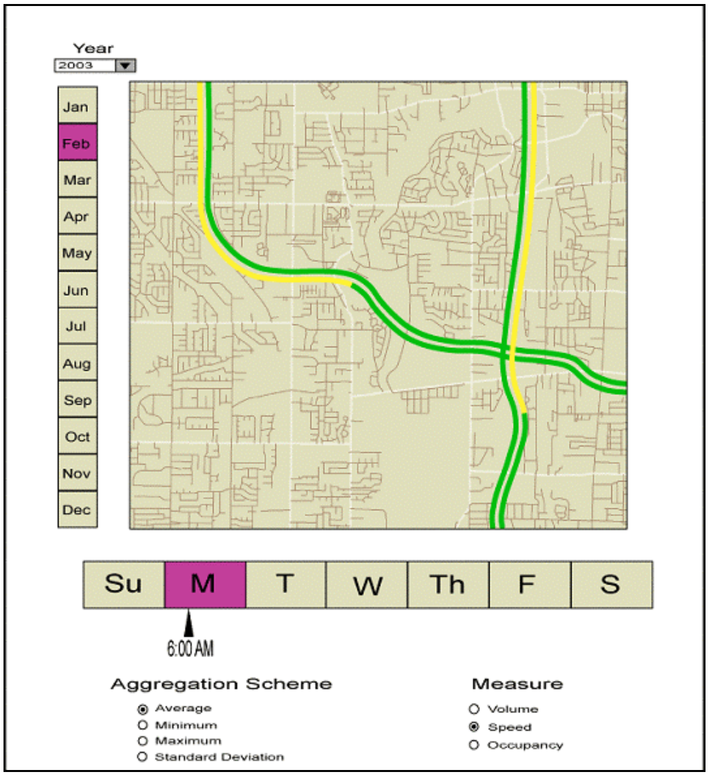


Figure 6.4 User Interface 1

The main frame of Figure 5.4 shows a zoomed-in portion of the freeway system. The combo box in the upper left corner allows year selection. The column of buttons on the left allows month selection. The bottom row of buttons allows selection of the day of the week. The arrow below "M" is a slider used to select a time of day. The radio buttons on the bottom select the aggregation scheme and the measure of interest. The mainframe is updated according to the selected parameters. This figure shows the average speed for Mondays in February 2003 at 6:00 AM.

The map display can be implemented using Scalable Vector Graphics (SVG). It is possible to view SVG documents with any web browser by downloading a SVG viewer. These viewers are available free of charge from Adobe and are approximately 2.25 MB. It is anticipated that the next generation of web browsers will contain support for SVG.

If a wide audience will not view the map, a customized interface can be developed with ArcGIS software. This interface can be installed on any computer on which ArcGIS is installed.

In Figure 6.5, a traffic profile window has been added to the map with "choose link" and "show parabola" buttons. The dashed line represents the selected link. The traffic profile represents measurements on the link across the day selected. These measurements could refer to volume, speed, or occupancy, depending on the interests of users. Figure 6.5 shows volume selected as the measure of interest.

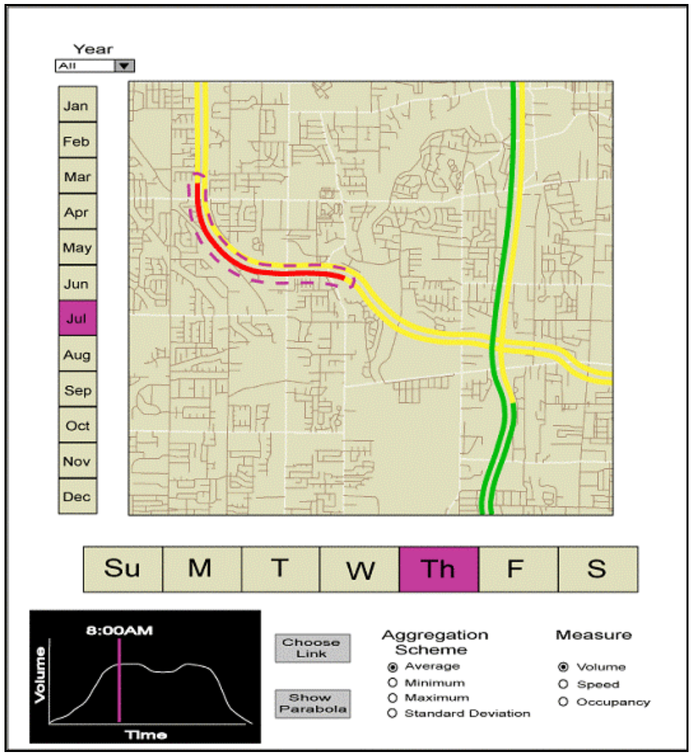


Figure 6.5 User Interface 2

In Figure 6.6, the "Show parabola" button is selected and another window has popped up with a traffic flow parabola. The light gray bar in the flow profile, when selected, allows users to view values in animation. The bar also can act as a slider control. When it is moved, the color attribute of the selected link and the position of the marker on the parabola change according to data retrieved for that particular time of day.

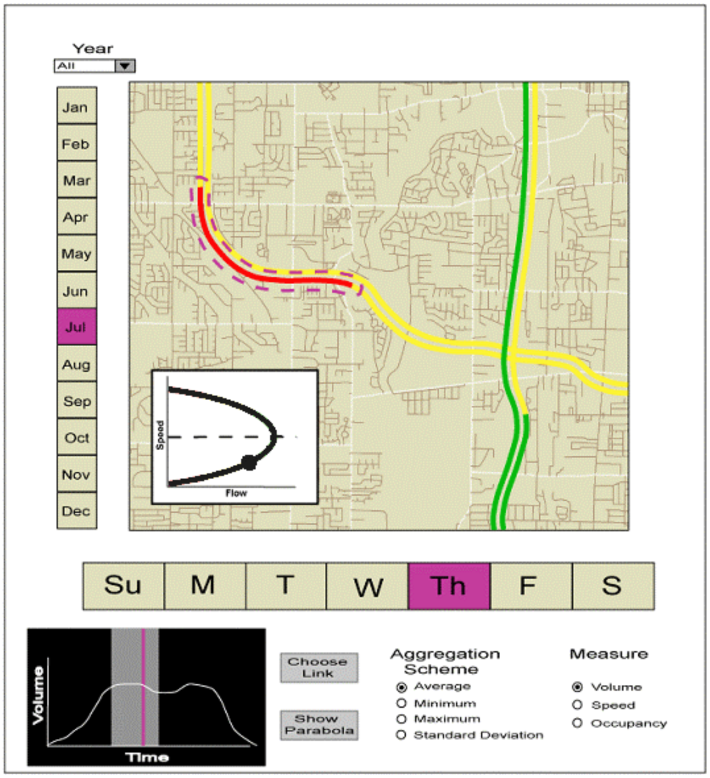


Figure 6.6 User Interface 3

The parabola and histogram tools could be implemented in SVG, Java, or in other programs.

6.6 Software/Hardware Requirements

The proposed TMS data management system consists of five major components:

- Windows 2000 advanced server
- SQL Server 2000 DBMS
- JavaServer Pages (JSP)
- Apache web server
- Web GIS package

The operating system for the client (Web Browser) must provide a graphical user interface, have strong industry support, promise longevity, and offer affordable licensing. The most probable client operating system candidate is Microsoft Windows Operating Systems (2000, XP).

The SQL Server 2000 DBMS or relational database is the key component of the system and stores the TMS data. The database also provides basic statistical and aggregation functions. External programs or scripts can be accessed via JSP. This is an Internet method of calling one program from another. JSP increases aggregation, statistical, and data manipulation capabilities beyond those available in the SQL Server 2000 software. Programs called from the JSP gateway are at the heart of the system as they provide links to all system elements.

Apache web server is a public domain Unix-based software package and is considered the world's most popular and secure web server. Various server-side applications can be developed by JSP on the web server to process requests from web users. Through the web server, users can easily communicate with DBMS. In the proposed system's architecture, one machine is dedicated to the database while the other handles web server functions.

Many commercial products integrate DBMS and geographic data to enable spatial queries and analysis. ESRI Internet Map Server provides a web-based map interface for transportation data.

7. DATA QUALITY CONTROL

ITS data is problematic. Invalid data do not accurately represent road conditions, but still are used to draw conclusions. According to nationwide research and practice review, data quality is one of the primary concerns of many archived ITS data users. Quality control procedures are especially critical to ITS data for several reasons:

- The large potential volume of ITS data makes it difficult to detect errors using traditional manual techniques.
- The continuous monitoring nature of ITS data implies that equipment errors and malfunctions are more likely to occur during operation than during periodic data collection efforts.
- Archived data users may have different (potentially more stringent) quality requirements than real time users of the same data.
-

However, little is known about how to identify and adjust questionable ITS data received from field equipment. An automated data analysis method must be built into the ITS data archiving system.

Monitoring speed, volume, and occupancy on freeways provides ITS data and facilitates transportation analysis. The data quality discussed in this research focuses on data received from TMS sites. This study recommends a comprehensive method for controlling the quality of archived TMS data. Figure 7.1 shows the procedure for performing data quality control.

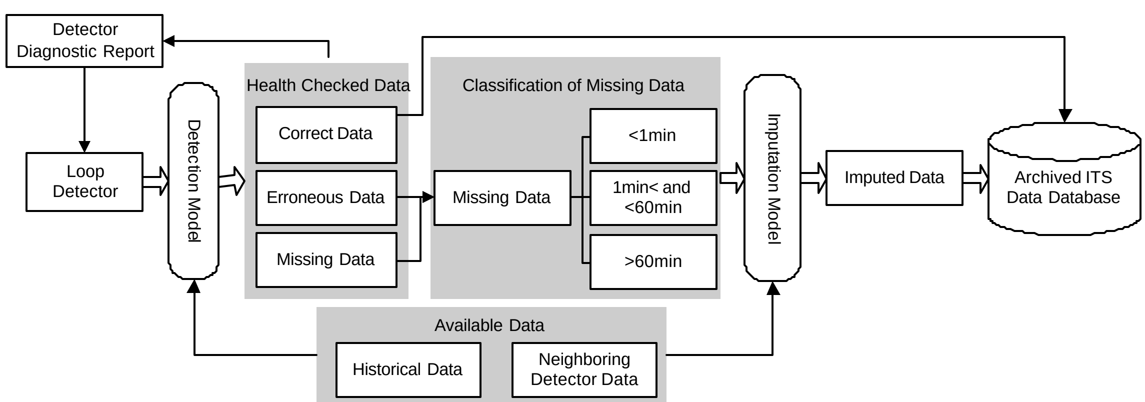


Figure 7.1 TMS Data Detection and Imputation Procedure

A data error-detecting model is applied in this procedure to detect possible TMS data errors. An imputation model recovers data gaps caused by erroneous and missing data. The complete set of data includes correct data and imputed data with flags. It is transferred into DBMS for on-line uses. The procedure produces detector diagnostic report that can be used for detector maintenance.

7.1 Error Patterns

Error in loop data systems prevents their effective use. Researchers define three types of data errors: suspect data, erroneous data, and missing data. Figure 7.2 illustrates an example of data error patterns occurring with TMS data. Figure 7.2 shows the data error types at Interstate 15 Northbound at 5800 South on February 24, 2002.

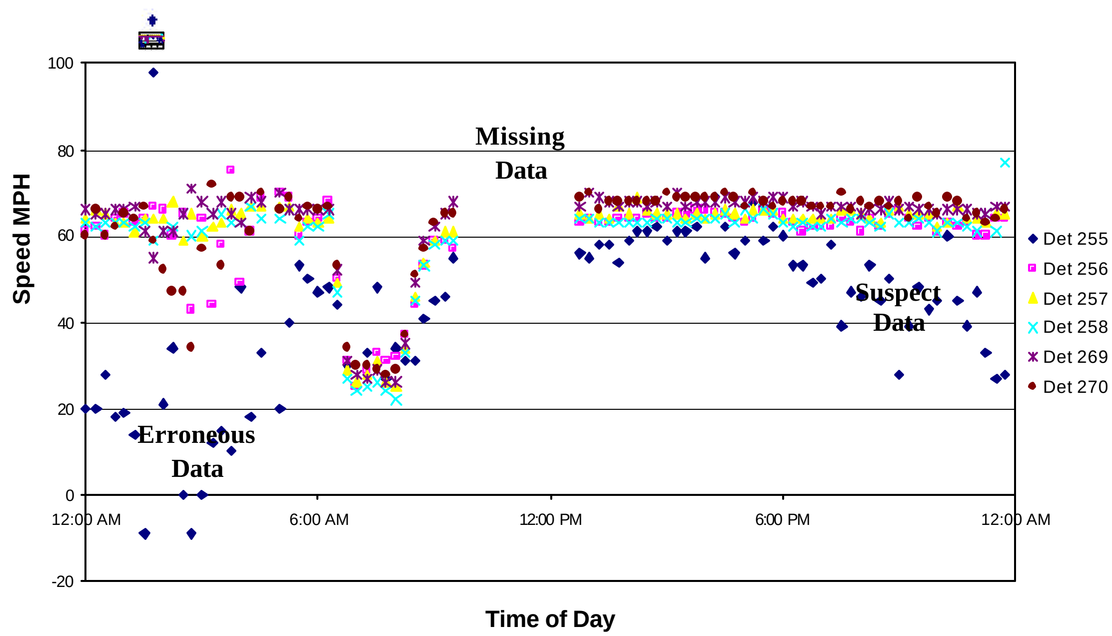


Figure 7.2 TMS Data Patterns Example

Erroneous data consists of unrealistic speeds, such as negative speeds and speeds beyond 100 miles per hour. Data gaps caused by detector malfunction or communication failure are defined as missing data. Suspect data are those that do not reflect the expected outcome. A comprehensive procedure that checks data quality, identifies and corrects erroneous data, and imputes missing data should be performed in the TMS data archiving system.

7.2 Error Detection

A data health check procedure determines that TMS data is accurate and complete. This procedure identifies errors that are unacceptable according to certain thresholds. Table 7.1 shows four error types.

Table 7.1 Data Error Types

Error Type	Description
1	Negative value (data error code)
2	Unrealistic speed or volume
3	Constant non-negative value
4	Missing Data (data not reported)

Figure 7.3 shows data health-checking results by day and different error types for 15-minute TMS data from April 1 through April 30, 2003. "Good data" has passed data error screening. These data are not necessarily correct, but have a high possibility of being accurate. On April 22 and April 23, a significant part of data was unreported. However, all other days show a constant percentage of erroneous data for each data error type. This indicates that data errors did not occur randomly. Some stations constantly reported unreliable traffic counts. Therefore, routine TMS inspection and maintenance should be performed to improve the data quality.

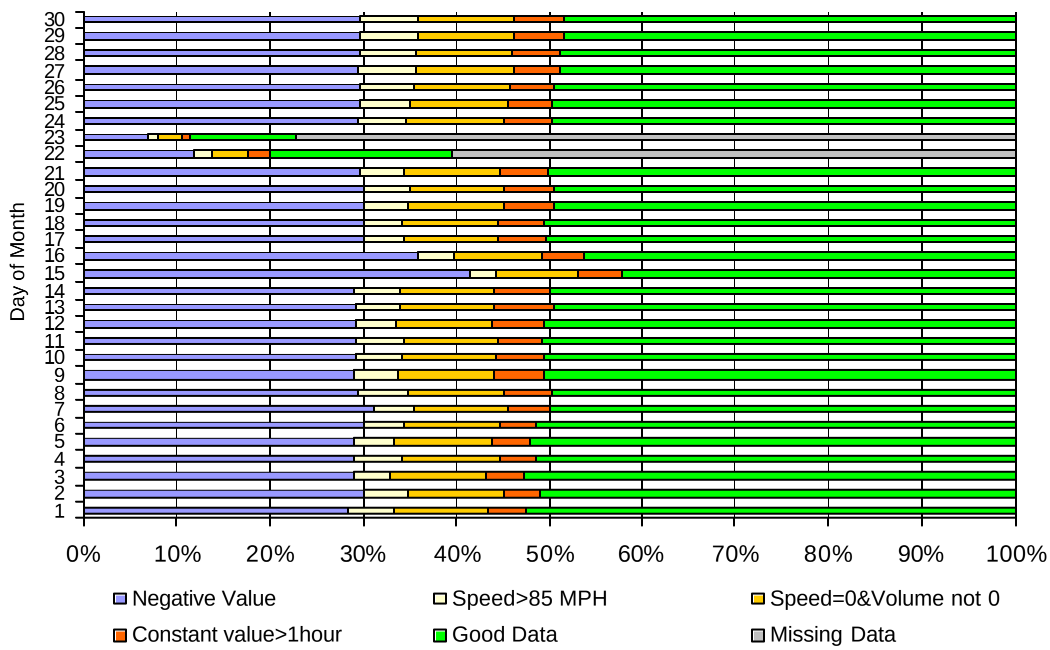


Figure 7.2 Variation of TMS Data Error Distribution

Statistics in Table 7.2 are based on recent 15-minute data throughout from April 2003. They give the current TMS data health status. Data error usually results from missing data, including unreported data and data coded by "-9." Communication and software failures and data malfunction can account for this error. Local detector controllers do not often perform data health checks. In the case of detector malfunction, they return records coded by "-9." Negative data is the primary type of data error. It comprises 28.8 percent of total data, because there are a significant number of malfunctions with TMS detectors. Unrealistic data errors are a result of inaccurate speed detection. Several stations produced inaccurate speed measures. Some inconsistent data also exist, such as a speed of zero, while volume is not zero. However, this case could occur when traffic is jammed due to congestion or accidents. This erroneous situation recurred on specific TMSs. Therefore, these TMSs were reporting unrealistic values. In addition, some TMSs constantly reported the same records. Speeds and volumes of zero represented 4.7 percent of total records.

Table 7.2 Statistical Analysis of TMS Data Errors

Error Types	Description	Possible Causes	Percentage of Errors
Missing Data	Missing time intervals	Communication Failure Malfunction, Installation	4.8%
Negative Value	Volume & Speed = -9	Malfunction Communication Failure	28.8%
Unrealistic Value	Avg. speed > 85 MPH	Calibration	4.6%
	Speed=0 but Volume not 0	Malfunction	9.7%
Constant Value	Constant value in a long time (constant 0 records)	Malfucntion	4.7%

Statistics based on 15-minute data from April 1 to April 30, 2003.

Table 7.3 classifies the performance of TMS sites. Only 31.8 percent of TMSs provided reliable data. The communication lines at some metered on-ramps perform the signal communications instead of carrying data from nearby TMS sites back to the TOC. This causes missing data at those sites.

Table 7.3 Performance Classification of TMS

Classificaiton of TMS		Percentage of # TMS
Performance	Error Range	
Reliable	0%-10%	31.8%
Questionable	10%-30%	8.9%
Malfunction	30%-70%	9.7%
Bad	70%-100%	49.6%

7.3 Error Imputation

Erroneous and missing data must be imputed. Cumulative statistics, such as VMT and VHT, are most affected by missing data. Measures, such as delay and travel time, are mistakenly reflected due to inaccurate speeds. The imputing procedure restores gaps caused by bad data. We recommend the models in Table 7.4 for imputing invalid data.

Table 7.4 Suggested TMS Data Error Imputation Model

Data Gap	Method Description
Data Gap<1min	Average of neighboring time data
1min<Gap<60mins	Primary: Weighted neighboring location detectors' data Secondary: Average of neighboring time data
Data Gap>60mins	Primary: Weighted neighboring location detectors' data Secondary: Time Series Secondary: Factor Approaches

Factor Approaches

Factor approaches may be the most popular data imputation and prediction methods. They develop a set of factors from an historical data set and apply these factors to new data. This allows for predictions. For example, a set of hourly, daily, and monthly factors can be developed based on data from reliable traffic counts. Missing traffic data then can be predicted by applying these factors to short-period traffic counts. These methods are simple and efficient.

Time Series Analysis

A time series is a chronological sequence of observations of a particular variable. Time series data are often examined to discover a pattern for inferring missing data. It assumes that historical value of a variable indicates its value in the future. Many techniques model univariate time series, such as exponential smoothing, Holt-Winters procedure, and Box-Jenkins procedure. Time series analysis methods utilize the correlation between historical data and missing current data.

Using Neighboring Detector Data

Figure 7.4 shows correlations among speeds, occupancies, and volumes of detectors in nearby locations. Therefore, measurements from one location can be used to estimate quantities at other locations. A more accurate estimate can be formed if all data from the neighboring loops is used in the estimation. This model uses current neighboring detector data. Therefore, good neighboring detectors must be in use. Secondary methods must be used when neighboring detectors are inadequate.

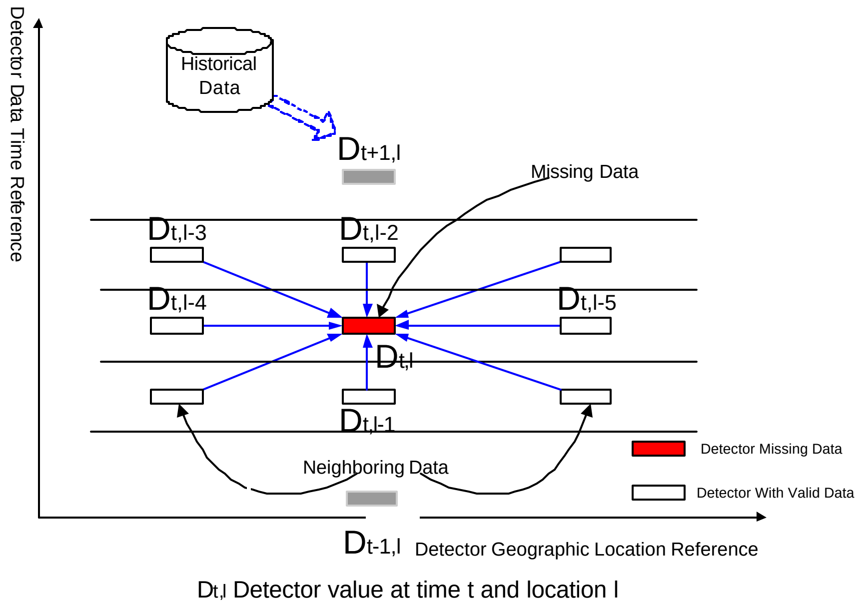


Figure 7.4 Illustration of Neighboring Detector

For a specific detector at time *t*, neighboring detector data includes neighboring time data and neighboring location data. These two types of data can be applied to two different imputation models. S traffic situation remains relatively stable over a short time period due to the continuity of traffic flow. Therefore, a short time period of missing traffic data can be substituted with an average of the same detector's data before and after the missing period. This equation is used to infer missing data:

$$D_{t,l} = (D_{t-1,l} + D_{t+1,1})/2$$

Where:

$D_{t,l}$ = Inferred missing detector value at time t and location l.

$D_{t-1,l}$ = Detector value at previous time interval $t-1$ and location l.

$D_{t-1,l}$ = Detector value at following time interval $t+1$ and location l.

Research by California's PeMS found that traffic data at a specific location is linearly correlated to traffic data at its neighboring locations. Therefore, missing detector values can be inferred from data of neighboring locations. This is shown in the following equation. A weighting factor scores how similar a detector value is to that of its neighboring location.

$$D_{t,l} = \sum (w_i * D_{t,l-i})$$

Where:

w_i = Weighting factor that detector $D_{t,l-i}$ contributes to $D_{t,l}$

$D_{t,l-i}$ = Detector value at time t and the ith neighboring detector of detector $D_{t,l}$

8. SYSTEM TRANSFORMATION PLAN

This section outlines the steps for implementing the TMS data archiving system. Many phases and activities overlap due to the project's dynamic nature.

8.1 Recommended Development Methodology

Creation of an efficient TMS data archiving system involves a series of developmental sub-projects. Therefore, an object-oriented development approach is a strategic option. The object-oriented approach builds reusable components. Initial object-oriented construction costs are higher, but returns on the investment follow. With each reuse of a component, quality and reliability improve and the effort required decreases.

8.2 Immediate Next Steps

The project's initial steps should include the following:

- Develop a detailed project management plan and infrastructure. This should include a staffing plan work breakdown structure for building the system and maintaining it.
- Build the project infrastructure by purchasing equipment, setting up space, and assembling the project staff.
- Verify the feasibility of the proposed architecture with a limited functionality prototype or a proof of concept. This proof of concept will verify that the proposed architecture can support transactional loads (100 users trying to access the same page that is loaded on the server) and connection requirements of the process. It will also confirm that the system integrations are achievable. At this point, the proof of concept will not need to demonstrate the system's user-level functionality or gain user acceptance for the interface.
- Begin building the detailed requirements document and functional design documents.

8.3 Step-by-Step Approach

The development effort will adhere to the following phases:

- Project Initiation
- Initiate Technology Evaluation and Selection Sub-Project
- Concept Definition
- Functional Design
- Development and Testing
- Deployment and Distribution
- Continued Development and Maintenance

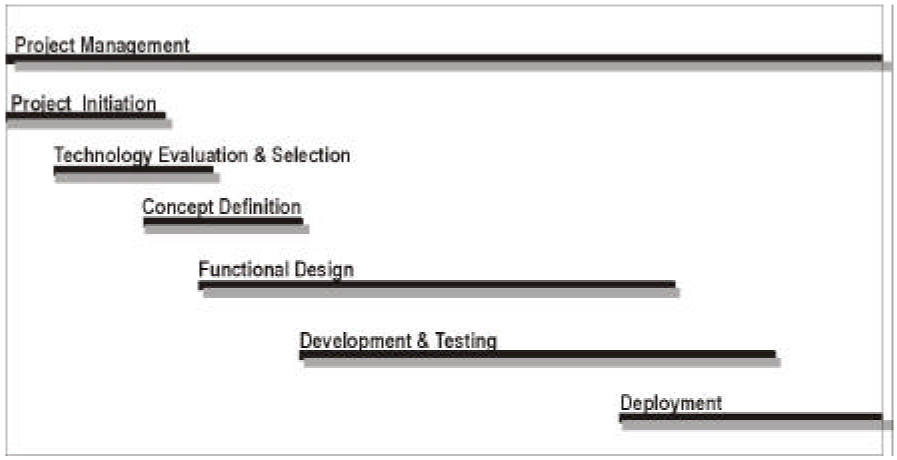


Figure 8.1 Transformation Map

8.3.1 Project Initiation

The initial project phase is administrative and time consuming. However, it sets the stage for future development activity. It establishes a clear plan, identifies who will do the work and how it will be done, and ensures that provisions are made for the necessary physical environments. The high level activities contained in this phase are:

- developing and implementing the project management environment
- documenting the project plan and vision
- documenting and initiating a risk management strategy

Risks could include work by graduate students unfamiliar with recent technologies. The learning curve would mitigate the risk.

- establishing the development process and standards like coding guidelines, etc.
- building a development team
- assembling and testing the Rapid Application Development (RAD) technical environment

8.3.2 Initiate Technology Evaluation and Selection

The Technology Evaluation and Selection phase is a sub-phase in the Concept Definition and Functional Design phases. The technology selected for this phase must meet analytical requirements that the proposed solution places on system integration.

This phase enables a non-standard approach. It is not a comprehensive requirement and design effort. However, it provides information necessary to pursue technology proof of concepts. If the first integration of technologies is inadequate, new selections should be made and tested until the proper integration is achieved. Specific technologies that need to be evaluated and integrated into a proof of concept are:

- Data conversion and loading tools
- Data browsing tools/development environments
- Data repository and management system (DBMS)
- The operating systems for the client and server (Operating systems)
- Hardware required to build the architecture

8.3.3 Concept Definition

Concept Definition lists the requirements for product development, deployment, and performance. It may include the following activities:

- gathering and documenting requirements (from users)
- evaluating technologies
- determining prototype requirements and strategy
- training development staff (graduate students)
- creating the deployment strategy and requirements after the system is built

8.3.4 Functional Design

In the Functional Design phase, system architects conceive and document the functioning of the system technology. They also study the look and feel of the system from a user's perspective. This phase requires significant user involvement. Functional Design activities include:

- designing the system architecture
- designing the program architecture
- creating a user interface design
- performing prototyping between the different layers of the system
- creating a release strategy (releasing parts of the application to the users to test it while it is being built)

8.3.5 Development and Testing

During the development and testing phase, programmers write code and integrate systems. Other critical activities in this phase include documentation, testing, and revision tracking.

The project will consist of at least three concurrent development efforts, one for each layer. The Presentation Layer (Client), the Server Layer (Server and DBMS), and the Transport Layer will be developed in parallel. Sample activities contained in this phase are:

- developing code
- performing revision control
- creating program documentation
- developing training manuals
- testing (unit, systems, integration, and user acceptance)
- releasing certification

8.3.6 Deployment and Distribution

The Deployment and Distribution phase trains users on how to start the new system. A messy project beginning can tarnish an otherwise successful project. Transitioning customers to the new system is challenging, therefore the system is released in phases. Deployment and Distribution involves implementing the training plan and release strategy.

8.3.7 Continuing Development and Maintenance

Rather than performing ad hoc bug fixes, it is more reliable, safe, and cost-effective to plan system releases. Once the system is in place and running, all issues are categorized as follows:

- New Feature Requirements: new business functionality to be incorporated into the system.
- Feature Enhancement: the current features may be incomplete, or a new user requirement may necessitate feature enhancement
- Error: a severe problem graded from one to five (one being highest priority and five being the lowest priority)
- Performance Issues

9. RISKS, RESOURCES, AND ESTIMATED COSTS

This section addresses project logistics such as success factors, risk factors, resources, and estimated costs.

9.1 Critical Success Factors and Risk Factors

Management support and commitment to the project is crucial. Building successful object-oriented RAD projects involves training, planning, personnel development, and software development. Leaders must coordinate these factors. Successful companies have cultivated flexibility as they have learned and changed their procedures to meet the requirements of the new system.

It is difficult for an organization to absorb and adapt to many new techniques, technologies, tools, systems, and processes at the same time. The following risk factors must be addressed in the early stages of project integration:

- building client/server applications for the first time
- hiring new people with skill sets that are new to UTL
- having the new people build object oriented applications
- having existing staff use, for the first time, object-oriented applications, object oriented development, new languages, new operating systems, and new database management systems

9.2 Resources

This section addresses the need for staff, space, equipment, and software for creating the project's infrastructure.

9.2.1 Staffing

New resources must be acquired for graphical user interface development and server application development. The following is a model for the team structure:

- Project Manager (1)
- Architect (1) (Consultant)
- Server Architect/Team Lead (1) (Grad Student)
- Client Architect/Team Lead (1) (Grad Student)
- Server Programmers (2) (Grad Students)
- Client Programmers (2) (Grad Students)
- Human Factors Engineering Expert (1)
- Database Architect (1) (Consultant)
- Tester (1) (Grad Student)
- Technical Writer (1) (Grad Student)

The system requires three staffing approaches:

1. Hire individuals on a contract basis - these people provide specialized services with a narrow application to the development effort. They will handle the intitial work overflow of the project. They will not be needed consistently after implementation. An example of a contracted hire is a human factors engineer.

2. Secure direct hires for positions of key, ongoing importance - direct hires will likely be senior systems architects, visual developers, and technical writers.

3. Train and develop interested existing staff - the existing Data Processing staff fits into this category. Even without a functional role change, the new system will mandate the use of new conversion and data processing tools.

9.2.2 Development Lab (Utah Traffic Lab)

A likely development environment is as follows:

- approximately 300 square feet of office space
- one NT server for active files, source control, and system documents
- one Web Server
- Internet connection
- A DBMS
- compilers and tools required for server application development
- libraries to build client/server interface
- version control system
- e-mail accounts for developers

9.3 Time and Cost Estimates

The sections below provide estimates for project staffing, costs, and time.

9.3.1 Staffing

A project of this level requires an average of three to four full-time developers, one part-time project manager, a DBA as a consultant, and various specialists.

9.3.2 Costs

Labor costs are expected to be between $360,000 and $450,000. The hardware and software costs are not included but are order of magnitude estimates. Order of magnitude estimates generally are based only on experience of the development team and have an accuracy level of –25 percent to +75 percent. A budget estimate is produced after the budgeting activities in the project initiation phase. This estimate has an accuracy of –10 percent to +25 percent. A definitive estimate is produced after completion of the project's initiation phase and a comprehensive work breakdown structure is produced. These estimates have an accuracy of –5 percent to +10 percent.

9.3.3 Time

It is estimated that it will take this team between 12 and 15 months to implement the system and add new functionality.

10. CONCLUSIONS AND RECOMMENDATIONS

10.1 Conclusions

ITSs produce massive amounts of operational data used in real time traffic control strategies. ITS data could have many additional applications in transportation operations, planning, reporting, and research. TMSs are a basic component of Utah's ITS. These stations report traffic volume, speed, and occupancy every 20 seconds. Presently, TMS data lacks an archival and analysis strategy. This keeps it from being widely utilized. However, agencies require this information for multiple purposes.

This research considered nationwide ITS uses to establish guidelines for implementing an automated TMS data collection, archival, and analysis system. Five components of this system were identified and examined: data aggregation, data quality control, data storage, database management system, and user interface. This research particularly investigated the accuracy of TMS data. The study also developed models to derive various performance measures from TMS data. In addition, a prototype was designed to show how the system processes raw TMS data and applies them to diverse applications. In conclusion, this study demonstrated how to archive and better utilize TMS data for specific uses.

10.2 Recommendations

Three options exist for implementing an automated data archiving and analysis system in Utah: developing a new system, expanding UDOT's current DOL software, or introducing California's PeMS to Utah. The strengths and weaknesses of the alternatives should be evaluated before selection in order to determine which one best suits the state's needs.

System Functions

A new system must be developed to meet proposed system requirements and specific data needs. This system would provide the most complete functions to satisfy local user needs. Both DOL and PeMS systems must be customized to perform the desired services.

The existing DOL performs well in traffic delay analysis. However, to meet functions required in the new system enhancements are necessary in data archiving, data quality control, and performance measures.

Although PeMS originally was developed for use by Caltrans, it was also designed to quickly adapt to other specific customer requirements. PeMS reports, options, and user interface can be customized to fit any organization's specific needs for access control, storage, and performance reporting.

Cost

The proposed system's complexity makes it difficult to accurately predict the cost of software development. However, it could be estimated by comparing it to similar systems. PeMS development received full support from Caltrans. It is estimated that millions of dollars were spent on system improvements during the past years. The University of Virginia and the Virginia DOT garnered a one million dollar grant to make vast amounts of archived traffic data available over the Internet. DOL has provided a framework and some functions regarding the data archiving system. However, a significant amount of development is still needed to enhance its performance.

PeMS prices include a one-time installation fee, a monthly operating fee, and an upgrade fee. In addition, system development presents a major cost. However, implementing PeMS costs much less than developing a new system or expanding the current system and provides the same functions that a new system would provide.

Speed of Implementation

If a system development project is aimed at modifying an existing system or creating a new system typically it is divided into the following sequential phases: analysis, design, production, implementation, and operation. This process of development is time-consuming and requires the collaboration of architect teams, analysts, programmers, testers, and users. PeMS took more than five years to develop. In general, it takes two years to create a new, fully functioning system. However, because UDOT already has a sophisticated infrastructure that is compatible with PeMS, PeMS is ready-to-use. From the perspective of users, UDOT can benefit from TMSs in a short time by deploying PeMS.

Expandability

UDOT is continuing to expand its TMSs and other ITS data to include incident data, weather data, and traffic data. Therefore, the potential for system expansion is important. To remain current, the system must be updated constantly. This would be time consuming and costly for a system developed by a team without consistent interest in ITS data archiving. Because a company specializing in PeMS development has been established, the PeMS will continue to evolve using the newest technologies and suggestions from different users.

Risk

PeMS development presents many risks. Technical issues are of primary concern. A number of people with expertise in transportation, statistics, and computer technology contributed to the successful development of PeMS. Collaborative effort is critical as it is difficult for an organization to acquire a diversity of skills. Time and financial cost also are concerns. The PeMS system is complex. Therefore, development of a mature system is costly. Most projects use more resources than planned, take more time to finish, and are less functional and durable than expected. PeMS, however, has been deployed and has been functioning successfully in California. By implementing PeMS, Utah will avoid risks associated with building a new system.

Recommended System Implementation

User expectations of a system influence the type of system they choose. If users think a current DOL system already meets their data needs, then DOL is a cost-effective option. However, if they want a system that can be used in purposes beyond traffic delay analysis they should consider enhancing their DOL system, building a new system, or using California's PeMS system. Table 10.1 summarizes the comparison of different system implementation approaches. PeMS is attractive because of its instant deployment capability, lower cost, and fewer risk. In addition, it provides the same capabilities as other options.

Table 10.1 Comparison of System Implementation Approaches

Comparison Criteria	Build New System	Expand DOL	Install PeMS
Cost	□	□	◪
Speed of Implementation	□	◪	■
System Functions	■	◪	◪
Expandability	◪	□	■
Risk	◪	◪	■
Overall Rating	◪	□	■
Recommendation	**Install PeMS**		

Note: ■ = Excellent; ◪ = Fair; □ = Poor

REFERENCES

1. Guidelines for Developing ITS Data Archiving Systems. Texas Transportation Institute, September 2001.

2. ITS Data Archiving Five-Year Program Description. U.S. Department of Transportation, March 2000.

3. Ralph Gillmann, Status of ITS Traffic Data for HPMS. The FHWA Office of Highway Policy Information, August 2002.

4. Pravin Varaiya, Freeway Performance Measurement System: Final Report. California PATH Working Paper, 2001.

5. ITS Data Archiving: Case Study Analyses of San Antonio TransGuide® Data. Texas Transportation Institute, August 1999.

6. Traffic Data Acquisition and Distribution (TDAD). Washington State Transportation Center, May 2002.

7. ITS Data Archiving: Summary Report. Texas Transportation Institute, 2000.

8. Joy Dahlgren, Renaldo C. Garcia, Completing the Circle: Using Archived Operations Data to Better Link Decisions to Performance. California PATH Research Report, 2001.

9. Lawrence Jesse Glazer, Roberto Cruz, Intelligent Transportation System at the 2002 Salt Lake City Olympic Games. Iteris, Inc., August 2002.

10. Nicholas J. Garber, Lester R. Hoel, Traffic and Highway Engineering, Third Edition.

11. Shrank, D Lomax, The 2001 Urban Mobility Report, Texas Transportation Institute. The Texas A & M University System, May 2001.

12. Rod E. Turochy, Brian L. Smith, Some of the Many Uses of Information Generated From Archived Traffic Data. Virginia Transportation Research Council, 2001.

13. Pravin Varaiya, Freeway Data Collection, Storage, Processing, and Use. Transportation Research Board Workshop: The Roadway INFOstructure, August 2002.

14. Chao Chen, Karl Petty, Freeway Performance Measurement System: Mining Loop Detector Data. The 80th Annual Meeting Transportation Research Board, Washington, D.C., July 2000.

15. Henry X. Liu, Rachel He, A Literature and Best Practices Scan: ITS Data Management and Archiving. University of Wisconsin at Madison, May 2002.

16. Traffic Monitoring Guide. U.S. Department of Transportation, May 1, 2001.

17. P. Hu, R. Goeltz, R. Schmoyer, Proof of Concept of ITS as An Alternative Data Resource: A Demonstration Project of Florida and New York Data. Federal Highway Administration, September 2001.

18. Gopal Krishna Narra, Peter Martin, Traffic Data Collection, Management and Storage. University of Utah, November 2001.

19. Shawn M. Turner, Its Data Management System: Year One Activities. Texas Transportation Institute, August 1997.

20. Haitham Al-Deek, Amr Abd-Elrahman, Evaluation Plan for The Conceptual Design of the Florida Transportation Data Warehouse. Florida Transportation System Institute, March 2002.

Appendix 1: Survey Form

To: Transportation Community and Agencies
From: Joe Perrin, University of Utah
Re: Transportation Data Needs Survey

The University of Utah Traffic Lab is building a blue print for a transportation database that will store and archive data to provide a wealth of information. UDOT's ITS applications, sensors and detectors are a potentially rich source of data about transportation system characteristics and performance. For example, the traffic monitoring stations along the freeways (known as TMS's) provide traffic volumes, speed and density. These devices are located each ½ mile, on each lane on freeways throughout the Salt Lake Valley.

We are investigating how the TMS data can support UDOT divisions and other agencies in their data needs. We need to find out who needs data, and what type: AADT, turning movements, speeds, travel time, vehicle occupancy, etc. We're compiling lists of data types that agencies use for evaluations and making decision. You may need data that cannot be collected, but as technology evolves, so do fresh opportunities. So we need to know your wants as well as your needs.

You are probably tired of surveys and questionnaires, but please spare us 5 minutes and jot down your ideas. You can e-mail the form to peng@trafficlab.utah.edu, or fax it to 801 585 5860. If you could do this by Dec 6th we'll be very grateful!

Thank you in advance. if you have nay questions, please contact Peng Wu at *peng@trafficlab.utah.edu* or (801) 585-5859; or. Joe Perrin at *perrin@civil.utah.edu* or (801) 585-1019.

Sincerely,

Joseph Perrin, PhD, PE, PTOE
Research Assistant Professor
University of Utah Trafficlab

Examples of data needs

Data	**Units**	**Time period**	**Purpose**	**Location**
Speed	mph	5 minute	Daily profiles	Freeways and Arterials
Volume	vehicles	15 minute	Capacity analysis	Arterials
Flow	vph	15-minute	Origin – destination assignments	Major roadways by individual block
AADT	veh	Daily	Planning model input / accident analysis	Roads classified as collectors and higher
Turning movement volumes	veh	Peak period (7-9 AM and 4-6 PM)	Intersection analysis	Major intersections
Please add your needs below				
Data	**Units**	**Time period**	**Purpose**	**Location**

Appendix 2: **Interview Table 1 Input from TOC/ UDOT Joe McBride**

Data	**Units**	**Time period**	**Purpose**	**Location**
Volume	Veh	5 minute	Determining if traffic could be diverted from freeway during and incident. Volume need in real time.	Arterials
Speed	Mph	15 min Average	Freeway contour congestion map	Freeway
Volume	Veh	Hourly	Evaluate capacity during construction projects.	Freeway and arterials
Turning movement volumes	Veh	Peak period (7-9 AM and 4-6 PM)	Intersection analysis and traffic signal timing plans	Major intersections

Appendix 2: Interview Table 2 Input from Planning / UDOT Gary Kuhl

Data	**Units**	**Time period**	**Purpose**	**Location**
Lane distribution	volume	hourly	HOV analysis, Pavement design	
Directional Split	%	AM/PM Peak hour	Capacity analysis	
Peak hour volumes				
Ramp Volumes				
Directional & combined Volumes				
Vehicle classification	% of volume		% trucks, by type or length	

Appendix 2: Interview Table 3 Input from Traffic Safety / UDOT Rob Clayton

Data	**Units**	**Time period**	**Purpose**	**Location**
Pk Hr Vol	Vph	15-min	Safety Studies; Traffic Studies	Collectors and higher
AADT	Veh	Daily	Safety Studies; Traffic Studies	Collectors and higher
AWDT	Veh	Daily	Safety Studies; Traffic Studies	Collectors and higher
Turning movement vols	Vph	15-min	Safety Studies; Traffic Studies	Collectors and higher
Speed	Mph	?	Safety Studies; Traffic Studies	Collectors and higher
Density	Vpm	15-min	Safety Studies; Traffic Studies	Interstate
Vehicle Classification	Veh	1-hour	Safety Studies; Traffic Studies	Collectors and higher
Lane Utilization	Veh/lane	1-hour	Safety Studies; Traffic Studies	Collectors and higher

Appendix 2: Interview Table 4 Input from Traffic Statistics / UDOT Tammy Kaeser

Data	Units	Time period	Purpose	Location
Volume	Veh	15 – minute	AADT calculation and other traffic statistics	All State, Federal-Aid Eligible, or other routes and ramps available
AADT / ADT	Veh	Daily	Traffic Statistic and Reporting Use – if this is an annual average the count must be documented as to where, when, and how it was done and the formula must be documented, factors documented, and they must follow TMG and AASHTO Guidelines – if this is a daily average then a 24 hour count and accompanying data on where, when, and how counted is all that is needed	All State, Federal-Aid Eligible, or other routes and ramps available
Class – Axle	Vehicle Class (FHWA –13 classes)	Hourly	AADT development and reporting needs – count must be documented as to where, when, and how it was done	All State, Federal-Aid Eligible, or other routes and ramps available
Class – Length	Vehicle Class by feet or meters	Hourly	AADT development and reporting needs – count must be documented as to where, when, and how it was done	All State, Federal-Aid Eligible, or other routes and ramps available
Weight – Axle	By Vehicle / tons	Hourly	AADT development and reporting needs – count must be documented as to where, when, and how it was done	All State, Federal-Aid Eligible, or other routes and ramps available

Appendix 2: Interview Table 5 Input from UDOT Blake Hansen

Data: I would like speed, volume, and occupancy.

Units: miles per hours, vehicles (denominator to be provided by the time frame - vehicles / 15 minutes or vehicles / hour), percent of time occupied.

Purpose: it depends on who is asking me for data. Some engineering purposes, some planning purposes, and some other analysis purposes.

Time Frame: I would like to have an adjustable time frame, and one that compiles the data over time.

For example, collect speeds, volumes, and occupancy over 15 minute periods by lane for 1 year. After 1 year, condense these down into 1 hour aggregates by station (not by lane), to manage the size of the database.

Some data may not generally be as "important" as other data (i.e. freeway ramps). Therefore, allow an option to 'turn on' the collection on a ramp to collect for a week or so, export it, then turn it off again. Provide an option to collect data in the smallest quantity possible (i.e. 1 second / 20 second - depending on the device for research purposes :) - but it is not necessary to collect them for all devices in this small of a bite all of the time. when averaging items (such as speed or occupancy), make them weighted averages - and provide the weighting factor for future calculations.

Location: Anywhere there are detectors: Intersections, Arterials, Freeway mainline, freeway ramps.

Appendix 2: Interview Table 6 Input from MAG Chad Worthen

1. Signal Coordination - you will be collecting speed and volumes on arterials. If we had a field for if the arterial had signal coordination, we could track the benefits that signal coordination could have on time, flow, and congestion relief.

2. Incidents - a field showing when/where incidents occur on freeways or major transportation facilities would give us an idea of how traffic patterns change with incidents. This could also help to quantify benefits from ITS type projects.

3. Ramp Metering - information collected on when/where ramp metering is implemented and how it impacts traffic flow would also be helpful.

I have also forwarded your email to Mike Brown & John Britting at WFRC. They are the ones mainly responsible for model development along the Wasatch Front and would be able to give good feedback on information that could be collected/maintained by UDOT's ATMS that would benefit model development.

Chad Worthen
Transportation Engineer
Mountainland Association of Governments

Appendix 2: Interview Table 7 Input from WFRC Wayne Bennion

Data	**Units**	**Time period**	**Purpose**	**Location**
TMS ID			Know location of data	
Average volume	Vehicles	15 minute (weekday, i.e. AWDT)	Validate models, support development of plans and programs	Freeways and Arterials
Average speed	Mph	15 minute (weekday)	Validate models, support development of plans and programs	Freeways and Arterials
Truck percentage	Percent	Hour (weekday)	Validate models, support development of plans and programs	Freeways and Arterials
Truck classification	Category	Hour (weekday)	Validate models, support development of plans and programs	Freeways and Arterials
Truck speed	Mph	Hour (weekday)	Validate models, support development of plans and programs	Freeways and Arterials
Day of week distribution factors for volume & speed	Percent	Hour (weekday)	Support development of plans and programs	Freeways and Arterials
Monthly distribution factors for volume & speed	Percent	Hour (weekday)	Support development of plans and programs	Freeways and Arterials
AADT/ AWDT conversion factors by season	Percent		Air quality analysis	Freeways and Arterials
Turning movement volumes	Vehicles	Peak periods (6 – 9 am & 3 – 6 pm)	Intersection analysis	Arterials intersecting with arterials

Appendix 2: Interview Table 8 Input from SL County Kevyn Smeltzer

I think the data you had listed in your survey form was all we would need. Most of the data we collect now is on residential streets so it would not be of any use in this case.

Appendix 2: Interview Table 9 Input from UTA Richard Hodges

Data	**Units**	**Time period**	**Purpose**	**Location**
Segment speed – selectable by street or road segment – custom definable	Mph/kph	15 min	Travel speed comparisons, schedule adherence investigations, route performance analysis	Arterials and lower
Incidents/accidents	Number and type	Daily	Same as above	All, especially arterials
Volume	Directional vph	Hourly	Same	Arterials and lower
Weather correlations	Events	Daily	Same	Same as above
Vehicle delay	Min by location	15 min	Same	Same as above
Traffic composition	Classified by type	Hourly	Same	Same as above

Appendix 2: Interview Table 10 Input from Research/ UDOT Stan Burns

Data	Units	Time period	Purpose	Location
Volume	vehicles	5 m	Daily volumes/ Peak hr	Freeways
Speed	mph	5 m	Contour maps	Freeways
Volume/Speed	vehicles	5 m	Delay linked to VMS	Freeways
AADT	veh	Daily	Pavement Management	Freeways and Arterials
Speed	mph	5 m	Travel Time(real vs. historical)	Freeways
Volume/Speed	vehicles	5 m	Operations - Snowplowing	Freeways

Appendix 3. Performance Measures Table

Volume Measures and Storage

Space Level / Time Level	Point		Link			Corridor			Freeway
	Detector	**Station**	**Lane**	**Direction**	**Road Width**	**Lane**	**Direction**	**Road Width**	
20 Seconds	LO	LO	TLQ	TLQ	TLQ				
5 Minutes	TL/LO	TL/LO	TLQ	TLQ	TLQ				
15 Minutes	TLQ	TLQ	TLQ	TLQ	TLQ				
Hour	TLQ	TLQ	TLQ	TLQ	TLQ				
Peak Periods	TLQ	TLQ	TLQ	TLQ	TLQ				
Day	TLQ	TLQ	TLQ	TLQ	TLQ				
Week	TLQ	TLQ	TLQ	TLQ	TLQ				
Month	LLS	LLS	TLQ	TLQ	TLQ				
Year	LLS	LLS	TLQ	TLQ	TLQ				

Long time off line storage (LO)
Temporary one year on Line storage (TL)
Temporary one year on Line Query (TLQ)
Long time on Line Storage (LLS)

Appendix 3:

Speed Measures and Storage

Space Level / Time Level	Point		Link			Corridor			Freeway
	Detector	**Station**	**Lane**	**Direction**	**Road Width**	**Lane**	**Direction**	**Road Width**	
20 Seconds	LO	LO	TLQ	TLQ	TLQ				
5 Minutes	TL/LO	TL/LO	TLQ	TLQ	TLQ				
15 Minutes	TLQ	TLQ	TLQ	TLQ	TLQ				
Hour	TLQ	TLQ	TLQ	TLQ	TLQ				
Peak Periods	TLQ	TLQ	TLQ	TLQ	TLQ				
Day									
Week									
Month									
Year									

Long time off line storage (LO)
Temporary one year on Line storage (TL)
Temporary one year on Line Query (TLQ)
Long time on Line Storage (LLS)

Appendix 3:

Occupancy Measures and Storage

Time Level \ Space Level	Point		Link			Corridor			Freeway
	Detector	**Station**	**Lane**	**Direction**	**Road Width**	**Lane**	**Direction**	**Road Width**	
20 Seconds	LO	LO							
5 Minutes	TL/LO	TL/LO							
15 Minutes	TLQ	TLQ							
Hour	TLQ	TLQ							
Peak Periods	TLQ	TLQ							
Day									
Week									
Month									
Year									

Long time off line storage (LO)
Temporary one year on Line storage (TL)
Temporary one year on Line Query (TLQ)
Long time on Line Storage (LLS)

Appendix 3:

Travel Time Measures and Storage

Space Level / Time Level	Point		Link			Corridor			Freeway
	Detector	Station	Lane	Direction	Road Width	Lane	Direction	Road Width	
20 Seconds									
5 Minutes									
15 Minutes									
Hour			TLQ	TLQ		TLQ	TLQ		
Peak Periods			TLQ	TLQ		TLQ	TLQ		
Day			TLQ	TLQ		TLQ	TLQ		
Week			TLQ	TLQ		TLQ	TLQ		
Month			TLQ	TLQ		TLQ	TLQ		
Year			TLQ	TLQ		TLQ	TLQ		

Long time off line storage (LO)
Temporary one year on Line storage (TL)
Temporary one year on Line Query (TLQ)
Long time on Line Storage (LLS)

Appendix 3:

Appendix 3:

Delay Measures and Storage

Time Level \ Space Level	Point		Link			Corridor			Freeway
	Detector	**Station**	**Lane**	**Direction**	**Road Width**	**Lane**	**Direction**	**Road Width**	
20 Seconds									
5 Minutes									
15 Minutes									
Hour			TLQ	TLQ	TLQ	TLQ	TLQ	TLQ	TLQ
Peak Periods			TLQ	TLQ	TLQ	TLQ	TLQ	TLQ	TLQ
Day			TLQ	TLQ	TLQ	TLQ	TLQ	TLQ	TLQ
Week			TLQ	TLQ	TLQ	TLQ	TLQ	TLQ	TLQ
Month			TLQ	TLQ	TLQ	TLQ	TLQ	TLQ	TLQ
Year			TLQ	TLQ	TLQ	TLQ	TLQ	TLQ	TLQ

Long time off line storage (LO)
Temporary one year on Line storage (TL)
Temporary one year on Line Query (TLQ)
Long time on Line Storage (LLS)

Appendix 3:

VMT Measures and Storage

Space Level / Time Level	Point		Link			Corridor			Freeway
	Detector	**Station**	**Lane**	**Direction**	**Road Width**	**Lane**	**Direction**	**Road Width**	
20 Seconds									
5 Minutes									
15 Minutes									
Hour									
Peak Periods									
Day						TLQ	TLQ	TLQ	TLQ
Week						TLQ	TLQ	TLQ	TLQ
Month						TLQ	TLQ	TLQ	TLQ
Year						LLS	LLS	LLS	LLS

Long time off line storage (LO)
Temporary one year on Line storage (TL)
Temporary one year on Line Query (TLQ)
Long time on Line Storage (LLS)

Appendix 3:

VHT Measures and Storage

Space Level / Time Level	Point		Link			Corridor			Freeway
	Detector	Station	Lane	Direction	Road Width	Lane	Direction	Road Width	
20 Seconds									
5 Minutes									
15 Minutes									
Hour									
Peak Periods									
Day						TLQ	TLQ	TLQ	TLQ
Week						TLQ	TLQ	TLQ	TLQ
Month						TLQ	TLQ	TLQ	TLQ
Year						LLS	LLS	LLS	LLS

Long time off line storage (LO)
Temporary one year on Line storage (TL)
Temporary one year on Line Query (TLQ)
Long time on Line Storage (LLS)