



U.S. Department
of Transportation

**Federal Highway
Administration**

ITS Data Archiving: Case Study Analyses of San Antonio TransGuide[®] Data

Report No. FHWA-PL-99-024

Final Report

August 1999



Prepared for

ITS Joint Program Office and the
Office of Highway Policy Information

Prepared by

Texas Transportation Institute
The Texas A&M University System

1. Report No. FHWA-PL-99-024		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle ITS Data Archiving: Case Study Analyses of San Antonio TransGuide® Data				5. Report Date August 1999	
				6. Performing Organization Code	
7. Author(s) Shawn M. Turner, William L. Eisele, Byron J. Gajewski, Luke P. Albert, and Robert J. Benz				8. Performing Organization Report No.	
9. Performing Organization Name and Address Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135				10. Work Unit No. (TRAIS)	
				11. Contract or Grant No.	
12. Sponsoring Agency Name and Address Federal Highway Administration ITS Joint Program Office (HOIT) and the Office of Highway Policy Information (HPPI) 400 Seventh Street, SW Washington, DC 20590				13. Type of Report and Period Covered Final: March 1998 - August 1999	
				14. Sponsoring Agency Code FHWA CO6489605 (Amend. 1)	
15. Supplementary Notes Research performed in cooperation with the Federal Highway Administration and the Texas Department of Transportation. Project COTR: Mr. Ralph Gillmann, HPPI Research Project Title: Retention, Management, and Sharing of ITS Data					
16. Abstract This report documents: 1) the state-of-the-practice in retaining and using archived ITS data, and 2) lessons learned from case study analyses of archived ITS data from TxDOT's TransGuide® center in San Antonio. The case study analyses of TransGuide® data were focused on data aggregation and data quality (i.e., data errors, missing data, and data accuracy). The state-of-the-practice findings indicate that many operations centers are becoming more interested in archiving ITS data. Several concerns have been expressed by potential users about ITS data quality and location referencing of ITS detector locations. Analyses of TransGuide® data resulted in statistical procedures that can be used to help guide decisions about appropriate data aggregation levels. The researchers also present innovative data archiving approaches that can help to address different data aggregation needs, including capabilities such as data sampling, on-demand archiving, and data broadcasts. In other analyses, the researchers found mixed results in comparing traffic volumes from TransGuide® detectors to other nearby TxDOT traffic counting stations and ground truth video. At one location, traffic volumes from both detectors were within 5 percent of ground truth, whereas detectors at the second location ranged between 12 to 38 percent less than ground truth. In analyses of a full month of data, the researchers also found that the level of missing data was nearly 20 percent, whereas basic quality control checks flagged only 1 percent of the data as suspect or erroneous.					
17. Key Words archived ITS data, data aggregation, quality control, ITS data accuracy, ITS data management			18. Distribution Statement No restrictions. This document is available to the public through NTIS: National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161		
19. Security Classif.(of this report) Unclassified		20. Security Classif.(of this page) Unclassified		21. No. of Pages 146	22. Price

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration (FHWA) or the Texas Department of Transportation (TxDOT). This report does not constitute a standard, specification, or regulation.

ACKNOWLEDGMENTS

The authors wish to acknowledge the assistance and support of the following persons:

- **Dr. Joseph Peters** in the ITS Joint Program Office and **Mr. Ralph Gillmann** in the Office of Highway Policy Information of FHWA, who supported this research effort;
- **Mr. Pat Irwin** and his **staff at TransGuide®**, who have cooperated with and supported TTI research (including ITS data archiving) for many years;
- **Mr. Ken Mora** in the Transportation Planning and Programming Division at TxDOT, who served as the TxDOT liaison for this FHWA-funded study;
- **Mr. Mark Hodges** and **Mr. Jeff Reding** in the Transportation Planning and Programming Division at TxDOT, who provided ATR traffic volume data and detector locations;
- **Mr. Christopher Poe**, director of the TransLink® ITS research program at TTI, who has supported ITS data research through the TransLink® program since 1996;
- **Mr. Darrell Borchardt** in the Houston office of TTI, who helped support this and other research by providing ITS-generated data from Houston;
- **Dr. Clifford Spiegelman** in the Statistics Department of Texas A&M University, who provided advice and statistical expertise on ITS data archiving issues;
- the numerous transportation professionals who took time out of busy schedules to share their ITS data archiving experiences with TTI researchers, as well as those who helped to improve this research report by providing valuable review comments.

TABLE OF CONTENTS

	PAGE
LIST OF FIGURES	vii
LIST OF TABLES	ix
LIST OF ACRONYMS	xi
EXECUTIVE SUMMARY	xiii
CHAPTER 1. INTRODUCTION	1
HISTORICAL PERSPECTIVE	2
ITS DATA ARCHIVING SYSTEM IMPLEMENTATION ISSUES	2
STUDY OBJECTIVE	4
REPORT OVERVIEW	4
CHAPTER 2. BACKGROUND AND STATE-OF-THE-PRACTICE	5
CASE STUDIES OF EXISTING ITS DATA ARCHIVING PRACTICES	5
Phoenix, Arizona	8
Los Angeles, California	9
Orange County, California	10
San Francisco, California	10
Atlanta, Georgia	10
Chicago, Illinois	10
Montgomery County, Maryland	11
Detroit, Michigan	11
Minneapolis-St. Paul, Minnesota	12
TRANSCOM, New York/New Jersey/Connecticut	12
Ft. Worth, Texas	13
Houston, Texas	14
San Antonio, Texas	16
Seattle, Washington	17
Toronto, Ontario, Canada	17
State of Virginia	18
North America Pre-clearance and Safety Systems	19
STATE-OF-THE-PRACTICE IN USING ARCHIVED ITS DATA	21
Metropolitan Planning Organizations / State Transportation Planners	23
Traffic Management Operators	34
Maintenance and Construction	35
Transportation Researchers	35
Private Sector	39

TABLE OF CONTENTS (Continued)

	PAGE
ADVANCING THE STATE-OF-THE-PRACTICE IN ITS DATA ARCHIVING	40
Development of the ADUS Architecture	41
Data Standards Development	44
ITS Data Registry	46
Volpe Center Case Studies	47
Oak Ridge National Lab/Florida DOT/New York DOT Demonstration Project .	47
Washington State	48
TransLink® ITS Research Program, Texas Transportation Institute	51
SUMMARY OF FINDINGS AND LESSONS LEARNED	53
CHAPTER 3. EXAMINATION OF ITS DATA ARCHIVING ISSUES	55
INCORPORATING DATA NEEDS INTO ITS DATA ARCHIVING	56
Process of Defining or Establishing Data Needs	56
Previous Work on Data Needs and User Requirements	60
Summary of Factors to Consider in Developing ITS Data	
Archiving User Requirements	67
AGGREGATION OF ITS TRAFFIC MONITORING DATA	68
Approach 1: Save Aggregated Data	68
Approach 2: Save All Data	68
Approach 3: Focus on Strategic Archiving, Build Advanced Capabilities	70
Data Sampling	71
Aggregation Based on Statistical Variability	72
Archiving on Demand	88
Data Broadcast	89
ITS DATA QUALITY CONTROL	90
Erroneous or Suspicious Data	90
Nature and Extent of Missing Data	96
Accuracy and Comparability of ITS Data	102
DATA STORAGE TOOLS AND ISSUES	114
Archived ITS Data Storage Tools	114
Privacy Issues	118
CHAPTER 4. FINDINGS AND CONCLUSIONS	119
STATE-OF-THE-PRACTICE IN RETAINING AND	
USING ARCHIVED ITS DATA	119
EXAMINATION OF ITS DATA ARCHIVING ISSUES	120
REFERENCES	123

LIST OF FIGURES

FIGURE		PAGE
1	Summary of Steps in Implementing ADMS	3
2	Proposed AZTech Data Server	8
3	Proposed Data Management System in Ft. Worth's TransVISION	14
4	Press Release Describing Advantage CVO Data Policy	20
5	Traffic Profile of SR-520 General Purpose Lanes	24
6	Estimated Average Weekday Travel Time for Eastbound SR-520	25
7	Estimated Weekday Volume, Speed, and Reliability Conditions for Westbound SR-520, 76 th Avenue N.E.	26
8	Average Morning Peak Hour Travel Times for Katy Freeway, 1994	27
9	Morning Peak Hour Travel Time Reliability	28
10	Average Weekday Traffic Volume Patterns Derived from ITS Data, Phoenix Urban Area	30
11	Monthly Traffic Variations for Freeways Using ITS Data, Phoenix Urban Area	31
12	Traffic Volume-Speed Relationship Using ITS Data	31
13	Illustration of Hypothesis that ITS Strategies are Most Effective in High-Variability Systems	32
14	Accident and Weather Scenario Mapping by Demand Level Using Archived ITS Data and Other Data Sources	33
15	National ITS Architecture	43
16	ITS Data Registry Functional Concepts	46
17	Main Interface of CDR Software	49
18	Schematic of TDAD System	50
19	Query Interface of DataLink System (http://vixen.cs.tamu.edu)	52
20	Process that Starts with Enumeration of Available ITS Data	57
21	Process that Starts with Stakeholder Data Needs	58
22	Iterative Process that Establishes Dialogue between Data Stakeholder Groups and Data Provider	59
23	Cross-Validated Mean Square Error Technique	74
24	F-Statistic Technique	78
25	Typical Speed Profile and Results of CVMSE Analysis for Optimal Aggregation Level	81
26	Typical Speed Profile and Results of F-Statistic Analysis for Optimal Aggregation Level	81
27	Illustration of Loss Profiles for the Cross-Validated Mean Square Error Technique ...	84
28	Range of Possible ITS Data Aggregation Solutions	85
29	Example of Statistical Process Control Guidelines	94
30	Identifying Detector-Specific Missing Data	98

LIST OF FIGURES (Continued)

FIGURE		PAGE
31	Identifying the Extent of Detector-Specific Missing Data	99
32	Identifying Server-Specific and System-Wide Missing Data	100
33	Location of ATR S184 and Nearby TransGuide® Loops	103
33	Location of ATR S185 and Nearby TransGuide® Loops	103
35	Comparison of Hourly Volumes from ATR S184 EB and Equivalent Upstream TransGuide® Detectors	107
36	Comparison of Hourly Volumes from ATR S184 EB and Equivalent Downstream TransGuide® Detectors	107
37	Comparison of Hourly Volumes from ATR S184 WB and Equivalent Upstream TransGuide® Detectors	108
38	Comparison of Hourly Volumes from ATR S184 WB and Equivalent Downstream TransGuide® Detectors	108
39	Comparison of Hourly Volumes from ATR S185 SB and Equivalent Upstream TransGuide® Detectors	109
40	Comparison of Hourly Volumes from ATR S185 SB and Equivalent Downstream TransGuide® Detectors	109
41	Comparison of Hourly Volumes from ATR S185 NB and Equivalent Upstream TransGuide® Detectors	110
42	Comparison of Hourly Volumes from ATR S185 NB and Equivalent Downstream TransGuide® Detectors	110
43	Comparison of Hourly Volumes from Upstream and Equivalent Downstream TransGuide® Detectors on I-10 East	111
44	Comparison of Hourly Volumes from Upstream and Equivalent Downstream TransGuide® Detectors on I-10 West	111
45	Comparison of Hourly Volumes from Upstream and Equivalent Downstream TransGuide® Detectors on I-37 South	112
46	Comparison of Hourly Volumes from Upstream and Equivalent Downstream TransGuide® Detectors on I-37 North	112

LIST OF TABLES

TABLE	PAGE
1	Contacts for Existing ITS Data Archiving Practices 6
2	TMCs and Agencies Contacted in First and Second Telephone Surveys 7
3	Example of User Requirements Matrix for Houston TranStar 15
4	Stakeholders and Example Applications for Archived ITS Data 21
5	Efforts to Advance the State-of-the-Practice in Archived ITS Data 40
6	Incorporation of the ADUS into the National ITS Architecture 42
7	Attributes of a Vehicle Detector Count Using ITE’s TMDD 44
8	Selected ITS Standards Relevant to ITS Data Archiving 45
9	Documentation of Data Needs in Transportation Decision Making 61
10	ITS Data Relevant for Archiving 62
11	Comparison of Approaches for Aggregating Archived ITS Data 69
12	Illustration of Every N th Day Sampling 71
13	Illustration of Concentrated Weeks Sampling 72
14	CVMSE and F-Statistic Algorithm Results, IH-10 West, Milepost 566.641 80
15	Example Matrix: Level of Aggregation for ITS Data Archiving 86
16	Quality Control Screening Scenarios and Corresponding Flag Values 92
17	Summary of Suspect or Erroneous TransGuide [®] Data, October 1998 93
18	Extent of Suspect or Erroneous TransGuide [®] Data for Individual Scenarios, October 1998 95
19	Typical Causes and Characteristics of Missing ITS Traffic Monitoring Data 96
20	Identification of Missing Data for San Antonio’s TransGuide [®] 97
21	Equations Used to Compare ATR Stations to TransGuide [®] Detectors 104
22	Results of Hourly Traffic Volume Comparisons Between ATR and TransGuide [®] Detectors, May 1998 106
23	Results of Ground-Truth Traffic Volume Comparisons, December 1998 113
24	Comparisons of Computer Storage Media and Devices 115

LIST OF ACRONYMS

AADT	annual average daily traffic volume
AASHTO	American Association of State Highway and Transportation Officials
ADS	AZTech data server (Maricopa County, Arizona)
ADMS	archived data management subsystem
ADUS	archived data user service
ANN	artificial neural network
ANOVA	analysis of variance
ASTM	American Society for Testing and Materials
ATMS	advanced transportation management system
ATR	automatic traffic recorder
AVI	automatic vehicle identification
AWDT	average weekday traffic volume
BTS	Bureau of Transportation Statistics
CAA	Clean Air Act Amendments
CAD	computer-aided dispatch
Caltrans	California Department of Transportation
CATS	Chicago Area Transportation Study (Illinois)
CCTV	closed circuit television
CD	compact disc
CMAQ	congestion mitigation/air quality
CMS	changeable message sign
CVMSE	cross-validated mean square error
CVO	commercial vehicle operations
DMS	dynamic message sign
DOT	department of transportation
DVD	digital video/versatile disk
FHWA	Federal Highway Administration
FOT	field operational test
GIS	geographic information system
HAR	highway advisory radio
HGAC	Houston-Galveston Area Council (Texas)
HOV	high-occupancy vehicle
IEEE	Institute of Electrical and Electronic Engineers
ITE	Institute of Transportation Engineers
ITS	intelligent transportation system
ITS	Institute of Transportation Studies (University of California at Berkeley)
KTC	Kentucky Transportation Center (University of Kentucky)
LCS	lane control signal

LIST OF ACRONYMS (Continued)

LCU	local controller unit
MDOT	Michigan Department of Transportation
M-NCPPC	Maryland-National Capital Park and Planning Commission
MITS	Michigan Intelligent Transportation System Center (Detroit)
Mn/DOT	Minnesota Department of Transportation
MOE	measure of effectiveness
MPO	metropolitan planning organization
NISS	National Institute of Statistical Sciences
NORPASS	North American Pre-Clearance and Safety System
OLAP	on-line analytical processing
OLTP	on-line transaction processing
RAID	redundant array of independent disks
RCOC	Road Commission of Oakland County (Michigan)
SDO	standard development organization
SNN	spectral-basis neural network
SQL	structured query language
SwRI	Southwest Research Institute (San Antonio, Texas)
TDAD	Traffic Data Acquisition and Distribution (University of Washington)
TLFD	trip length frequency distribution
TMC	transportation management center
TMDD	Traffic Management Data Dictionary
TP&P	Transportation Planning and Programming Division (TxDOT)
TRAC	Washington State Transportation Center (University of Washington)
TRANSCOM	Transportation Operations Coordinating Committee (New York/New Jersey/Connecticut)
TransGuide	Transportation Guidance System (San Antonio, Texas)
TRANSMIT	TRANSCOM's System for Managing Incidents and Traffic
TransVISION	Transportation Vehicular Information and Surveillance on an Intelligent highway Optical Network (Ft. Worth, Texas)
TSC	Traffic Systems Center (Chicago, Illinois)
TSM	transportation system management
TSMC	Transportation Systems Management Center (Seattle, Washington)
TTI	Texas Transportation Institute (Texas A&M University System)
TxDOT	Texas Department of Transportation
VDOT	Virginia Department of Transportation
VMT	vehicle-miles of travel
WIM	weigh-in-motion
WSDOT	Washington State Department of Transportation

EXECUTIVE SUMMARY

The primary objective of this research study was to develop guidance for implementing (intelligent transportation system) ITS data archiving systems by identifying and examining selected implementation issues. To accomplish this objective, Texas Transportation Institute (TTI) researchers performed case study analyses using archived loop detector data from the TransGuide® system in San Antonio, Texas. The selected implementation issues and associated questions examined in this study include:

- **data aggregation** - What is the desired aggregation level at which to retain traffic monitoring data? What are the tradeoffs in saving data at different levels? Are there other data archiving strategies besides simply saving all data at a single aggregation level?
- **data quality** - What types of data quality should I be concerned about (i.e., erroneous or suspect data, missing data, and data accuracy)? How do I address these concerns with quality control procedures?

The researchers also gathered and synthesized information about existing ITS data archiving practices, the most common applications of archived ITS data, and data storage tools and issues.

ITS DATA ARCHIVING PRACTICES

The research team surveyed fifteen traffic management centers (TMCs) in the spring of 1999 to gather information on current ITS data archiving practices. Information on the current uses of archived ITS data was also gathered from transportation agencies in these same areas. The surveys indicated that nearly all TMCs were saving ITS-generated data, but most were at different stages in making the archived data accessible to a wide variety of users. Several of the TMCs interviewed were developing compact disc (CD) or Internet-based archived data management systems to reduce the problems associated with archived data on magnetic tape cartridges. User requirement studies were being conducted in several areas to identify the users and uses of the archived data. Several interviewees mentioned data quality as a major issue in ITS data archiving.

Many of the uses of archived ITS data to date have been for transportation planning or research. These planning uses have primarily consisted of congestion management systems, performance measurement, model calibration/validation, and basic traffic counts and factors. The applications for research have focused on model development and traffic flow theory. A few of the traffic management operators are using archived ITS data to refine operating procedures and perform annual evaluations.

AGGREGATION OF ITS TRAFFIC MONITORING DATA

Because ITS data are collected at frequent intervals (typically 20 to 30 seconds) at many locations, data aggregation (to 5 or 15 minutes) is seen as a way to reduce data storage and make ITS data less cumbersome for ordinary data users. The research team developed two statistical methods for determining optimal data aggregation intervals based upon statistical variability. The two statistical methods yielded similar results: little or no aggregation (one minute or less) during congested periods, hourly aggregation during low traffic periods, and moderate aggregation (five or fifteen minutes) during transition periods.

The authors acknowledge that there are a wide range of possible aggregation solutions based upon these statistical techniques (Figure S-1). Ultimately, the aggregation solutions may be driven by non-statistical parameters such as cost (e.g., "how much do we/the market value the data?"), ease of implementation, system interface requirements, and other constraints. Some may only desire a simple ITS data archiving solution for existing data needs, thereby negating the need for using statistical techniques. Others may desire a single fixed aggregation level throughout the day, in which the statistical techniques or their results could be used to examine the tradeoffs between optimal aggregation levels and existing constraints. Still others may wish to implement dynamic aggregation levels in ITS data archiving, in which case the statistical techniques described here could be programmed and automated within the data archiving system.

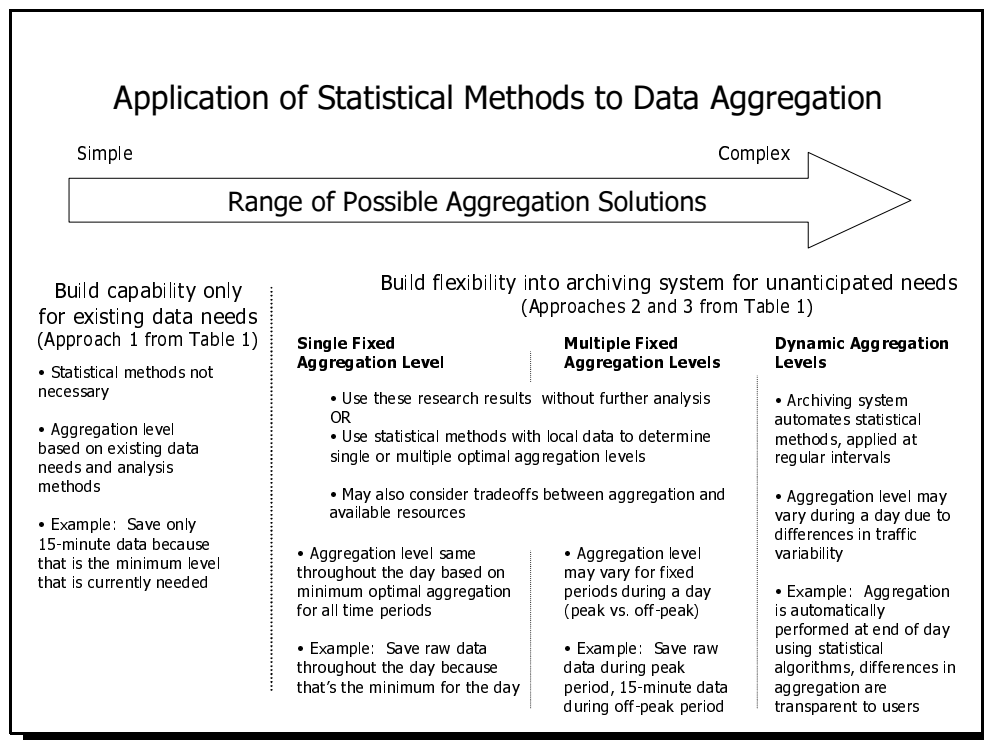


Figure S-1. Range of Possible ITS Data Aggregation Solutions

QUALITY CONTROL FOR ITS DATA

The quality control efforts in this study were focused on a month of ITS traffic monitoring data (collected by inductance loop detectors) from San Antonio's TransGuide® system, and analyzed three basic attributes of quality control:

- suspicious or erroneous data;
- nature and extent of missing data; and
- accuracy and comparability of ITS data to similar data sources.

Suspicious or Erroneous Data

The analysis of suspicious or erroneous data found that about 1 percent of the archived ITS data could be labeled as "suspect" according to current practices for identifying ITS data errors. The algorithms used to identify suspect or erroneous data were aimed at identifying blatant data errors, so more stringent quality control rules would result in a higher percent of the data being labeled as "suspect." The research team recommended the development of more advanced quality control procedures for archived ITS data. These advanced quality control procedures should examine each data record in relation to other data records in close proximity in both time and space. Traffic patterns and profiles from adjacent lanes can also be compared to ensure consistency and reliability. More advanced quality control procedures could also compare each data record to historical patterns and trends for average values and typical ranges of variability at that location.

Nature and Extent of Missing Data

Although the percent of archived data labeled as "suspect" was quite small, missing data records did account for a significant portion of the data (about 22.5 percent). Missing ITS data is a result of roadway construction activities, equipment failures, disruptions in communications, or computer hardware and software failures. Given the continuous operation of most ITS traffic monitoring devices, missing data is almost inevitable. In San Antonio, this missing data typically ranged from 5 to 25 percent for most days. Because of excellent loop maintenance programs in San Antonio, only about 5 to 15 percent of this missing data can be accounted for by loop detector failure. The remaining 10 or more percent can most likely be explained by central traffic management or data archiving system failures.

Accuracy of ITS Data

In initial conversations with TxDOT's Transportation Planning and Programming (TP&P) Division about using San Antonio's TransGuide® data, statewide transportation planners named data quality as their most pressing concern. More specifically, the statewide planners were most interested in how TransGuide® data compared to data collected at nearby automated traffic recorder (ATR) stations. To answer these questions about data comparability, TTI compared a full month of TransGuide® traffic volume data to hourly traffic volume data collected at two

ATR stations that were located along freeways instrumented in Phase One of TransGuide®. Aside from the comparability of TransGuide® traffic volumes to ATR traffic volumes, TTI researchers were concerned about the accuracy of both types of traffic monitoring equipment to "ground truth." To answer questions about how well these two data sources compared to "ground truth," TTI researchers used the TransGuide® surveillance cameras to record actual traffic conditions.

The analysis of TransGuide®, ATR, and ground truth traffic volume data found a wide range of comparability between the ATR and TransGuide® data. At the IH-10 location, the ATR and TransGuide® traffic volumes compared favorably, typically falling within 10 percent of one another. However, the IH-37 location showed large differences in traffic volumes, ranging from 18 to 39 percent in directly comparing ATR and TransGuide® detectors. The researchers also found mixed results in comparing the collected traffic volumes to ground truth. At the IH-10 location, both the ATR and TransGuide® traffic volumes were within 5 percent on average of ground truth volumes, with no clear difference between ATR and TransGuide® detectors. At the IH-37 location, however, the ATR detectors were slightly more accurate (within 4 and 11 percent on average) than the TransGuide® detectors (within 13 to 38 percent on average).

CONCLUSIONS AND RECOMMENDATIONS

Most Centers Have ITS Data Archiving on Their "Radar" - The research team found that most operations and management centers contacted were at least considering ITS data archiving, if not actively developing plans for archived data systems. Because of this widespread interest in ITS data archiving, ADUS deployment guidance will be imperative in the further development of integrated data systems. A common assertion is that many centers have not been archiving ITS data; however, the research team found in 1997 that 12 of 15 centers were archiving data, but that the data were either not readily accessible or usable (stored on proprietary magnetic tape cartridges) or not widely distributed within the metropolitan area.

Access to Archived ITS Data Creates Opportunities - The research team found the most widespread use of archived ITS data was in locations where the data were easily or publicly accessible. In these locations, user groups outside of the operations center were able to develop data extraction and analysis tools because the archived ITS data were easily accessible. The logical conclusion is that in areas where the operations center may not be able to develop systems that support ADUS, these centers should at least provide easy access to the data so that other user groups can develop systems that support ADUS functions. Examples of easy access to archived ITS data include distribution via the Internet or CDs.

Planners and Researchers Have Been Primary Archived Data Users to Date - The research team found that the two most common user groups have been researchers and planners, with applications ranging from basic traffic statistics to advanced model and algorithm development. In the few locations where traffic management operators were using archived ITS data, they were realizing benefits in improving traffic operations and management, as well as an ability to quantify these benefits. Private sector users that add value to archived ITS data are emerging

users, although the business models are not yet clearly defined.

EXAMINATION OF ITS DATA ARCHIVING ISSUES

Need to Establish Ongoing Dialogue Between Data Providers and Data Users - The best synergy for archiving ITS data occurs when there is an ongoing dialogue between data providers (e.g., operations centers) and data users. An ongoing dialogue will help in the following ways: 1) data users may better understand the available ITS data and its intricacies; 2) data providers may better understand the needs of data users; and 3) data providers and data users may work together in establishing ITS detector designs that meet the data needs of many groups.

Consider Innovative Archiving Approaches to Address Different Data Aggregation Needs - To date, there have been two basic suggested approaches to archiving ITS data: 1) save aggregated data to meet existing needs in a cost-effective manner; 2) save raw, disaggregate data to maintain flexibility of analyses and data exploration. The research team suggests a third approach, in that basic aggregated data are saved for current needs, yet innovative archiving capabilities exist that can provide advanced data users access to raw, disaggregate data. The innovative capabilities discussed in this report include raw data sampling, aggregation based upon statistical variability, on-demand archiving, and data broadcasts.

Need Improved Quality Control Procedures - These advanced (and ideally automated) procedures could compare current data at a given location to upstream and downstream detectors, as well as examining historical values, trends, and patterns. Quality control procedures to establish absolute data accuracy ("ground truthing") should be also be included. The current consensus is that archived ITS data not passing quality control checks should be flagged as such, with users making the choice of imputed or replacement values appropriate for their analyses.

Account for Missing Data in Data Archiving System Designs - It is critical that ITS data archiving systems be able to reflect missing data in summary or aggregate statistics. Cumulative statistics, such as vehicle-miles of travel (VMT), are most affected by missing data and will likely require data users and system designers to collaborate on how to account for the missing data. As with quality control checks, the consensus is that missing data be flagged as such, with users making the choice of appropriate imputed or replacement values.

CHAPTER 1. INTRODUCTION



CHAPTER OVERVIEW

☞	Historical Perspective	Provides an historical perspective on archiving of operational data, summarizes problems noted in previous work.
☞	ITS Data Archiving System Implementation Issues	Lists implementation issues associated with ITS data archiving, suggests steps in implementing data archiving.
☞	Study Objective	Describes objectives of research study.
☞	Report Organization	Summarizes report organization.

With the advent of the archived data user service (ADUS) in 1998, the National ITS Architecture officially embraced the concept of saving (a.k.a. retaining, archiving) real-time, operational data for other non-real-time applications, such as transportation planning, safety analyses, and research. Intelligent transportation system (ITS) applications and their respective sensors and detectors are a potentially rich source of data about transportation system performance and characteristics. Examples of the most common data elements potentially available from ITS include:

- **traffic monitoring and detection systems** - vehicle volume, speed, travel time, classification, weight, and trajectories;
- **traffic control systems** - time and location of traffic control actions (e.g., ramp metering, traffic signal control, lane control signals, message board content);
- **incident and emergency management systems** - location, cause, extent, and time history of roadway incident/emergency detection and clearance; and
- **advanced public transit systems** - transit vehicle boardings by time and location, vehicle trajectories, origins and destinations.

In this report, ITS data refers to data that is typically collected by and/or generated for ITS applications. ITS data archiving refers to the systematic retention of data that is generated by ITS applications. The philosophy of ADUS encompasses more than just ITS data archiving, however, and can include data extraction, manipulation and analysis tools as well.

HISTORICAL PERSPECTIVE

Although the practice of saving or archiving operational data from traffic control sensors and detectors is not a new concept, national interest in ITS data archiving has grown significantly since 1997, due mostly to the formation of ADUS. As early as the 1970s, the Illinois Department of Transportation (DOT) was saving aggregated loop detector data in Chicago to report “minute-miles” of congestion (1). Similarly in 1968, the Texas Highway Department and the Texas Transportation Institute (TTI) were using an IBM 1800 computer to save and analyze loop detector data along the Gulf Freeway in Houston (2). The archived Houston data were used to support level of service and merging research studies, as well as to demonstrate and quantify the effects of incidents on the freeway corridor. The Washington State DOT (WSDOT) and the University of Washington have been archiving loop detector data from Seattle’s freeway traffic management system since 1981, with researchers and planning agencies being primary users (3). Loop detector data from Highway 401 in Toronto, Ontario (Canada), also has been used extensively since the 1980s for traffic flow theory and capacity research.

To date, however, ITS and operational data archiving has mostly been informal arrangements with a single data user group (e.g., researchers, planners) who were able to work with these large data sets. Numerous problems surfaced in this early work with archived data:

- proprietary data formats and data storage devices (e.g., magnetic tape cartridges) hindered archived data distribution;
- distributing archived data to users sometimes placed an unreasonable burden on operations personnel;
- processing and analyzing the large quantities of data in a mainframe computer environment was not user-friendly; and
- missing and erroneous data were evident throughout the archived data.

ITS DATA ARCHIVING SYSTEM IMPLEMENTATION ISSUES

The ADUS addition to the National ITS Architecture, set to be released later in 1999, will provide an overall framework for designing archived data management subsystems (ADMS) and archived data user systems. Some of the problems described earlier may be lessened by adoption of the ADMS architecture and associated data standards. However, there are numerous technical and institutional issues that agencies may still have to address in implementing ADMS (4):

1. development, operation, and maintenance costs
2. system access
3. ownership

4. data quality
5. data management
6. data and communications standards
7. privacy concerns
8. data analysis
9. coordination with other data collection efforts
10. liability
11. confidentiality of privately collected data
12. incremental and uncoordinated ITS deployments
13. retrofitting vs. new development of systems
14. data not defined by the National ITS Architecture
15. conformance with metric conversion standards
16. training and outreach

Figure 1 shows important steps in the implementation of ADMS. The figure illustrates that both institutional arrangements (i.e., “Data Partnerships”) and technical details (i.e., “System Design”) are necessary in deploying an effective ADMS.

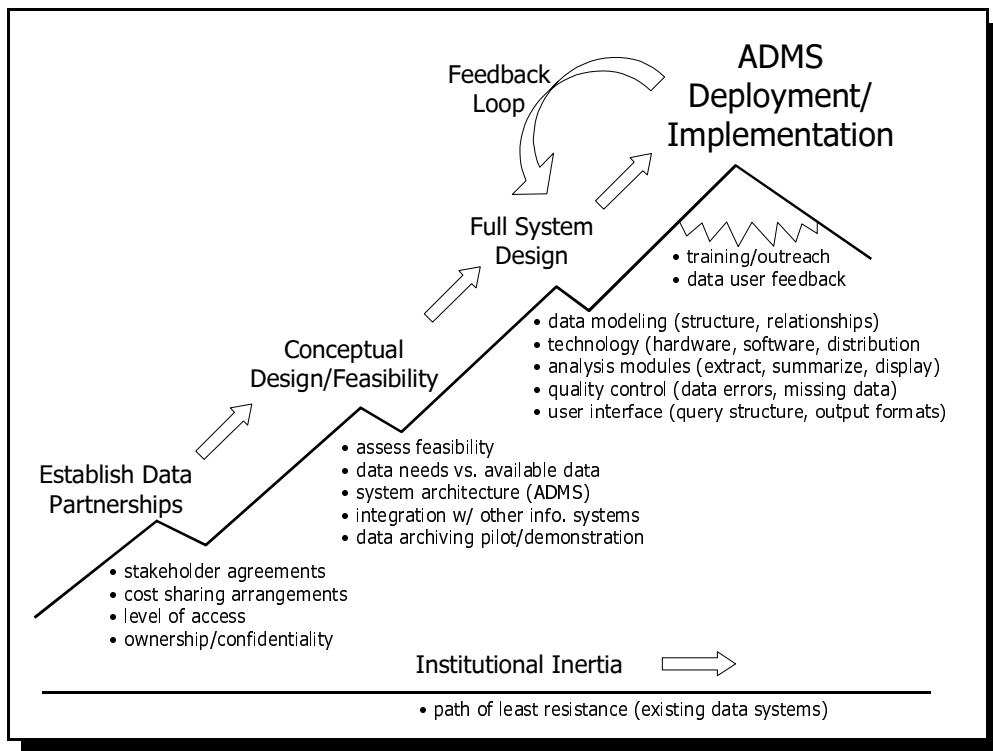


Figure 1. Summary of Steps in Implementing ADMS

STUDY OBJECTIVE

The primary objective of this research study was to develop guidance for implementing ITS data archiving systems by identifying and examining selected implementation issues. To accomplish this objective, TTI researchers performed case study analyses using archived loop detector data from the TransGuide® system in San Antonio, Texas. The selected implementation issues and associated questions examined in this study include:

- **data management** - What is the desired aggregation level at which to retain traffic monitoring data? What are the tradeoffs in saving data at different levels? Are there other data archiving strategies besides simply saving all data at a single aggregation level?
- **data quality** - What types of data quality should I be concerned about (i.e., erroneous or suspect data, missing data, and data accuracy)? How do I address these concerns with quality control procedures?

The researchers also gathered and synthesized information about existing ITS data archiving practices, the most common applications of archived ITS data, and data storage tools and issues.

REPORT OVERVIEW

This report is organized into the following four chapters:

- **Chapter 1. Introduction** - introduces ITS data archiving, implementation issues, and study objective;
- **Chapter 2. Background and State-of-the-Practice** - summarizes existing ITS data archiving practices, describes and provides examples of common applications for archived ITS data, and discusses efforts to advance the state-of-the-practice;
- **Chapter 3. Examination of ITS Data Archiving Issues** - describes exploratory analyses conducted to examine implementation issues such as data management and aggregation, data quality, and data storage tools; and.
- **Chapter 4. Findings and Conclusions** - summarizes major findings and lessons learned related to selected ITS data archiving issues.

CHAPTER 2. BACKGROUND AND STATE-OF-THE-PRACTICE



CHAPTER OVERVIEW

☞	Case Studies of Existing ITS Data Archiving Practices	Describes data archiving practices at several operations or transportation management centers.
☞	State-of-the-Practice in Using Archived ITS Data	Summarizes and provides examples of the most common applications of archived ITS data.
☞	Advancing the State-of-the-Practice in ITS Data Archiving	Presents National ITS Architecture development as well as research efforts aimed at advancing ITS data archiving practices.
☞	Summary of Findings and Lessons Learned	Summarizes findings in regard to state-of-the-practice in ITS data archiving.

CASE STUDIES OF EXISTING ITS DATA ARCHIVING PRACTICES

This section contains information on the state-of-the-practice in ITS data archiving. These experiences are drawn from transportation management centers (TMCs) or other operations centers (e.g., commercial vehicle operations (CVO) at weigh stations) where real-time traffic monitoring data are collected. Table 1 contains the contacts from which this ITS data archiving information was obtained.

Previous TTI research documents the existing practices of retaining ITS data for transportation analyses (5,6). The research found that the uses and users of archived ITS data appeared to be relatively consistent at the TMCs that were surveyed. They include requests for traffic counts to estimate construction impacts, requests for planning and modeling uses, and researcher requests. While the different TMCs have similar uses and user requests, the data storage and aggregation practices were not consistent. Of the 15 TMCs interviewed in the spring of 1997, three were not saving any of the ITS data. Nine of the TMCs saved traffic data at intervals of one minute or less, and the remaining three TMCs saved the data at an aggregation level between 5 and 15 minutes.

As part of this study, the research team re-visited the list of TMCs from the previous research to identify any changes in ITS data archiving strategies. Table 2 summarizes the locations that were contacted in the first and second set of telephone surveys. In this second follow-up survey, it was found that the TMC operators and managers are beginning to recognize the value of ITS data archiving for secondary uses and are interested in saving, formatting, and aggregating data to

facilitate other uses beyond the traditional real-time needs. There appears to be a stronger interest in interdisciplinary coordination (i.e., operations staff working more closely with planning personnel), although financial allocations are often the leading hindrance to such efforts. The following paragraphs describe ITS data archiving and retention practices at these selected TMCs.

Table 1. Contacts for Existing ITS Data Archiving Practices

Location	Agency	Contact Person and Phone Number
Phoenix, Arizona	Maricopa County DOT	David Wolfson, (602) 506-6950
	Maricopa Association of Governments	Mark Schlappi, (602) 254-6300
Los Angeles, California	Caltrans	David Lau, (213) 897-0391
Orange County, California	Caltrans	Omid Segal, (949) 724-2467
San Francisco, California	Caltrans	Dick Fahey, (510) 286-5761
Atlanta, Georgia	Georgia DOT	Dennis Reynolds, (404) 635-8027
Chicago, Illinois	Illinois DOT	Tony Cioffi, (708) 524-2145
Montgomery County, Maryland	Montgomery County DOT	Bruce Mangum, (301) 217-2197
	Maryland National Capital Park and Planning Commission	Don Ostrander, (301) 495-2184
Detroit, Michigan	Michigan DOT	Arvyd Satraitis, (313) 256-9800
Minneapolis-St. Paul, Minnesota	Minnesota DOT	James Aswegan, (612) 373-2761
Jersey City, New Jersey	TRANSCOM	Sanjay Patel, (201) 963-4033
Fort Worth, Texas	Texas DOT	Abed Abukar, (817) 370-6621
Houston, Texas	Texas DOT	Cindy Gloyna, (713) 802-5147
San Antonio, Texas	Texas DOT	Pat Irwin, (210) 731-5249
Seattle, Washington	Washington DOT	Mahrokh Arefi, (206) 440-4462
Toronto, Ontario, Canada	Ontario Ministry of Transportation	David Tsui, (416) 235-3538
Lexington, Kentucky	NORPASS, Kentucky Transportation Center	Jennifer Walton, (606) 257-4513

Table 2. TMCs and Agencies Contacted in First and Second Telephone Surveys

Geographic Location	Contacted in First Survey (Spring 1997)	Contacted in Second Survey (Spring 1999)
Phoenix, Arizona	Arizona DOT Traffic Operations Center	Maricopa County DOT Maricopa Association of Governments
Los Angeles, California	Los Angeles (Caltrans District 7) Transportation Management Center	Los Angeles (Caltrans District 7) Transportation Management Center
Orange County, California	—	Orange County (Caltrans District 12) Transportation Management Center
San Francisco, California	San Francisco Bay Area TMC	San Francisco Bay Area TMC
Atlanta, Georgia	Georgia DOT ATMS	Georgia DOT ATMS
Chicago, Illinois	Traffic Systems Center	Traffic Systems Center
Montgomery County, Maryland	Montgomery County TMC	Montgomery County TMC Maryland National Capital Park and Planning Commission
Detroit, Michigan	Michigan ITS Center	Michigan ITS Center
Minneapolis-St. Paul, Minnesota	Mn/DOT TMC	Mn/DOT TMC
Jersey City, New Jersey	TRANSCOM	TRANSCOM
Long Island, New York	INFORM	—
New York City, New York	MetroCommute	—
Ft. Worth, Texas	—	TransVISION
Houston, Texas	TranStar	TranStar
San Antonio, Texas	TransGuide	TransGuide
Seattle, Washington	WSDOT TSMC	WSDOT TSMC
Toronto, Ontario, Canada	COMPASS	COMPASS
Lexington, Kentucky	—	NORPASS (formerly Advantage CVO) (Kentucky Transportation Center)

Note: “—” indicates that personnel at this location were not contacted during the particular survey.

Phoenix, Arizona

The Arizona DOT (ADOT) Traffic Operations Center in Phoenix began operation in September 1995. All of the 20-second loop detector data that comes into the Center are saved. The loop data includes volume, occupancy, and an estimate of speed from the single-loop detectors in the field. Five minute summaries of the data are also created and saved. Previous telephone interviews with TMC personnel indicated that data storage may become a significant concern, and due to the large size of data files used for analysis, it is often difficult to manage the data sets with most desktop spreadsheet and database software.

Efforts at the Maricopa County DOT, in cooperation with the Arizona DOT, are developing a system to meet data archiving and access needs (7). The system is called the AZTech Data Server (ADS), and is shown in Figure 2. The formatted records at the top of the figure represent data inputs from different sources such as incident data or loop detector data. The data spooler/logger provides the quality control and screening of the data as it is received from the different sources and then logs the data. Data are then provided to ETAK, which is currently developing their traveler information service for the public, the ADS, and to other agencies that have a data feed. As shown in Figure 2, it is within the ADS that user-defined queries of the data can be performed for both on-line and off-line applications.

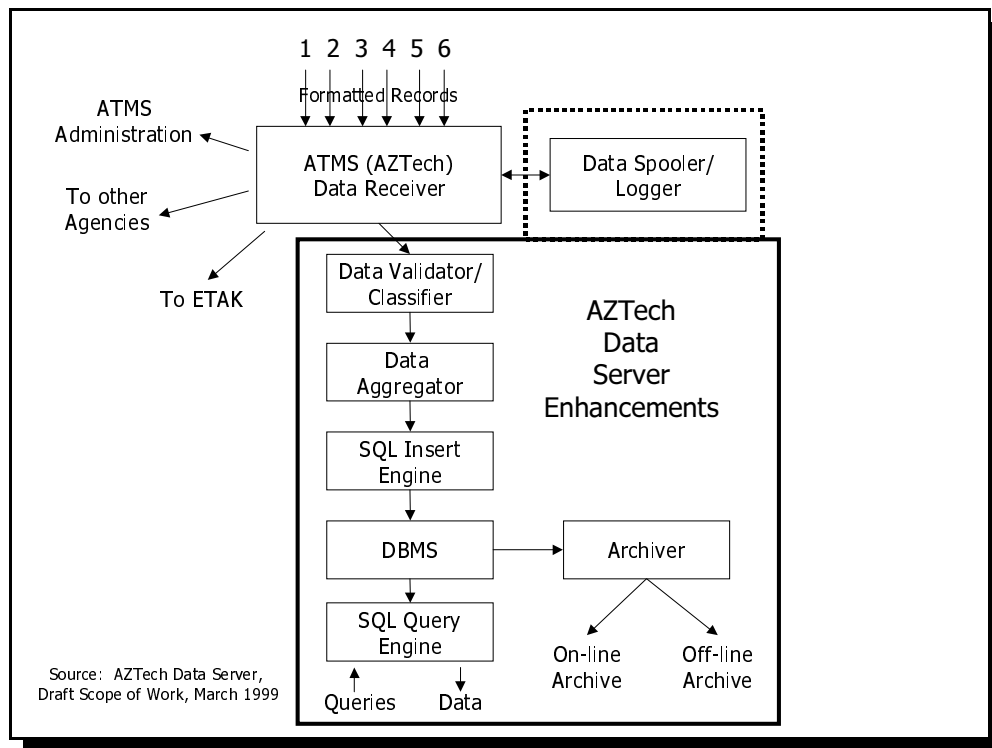


Figure 2. Proposed AZTech Data Server

While loop detector and incident data are currently the primary data inputs, it is anticipated that many other sources will also become available. The objective of the project is to “enhance the ADS to provide and maintain valid, classified ITS-derived data for use in planning, modeling, and real-time operation applications.” Ultimately, the ADS will provide an interface for all data that enter the system. The project is funded at \$500,000 through congestion mitigation and air quality (CMAQ) program funds and is scheduled for completion in September 2000.

The first steps of the project are underway. These include evaluating available user services, data screening and formatting issues, and appropriate validation and quality control checks. It is anticipated that different users will accept differing levels of data validity, and the system developers are considering rule-based methods (with thresholds defined by users) for screening the loop detector data. These data validity rules may be based upon Boolean operators on volume, occupancy, and speed for the data of interest to the user, or they may also allow the user to select statistical validity thresholds (e.g., 30 percent of the weekly average).

The Maricopa Association of Governments is performing a congestion study in the region utilizing data from the ADOT traffic operations center. Speed, volume, and vehicle classification data were desired on freeways and arterial streets. Aerial photographs were used to obtain vehicle density on freeways and queue lengths on arterial streets. One year of 5-minute traffic volume data were obtained from the ADOT center. Video was used to obtain classification data. Fifteen-minute volumes were recorded from the video data, and it was compared to the ADOT volume data. It was found that sometimes the data matched well, and other times the ADOT loop detectors were not operating or the detector volumes were low. The Association has been in contact with ADOT about these data quality issues and was informed that ADOT would take efforts to maintain the loop detectors.

Los Angeles, California

The Los Angeles (Caltrans District 7) TMC is in the process of developing a new system for archiving and accessing ITS data. The TMC currently polls their loop detectors every 30 seconds. Prior to the development of the new archiving system, the TMC saved three days of 30-second data and four days of five-minute summaries into temporary storage. The 30-second data have been permanently archived to magnetic tape cartridges since the TMC was opened, and Caltrans had developed special software that was necessary to retrieve this 30-second data from their mainframe computer.

The new system will provide on-line access to 13 months of data through a relational database for personnel in the Center. Beyond the 13-month time period, the data are planned to be archived to magnetic tape cartridges. In the future, as part of the new system development, it is anticipated that archiving to magnetic tape cartridges will be replaced by archiving the data to compact disc (CD). Future plans also include evaluating different aggregation levels for a given CD (e.g., having a CD for daily summaries, one for 15-minute data, etc.). This CD system would

act as a CD data clearinghouse for interested users. If successful, the system being developed in Los Angeles will serve as the prototype for eventual implementation in all Caltrans TMCs.

Orange County, California

TMC personnel in Los Angeles are developing a new system that will be used at all TMC locations in California, including Orange County. It is their intent to save not only the 30-second data, but also 5-minute, hourly, daily, and monthly data on CD. In Orange County, volume and occupancy data are directly measured from the single loop detectors and speed is estimated.

San Francisco, California

There were not significant changes in the data archiving and management strategies at the San Francisco Bay Area TMC since the surveys were performed in the Spring of 1997. The TMC in San Francisco does not archive data at this time. Personnel at the TMC do recognize the importance of the data for future transportation applications, and they are interested in data archiving activities. They have recently submitted a proposal to use State Planning and Research Funds to evaluate archiving activities and automated techniques.

Atlanta, Georgia

Data archiving practices at the Georgia DOT Advanced Traffic Management System (ATMS) have not changed since the original survey was performed in the Spring of 1997. Volume and occupancy, along with an estimate of speed, are sent from video detection cameras to the ATMS every 20 seconds. The data are not currently saved. Evaluating data archiving issues is of interest to personnel at the ATMS, and efforts to evaluate these considerations will be investigated at a later date.

Chicago, Illinois

The Illinois Traffic Systems Center (TSC) is another TMC within which archiving and data retrieval issues are being strongly re-evaluated. The Center is currently obtaining volume and occupancy data and estimating speeds from approximately 2,000 single-loop detectors that cover 210 km (130 miles) of freeways in the area. The detectors are located on all lanes every 4.8 km (3 miles) and at the entrance and exit ramp locations. At approximately 0.8 km (0.5 mile) intervals, there are detectors in the center lane from which lane occupancy is obtained and travel time is estimated. The 20-second data are aggregated to the five-minute level, and only the travel time and occupancy data are saved at this level from 5 to 10 a.m. and 2 to 7 p.m. Hourly volume data across all lanes are saved from the detector stations spaced at three miles. Both the five-minute and hourly data are permanently archived to magnetic tape cartridges.

The Center personnel have an interest in using the available data for many secondary uses and want to be able to support the demand. Primary users of the data include in-house requests for

lane closures and performance evaluation, the local metropolitan planning organization (MPO), media, and researchers. They have recently hired a consultant to evaluate the data retention and archiving efforts in the Center. Many issues are being considered in this effort including determining what data should be saved, at what level it should be saved, what medium should be used for storage, and what type of automation should be developed. It is anticipated that about one year of data is desired on-line, and a database management system that saves the data to CD, rather than tape cartridges, is also desired. One of the first tasks for the contracted consultant is to evaluate the potential uses and users of the data.

Montgomery County, Maryland

The TMC in Montgomery County, Maryland collects one-minute volume and speed data from loop detectors, then aggregates to the five-minute level before saving. The loop detectors are located on arterial streets downstream of signalized intersections. This information is used to assist in determining offsets and splits for signal timings along the arterial streets.

The Center personnel also have an interest in developing a more strategic ITS data archiving plan. They are currently working on a data archiving effort funded at \$100,000 to identify users and uses of the data, a least common denominator of data aggregation to satisfy these uses, and a method or system to query data of interest about a particular roadway link. Work is being performed to supply data to the local Maryland-National Capital Park and Planning Commission (M-NCPPC). This data application is discussed on page 29 of this report.

Detroit, Michigan

The Michigan ITS (MITS) Center is another TMC that is developing standards on an aggregation level and determining an on-line data access strategy. The Center contains two loop detector systems. The older system covers 52 km (32 miles) of freeway with single loops located at nominal 0.54 km (0.33 mile) spacings. The detectors are scanned by the central traffic control system every 10 milliseconds and the data are aggregated to hourly lane volumes, which are saved to magnetic tape. The newer system covers 240 km (150 miles) of freeway with double-loop detectors (i.e., double loops provide spot speed estimates by using the time difference of arrival) at nominal 3.2 km (2 mile) spacings. These data are aggregated in the field at 20-second intervals and then sent to the Center. The data are then aggregated up to one-minute and the volume, occupancy, and speed are saved. Personnel at the Center hope to eventually fill in additional dual-loop detectors between the existing loops in the newer system. The Center keeps one week of data on-line at the one-minute aggregation level. Data older than one week are saved to magnetic tape cartridges.

Personnel at the MITS Center are interested in developing better access to archived data for a variety of users and uses. They are evaluating the possibility of a common aggregation level to satisfy all users as well as trying to develop a more efficient on-line data access strategy. Center personnel commented that this is difficult since different users inevitably require different types

of data at different aggregation levels. They would also like to incorporate their changeable message sign (CMS) log into this data set. The CMS log includes data such as the time a message was posted or removed, as well as the message content for a given incident. Personnel at the Center also expressed the desire to integrate information that comes into the Center through public telephone and the highway advisory radio (HAR) into the database. They will likely consider CD or digital video/versatile disk (DVD) for archiving rather than the magnetic tape cartridges that are currently used.

Loop detector failures are another concern at MITS. In the hourly volume summaries that are saved, a measure of the “lane operability” is also saved. The measure is the loop-minutes of failure, which is calculated as the number of loops failing times the number of minutes failing. When making temporal or spatial comparisons, this measure provides some insight into the extent to which a loop detector(s) may not be reporting. For example, if the loop operability equals nine, this could mean that three loops are out for three minutes, or one loop is out for nine minutes. This does not evaluate the extent that volumes may be reading too high or too low, and these data quality issues are also a concern at the Center. To date, the primary users of the archived data are in-house for freeway management.

Minneapolis-St. Paul, Minnesota

The Mn/DOT TMC monitors over 85 percent of the freeways in the Minneapolis-St. Paul metropolitan area. The most significant source of ITS data is from 3,800 loop detectors. The detectors provide volume and occupancy. The speeds can be estimated based on an assumed vehicle length. These data are obtained every 30 seconds and every five minutes. The five-minute data are aggregated and saved. Thirty-second station data is stored in a binary day file and is available upon request. Daily five-minute information is stored to a CD and about six months of data can be saved to one CD. The process is automated and the data have been saved since a computer conversion in 1993. To date, only the loop data have been archived; however, there is interest in eventually saving ramp metering data (e.g., vehicle release rates) as well.

Most requests for the archived data are from the DOT for traffic analysis, construction impact determination, and planning applications. Researchers from local universities often request the data as well. Another significant use of the data is for the evaluation of ramp metering strategies. This use has fostered an interest in a more detailed study and calibration of the loops. This effort is explained in greater detail on page 34 of this report.

TRANSCOM, New York/New Jersey/Connecticut

TRANSCOM (the Transportation Operations Coordinating Committee) opened in 1986 to coordinate traffic management in New York, New Jersey, and Connecticut. TRANSCOM’s System for Managing Incidents and Traffic (TRANSMIT) utilizes vehicles with automatic vehicle identification (AVI) transponders on the vehicles. Data archiving practices at TRANSCOM had not changed since they were contacted as part of the original survey in 1997.

Raw tag reads for each vehicle are not saved, and software at the TMC saves the data into 15-minute periods.

Ft. Worth, Texas

The TransVISION TMC in Fort Worth, Texas is scheduled to open by September 1999. The TransVISION system is currently under development, and many questions are still being addressed. These questions include (8):

- What type of data is available?
- Do we need to save it all?
- Will it add value to our TMC system operations?
- In what form should that data be saved?
- How do we share this data with other agencies in our area, including the City of Arlington, the Dallas District, and others?

The ultimate system design for TransVISION includes a relational database for on-line storage of archived ITS data; however, the funds are not yet available for building this data archiving component. The data anticipated in the center include traffic incident information, ramp meter timing, dynamic message sign (DMS) stored messages, lane control signal (LCS) patterns, geographic information system (GIS) data, traffic signal timing, closed circuit television (CCTV) control functions, electronic work orders, detection system data, road construction, maintenance, and special events data. These data will be organized and logged around four areas including incident data, closure data, operator data, and detector data.

There are five major subsystems to the TransVISION center: communications, video, dynamic traffic management, data management, and the operating platform. The data management subsystem architecture is shown in Figure 3. The subsystem includes both an on-line transaction processing (OLTP) server and an on-line analytical processing (OLAP) database server. The OLTP server provides data storage for the most recent month. At periodic intervals, data will be replicated (i.e., duplicated and verified) from the OLTP server to the OLAP server. After about one year, the data from the OLAP server will be archived to permanent storage media, such as a CD. As additional financial resources become available, TransVISION personnel would like to expand the system beyond the 30-day limitation within the OLTP server.

The OLTP server manages the relational database and the real-time operations. The most recent information is held in the OLTP, and this information is updated in real-time. The OLTP server will provide short-term storage of about one month of data. With more than 2,000 loops, 100 bytes per loop, and 1 record every 50 seconds, the data load in the OLTP is about 10.3 gigabytes per month. When implemented, the OLAP server will interact automatically with the OLTP server to perform data replication (i.e., duplication and verification). The OLAP database will also provide for queries of the data that are available for the last year, and the database will permanently archive data to either CD or magnetic tape cartridges.

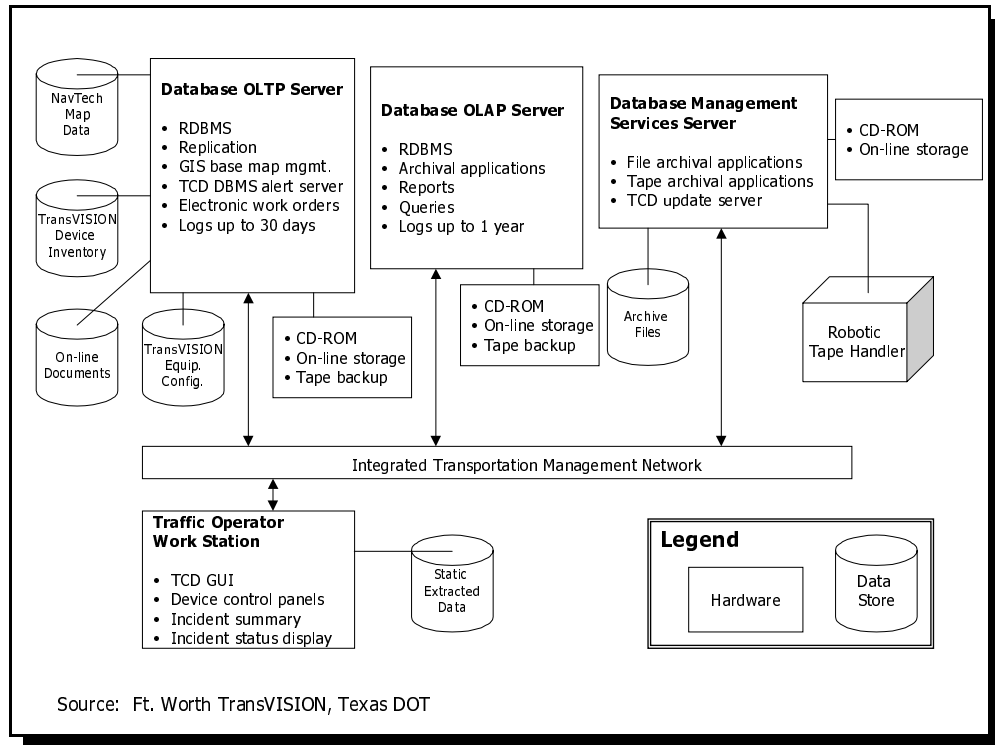


Figure 3. Proposed Data Management System in Ft. Worth's TransVISION

Many issues are still being considered in the development of the TransVISION data archiving system. These include:

- Is the current design direction adequate?
- Does new technology cause a change in direction for archived storage?
- How much do we aggregate the data and how long do we keep it?
- Which specific data will other agencies be interested in?
- How can others connect to this data considering security, system load, etc.?

Houston, Texas

The Houston TranStar Center provides traffic management for the Houston region's freeway system. Inductance loops and video detection units are used on a portion of the freeway system, but the data from these devices are not archived. Houston also has an extensive AVI system in operation. Over 300,000 AVI tags are in circulation in the metropolitan area, and tag reads are sent to the Center in real-time. The AVI travel time data are stored in 15-minute summaries for future use. In addition, AVI data have been used for quantifying the benefits of ramp metering and high-occupancy vehicle (HOV) lanes.

In a recent ITS data archiving workshop in Texas, information resource personnel for TxDOT's Houston District and TranStar discussed current data warehousing development activities (9). The information resource manager discussed how the statewide ITS integrator, Southwest Research Institute (SwRI), can help TranStar define user requirements for the development of an enterprise data management system. Approximately \$1.2 million is currently estimated for this data warehousing system, and one terabyte of data storage is being planned.

The information resource administrator noted that some data warehouse efforts have failed because the system was not designed to meet end user requirements. Therefore, the determination of end user requirements by SwRI is an important component of the project effort. Table 3 illustrates an example of a matrix of user requirements that will be developed for TranStar's data warehouse.

Table 3. Example of User Requirements Matrix for Houston TranStar

User Requirements	Functions	Agency	Type of Data
<ul style="list-style-type: none"> • List of data elements based on user requirements 	<ul style="list-style-type: none"> • Planning • Construction • Operation • Maintenance 	<ul style="list-style-type: none"> • City • County • TxDOT • Transit • MPO 	<ul style="list-style-type: none"> • Planned or Proposed
<p>User Requirements Questions:</p> <ul style="list-style-type: none"> • Who is the end user? • Where are they? • What data do they need? • How are they using it? • Does it meet their needs? • Is it still needed? 			

Source: adapted from Wegmann and Gloyna 1998 (9)

As a precursor to the enterprise-wide data warehousing plans, the Houston-Galveston Area Council (HGAC) is working with Houston TranStar on a pilot project that will provide HGAC with archived ITS data for their planning and modeling applications. The project, which was funded through the ITS Priority Corridors program, is focused on archiving loop detector data from selected locations in Houston. This pilot effort began in late 1998.

San Antonio, Texas

Phase One of the TransGuide Center, complete since 1995, has double-loop detector coverage over 42 km (26 miles) of freeways at nominal 0.8 km (0.5 mile) spacings. TransGuide also has an AVI system that covers about 98 miles of freeways and limited access highways.

TransGuide computer servers poll the local loop controllers every 20 seconds, and this 20-second data is made available over the Internet (<http://www.transguide.dot.state.tx.us/statistics.html>) at the end of every day. Providing the data via the Internet reduces the time spent distributing data that is requested. The Internet site typically contains the most recent month of loop detector data. Data are archived to tape, but it is anticipated that storage will be moving to CD in the near future.

Data that are collected at TransGuide can be categorized into five types (10) as described below, with the operational data being the most common type made publicly available:

- **operational data** - real-time traffic data from the loop detector and AVI systems, scenario execution logs, incident information, and the database of nominal travel times on non-instrumented routes.
- **real-time procedural data** - user access logs, system accounting records, and resource utilization logs.
- **performance data** - system error logs, system access logs, and database performance monitoring logs.
- **system data** - system data include the source code, executable programs, design documents, equipment configurations, and network management databases.
- **administrative data** - operator logs, equipment maintenance records, and system configuration data.

TransGuide is committed to providing the data that are available through the TMC in an accessible and efficient manner as encouraged in TxDOT's Information Sharing Policy.

Data from the TransGuide system have been used by TTI researchers for the development of an ITS data management system (DataLink) (5,6). DataLink provides a user-friendly interface capable of querying the TransGuide data for different roadway facilities, time periods, and aggregation levels. Such a system provides a common repository of information for individuals that may want to utilize the information for different applications (e.g., operations, incident management, planning). Researchers at TTI are also using incident detection and response logs from the TransGuide system to evaluate their incident management program.

Seattle, Washington

The data archiving strategies at the WSDOT TMC in the Puget Sound region have not changed since personnel were contacted for the first survey in 1997. Volume, occupancy, and an estimate of speed are collected and sent to the ATMS every 20 seconds. The loop detector data are stored at the five-minute aggregation level and it currently requires four CDs to save one year of five-minute data.

Toronto, Ontario, Canada

COMPASS is a traffic management system that monitors portions of Highway 401 in Toronto, Canada. The center monitors over 2,800 loops as well as numerous video surveillance cameras. The loop detectors are polled every 20 seconds and report volume, occupancy, and an average speed from an assumed vehicle length for the single-loop detector stations. Whenever maintenance is required of the highway pavement, double-loop detectors are being installed to replace the single-loop detectors.

The 20-second data are aggregated to five-minute, 15-minute, one hour, daily, and monthly time periods. The TMC archives speed, volume, and occupancy data for 20-second and five-minute time increments. For data summaries of 15 minutes or more, only volume data are saved. The five-minute time increment was selected because it appeared to provide a convenient time increment for many users. Common archived data users and uses include in-house requests and researchers, who desire 20-second data for simulation and algorithms. In-house requests account for about 60 percent of all data requests (e.g., traffic forecasts, roadway impact analyses).

There are other systems in operation in the Toronto area as well. The COMPASS system is along Highway 401 as described above. Two additional COMPASS systems are along the Queen Elizabeth Way. One TMC is located in Mississauga and the other is in Burlington that monitors the skyway bridge. The latter two systems collect data at 30-seconds and the data are archived at the same aggregation levels as the Highway 401 COMPASS system described above.

The data have traditionally been archived on 8-mm magnetic tape cartridges. The tapes require special hardware and software to download the data, and data users found it relatively difficult to obtain historical data from this system. Further, over extended periods of time (i.e., greater than about five years), the tapes begin to lose reliability and potentially data.

Since about 1997, the COMPASS system has been saving the data to CD, and they have been developing software to extracting data from a CD. The CD itself is being developed to contain software that will create and display tabular summaries of interest to the user. COMPASS personnel are currently assessing the appropriate and most useful format and aggregation level for the CDs for a variety of users and uses (i.e., should the CD contain only 20-second or hourly data?). This decision will likely be made depending upon the demand for the CDs that are being

developed at the different aggregation levels. One other difficulty of the CD application is that it operates slowly.

The sharing of archived ITS data to numerous widespread users is a relatively new effort at COMPASS and they are in the process of developing information use and sharing policies. The experience at COMPASS illustrates the interest in designing systems that promote more efficient storage and retrieval of archived ITS data.

State of Virginia

The Virginia DOT (VDOT) is developing an enterprise data warehouse system that will integrate the nine legacy systems within the Department (11). One of the many data sources being considered in this integration are archived ITS data. The intent of the data integration is to bring all of the different systems together into one unified data system. The nine legacy systems of the Department operate on four different database systems. Adding to the complexity, there are three hardware platforms and three operating systems in use as well. The system under development at VDOT will allow for the screening, quality control, and summarization of the data as it is input into the data warehouse. The enterprise data model describes the planning data that are the most common throughout the agency and serves as the foundation for the data integration effort. Fully understanding the data uses prior to system design is important.

VDOT stresses the importance of integrating the databases that are currently available for different uses. Without full integration, and appropriate connections between the different databases, “data islands” may occur. These are databases that cannot be accessed through the integrated system and are left out of the larger archived data system. This is especially important to ensure that database systems provide regional integration and can share information.

Another observation that is clear with the VDOT experience is that the Department has assigned significant resources for the data warehouse project. This demonstrates the importance of these growing concerns and the recognition that providing archived ITS data and information through an integrated system is a necessary role of transportation agencies. VDOT has assembled a collective group of representatives collectively known as the MIS 2000 committee that is assigned to act as a functional enterprise level authority for different functional areas and different districts. By functioning across different divisions and districts, it can also help in strategic systems planning efforts. In addition, a configuration control board has been established that is composed of personnel from both VDOT and TransCore to provide input into technical aspects of the data warehouse. Data custodians have also been assigned to monitor data for the data warehouse that is specific to a transportation subject matter. Finally, a data management division has been staffed with data management professionals to provide leadership on the technical aspects of the data management. These groups all provide valuable insight into the development and continued success of a data management system that incorporates data from many sources into one comprehensive data warehouse.

North America Pre-clearance and Safety Systems

The NORPASS (North America Pre-clearance and Safety Systems) program has been recently established to assist in the interoperability of safe and efficient pre-clearance of trucks at weigh/inspection stations. NORPASS merges the CVO project based at the Kentucky Transportation Center (Advantage CVO) and the Multi-jurisdictional Automated Preclearance System (MAPS) programs as well as other states and provinces. CVO provides a unique ITS data source that could have many secondary applications (12). However, there are considerable privacy issues that arise when using WIM and AVI transponders. Truckers are concerned that the data can potentially be saved with such a system and used against them for enforcement. Therefore, privacy concerns are high in the use of CVO data. In fact, one of the largest issues facing the electronic screening programs is ensuring that the operations do not provide stricter enforcement on vehicles equipped with transponders and to ensure that the carriers are kept anonymous. To assist in relieving such concerns, Advantage CVO prepared a data policy, which is illustrated in Figure 4.

Available CVO data typically include the carrier and truck identification, trip history, weight, and vehicle classification. From adjacent weigh-in-motion (WIM) detectors, data such as the date and time stamp, axle and gross weight, vehicle classification, and speed are collected. As stated in the policy, Advantage CVO does not retain any carrier-specific or vehicle-specific data for over 24 hours. One common concern among truckers is that the date/time stamp will not be consistent with their log books. The second concern is that the speed measured at consecutive WIM sites will be used for speed enforcement. However, these data are not saved for over 24 hours or used for enforcement purposes. There is not as much concern over the actual weight data that are recorded because that is what is traditionally recorded at the weigh stations and is understood by the truckers. The ITS America CVO Policy Subcommittee has adopted “Fair Information Principles for ITS/CVO” and the “ITS/CVO Interoperability Guiding Principles” that describe their position on secondary uses of CVO data (available at <http://www.itsa.org/cvo.html>).

PRESS RELEASE

Tuesday, August 25, 1998



Advantage CVO adopts data policy to ensure continued protection of motor carrier privacy.

On July 29, 1998, the Advantage CVO Policy Committee approved an **Operations Center Data Policy** which specifies procedures for the collection, handling and distribution of electronic data. Under this policy, no carrier-specific or vehicle-specific data will be retained any longer than 24 hours, either at weigh stations or at the Advantage Operations Center. Participating state agencies will be provided with summary data that is equivalent in nature to the data collected manually at that state's weigh stations. No additional data will be collected as a result of a carrier participating in electronic screening.

Since Advantage CVO does not charge transaction fees for participating carriers, there is no need to collect and retain transaction data for billing purposes. This allows the Operations Center to discard data on a daily basis, or even more frequently if it becomes necessary.

In the four years since Advantage CVO first began operations, the system has granted over half a million green lights (i.e., weigh station bypasses) to participating motor carriers on Interstate-75 and Canadian Route 401. During that time, there has never been even a single instance of transaction data collected by Advantage CVO being used for enforcement purposes. Even with this outstanding track record, the Advantage CVO Policy Committee chose to implement a formal policy to reassure participating carriers that data will not be used inappropriately in the future.

The Advantage CVO Policy Committee includes representatives from State/Provincial agencies in Ontario, Michigan, Ohio, Kentucky, Georgia, and Florida, as well as from the National Private Truck Council and several major trucking companies.

For more information contact:

Joe Crabtree, Kentucky Transportation Center

606-257-4508, fax 606-257-1815, e-mail crabtree@engr.uky.edu

Figure 4. Press Release Describing Advantage CVO Data Policy

STATE-OF-THE-PRACTICE IN USING ARCHIVED ITS DATA

The previous section described existing ITS data archiving practices at selected TMCs and other operations centers. Personnel from many of the TMCs indicated that they receive data requests for a variety of transportation applications. There are numerous stakeholder groups for ITS data archiving because of the utility and variety of ITS-generated data items. Table 4 summarizes the key stakeholder groups and their primary applications for archived ITS data. These stakeholders and their applications may help in determining appropriate designs for ITS data archiving systems. This section describes several of the stakeholders shown in Table 4 in further detail, including discussion of the following:

- typical data elements of interest;
- spatial/temporal aggregation levels of interest;
- additional data characteristics of importance; and
- example applications.

The applications highlighted in this section represent the most common applications of archived ITS data to date; therefore, example applications are illustrated for a few of many several stakeholder groups.

Table 4. Stakeholders and Example Applications for Archived ITS Data

Stakeholder Group	Primary Transportation-Related Functions	Example Applications
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	<ul style="list-style-type: none"> • congestion monitoring • link speeds for TDF 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)	<ul style="list-style-type: none"> • pre-planned control strategies (ramp metering and signal timing) • highway capacity analysis • saturation flow rate determination • microscopic traffic simulation <ul style="list-style-type: none"> -- historical -- short-term prediction of traffic conditions • dynamic traffic assignment • incident management • congestion pricing operations • evaluation and performance monitoring

Table 4. Stakeholders and Example Applications for Archived ITS Data (Cont.)

Stakeholder Group	Primary Transportation-Related Functions	Example Applications
Transit operators	Day-to-day transit operations: scheduling, route delineation, fare pricing, vehicle maintenance; transit management systems; evaluation and planning	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • emission rate modeling • urban airshed modeling
MPO/state freight and intermodal planners	Planning for intermodal freight transfer and port facilities	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • HazMat response and enforcement • congestion management • 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	<ul style="list-style-type: none"> • 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	<ul style="list-style-type: none"> • 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: Margiotta 1998 (4), pp. 4-5.

Metropolitan Planning Organizations / State Transportation Planners

The first stakeholder group identified in Table 4 are MPOs and state transportation planners. These users commonly require ITS data for uses such as congestion monitoring, system performance evaluation, determination of average traffic statistics, and multi-modal demand estimation. Transportation planning applications often require data at a relatively aggregate level. For example, hourly or even daily summaries of volume data and hourly summaries of speed data are used for long-range planning, system performance evaluation, or congestion monitoring. Most planning applications typically do not require data at a more disaggregate level than 15 minutes. Many planning applications (e.g., congestion management, system performance evaluation) require data from several different segments and locations throughout a region. These applications generally require daily summary statistics; however, peak hours and peak periods are also often of interest. The remainder of this section will describe applications of archived ITS data for planning applications.

Example: Freeway Performance Evaluation in Puget Sound, Washington

Researchers at the Washington State Transportation Center (TRAC), in cooperation with WSDOT, have prepared a report on facility usage and freeway performance in the central Puget Sound area using 1997 archived ITS data (13). Data for the study were obtained from the WSDOT FLOW system inductance loop detectors. These detectors are placed at approximately 0.8 km (0.5 mile) spacings and obtain volume and occupancy data. Speeds are estimated from the single detectors. These data are aggregated at five-minute intervals and archived to CD.

The results of the WSDOT FLOW evaluation produce valuable information for system performance and congestion management by monitoring traffic volumes throughout the network and reporting vehicle speeds and when and where congestion is occurring. Facility usage data are presented for various facilities for the peak period and peak hour. The presentation methods in the evaluation were designed to illustrate temporal and spatial aspects of congestion for non-technical audiences (e.g., policy makers, legislators). For example, Figure 5 presents the congestion level in each direction of travel along SR-520.

Figure 6 presents the results of average facility travel times for a given hypothetical travel route on the general-purpose lanes. The figure shows the average general-purpose travel time, the 90th percentile general-purpose travel time, and the percent of time the speed is less than 45 mph (66 km/hr) for a given trip start time (congestion frequency). Figure 7 shows a similar graphic that presents the congestion frequency by time of day along with the corresponding traffic volumes. The line representing the volume is presented in colors that indicate the speed of the traffic as well.

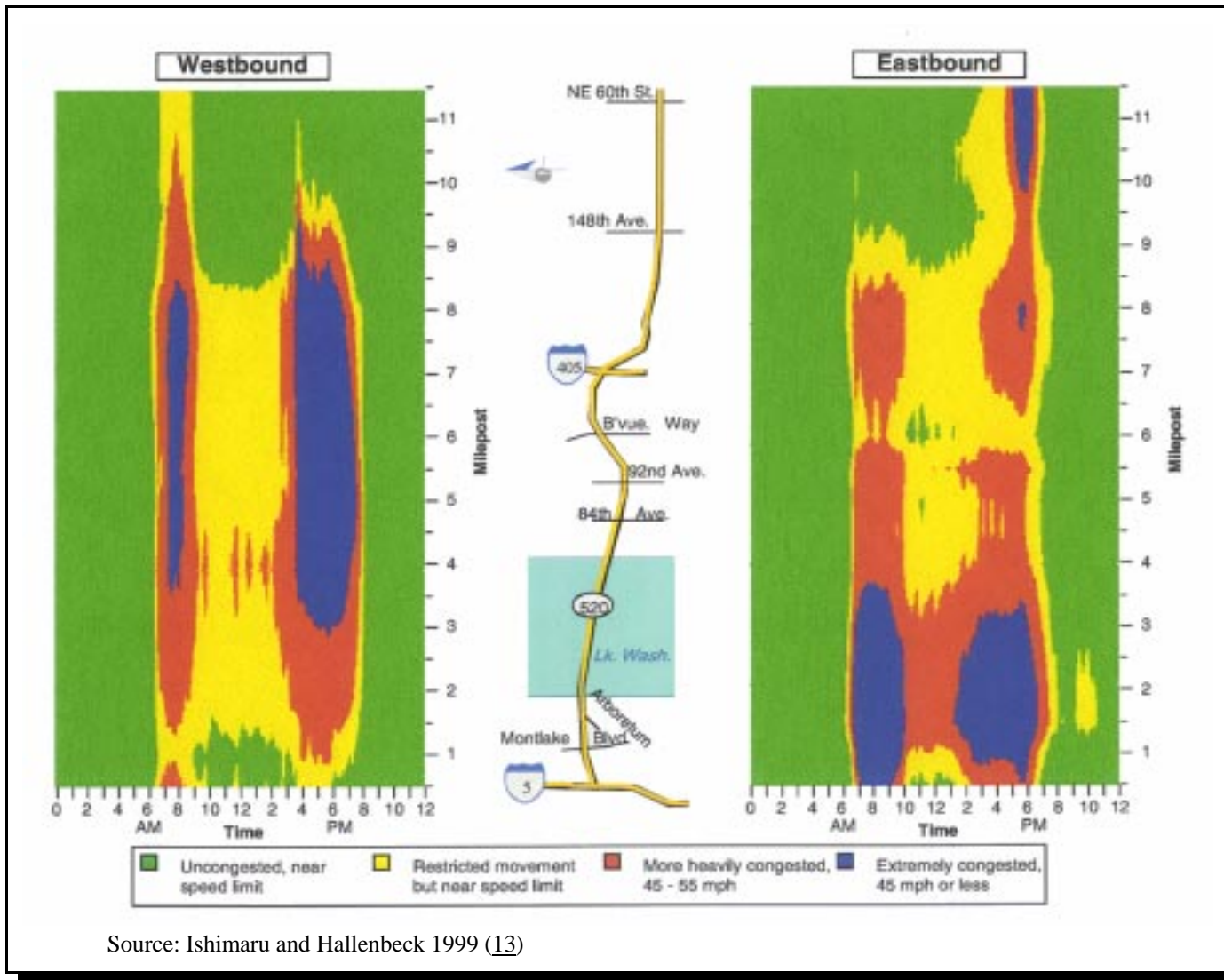
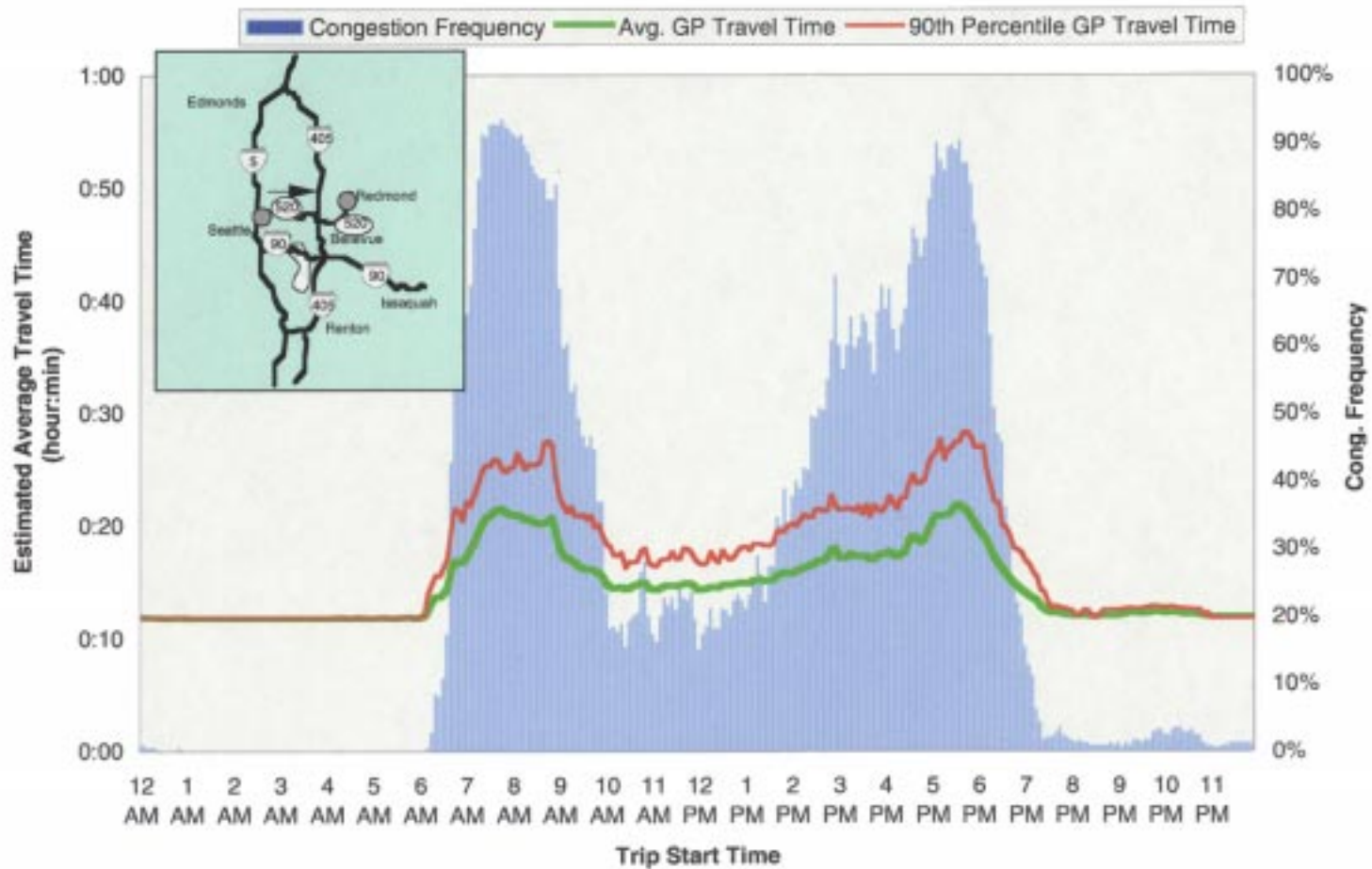


Figure 5. Traffic Profile of SR-520 General Purpose Lanes



Source: Ishimaru and Hallenbeck 1999 (13)

Figure 6. Estimated Average Weekday Travel Time for Eastbound SR-520

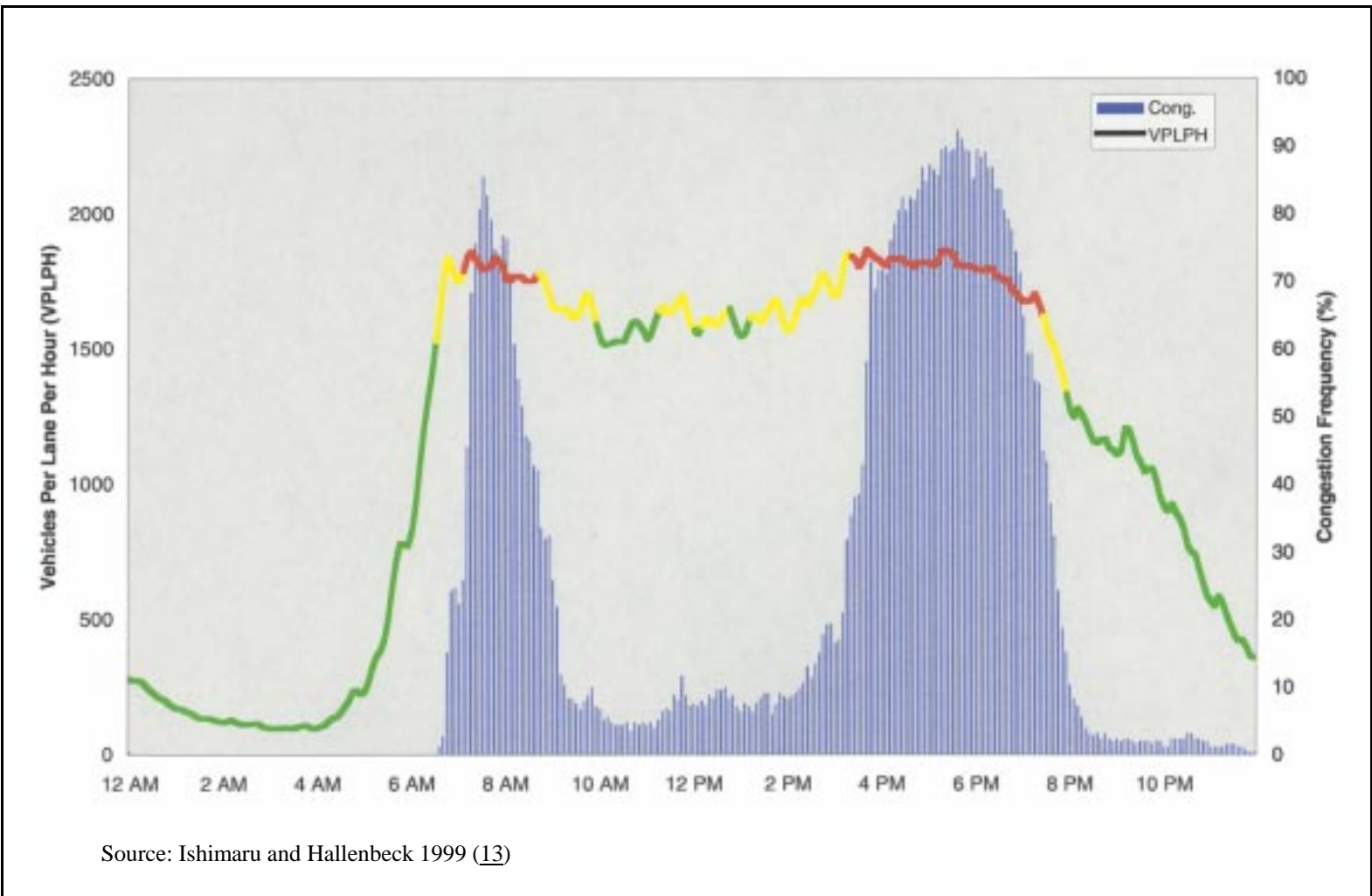


Figure 7. Estimated Weekday Volume, Speed, and Reliability Conditions for Westbound SR-520, 76th Avenue N.E.

Example: Evaluation of Houston HOV Lanes

TTI researchers evaluated three HOV corridors in the Houston area using archived AVI data from 1994 (14,15). Eight months of the AVI probe vehicle data were used to determine travel time savings and reliability along three major freeway and HOV corridors.

The results of the travel time analysis indicated that the HOV lanes provided a substantial travel time savings over the adjacent freeway general-purpose lanes. Savings were proportional to the level of congestion being experienced in the adjacent freeway general-purpose lanes. The extensive data set available with the AVI traffic monitoring system provided detailed information about day-to-day speeds and variability of speeds. For example, Figure 8 shows large variations in daily peak hour travel times for the freeway general-purpose lanes with small variations on the HOV lane. Travel time variability was evaluated with measures of standard deviation and confidence intervals. For example, Figure 9 supports one of the asserted benefits of HOV lanes, in that they provide higher levels of travel time reliability than freeway general-purpose lanes. The results of this study provided valuable information for HOV system performance measurement.

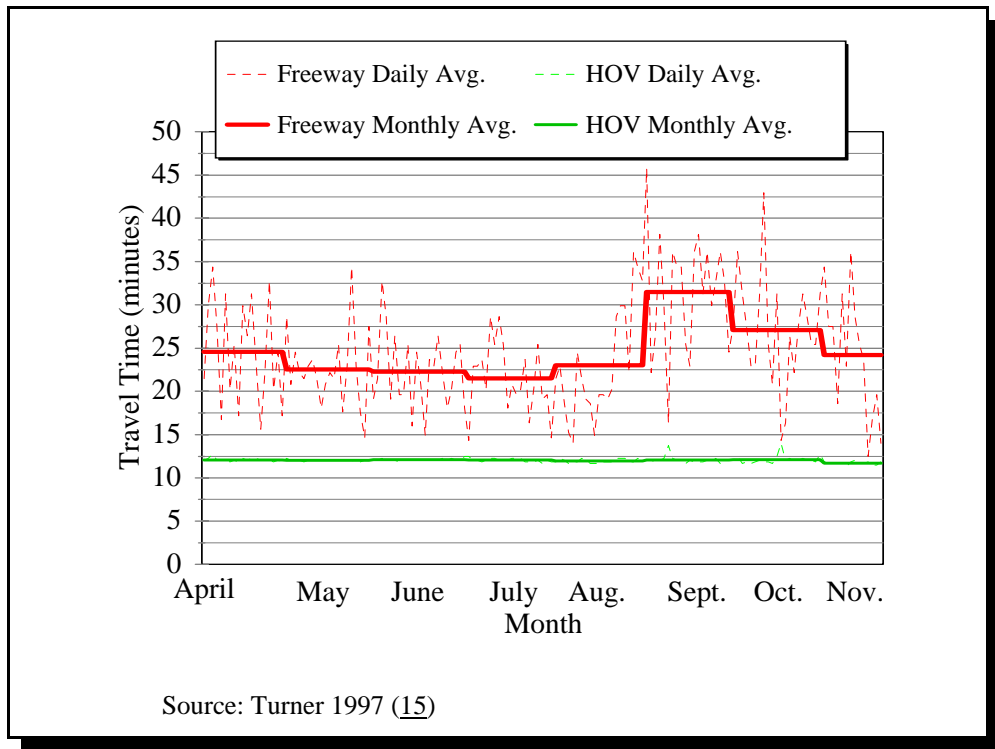


Figure 8. Average Morning Peak Hour Travel Times for Katy Freeway, 1994

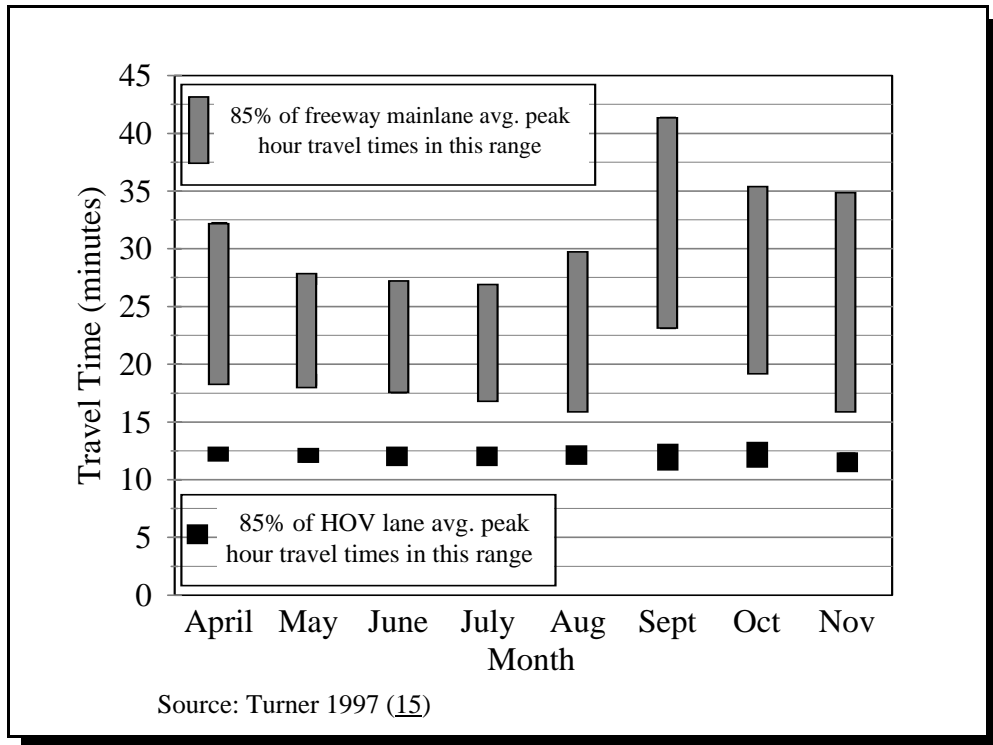


Figure 9. Morning Peak Hour Travel Time Reliability for Katy Freeway, 1994

Example: National Traffic Performance Indicators

Research sponsored by the Bureau of Transportation Statistics (BTS) is being performed at the Oak Ridge National Laboratory to develop real-time traffic indicators using data collected by ITS sensors (16). An Internet site documents this research effort at <http://micron4.ed.ornl.gov/>. This research effort is exploring the possibility of using data available from TMCs throughout the country to develop national performance indicators in real-time. The research is using total vehicle miles of travel (VMT) and total vehicle-hours of delay as performance measures. An automated process has been developed to obtain and process ITS traffic condition data from four cities in the U.S.: San Diego, California; Milwaukee, Wisconsin; San Antonio, Texas; and Seattle, Washington. The authors have noted many useful results (16):

- The minute-by-minute data for individual cities reveal that delay is highly sensitive to small changes in system average speeds;
- The individual cities' system-wide travel and congestion indices show well defined daily and weekly patterns;

- There is a need for greater consistency in definitions, methods, and conventions for ITS data;
- There is a need for further addressing the following technical issues: error handling, data capture, statistical inference, and data reduction and storage; and
- There are many possible uses of a developed national set of indicators, such as quantifying congestion trends and monitoring other congestion attributes like delay per VMT or throughput per lane-mile.

Example: Traffic Statistics in Chicago, Illinois

Planners at the Chicago Area Transportation Study (CATS) have been working with personnel at the Illinois TSC to use 1995 archived ITS data for planning applications (17). CATS has used the data available from the TSC to develop a travel atlas for 1995 for the Chicago metropolitan area. The document provides a valuable resource to planners for information such as annual average daily traffic (AADT) volumes. Additional information is illustrated in the document, including monthly seasonal factors, day-of-the-week traffic variation, holiday travel rates, hourly variation of traffic, and estimated travel times and speeds.

Example: Traffic Statistics and Congestion Monitoring, Montgomery County, Maryland

Personnel at the Montgomery County TMC collect arterial street traffic volume and speed data and are working with staff at the Maryland-National Capital Park and Planning Commission (M-NCPPC) to supply them with traffic volume data. The driving force for such an effort is limited budget allocations for data collection programs. Planning personnel see the use of archived ITS data as an opportunity to save money in the data collection effort. Increasingly, there appears to be more interest from operations personnel in developing working relationships to facilitate the coordination of such efforts.

The M-NCPPC has worked closely with TMC staff that are performing data archiving activities. M-NCPPC planners view the archived ITS data as an opportunity to obtain traffic volume data for primary arterial facilities in the area at a reduced cost and increased safety. Funding has been secured and personnel are developing a prototype interface for displaying volumes and speeds near intersections. The prototype interface is anticipated to be completed in April 1999. Obtaining traffic volume data from the Center has saved considerable financial resources for M-NCPPC.

One drawback that was noted is the accuracy of the speed data from the arterial street detectors. Since the loop detectors are located downstream of signalized intersections, there is concern about the accuracy of the speed estimates. The intent is to use the hourly and daily data summaries to evaluate daily and seasonal variations in traffic characteristics. There is also

interest in evaluating more disaggregate data levels since the data are available for closer study of peaking characteristics. Although the prototype interface is anticipated in April 1999, this effort is relatively new and there appears to be a rather large learning curve for the relational database used to store the data.

Another concern is precise geographic location references or the Montgomery County TMC's loop detectors. For example, the Center has the detectors and intersections numbered, but to match this information with a map, some type of referencing must be performed. It is anticipated that software will need to be developed to determine the grid coordinates on a map to refer to the detector and intersection locations.

Example: Traffic Statistics and Patterns, Phoenix, Arizona

Transportation planners at the Maricopa Association of Governments are using archived ITS data provided by the ADOT Traffic Operations Center. The planners are using the archived ITS data to study traffic patterns, characteristics, and trends in the Phoenix area. For example, Figure 10 illustrates average weekday traffic patterns by hour of the day. Figure 11 illustrates monthly traffic variation for freeways using ITS data. Figure 12 displays a traffic speed-volume relationship at a given freeway detector location.

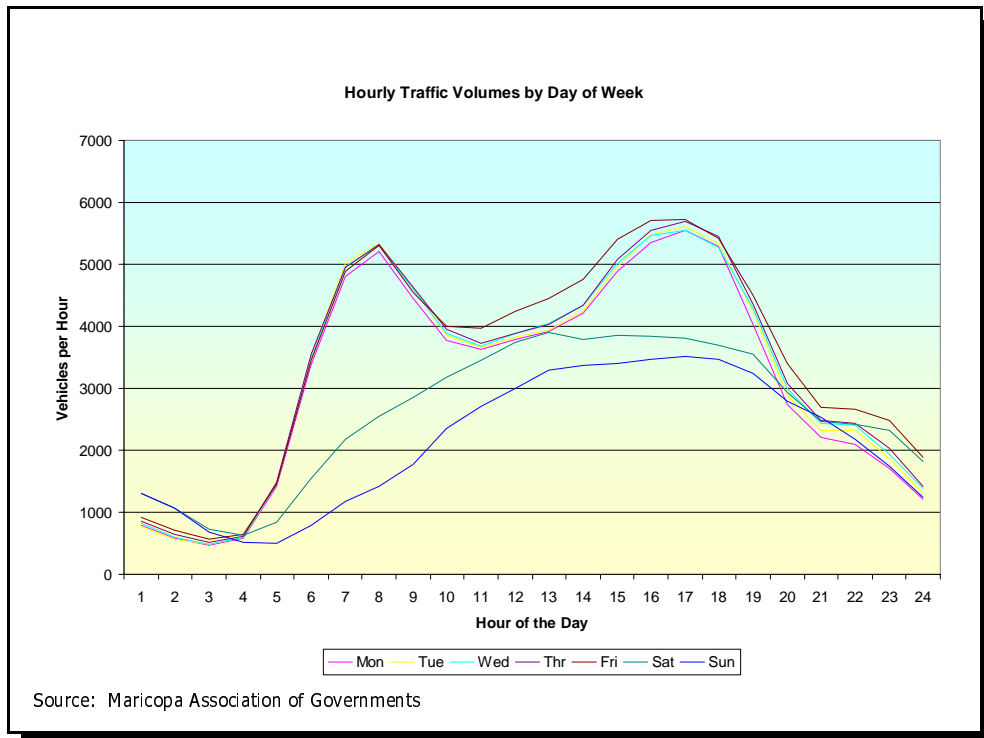


Figure 10. Average Weekday Traffic Volume Patterns Derived from ITS Data, Phoenix Urban Area

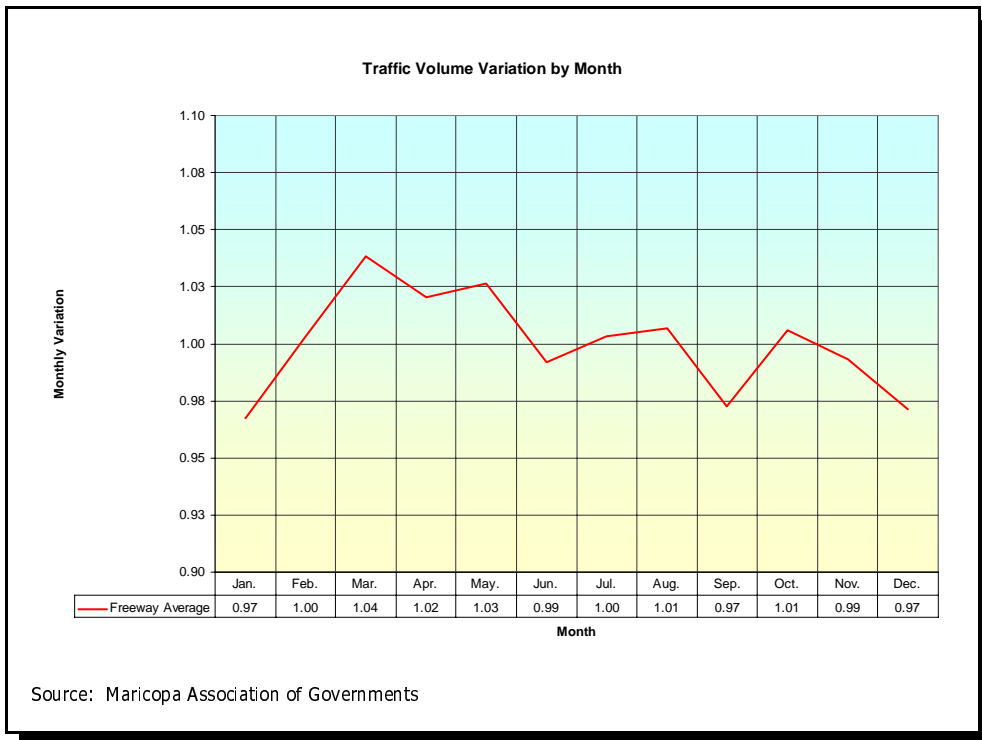


Figure 11. Monthly Traffic Variations for Freeways Using ITS Data, Phoenix Urban Area

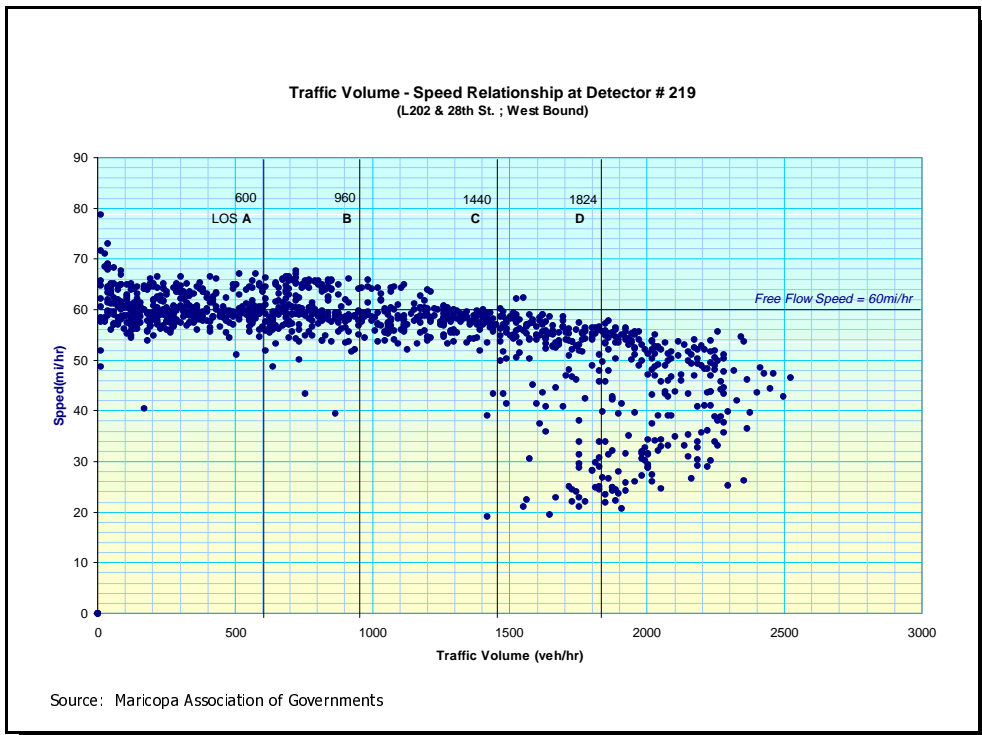


Figure 12. Traffic Volume-Speed Relationship Using ITS Data

Example: Regional ITS Planning, Puget Sound, Washington

In regional ITS planning efforts in the Seattle area, researchers at Mitretek Systems have developed scenarios to produce probability maps of the variability in roadway capacity from incidents and extreme weather for a range of demands (18). The hypothesis of the work is that ITS strategies are most effective during high-variability situations (Figure 13). Several data sources were used in the case study analysis in the Seattle area, including archived ITS data from the regional TMC. Additional data on accidents, weather, and incidents were also gathered. A scenario hierarchy was developed that began by separating a condition into an event/non-event, then by the weather condition, incident condition, volume ratio, and then accidents. This approach provides a mapping of incident and weather scenarios by demand level as shown in Figure 14. This approach also provides more insight than traditional travel demand forecasting techniques by considering how ITS may affect travel.

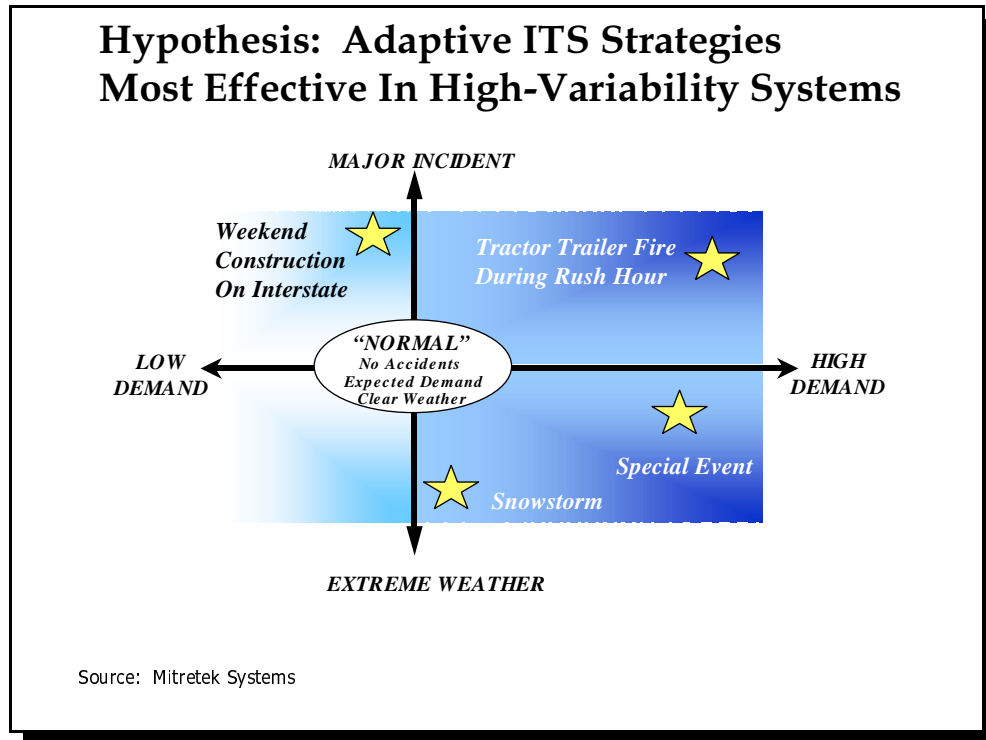


Figure 13. Illustration of Hypothesis that ITS Strategies are Most Effective in High-Variability Systems

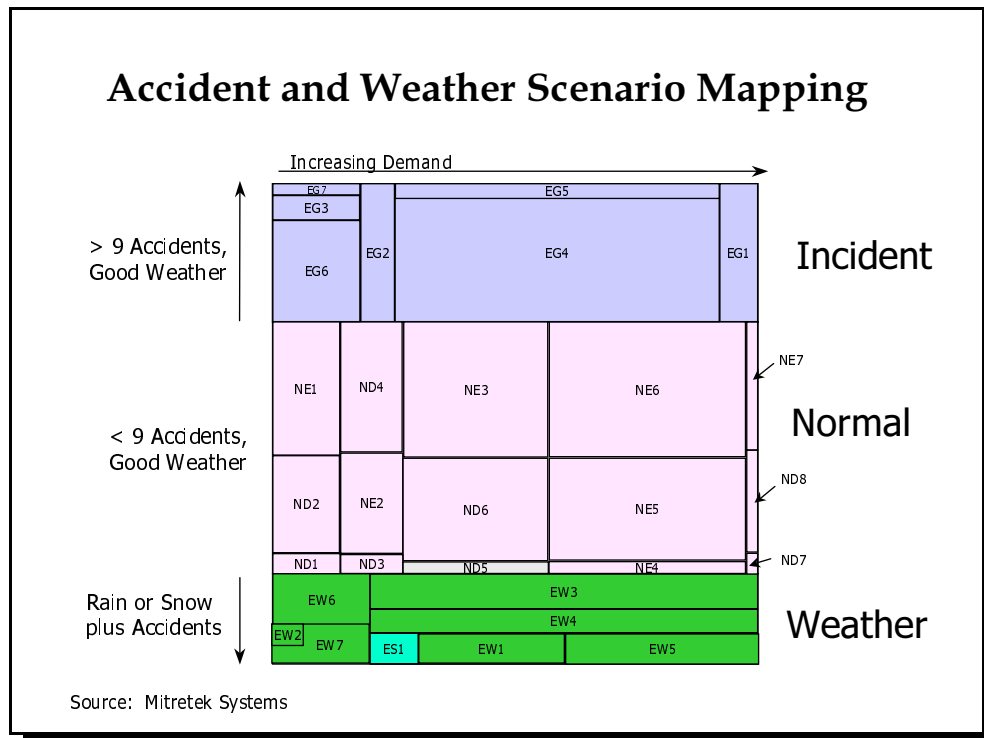


Figure 14. Accident and Weather Scenario Mapping by Demand Level Using Archived ITS Data and Other Data Sources

Example: Alternatives Analysis and Model Calibration, Houston, Texas

Research at TTI has evaluated the benefits of HOV lanes relative to other transportation improvements through a range of congestion levels (14). Five different alternatives were evaluated: 1) no-build; 2) add an HOV lane; 3) add a general-purpose lane in each direction; 4) add two general-purpose lanes in each direction; and 5) implement transportation system management (TSM) practices to increase the freeway capacity by ten percent. The measures of effectiveness (MOE) that were evaluated included speed, delay, emissions, and fuel consumption. Archived AVI probe vehicle data were aggregated into 15-minute time periods for calibration of baseline FREQ10PL models. Once the baseline models were calibrated, the various alternatives were analyzed.

Additionally, TTI researchers used the same 1994 AVI travel time data set to analyze variability to determine required sample sizes (19). The study found that existing sample sizes, or the number of AVI-equipped vehicles, varied by location and proximity to Houston’s toll facilities. In most 15-minute periods during the peak hour, about 15 to 20 probe vehicles reported travel times on the freeway general-purpose lanes. On the HOV lanes, about 5 to 10 probe vehicles reported travel times during most 15-minute periods in the peak hour. The study concluded that the existing probe vehicle system was providing reliable peak period travel speed information.

Traffic Management Operators

Traffic management operators are the second stakeholder group identified in Table 4. These stakeholders are concerned primarily with the day-to-day operations of TMCs. Archived data applications of interest to traffic management operators include developing traffic control strategies, performing highway capacity analyses, incident management, and performance evaluation and monitoring. Data requirements for these analyses are often at a more disaggregate level than the transportation planning applications discussed in the previous section. Data are often desired at aggregation levels of 1, 5, or 15 minutes. Some applications that benefit traffic management operators (e.g., incident detection algorithm development) may require data at intervals of less than 1 minute. Typical applications evaluating highway capacity or ramp-metering strategies may utilize data at the 5-minute aggregation level. These applications often require data for various segments and time periods.

Example: Ramp Metering Evaluation, Mn/DOT TMC

Personnel at the Mn/DOT TMC monitor an extensive ramp metering system that is an integral component of traffic operations. TMC personnel had a growing concern that the ramp detectors in the system were not providing reliable data. Therefore, calibration of the loops was performed to ensure that accurate data were being used as input to the ramp metering system. The occupancy and volume of detector stations were plotted to evaluate the resulting speed estimate provided by the loop data from an estimate of vehicle length. The data were evaluated during relatively free-flow traffic conditions when there are more accurate estimates of the speed along the particular segment. The loop detector data can then be calibrated accordingly with the resulting calibration factor for the particular detector.

The archived ITS data has proved valuable for improving ramp metering operations. Due to the importance of the ramp metering system and the accuracy of the loop data as input to such a system, the work is being further pursued. After reviewing the archived ITS data, TMC staff have realized that even the addition of as little as 5 to 10 vehicles in a 5-minute period, or a change of as little as 2 percent occupancy, can have a dramatic impact on traffic operations.

An issue that was noted during conversations was the geographic location referencing for the detectors. They are in the process of developing a GIS database to establish more accurate, easily-defined detector station locations.

Example: Ramp Metering and Incident Management Evaluation in Texas

TTI researchers assisted TxDOT in the evaluation of ramp metering on the Katy Freeway (I-10) in Houston, Texas through the use of 15-minute summaries of archived AVI data. The study evaluated travel times along several freeway corridors that contain ramp metering. This work provided insight to traffic management operators that the ramp-metering strategies were successful and the ramp metering system was expanded (20).

Researchers at TTI are also using the incident detection logs from the TransGuide[®] system in San Antonio to evaluate incident time events such as when certain types of incidents/events occur, how long it takes to execute a response, when incidents are responded to, and how long it takes to clear the incident. The archived incident log data provide time stamps and location references of all incident-related events. The results of this research will be valuable in assisting transportation management operators in effectively responding to incidents on the San Antonio freeway system.

Maintenance and Construction

Maintenance and construction stakeholders require data for pavement maintenance, roadside maintenance, or lane closure scheduling. Lane closure scheduling was one of the applications most frequently mentioned in the previously discussed surveys. Hourly traffic volume and/or speed data are often required to determine when lanes should be closed and re-opened to cause the smallest impact to motorists. Hourly traffic volumes are also of value for pavement and bridge maintenance to determine the amount of infrastructure use to estimate maintenance needs.

Transportation Researchers

Transportation researchers typically desire the most disaggregate data of all the stakeholders identified in Table 4. They will often use data at the most disaggregate form available (i.e., 20-second or individual probe vehicle data). Sample applications include forecast and simulation model development, theoretical analysis of highway capacity, and incident detection modeling. Advanced data users such as transportation researchers generally have a significant amount of computing power and data manipulation tools at their disposal to assist them in analysis of large data sets. Data are often desired over many segments and different time periods for research purposes for close study of temporal and spatial changes.

Example: Traffic Flow Characteristics Research, California

Research performed at the Institute of Transportation Studies (ITS) at the University of California at Berkeley has used ITS data supplied by loop detectors to evaluate traffic features of freeway bottlenecks on two freeways in Toronto, Ontario, Canada (21). Speed, occupancy, and volume data were collected every 30 seconds along the Queen Elizabeth Way and the same data were collected every 20 seconds along the Gardiner Expressway. The study evaluated several freeway bottleneck measures including the flow immediately prior to queue, discharge rate immediately following the queue, recovery discharge rate, average discharge rate, and the percent difference between the flow immediately prior to the queue and the average discharge rate. Findings of interest include the following (21):

- Flow can drop substantially following the formation of an upstream queue;

- The discharge flows in active bottlenecks exhibit near-stationary patterns that (slowly) alternate about a constant rate;
- A bottleneck's long-run discharge rates are consistent from day to day, while other flow features exhibit daily variation; and
- Bottlenecks occur at fixed (i.e., reproducible) locations.

The research team further discovered that the difference between the flow immediately prior to the queue formation and the average discharge rate can be as high as 10 percent. The research team indicated that many of the effects of the freeway bottlenecks are presented and that the actual causes are being studied. This includes the reasons for the driver behavior that caused the relatively high vehicle flows to cause the bottlenecks or the reasons the queues were always observed at about one kilometer (0.6 mile) downstream or more from the on-ramp of interest. Understanding the driver behavior will allow for better decision-making for traffic management operations such as ramp metering and changeable message signs. Finally, the research team anticipates the evaluation of freeway segments in the U.S. to determine the similarity between U.S. freeways and those studied in Toronto.

Example: Development of Congestion Prediction Model

The National Institute of Statistical Sciences (NISS) and Bell Laboratories has performed research using loop detector data for predicting congestion in real-time (22). This research is valuable for developing models that may predict congestion so that traffic operators can make traffic management changes (i.e., altering ramp-metering strategies or changeable message signs appropriately). Twenty-second data were obtained from the Seattle area, which include traffic volume and lane occupancy obtained from the single detectors. The data were aggregated to the one-minute level and breakdown was defined as, “the first of five consecutive one-minute intervals in which the estimated speed over the detector at that location is less than 48 km/hr (30 mph). Similarly, a breakdown ends with the first of five consecutive one-minute intervals in which speeds exceed 48 km/hr (30 mph)” (22). The study corridor was an 18.8 km (11.7 mile) section of I-5 in Seattle with 24 detectors in the northbound direction and 23 detectors in the southbound direction. Speeds were estimated from the occupancy and volume data using an assumed vehicle length, and the data were averaged across lanes at each detector station.

A tree-structured CART classification was used in the study to predict a breakdown or no breakdown given the set of predictor variables. The predictor variables are a function of the volume, occupancy, and speed data. The conclusions of the study were as follows (22):

- Data quality is a significant concern and about 40 percent of the data were discarded;

- CART was found to be a useful tool for constructing models to predict breakdown based upon speed and occupancy;
- Using more than one detector station increases the predictability of breakdown further into the future;
- There was a gain in predictability by using the finer detail of the one-minute data over the 5-minute data in both the training data set and the test set with respect to detection rate, although the difference averaged only about three percent.

Example: Development of Travel Time Prediction Models, Texas

Research underway at TTI has been using archived AVI probe vehicle data from Houston TranStar for both travel time forecasting and origin-destination estimation. One research effort evaluated the use of an artificial neural network (ANN) for use in estimating travel times in the near future (23). Five-minute time periods were used for the prediction. The ANN model was used since it allows for the use of non-linear relationships for travel time prediction that include both spatial and temporal components. When predicting one or two periods into the future (i.e., 5 to 10 minutes), the ANN model that considered only the previous travel time information from that link gave the best estimates. When predicting three to five time periods into the future (i.e., 15 to 25 minutes), the ANN model that considered both spatial and temporal inputs including the travel time estimates from the upstream and downstream links, as well as the link of interest, provided the best prediction ability.

Similar research has also compared the ANN modeling with other methods of travel time prediction (24,25). The first step of the modeling methodology was to group the data into similar clusters with an unsupervised clustering technique. The second step was to calibrate and test the ANN model with actual travel time data obtained from the Houston TranStar AVI system. It was found that the ANN provided the best results over several other existing methods of estimating link travel times including a Kalman filtering model, an exponential smoothing model, using historical profiles, or using a real-time profile. The latter methods of using real-time profiles are those often employed by most advanced transportation management systems.

Continued research has also evaluated the use of spectral basis neural networks (SNNs) (26). This research has found that SNN's direct forecasting approach provides better travel time prediction results than the other models studied. Further, this research effort also evaluated the correlation of travel time along a route. It was found, based on both theoretical and empirical results, that the correlation between link travel times affects the accuracy of the resulting route travel time. An interesting result is that the error in link travel times are not additive and the error associated with the route travel times may actually be lower than the error in link travel times.

Example: Estimation of Origin-Destination Matrices, Texas

Archived AVI probe vehicle data available from the Houston TranStar system are also being used for origin-destination estimation (27). The basic premise of this research is that while the AVI data has traditionally been used for real-time link speeds, it also provides a sample of point-to-point trip estimates as well. A computer program has been developed that can trace individual vehicles from the start of their journey on the Houston freeway system to their destination. This information can then be used to derive a sample origin-destination matrix, which is then used to estimate the complete freeway origin-destination matrix. The research found that trip length frequency distributions (TLFDs) vary by time of day and day of the week, although there was similarity among weekdays. The information provided by the research could be used for such applications as estimating parameters for highway-based trip distribution models (e.g., gravity models) or validating planning models with information about link or route travel times as well as freeway trip volumes. Essentially, the matrix of information can provide an inexpensive cordon count of spatial and temporal information in the network. Future research in this area will evaluate using the sampling of route choices, the estimation of link choice proportions, methods for relating the AVI origin-destination methods to the network origin-destination pairs, developing origin-destination matrix data for different time periods, and the real-time application of origin-destination for traffic operations. The potential for combining real-time traffic volume counts with the AVI origin-destination to forecast real-time origin-destination matrices on the I-10 corridor in Houston is also being studied.

Example: Traffic Flow Theory Research, Ontario

Research has been performed at McMaster University in Hamilton, Ontario, Canada to study speed-flow relationships during congested freeway conditions (28). This research is valuable to operations personnel as well as planners that desire a better understanding of the speed-flow relationship under congested conditions. Loop detector data (speed, volume, occupancy) along the Gardiner Expressway and Highway 401 were used in the study. Data were aggregated to the five-minute level for analysis and suspect data were removed from the data set. In addition, data that represented congested conditions was selected from the data available (i.e., data from transitions between congested and uncongested were removed from the data set). The analysis tested many different models including a cubic equation, power function, and an exponential function, and the R^2 values for the equations ranged from 0.75 to 0.78. The research also addressed questions such as whether the equation should be forced through the origin, the influence of construction data, and whether different freeways require their own functions. The research findings were as follows (28):

- It is necessary to have data from a range of flows to fit the curve for the congested portion of the speed-flow curve;
- It is important to remove data from transitions between the congested and uncongested data regimes;

- The relationship for construction periods seems to be different than when construction is not present;
- There may likely be a difference in the speed-flow relationship for different freeways due to different lane widths or other geometric factors;
- Downstream operating conditions affect the speed-flow relationship within congested periods; and
- Several sites may likely need to be combined for development of the speed-flow curves during congested periods, yet care must be taken in determining the sites and conditions that may be combined.

Private Sector

The private sector is also a significant stakeholder for archived ITS data. Many private sector companies are developing business models to add value to the data that are available through ITS and then reselling the value-added data. One business model includes the private sector receiving the raw archived data from the public sector agencies that collect the data from publicly-owned infrastructure. The private sector then reduces the large data sets to useful information for targeted users (e.g., state transportation planners) in return for a fee. Another business model has private sector (e.g., telecommunications companies) collecting data from privately-owned infrastructure, then selling the real-time and historical archived data to various public sector markets. There are many issues related to ownership of the data, intellectual property, and privacy concerns, to name only a few, that must be addressed when considering the private sector stakeholder.

ADVANCING THE STATE-OF-THE-PRACTICE IN ITS DATA ARCHIVING

This section describes activities and efforts focused on advancing the state-of-the-practice in the archiving or use of ITS data (Table 5). Several of these activities are concerned with development of the ADUS in the National ITS Architecture and related data standards. Several researchers are examining issues that affect ITS data archiving practices or data applications. It is clear from the discussion that many technical and institutional issues remain to be addressed before the retention and use of archived ITS data becomes standard operating procedure.

Table 5. Efforts to Advance the State-of-the-Practice in Archived ITS Data

Effort	Anticipated Contribution to ITS Data Archiving	Resources
Development of ADUS Architecture	Provide framework for deploying ADUS applications.	<i>ITS as a Data Resource: Preliminary Requirements for a User Service (4)</i> and National ITS Architecture, Version 3.0 (to include ADUS architecture)
Data Standards	Provide data dictionaries and meta attributes that provide capability for unambiguous exchange and storage of ITS data.	Standards documentation at http://www.its.dot.gov/standard/itssum.pdf (also see Table 8)
ITS Data Registry	Provide centralized clearinghouse for all data element descriptions and associated meta attributes.	Preliminary documentation at http://grouper.ieee.org/groups/scc32/datasetreg/DRFOP8.DOC .
Volpe Center Case Studies	Communicate lessons learned, opportunities, and guidance on the use of ITS data with case studies in four cities (Chicago, Detroit, Minneapolis, and San Antonio).	Projected deliverables include technical memoranda, four case study reports, and a final report. Final report publicly available later in 1999.
Oak Ridge National Lab/Florida DOT/New York DOT Demonstration Project	Identify institutional, technological, communication, data, and logistical issues that need to be addressed if and/or when ITS data is used for traffic monitoring needs.	Projected deliverables include technical memoranda and task reports. Expected study completion date in mid-2000.
Washington State	Develop a methodology, detailed procedures, and tools for conducting ongoing performance evaluations of the FLOW traffic management system.	<i>Central Puget Sound Freeway Network Usage and Performance (13)</i>
Texas Transportation Institute, TransLink® ITS Research Program	Development, maintenance, and expansion of DataLink, an on-line prototype ITS data management system; helped sponsor Texas ITS data archiving workshop; investigating performance measures using archived ITS data.	<i>ITS Data Management System: Year One Activities (5)</i> ; DataLink web site (http://vixen.cs.tamu.edu); and <i>Proceedings from the Texas ITS Data Uses and Archiving Workshop (29)</i>

Development of the ADUS Architecture

The ADUS is currently being defined and developed as a component of the National ITS Architecture, the purpose of which is to serve as a guide in the development of an integrated, multi-modal intelligent transportation system. It is envisioned that the ADUS architecture will provide a framework in which transportation information collected by ITS could be made available to a wide variety of stakeholders for data analysis and exploration. The following paragraphs describe the development of the ADUS architecture (also see Table 6).

The first organized efforts to create the ADUS began in 1997, subsequent to the first release (Version 1.0) of the National ITS Architecture. At the time, the Architecture did include a planning subsystem (Figure 15), which had been designed mainly as a “place holder” and had no user service to support. Because user services describe ITS functions from a user’s perspective, it was incumbent upon the planning subsystem users and stakeholders to formally request the inclusion of a data user service in the Architecture. FHWA’s Office of Highway Information Management (now Office of Highway Policy Information) led this effort on behalf of the data stakeholders, with ITS America providing support in assembling stakeholders for early discussions.

Based upon several stakeholder meetings held in 1998, an addendum to the ITS program plan and a set of user service requirements were developed to guide the creation and development of ADUS, the 31st user service. In January 1998, numerous stakeholder groups gathered to discuss the desirability of and issues associated with creating ADUS at the “ITS as a Data Resource” workshop held in Washington, D.C. An ADUS resource document was created from this initial meeting that contained preliminary user requirements (4). Stakeholder comments and input from a subsequent workshop in Seattle in July 1998 were used to finalize the ADUS addendum to the ITS Program Plan and the User Service Requirements. By late 1998, the U.S. DOT had begun work on developing the ADUS component of the National ITS Architecture.

At the time of this writing, the details of the ADUS architecture are being developed by U.S. DOT and its contractors in conjunction with numerous stakeholders at the local, state, and federal levels. The ADUS revision to the Architecture is scheduled for completion in August 1999, and once revised, will be available as part of the National ITS Architecture documentation (Version 3.0 or later). More information on the ADUS revision can be found at the following web sites:

- ITS American ADUS Resource Page: <http://www.itsa.org>
- U.S DOT National ITS Architecture: <http://www.its.dot.gov/arch/>
- Odetics Browsable National ITS Architecture:
<http://www.odetics.com/itsarch/>

Table 6. Incorporation of the ADUS into the National ITS Architecture

Action	Date
Champion for ADUS emerges	1997
Initial stakeholder meeting	January 1998
Position paper with preliminary requirements	April 1998
Second stakeholder meeting	July 1998
Addendum to the <i>ITS Program Plan</i>	September 1998
<i>User Service Requirements</i>	October 1998
Architecture Revision begins	December 1998
Subsequent stakeholder meetings	January 1999 March 1999 July 1999
Architecture revision complete	approx. September 1999

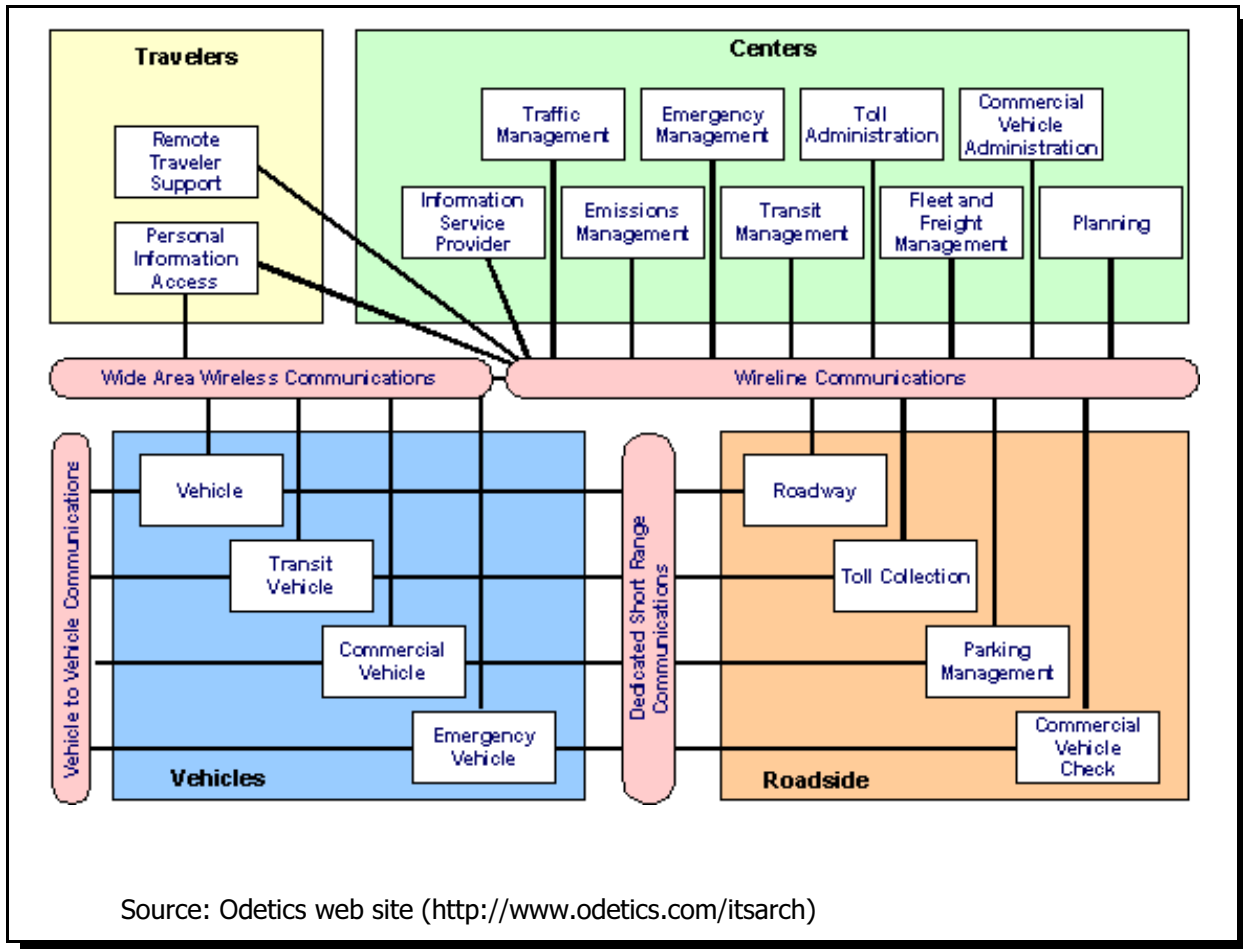


Figure 15. National ITS Architecture

Data Standards Development

Numerous transportation data standards are being developed to ensure interoperability in the implementation of ITS technologies, many of which are relevant to the ADUS. When implemented, these standards prescribe the attributes and format of data elements collected, communicated, or stored by ITS applications or components. Using the Institute of Transportation Engineer's (ITE's) Traffic Management Data Dictionary (TMDD) standards as an example, a vehicle count from a traffic detector has numerous attributes as shown in Table 7.

Table 7. Attributes of a Vehicle Detector Count Using ITE's TMDD

Data Attribute	Data Attribute Value
Descriptive Name:	DETECTOR_VehicleCount_quantity
Descriptive Name Context:	Manage Traffic
Definition:	The number of vehicles detected by a detector during a specific time period.
Class Name:	Traffic Detectors
Classification Scheme Name:	IEEE P1489, Annex B
Classification Scheme:	19971009, V0.0.7
Keywords:	Detector Vehicle count
Related Data Concept:	LINK_Volume_rate, DETECTOR_IdNumber_number
Relationship Type:	Possible identifier for a link volume or a detector count
ASN1 Name:	Detector-vehicle-count
ANS1 Data Type:	Integer
Representation Class Term:	Quantity
Value Domain:	SI 10-1997; vehicles
Valid Value Range:	0 to 100000
Valid Value List:	Not Applicable
Valid Value Rule:	Not Applicable
Internal Representation Layout:	999999
Internal Layout Maximum Size:	SIZE(32)
Internal Layout Minimum Size:	SIZE(32)
Remarks:	V1.1 - Revised Definition and Permissible Data Element Values.
Data Concept Identifier:	3515
Data Concept Version:	V1.3
Submitter Organization Name:	TMDD
Last Change:	19980610

Source: ITE TMDD Section 3, Version 1.3, June 1998

Table 8 lists selected data standards that are most relevant to the development of ADUS applications. The ITS standards setting effort has been industry-driven, with several standards development organizations (SDOs) leading various data standardization efforts. With numerous data standards efforts being led by several different SDOs, there is a potential for overlaps and/or conflicts between the various data standards efforts. The ITS Data Registry has been suggested as a solution to potential overlaps and/or conflicts between different data standards or dictionaries.

Table 8. Selected ITS Standards Relevant to ITS Data Archiving

Title of Standard	Standard Description	Lead SDO	For More Information
Advanced Traffic Management System (ATMS) Data Dictionary (TMDD): Sections 1 and 2 (Links/Nodes/Events) [TM 1.01]	A functional-level data dictionary for traffic management applications. Describes and standardizes roadway links and nodes in accordance with location-referencing message standards data elements for incidents and traffic-disruptive roadway events.	ITE	ITE Web Site: http://www.ite.org/standards/index.html
ATMS Data Dictionary (TMDD): Sections 1 and 2 (DMS/Video/Control/Etc.) [TM 1.02]	A functional-level data dictionary for traffic management applications. Includes data elements for traffic control, traffic detectors, actuated signal controllers, traffic modeling, vehicle probes, ramp metering data, dynamic message signs, video and camera control, parking management and weather stations.	ITE	ITE Web Site: http://www.ite.org/standards/index.html
NTCIP Data Collection and Monitoring Devices [AASHTO TS 3.DCM]	Specifies object definitions that may be supported by data collection and monitoring devices, such as roadway loop detectors.	AASHTO	NTCIP Web Site: http://www.ntcip.org
NTCIP Object Definitions for Transportation Sensor Systems [AASHTO TS 3.TSS]	Object definitions that are specific to and guide the data exchange content between advanced sensors and other devices in an NTCIP network. Advanced sensors may include video-based detection sensors, inductive loop detectors, sonic detectors, infrared detectors, infrared detectors and microwave/radar detectors.	AASHTO	NTCIP Web Site: http://www.ntcip.org
Template for ITS Message Sets [IEEE P1488]	Describes the structure and content of message sets for exchange between traffic centers, emergency management centers, and traveler information systems in a consistent and uniform manner.	IEEE	IEEE Web Site: http://www.ieee.org
Standard for Data Dictionaries for ITS [IEEE P1489]	Specifies a common set of meta entities and meta attributes for ITS data dictionaries, as well as associated conventions and schema, that enable describing, standardizing, and managing all ITS data.	IEEE	IEEE Web Site: http://www.ieee.org

Source: <http://www.its.dot.gov/standard/itssum.pdf>, Feb. 12, 1999.

ITS Data Registry

Sponsored by the Institute of Electrical and Electronic Engineers (IEEE), the ITS Data Registry is envisioned to be a central repository for all ITS data element definitions (30). The Data Registry will not replace any standard setting efforts, but will supplement these efforts by serving as an arbitrator of overlaps and/or conflicts in data element definitions. The Data Registry will classify all submitted data elements into one of seven quality status categories (Figure 16).

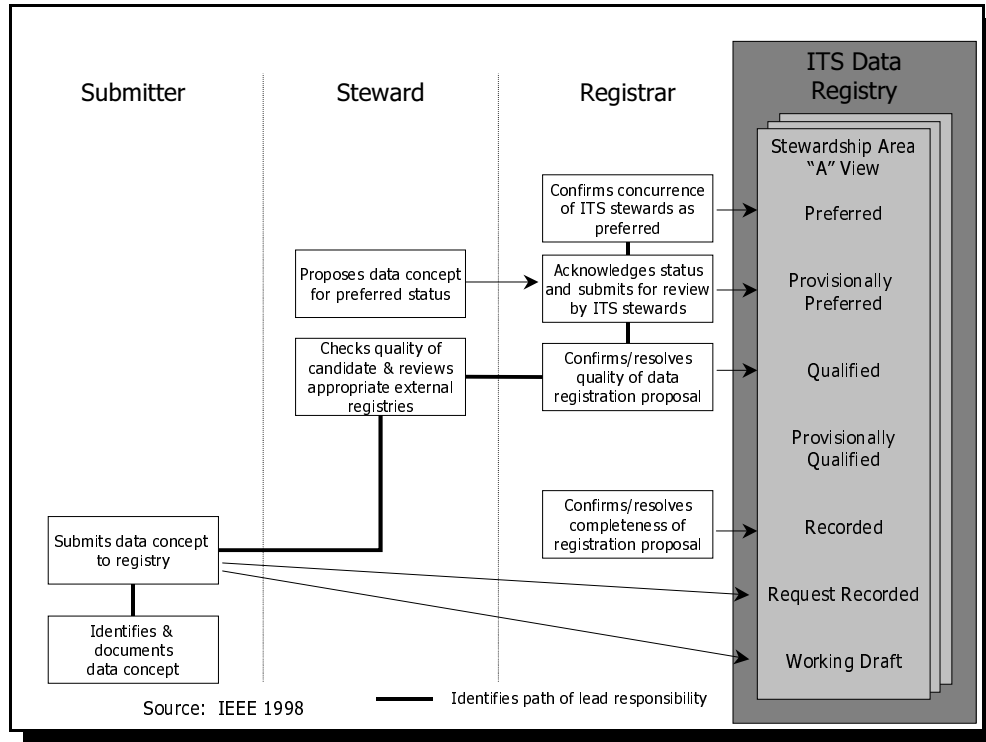


Figure 16. ITS Data Registry Functional Concepts

At the time of this writing, only the functional concepts related to the ITS data registry have been defined (31). The operation of the ITS Data Registry has not officially started, nor has the prototyping of a registry system been initiated. However, small-scale testing of the ITS Data Registry is expected in the near future. Access to data elements in the ITS Data Registry will likely be provided on a per agency basis, in which a single fee covers access to all Registry data elements for an entire agency.

Volpe Center Case Studies

Researchers at the Volpe Center are performing case studies on the use of ITS data for measuring transportation system performance and other planning applications (32). The case studies are being performed in four cities:

- Chicago, Illinois;
- Detroit, Michigan;
- Minneapolis, Minnesota; and
- San Antonio, Texas.

The research study has the following five objectives:

1. Provide an overview and inventory of current uses of data on both passenger and commercial vehicle use of the transportation system;
2. Identify system performance data needs and the requirements to meet those needs;
3. Identify opportunities for the full and effective use of ITS-generated data in transportation planning and system management;
4. Perform cross-cutting analysis to identify how ITS-generated data can contribute to better decision making and can be used to improve the metropolitan planning process nationwide; and
5. Communicate lessons learned, opportunities, and guidance to state and local transportation professionals on the effective use of ITS-generated data.

The case studies were completed in 1998 and a draft project report has been prepared. It is expected that the project report will be publicly available later in 1999.

Oak Ridge National Lab/Florida DOT/New York DOT Demonstration Project

Researchers at the Oak Ridge National Laboratory have initiated a study with the Florida and New York DOTs to assess the feasibility of using ITS data in traditional traffic monitoring programs (33). The demonstration project started in November 1998 and is projected for completion in 18 months. The researchers are attempting to address the following questions:

- Are ITS-generated data able to provide or estimate needed traffic monitoring information?
- If so, how close are the ITS-based estimates to those based on traditional data collection efforts?
- If so, what are the institutional, technological, communication, data, and logistical issues that need to be addressed? And how might they be addressed?

- When fully operational, can ITS-generated data be integrated with the more traditional data collection efforts?

The research study consists of the following three tasks:

1. Identify Data Gaps and Survey Lessons Learned
 - Inventory data collected under ITS deployments
 - Evaluate the adequacy of ITS-generated data
 - Identify barriers in using ITS-generated data
 - Characterize data gaps
 - Survey lessons learned
2. Develop, Instrument, and Test Procedure to Estimate Traffic Parameters
 - Develop schematic to link ITS-generated data and information needs
 - Estimate regional traffic parameters based on ITS-generated data
 - Quantify the precision of ITS-based traffic data estimates
 - Develop a communications procedure
 - Summarize results and document final procedure
3. Recommend Future Activities

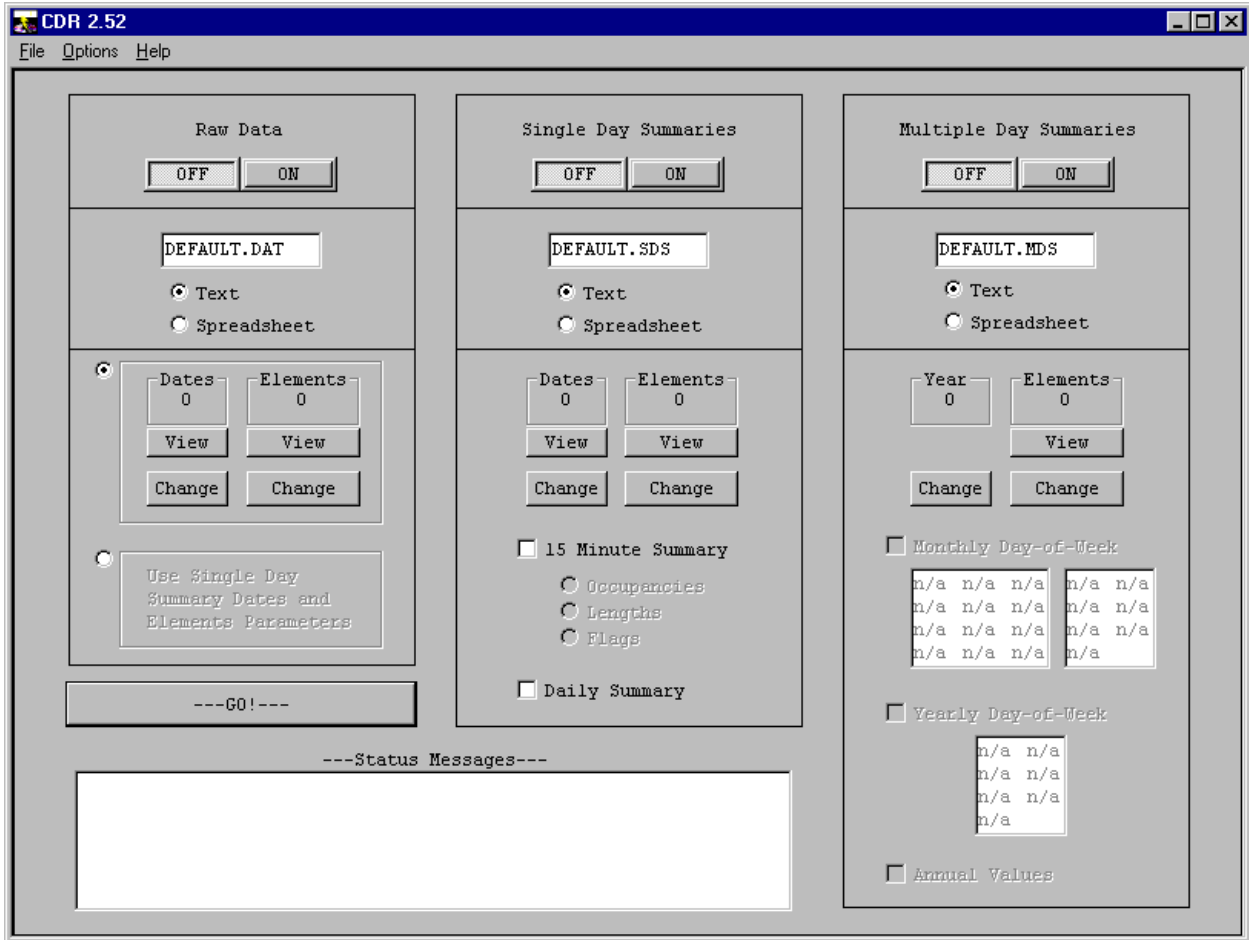
Washington State

Researchers at TRAC are working with WSDOT in demonstrating the usefulness and extending the applications of archived ITS data. The FLOW Evaluation Design study had the following objectives ([13,34](#)):

1. Develop a methodology, framework, and detailed procedures for conducting an ongoing series of evaluations of the performance and effects of the FLOW traffic management system now in operation on Puget Sound area freeways;
2. Create tools for performing those evaluations; and
3. Supplement earlier evaluation data with updated results by using the developed framework to evaluate selected portions of the FLOW system.

The FLOW Evaluation study created several analytical tools that are now being used by TRAC and WSDOT for reducing and analyzing archived ITS data. The regional TMC in WSDOT's Northwest Region (Seattle urban area and surrounding environs) currently archives loop detector data to CD at five-minute intervals (see previous section for discussion of TMC capabilities and practices). TRAC researchers have developed PC-based CD retrieval (CDR) software to retrieve and aggregate the five-minute loop data to local or network disk drives. Figure 17 shows the

graphical user interface for the publicly available CDR software. Another software program developed in this study, CD Analyst, uses spreadsheet macros to analyze the loop data as well as develop graphical presentations.

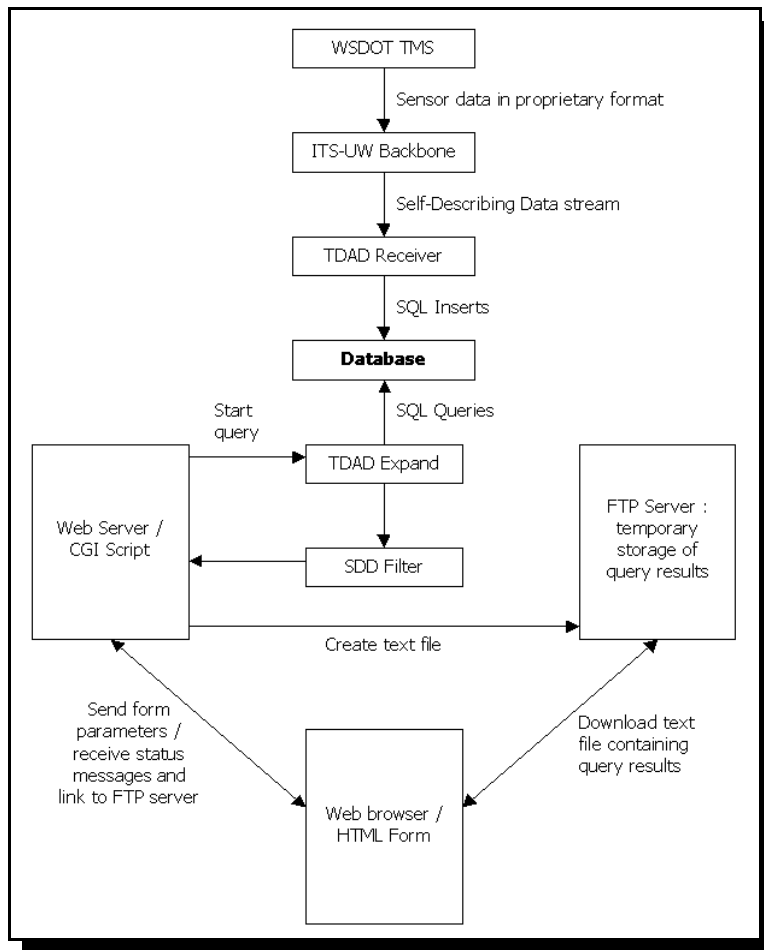


Source: CDR Software v. 2.52, Washington State DOT

Figure 17. Main Interface of CDR Software

TRAC researchers and WSDOT engineers and planners use the CDR and CD Analyst software tools to analyze archived ITS data. TRAC researchers are supplying the extracted archived data for policy development as well as the measurement and evaluation of freeway system performance. WSDOT engineers and planners use the archived data to generate historic traffic volumes and trends for use in numerous planning and design applications.

University of Washington researchers are also developing a web-based ITS data mining application in the Traffic Data Acquisition and Distribution (TDAD) study (35). This prototype application, located at http://www.its.washington.edu/tdad/tdad_top.html, has a map-based interface that is used to select loop detector location(s) of interest. At this time, the raw, 20-second loop data is being used to populate the TDAD database, and approximately six months of data have been accumulated. Figure 18 shows a schematic of the TDAD system database.



Source: <http://www.its.washington.edu/tdad/tdad.gif>,
University of Washington

Figure 18. Schematic of TDAD System

TransLink® ITS Research Program, Texas Transportation Institute

The TransLink® ITS Research Program at TTI has sponsored an “ITS Data Management System” focus area since 1997. Early work in this area was focused on the development of DataLink, a prototype on-line ITS data management system (5,6). The DataLink system warehouses loop detector data from San Antonio’s TransGuide® system in a 40+ gigabyte relational database. The loop detector data are aggregated from 20-second intervals to five-minute increments, mainly because of user data requirements as well as database performance. Users of the DataLink system can perform ad hoc queries within a web browser interface (Figure 19), eliminating the need to know structured query language (SQL). To date, the DataLink system has been storing data on-line since late 1997, and researchers are examining data aging issues related to permanent archiving of a subset of the five-minute data.

Ongoing work in this TransLink® focus area includes the following:

- **DataLink maintenance and refinement** - The DataLink system has been beta tested by TxDOT planners, and this feedback will be used to improve the user interface and query capabilities. The research team is also working on integrating incident information into the loop detector database.
- **Texas ITS data archiving workshop** - In cooperation with TxDOT and ITS Texas, TransLink helped to sponsor an ITS data archiving workshop in Texas (29). Participants included various agency personnel involved in operations, planning, safety, air quality, and research.
- **Performance measures from ITS data** - Researchers are developing a methodology and examples of using ITS data to support a range of transportation system performance measures.

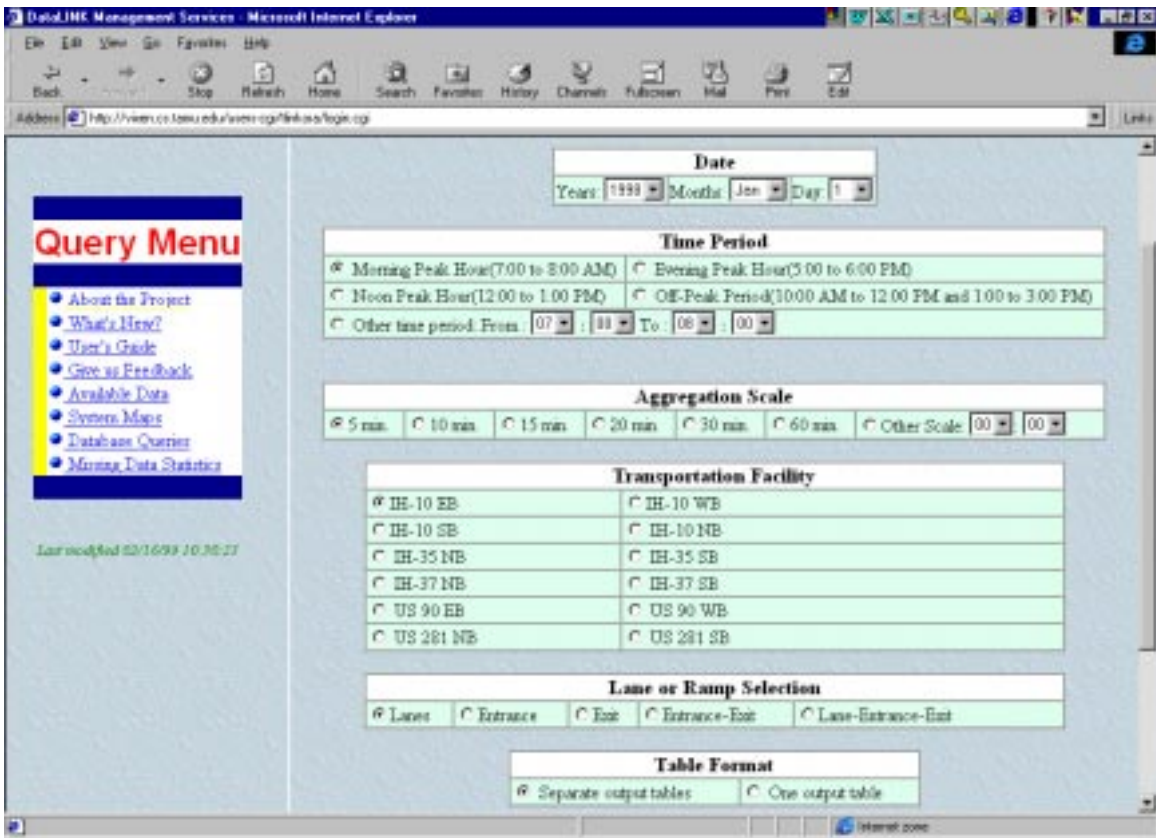


Figure 19. Query Interface of DataLink System (<http://vixen.cs.tamu.edu>)

SUMMARY OF FINDINGS AND LESSONS LEARNED

The research team summarizes the state-of-the-practice in ITS data archiving as follows:

- **Improved Interdisciplinary Coordination** - It appears that many TMC personnel are more interested in supporting the data needs of secondary users. For example, there appears to be increased support by operations personnel to make data accessible and in the correct format for planning applications. Although many agencies are still investigating appropriate formats, storage media, and interfaces, the coordination has improved.
- **Similar Concerns Among Practitioners** - There was strong interest in and support for the work being performed and the questions being raised during the telephone conversations. Nearly all of the experiences indicate similar interests and concerns in different regions. Common questions include:
 - Who are the users of the data?
 - What are the data uses?
 - What data should be kept?
 - What aggregation level should be selected?
 - What type of interface should be developed?
 - How can the system be automated?
 - What is the appropriate data format?

In addition, concerns were expressed related to the steep learning curve of data management systems (e.g., relational databases) by both operations personnel and other data users (e.g., planning staff).

- **Data Quality Concerns** - Many individuals noted concern over the quality of the data itself. Some of the concern may be attributed to the fact that these large ITS data sets are new to many data users, thus there is some unfamiliarity with the inherent quality of the data. The concern with data quality also may be because, in some cases, only minimal error detection is performed as the data is being collected. Many expressed that more work needs to be performed in regard to data quality.
- **Need for Geographic Location Referencing** - It was often noted that referencing the actual data obtained from the TMC with a location on a roadway segment was difficult. Secondary users of the data often have to develop equivalency matrices to compare several location referencing schemes.
- **Easy Data Access is Important** - It appears that as easy access to archived ITS data is provided, most data users create or use existing software to manipulate and

analyze the large data sets to meet their needs. Although there may be a relatively steep learning curve for developing or using such software and techniques, most dedicated users will often find a way to meet their needs. Many TMCs are considering the distribution of data via the Internet or CD, as opposed to traditional practices of archiving data onto magnetic tape cartridges or off-line storage devices.

- **Operations Personnel Receiving Benefits** - One of the primary driving forces in the improved interdisciplinary coordination is likely the fact that operations personnel are receiving benefits and improvements in how they perform their jobs. For example, when research and other activities provide valuable insight about when or how a changeable message sign or ramp-metering operations should be altered, operators see the benefit for such activities.
- **“One Size Does Not Fit All” Regarding Data Aggregation** - There does not appear to be a least common denominator of data aggregation that is significantly favored. Many of the applications provided here span many stakeholders including planning, operations, and research, and they use many different levels of aggregation. In fact, for many of the research applications, the research teams often express their interest in performing similar analysis at different aggregation levels.

Click on the following box to go to:

Part 2 of 3, which contains:

Chapter 3. ITS Data Issues and Analyses

Establishing Data Needs

Data Aggregation

Click on the following box to go to:

Part 3 of 3, which contains

Quality Control

Data Storage Tools

Chapter 4. Findings and Conclusions

State-of-the-Practice

Data Issues and Analyses

References