A Literature and Best Practices Scan: ITS Data Management and Archiving

Project identification number 0092-02-11

Final Report

Henry X. Liu*
Rachel He**
Yang Tao
Bin Ran
University of Wisconsin at Madison

Submitted to the Wisconsin Department of Transportation May 2002

* Currently at University of California at Irvine
** Currently at Princeton University

NOTICE:

This research was funded by the Wisconsin Council on Research of the Wisconsin Department of Transportation and the Federal Highway Administration under Project #SPR-0092-02-11. 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 of the Wisconsin Department of Transportation or the Federal Highway Administration at the time of publication.

Technical Report Documentation Page

1. Report No.	2. Government Accession No	3. Recipient's Catalog No

4. Title and Subtitle	5. Report Date
A Literature and Best Practices Scan: ITS Data Management and Archiving	May 2002 6. Performing Organization Code
7. Authors Henry X. Liu, Rachel He, Yang Tao, and Bin Ran	8. Performing Organization Report No.
9. Performing Organization Name and Address University of Wisconsin at Madison 1415 Engineering Drive Madison, WI 53706	10. Work Unit No. (TRAIS) 11. Contract or Grant No. 0092-02-11
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

To manage and operate the transportation system to exceed customer expectations, data is needed on the structure, status, use, and behavior of the entire transportation system. This "infostructure" serves the needs of the people and organizations that operate and use the system. There are at least three major aspects to fulfilling data needs:

- Understanding what data is needed by the full range of stakeholders
- Creating, operating, maintaining, and updating the mechanisms that will gather, analyze, coordinate, and store the data
- Providing convenient, timely, and affordable access to the data to both operators and users of the transportation system

This report aims to provide the literature review and best practice scan for the Intelligent Transportation Systems (ITS) data management and archiving, and articulate the relevance of good and complete data collection to transportation operations and management. The report provides examples of the types of users that can benefit from data, what types of data they would like and what benefits can accrue if these users obtain the quality data they seek. The report overviews the current data collection and management environment, characterizing both the nature and extent of current data collection and the techniques available – or soon to be available – for improved data management and archiving. Finally, this report provides recommendations on what can be done to improve data management to enable increased effectiveness of transportation operations and system usage.

17. Key Words	18. Distribution Statement
ITS, data archiving, data management, literature scan, best practices.	No restriction. This document is available to the public through the National Technical Information Service 5285 Port Royal Road Springfield VA 22161

18. Security Classif.(of this report)	19. Security Classif. (of this page)	20. No. of Pages	21. Price
Unclassified	Unclassified		

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Table of Contents

Executive Summary

To manage and operate the transportation system to exceed customer expectations, data is needed on the structure, status, use, and behavior of the entire transportation system. This "infostructure" serves the needs of the people and organizations that operate and use the system. There are at least three major aspects to fulfilling data needs:

- Understanding what data is needed by the full range of stakeholders
- Creating, operating, maintaining, and updating the mechanisms that will gather, analyze, coordinate, and store the data
- Providing convenient, timely, and affordable access to the data to both operators and users of the transportation system

This report aims to provide the literature review and best practice scan for the Intelligent Transportation Systems (ITS) data management and archiving, and articulate the relevance of good and complete data collection to transportation operations and management. The report provides examples of the types of users that can benefit from data, what types of data they would like and what benefits can accrue if these users obtain the quality data they seek. The report overviews the current data collection and management environment, characterizing both the nature and extent of current data collection and the techniques available – or soon to be available – for improved data management and archiving. Specifically, the following state-of-the-practice in ITS data archiving can be summarized:

(a) Easy access to archived ITS data creates opportunities;
(b) Planners and researchers have been primary archived data users to date;
(c) Similar concerns are shared among practitioners;
(d) There is improved interdisciplinary coordination but ongoing dialogue is needed;
(e) There are concerns about data quality and geographic location referencing; and
(f) There are concerns about different data aggregation levels.

Subsequently, various ITS data management and archiving business and organizational issues are reviewed. The performance measure is emphasized as the key in corporate decision and investment decision making process. Finally, this report provides recommendations on what can be done to improve data management to enable increased effectiveness of transportation operations and system usage. Specifically, the report proposes the goals and objectives for developing a Wisconsin ITS data management and archiving model. Various types of system requirements and customization requirements are presented for Wisconsin applications. A preliminary system architecture is also proposed, including data storage, database construction, access to data, and easy-to-use query interface. The system implementation and data mining model are also discussed.

1. Introduction

1.1 Problem Statement

ITS stands for "Intelligent Transportation Systems." For existing ITS projects and facilities, data are generated daily and need to be effectively archived. In addition, WisDOT anticipates that more ITS projects will be implemented in Wisconsin over the next few years, and there is an increasing need for managing, archiving and disseminating ITS data. Further delay in the development of an effective data-management and archiving process(es) will waste data, preclude opportunities for financially beneficial public-private partnerships, and further frustrate efforts to evaluate ITS projects and programs.

1.2 Research Objectives

The research team will conduct the "Literature and Best Practices Scan," working closely with the research team of the project titled "A Literature and Best Practices Scan: Perspectives and Expectations of Drivers." This project follows a three-step approach:

? *Step I* – Scan for resources, establish context, identify critical information;

? *Step II* – Prioritize resources, analyze pertinent information and identify information gaps; and

? *Step III* – Synthesize best information and recommend next steps (for example, since many other State DOTs may be in similar circumstances, a follow-up pooled-fund proposal may be appropriate, with Wisconsin as a lead state).

1.3 General Data Management Practices

Data warehousing provides an information architecture that serves as the enterprise-wide source of data for performance analysis and organizational reporting. Figure 1.1 is an example of an integrated information architecture that includes the operational and warehouse environments. Data warehousing architectures typically support strategic decision-making processes, and the queries issued to the data are often of an exploratory and ad hoc nature. A data warehouse is simply a single, complete, and consistent store of data obtained from a variety of sources and made available to the end users in a way they can understand and use in a business context (Devlin, 1997). Data Warehouse feeds the subject-oriented data marts. A data mart is a scaled-down version of a data warehouse, and typically contains between 5 and 15 GB of data, whereas a data warehouse typically contains more than 30 GB of data. The Meta Group defines a data mart as a subject-oriented equivalent of a data warehouse, typically a single functional area, which is tied to one or more specific applications. Data Base Associates International defines a data mart as an application-oriented subset of corporate wide information that departmental users use to make better-informed decisions. The Patricia Seybold Group defines a data mart as a highly focused information resource used to address a single business problem (Douglas, 1997).

The definition outlined for the national ITS architecture explicitly mentions the importance of integrated transportation systems and the interoperability via standard interfaces across the nation. The national ITS architecture is the framework of interconnected subsystems which together provide the ITS user services through allocated functionality and the defined interfaces. This architecture must be open and flexible to prevent unnecessary restriction to implementation choice and to accommodate the varied needs of the public and private sectors (USDOT, 1996), further bolstering the implementation of the data warehouse and data marts. The importance of data warehouse and data marts in the integrated ITS architecture has also been detailed in the paper by Scherer and Smith (1999). They explicitly highlight the importance of data warehousing and data mining for improved transportation operations and management. According to Scherer and Smith, the development of ITS systems that operate well independently is not sufficient. To realize the true potential of ITS, systems such as freeway management systems, transit scheduling systems, signal control systems, emergency computer aided systems, and even construction scheduling systems must be integrated to allow for coordinated regional decision-making. A recent study by IDC, which used information gathered from more than 60 organizations that had implemented data warehousing, found that data warehousing generated an average return on investment (ROI) of 401 percent over three years (Corey et al, 1998).

Hence, the development of an enterprise data warehouse and data mart to manage the ITS data, thereby providing the end users access to the data for online analytical processing (OLAP) and data mining for decision support, is crucial to the full utilization of ITS functionalities. OLAP offers high-performance access to large amount of summarized data for complex multidimensional analysis and easy reporting (http:\\www.sas.com\software\olap). OLAP is a powerful data analysis method for multi-dimensional analysis of data warehouses (Chaudhuri and Dayal, 1997). Data mining is defined as the process of detecting hidden patterns of correlated influence within very large volumes of data. The data to be mined frequently is first extracted from an enterprise data warehouse into a data mart. Another aspect of data mining is data visualization. Data visualization is often used along with data mining tools (Edelstein, 1998). Data visualization is graphical representation of corporate data and when used in conjunction with enterprise Intranet provides end users with enhanced capabilities for decision support. An Intranet is an organized, graphical interface to the organization's information. The two technologies are a perfect match (Youngworth, 1998).

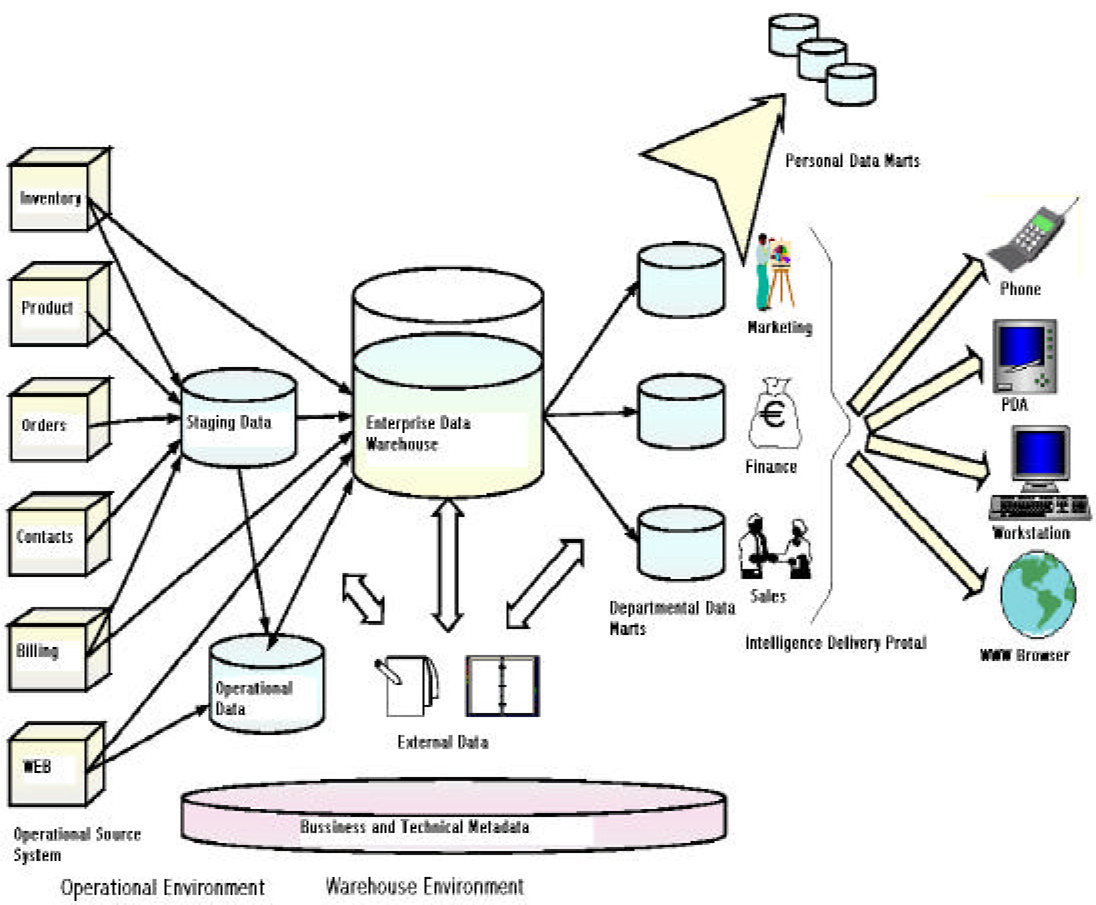

Figure 1.1: Example of an Integrated Information Architecture Including the Operational and Warehouse Environments

2. Existing ITS Data Archiving Systems: A State-of-the-Practice Scan

2.1 Data Sources

2.1.1 ITS Data Sources

Intelligent transportation system (ITS) applications and their sensors and detectors are a potentially rich source of data about transportation system performance and characteristics. Increasing deployment of ITS throughout the nation has brought an awareness that ITS data offer great promise for uses beyond the execution of ITS control strategies. Usually, ITS data refer to data that are typically collected by and/or generated for ITS applications. The most common data sources potentially available from ITS include:

• Traffic Surveillance Data
 - Representative data elements: vehicle volume, speed, travel time, classification, weight, and trajectories;

• Traffic Control Data
 - Representative data elements: time and location of traffic control actions (e.g., ramp metering, traffic signal control, lane control signals, message board content);

• Incident and Emergency Management
 - Representative data elements: location, cause, extent, time history, detection and clearance of roadway incident/emergency;

• Public Transit Data
 - Representative data elements: transit vehicle boardings by time and location, vehicle trajectories, origins and destinations;

• Crash Data
 - Representative data elements: location, time, cause, death, injury and clearance;

• Commercial Vehicle Operations Data
 – Representative data elements: cargo type, carrier, O/D, route and time; and

• Environmental and Weather Data
 – Representative data elements: location, time, precipitation, temperature and wind conditions.

Table 2.1 presents a more thorough description of the above 7 ITS data sources, their representative data elements and their potential uses. In the table, the data sources correspond to subsystems in the National ITS Architecture, but could be expanded to account for data sources not previously identified by that effort. (Some ITS deployments preceded the release of the National ITS Architecture.) The data elements roughly correspond to those in the National ITS Architecture's data dictionary and can be used as a starting point for verifying that the National ITS Architecture accounts for all relevant data flows.

Table 2.1 ITS Data Sources

ITS data sources	Sub-areas	Primary data elements	Typical collection equipment	Spatial coverage	Temporal coverage
Traffic Surveillance Data	Freeway traffic flow surveillance data	· Volume · Speed · Occupancy	· Loop detectors · Video imaging · Acoustic · Radar /microwave	Usually spaced at less than 1 mile; by lane	Sensors report at 20-60 second intervals
		· Vehicle classification · Vehicle weight	· Loop detectors · WIM equipment · Video imaging · Acoustic	Usually 50-100 per state; by lane	Usually hourly
	Arterial traffic flow surveillance data	· Volume · Speed · Occupancy	· Loop detectors · Video imaging · Acoustic · Acoustic microwave	Usually midblock at selected locations only ("system detectors")	Sensors report at 20-60 second intervals
	Visual and video surveillance data	· Time · Location · Queue length · Vehicle trajectories · Vehicle classification · Vehicle occupancy	·CCTV · Aerial videos · Image processing technology	Selected locations	Usually full-time
Traffic Control Data	Traffic signal phasing and offsets	· Begin time · End time · Location · Up/down-stream offsets	Field controllers	At traffic control devices only	Usually full-time
	Ramp meter and traffic signal preemptions	· Time of preemption · Location	Field controllers	At traffic control devices only	Usually full-time

	Ramp meter and traffic signal cycle lengths	· Begin time · End time · Location · Cycle length	Field controllers	At traffic control devices only	Usually full-time
Incident and Emergency Management	Incident logs	· Location · Begin, notification, dispatch, arrive, clear, departure times · Type · Extent (blockage) · HazMat · Police accident report reference · Cause	· CAD · Computer-driven logs	Extent of Incident Management Program	Extent of Incident Management Program
	Emergency vehicle dispatch records	· Time · O/D · Route · Notification, arrive, scene, leave times	CAD	Usually areawide	Usually full-time
	Emergency Vehicle locations	· Vehicle type · Time · Location · Response type	Automatic Vehicle Identification (AVI) or GPS equipment	Usually areawide	Usually full-time
	Construction and work zone identification	· Location · Date · Time · Lanes/shoulders blocked	TMC software		
Public Transit Data	Transit usage	· Vehicle boardings (by time and location) · Station origin and destination (O/D) · Para transit O/D	Electronic fare payment systems	Transit routes	Usually full-time
	Transit route deviations and advisories	· Route number · Time of advisory · Route segments taken	TMC software	Transit routes	Usually full-time

Crash Data	Highway crash data	· Location · Time · Vehicle type · Police accident report reference · Cause	· CAD · Computer-driven logs	Extent of Incident Management Program	Extent of Incident Management Program
	Train arrivals at Highway Rail Intersections	· Location · Begin time · End time	Field controllers	At instrumented HRIs	Usually full-time
Commercial Vehicle Operations Data	HazMat cargo identifiers	· Type ·Container/package · Route · Time	CVO systems	At reader and Sensor locations	Usually full-time
	Fleet Activity Reports	· Carrier · Citations · Accidents · Inspection results	CVO inspections	N/A	Usually Summarized annually
	Cargo identification	· Cargo type · O/D	CVO systems	At reader and Sensor locations	Usually full-time
	Border crossings	· Counts by vehicle type · Cargo type · O/D	CVO systems	At reader and sensor locations	Usually full-time
	On-board safety data	· Vehicle type · Cumulative mileage · Driver log (hours of service) · Subsystem status (e.g., brakes)	CVO systems	At reader and sensor locations	Usually full-time
Environmental and Weather Data	Emissions Management System	· Time · Location · Pollutant concentrations · Wind conditions	Specialized sensors	Sensor locations	Usually full-time
	Weather data	· Location · Time · Precipitation · Temperature · Wind conditions	Environmental sensors	At sensor locations	Usually full-time
	Location referencing data	Special case; pertains to all location references in ITS and planning			

	Probe data	· Vehicle ID · Segment location · Travel time	· Probe readers and vehicle tags · GPS on vehicles	GPS is areawide; readers restricted to highway locations	Usually full-time
	VMS messages	· VMS location · Time of msg · Mug content	TMC software	VMS locations	Hours of TMC operation
	Vehicle trajectories	· Location (route) · Time · Speed · Acceleration · Headway	· AVI or GPS equipment · Advanced video image processing	AVI restricted to reader locations; GPS is areawide	1-10 second intervals
	TMC and Information Service Provider generated route guidance	· Time/date · O/D · Route segments · Estimated travel time	TMC /Information Service Provider software	Usually areawide	Hours of TMC operation
	Parking and roadway (congestion) pricing changes	· Time/date · Route segment/lot ID · New price	TMC software	Facilities subject to variable pricing	Hours of TMC operation

2.1.2 Benefits of Archived ITS Data

In many cases, ITS-generated data are similar to data traditionally collected for these applications, but they are collected continuously and are much more voluminous in quantity and temporal coverage. ITS data have many benefits than traditionally collected data.

- The continuous nature of most data generated by ITS removes temporal sampling bias from estimates and allows the study of variability. Nearly all of the data currently collected for planning, operations, administration, and research applications are through the use of sample surveys (e.g., household travel surveys, short-duration traffic counts). Although attempts are made to adjust or expand the sample, the procedures are imperfect. With continuous ITS data, there is no need to perform adjustments to control sample bias. (Equipment or nonresponse errors are still present, though). Further, continuous ITS data allows the direct study of variability, which is becoming an important factor in the study of personal travel habits and the effect of extreme events (e.g., days with very high volumes).

- The detailed data needed to meet emerging requirements and for input to new modeling procedures can be provided by ITS data sources. The next generation of Travel Demand Forecasting (TDF) models (e.g., TRANSIMS) and air quality models (modal emission models) will operate at a much higher level of granularity than existing models. These need more detailed data than what is now collected. Traditional data sources are barely adequate for existing models -- there is little doubt that they will be incapable of supporting the next generation of models. Much data generated by ITS are collected at the levels of detail necessary to support these models. For example, roadway surveillance data (volumes, speeds, and occupancies) are typically reported every 20 seconds and GPS-instrumented vehicles can report positions and activity at time intervals as small as one second. Also, GPS-derived locations can pinpoint incident locations to within a few meters. This level of detail will be required for the input and calibration data used by the new models. Finally, as data generated by ITS are used more frequently for non-real-time purposes, it is likely that additional uses not currently foreseen will emerge. In addition, data on activity patterns and how travelers respond to system condition will be important for the next generation of models.

- As the focus of transportation policy shifts away from large-scale, long-range capital improvements and toward better management of existing facilities, ITS-generated data can support the creation and use of the system performance measures that are required to meet this new paradigm. The specification of transportation management systems in ISTEA has created the need for more intense system performance monitoring than current data can adequately support. Further, planners, operators, and administrators are increasingly being required to shorten their planning horizons. System performance measures provide objective feedback to transportation professionals on the effectiveness of programs and improvements, and also provide a common basis for comparing different jurisdictions. This kind of feedback is extremely important as the focus shifts to short-term management strategies. However, it is clear that data with higher resolution and accuracy than have been traditionally collected are required to support the use of system performance measures.

2.2 Data Archiving and Management

2.2.1 Data Archiving Practices

In September 1999, the Archived Data User Service (ADUS) was formally incorporated into the National ITS Architecture. ADUS, in the Architecture, identifies the sources of ITS data, logical data flows from the sources to the archives, functions required for archiving, different market packages and ways to implement ADUS, potential users and issues. With the advent of the ADUS, the National ITS Architecture officially embraced the concept of saving (retaining and archiving) real-time, operational data for other non-real-time applications, such as transportation planning, safety analyses, and research. That archived ITS-generated data is of great value, which is apparent in the number of transportation agencies that have already developed archives.

Table 2.2 Current ITS Data Archiving Practices

Location	Agency	Types of ITS Data Archived
Phoenix, Arizona	Maricopa County DOT & Maricopa Association of Governments	Loop detector data from freeways and arterials (plans underway to archive all "relevant" data used by the Traffic Operations Center)
Los Angeles, California	Caltrans	Freeway loop detector data
Orange County, California	Caltrans	Arterial loop detector data
Berkeley, California (PeMS)	Caltrans & PATH	Freeway loop detector data
Chicago, Illinois	Illinois DOT	Loop detector data from selected freeways
Lexington, Kentucky	NORPASS, Kentucky Transportation Center	CVO (commercial vehicle operations) data, WIM data
Montgomery County, Maryland	Montgomery County DOT & Maryland National Capital Park and Planning Commission	Loop detector data from selected arterials
Detroit, Michigan	Michigan DOT	Loop detector data from freeways
Minneapolis-St Paul, Minnesota	Minnesota DOT	Loop detector data from freeways
TRANSCOM, New York/New Jersey/Connecticut	TRANSCOM	Travel times derived from AVI-equipped vehicles
Houston, TX	Texas DOT	Travel times derived from AVI-equipped vehicles
San Antonio, TX	Texas DOT	Loop detector data from freeways, travel times derived from AVI-equipped vehicles, and incident management data
Seattle, WA	Washington DOT	Loop detector data from freeways
Milwaukee, WI (MONITOR)	Wisconsin DOT	Freeway data collected by electronic detectors, closed circuit television cameras, ramp meters and variable messages signs

This section contains information on the state-of-the-practice in ITS data archiving and management. 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 2.2 summarizes the locations, implementation agencies, and the archived ITS data types of various ITS data archiving practices.

2.2.2 Common Themes Among the Practices

Many existing traffic management centers are either currently archiving or plan to archive traffic surveillance data. However, archiving is often done on an informal basis with no storage guidelines and limited access capabilities. In some of these areas, where data are stored but are relatively inaccessible, data users have developed the tools required to access the data.

Traffic surveillance data are the primary type of data being archived. Since the study focused on traffic management centers, this result is not surprising. However, many transit systems that have deployed electronic fare payment and automatic vehicle location systems also routinely archive these data (including ridership counts by route segment and time of day, and station origin-destination patterns).

Data quality and location referencing are the two most common problems encountered by archived ITS data users. There is no clear consensus on to what level the data should be aggregated.

Researchers and planners have been the primary users of archived ITS traffic data to date. Transit planners, for instance, can use Automatic Passenger Counters (APC) to track where and when passengers come onto and exit the transit system. APC linked with Automatic Vehicle Location Systems can reduce the amount of person hours required to collect and properly display this information. There may be many other potential users of the data that have yet to be active users of archived ITS data such as safety analyzers.

2.2.3 Representative Case Study: PeMS

PeMS (***Pe***rformance ***M***easurement ***S***ystem) is a freeway performance measurement system that extracts information from real time and historical data. It is sponsored by Caltrans and developed by PATH. Based on the common sense belief that the best way to improve the freeway system is to enable people to easily figure out how the system is performing every day, PeMS obtains 30-second loop detector data in real time and presents information in various forms to assist managers, traffic engineers, planners, freeway users, researchers, and value added resellers or Travel Information Service Providers.

PeMS Architecture

Communication Architecture

Figure 2.1 shows the communication architecture.

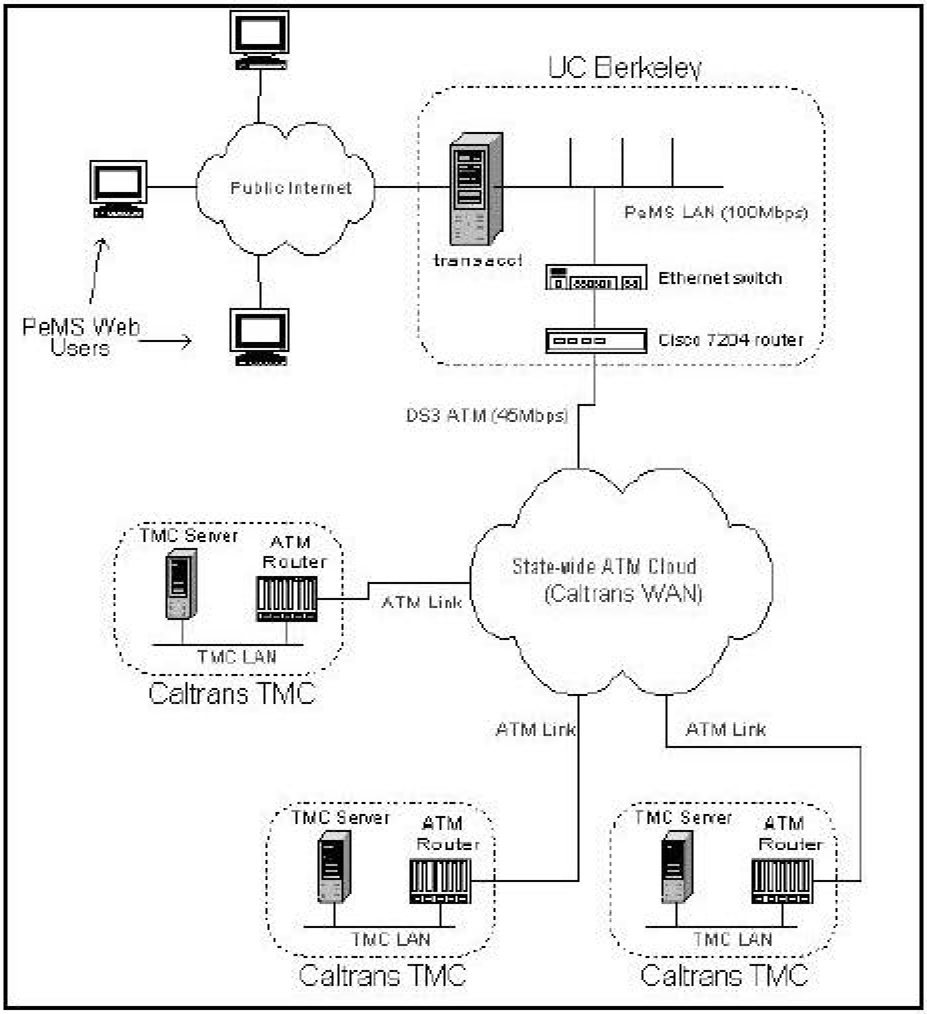

Figure 2.1 PeMS Communication Architecture

The PeMS computer, *transacct*, is a four-processor SUN 450 workstation located at the University of California at Berkeley. Users access PeMS over the Internet. *Transacct* also has a 45 Mbps link to the Caltrans ATM wide area network (WAN). The WAN is used to transfer data from districts to PeMS.

An individual Caltrans district is connected to PeMS over a permanent ATM virtual circuit. To establish such a circuit, the routing tables at the two ends must be configured. The configuration is done remotely from Caltrans Headquarters in Sacramento.

A district TMC and PeMS collect data as follows. A “front end processor (FEP)” at the

TMC receives data from freeway loops every 30 seconds. The FEP formats these data and writes them into the TMC database as well as into the PeMS database.

PeMS maintains a separate instance of the database for each district. Although the table formats vary slightly across districts, they are stored in PeMS in a uniform format so that the same software works for all districts.

Software Architecture
The software is organized in three layers. At the bottom is database administration. The work is standard but highly specialized: disk management, crash recovery, table configuration. Many parameters must be tuned to improve database performance. The top layer comprises applications that are described in the next section. The middle layer comprises software that works on the data as they arrive in real time, and

- aggregates 30-second values of flow and occupancy to lane-by-lane, 5-minute values;
- calculates the *g*-factor of each loop;
- uses the g-factor to calculate the speed for each lane;
- aggregates the lane-by-lane value of flow, occupancy, and speed across all lanes at each detector station; at this point, PeMS has flow, occupancy, speed, and travel time for each 5-minute interval for each detector station (one station typically serves the detectors in all the lanes at one location); and
- computes the basic performance measures.

Uses of PeMS

PeMS is a useful system. Different users run various applications through their Web browsers. Authorized users may directly query the database and develop custom applications.

Transportation Managers
By pulling up PeMS on the Web browser, a transportation manager at a TMC can compare the performance of the freeways with previous days. He/she can also compare the performance situation among the freeways, and allocate resources towards improving the worst performers.

Transportation Engineer
PeMS creates a map of the entire freeway network for each 5-minute interval in which each link is colored according to speed or any of the other computed averages. An “animation” application plays back these maps in sequence over any time interval. Transportation engineers can use this application to visualize the behavior of the network on any day. PeMS can also generate and display a contour plot of speed for a 24-hour period of time over a certain section of road. Comparing to the contour plots at the same location of the historical data, transportation engineers can search for some anomalous situation. If there is some anomaly, they can conclude that the loop detector located nearby may not work properly.

Traveler

PeMS can provide travelers travel time estimates and shortest paths for their trips. Travelers can bring up the district freeway map on their Web browsers, and select origins and destinations. PeMS decorates the map with two shortest routes, depending on whether or not the traveler can use the High Occupancy Vehicle (HOV) lanes, and a caption contains the corresponding travel times.

VARs—Value Added Resellers (Travel Information Service Providers)

VARs are businesses that package travel time information with other location-dependent services. They can access and process PeMS data and send real time information to the users during their trips. PeMS has developed a web-enabled cell phone prototype, which VARs can use.

Planner/Researcher

Transportation planners and researchers can make a lot of use of PeMS, such as using PeMS to analyze delay for any section of freeway and using PeMS to observe congestions developing in real time.

Features of PeMS

- PeMS is a low-cost system. It uses commercial-of-the-shelf products for communication and computation. Detector data are retrieved over the Caltrans ATM wide area network to which all districts are connected. The 45 Mbps link connecting PeMS to this network costs $2000/month. A copy of PeMS would cost about $300,000.

- PeMS is easy to use. Built-in applications are accessed through a Web browser. Custom applications can work directly with the database. PeMS brings large benefits. Caltrans managers can instantaneously obtain a uniform, and comprehensive assessment of the performance of their freeways. Traffic engineers can base their operational decisions on knowledge of the current state of the freeway network. Planners can determine whether congestion bottlenecks can be alleviated by improving operations or by minor capital improvements. Traffic control equipment (ramp-metering and changeable message signs) can be optimally placed and evaluated. Travelers can obtain the current shortest route and travel time estimates. PeMS can serve to guide and assess deployment of intelligent transportation systems (ITS).

- The PeMS software architecture is modular and open. A new district can be added online with six person-weeks of effort, with no disruption of the district's TMC (Traffic Management Center). Data from new loops can be incorporated as they are deployed. New applications are added as need arises. Although it is a prototype, it can serve as the blueprint for a 7x24 production system.

- PeMS is a key element of the Caltrans performance evaluation system. This system would help managers, planners, and engineers accurately estimate current performance; discover locations where improvements are likely to be most effective; evaluate in advance the benefits of suggested investments (ramp-metering, changeable message signs, freeway service polls, etc); and, after those investments are made, measure the resulting benefits. Such a system should be part of the daily operations, just as production, cost, sales and revenue figures are essential in the daily operations of a private corporation.

2.3 Data Utilization and Mining

Because of the utility and variety of ITS-generated data items, they can be used in traffic operation, transportation planning, traveler information system, and academic research. Many of the TMCs received data requests for a variety of transportation applications. Several stakeholder groups have been identified as having an interest in the use of archived ITS data. Table 2.3 introduces these stakeholder groups along with their primary functions and example applications.

Table 2.3 Stakeholders and Example Applications for Archived ITS Data

	Stakeholder Group	**Primary Transportation-Related Functions**	**Example Applications**
Traffic Operation	Traffic management operators	Day-to-day operations of deployed ITS (e.g., Traffic Management Centers, Incident Management Programs)	• Pre-planned control strategies (ramp metering and signal timing) • Highway capacity analysis • Saturation flow rate determination • Microscopic traffic simulation -- Historical -- Short-term prediction of traffic conditions • Dynamic traffic assignment • Incident management • Congestion pricing operations • Evaluation and performance monitoring
	Transit operators	Day-to-day transit operations: scheduling, route delineation, fare pricing, vehicle maintenance; transit management systems; evaluation and planning	• Capital planning and budgeting • Corridor analysis planning • Financial planning • Maintenance planning • Market research • Operations/service planning • Performance analysis planning • Strategic/business planning

	Construction maintenance personnel	Planning for the rehabilitation and replacement of pavements, bridges, and roadside appurtenances; scheduling of maintenance activities	• Pavement design (loadings based on ESALs) • Bridge design (loadings from the "bridge formula") • Pavement and bridge performance models • Construction and maintenance scheduling
	Commercial vehicle enforcement personnel	Accident investigations; enforcement of commercial vehicle regulations	• HazMat response and enforcement • Congestion management • Intermodal access • Truck route designation and maintenance • Truck safety mitigation • Economic development
	Emergency management services (local police, fire, and emergency medical)	Response to transportation incidents; accident investigations	• Labor and patrol planning • Route planning for emergency response • Emergency response time planning • Crash data collection
	Transportation system monitoring personal	Data collection related to system conditions and performance	Provide data for other stakeholders: • Traffic counts and travel times for planners (AADT, K- and D-factor estimation; temporal distributions) • Truck weights for maintenance personnel • Performance metrics for administrators
Transportation Planning	MPO and state transportation planners	MPO and state identifying multimodal passenger transportation improvements (long and short-range); congestion management; air quality planning; develop and maintain forecasting and simulation models	• Congestion monitoring • Link speeds for 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
	MPO/state freight and intermodal planners	Planning for intermodal freight transfer and port facilities	• Truck flow patterns (demand by origins and destinations) • HazMat and other commodity flow patterns
	Safety planners and administrators	Identifying countermeasures for general safety problems or hotspots	• Safety reviews of proposed projects • High crash location analysis • Generalized safety relationships for vehicle and highway design • Countermeasure effectiveness (specific geometric and vehicle strategies) • Safety policy effectiveness

	Land use regulation and growth management planners	Development and monitoring of ordinances related to land development	• Zoning regulations • Comprehensive plan development • Impact fees • Taxation policies
	Air quality analysts	Regional air quality monitoring; transportation plan conformity with air quality standards and goals	• Emission rate modeling • Urban airshed modeling
Traveler Information System	Federal government	Maintain national databases related to traffic operation, safety, transit, freight/CVO, etc.	• HPMS •FARS •NTB
	Private sector users	Provision of traffic condition data and route guidance (Information Service Providers); commercial trip planning to avoid congestion (carriers)	
Academic Research	Transportation researchers	Development of forecasting and simulation models and other analytic methods; improvements in data collection practices	• Car-following and traffic flow theory development • Urban travel activity analysis

2.3.1 Traffic Operation

Stakeholders in traffic operation 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 1: 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.

Example 2: 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.

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.

2.3.2 Transportation Planning

Transportation planners 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 1: 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.

Example 2: 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. 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 3: 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. The hypothesis of the work is that ITS strategies are most effective during high-variability situations. 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. This approach also provides more insight than traditional travel demand forecasting techniques by considering how ITS may affect travel.

2.3.3 Traveler Information System

Now traveler information systems mainly cover only small area services. They usually provide the real-time road information by radio, television and internet.

Example: MONITOR - Milwaukee Freeway Traffic Management Center, Wisconsin

MONITOR is designed to improve the safety and efficiency of the Milwaukee freeway system by reducing incidents and relieving traffic congestion. Milwaukee Freeway Traffic Management Center collects data through the use of electronic detectors, closed circuit television cameras, ramp meters and variable messages signs. All this information is collected and sent to the Traffic Operations Center in downtown Milwaukee. In the control room, operators alert media and the public of any problems on the freeway system by providing information to radio, television and news service organizations.

2.3.4 Academic Research

Transportation researchers typically desire the most disaggregate data of all the stakeholders identified in Table 2.4. 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 1: 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. 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:

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

• The discharge flows in active bottlenecks exhibit near-stationary patterns that (slowly) alternate at 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 could 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 2: 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. 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. 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, and 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). 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.

2.4 Wisconsin Statewide ATIS Study and ITS Architecture Review

2.4.1 Wisconsin Statewide ATIS Study

Current ATIS Related Projects in Wisconsin

The existing and planned ITS deployment in Wisconsin, such as the installation of surveillance systems on I-94, I-90 and I-39 (traffic detectors and CCTV), will allow the transportation districts and other Wisconsin agencies to assess the real-time status of traffic along these important corridors. Many of the existing and planned ITS projects have ATIS components. Table 2.4 presents the inventory of the current ATIS Related Projects in Wisconsin.

Table 2.4. Inventory of ATIS Related Projects in Wisconsin

Project	**Traveler Information Provided**	**Covered Geographic Area**
Communication and Data Systems Infrastructure Project (CDSI)	All member organizations will be given access to data and control flows available	Southeastern Wisconsin
Gateway	Road surface conditions, travel times, congestion, incidents, detours, road closures	Gary-Chicago-Milwaukee Corridor
Monitor	Road surface conditions, travel times, congestion, incidents, detours, closures, park and ride locations, event parking and information	Milwaukee
Integrated Corridor Operations Plan (ICOP)	Congestion levels, closures/alternate routes, transit scheduling	Southeastern Wisconsin
Dane County Incident Management Systems	Incidents	Beltline and interstate system in Dane County
Southeastern Wisconsin Traffic Incident Management Enhancement Program (TIME)	Travel times, congestion levels, incidents, closures/alternate routes, event parking and information	Freeway system in southeastern Wisconsin
I-39 Corridor Study	Road surface conditions, weather conditions, closures/alternate routes	I-39 corridor
WisDOT Central Office Transit Initiatives	Road surface conditions, transit scheduling, in-vehicle road guidance	Cities, or regions, with transit or paratransit services
Wisconsin Road Conditions 800 Number	Seasonal construction information, road closures, weight restrictions, winter road conditions	Statewide
Road/ Weather Information System (R/WIS) Program	Road surface conditions, weather conditions, airport and parking information	52 locations throughout Wisconsin
Foretell	Road surface conditions, weather conditions	Statewide

I-90/I-94 Highway Advisory Radio (HAR)	Road surface conditions, detours, closures, alternate routes	In spot areas along I-90 and I-94
Rest Area Traveler Information Monitors	Road surface conditions, weather conditions	19 rest areas and 8 traveler information center on various interstates in Wisconsin
WisDOT Internet Home Page	Road surface conditions, travel times, congestion levels, incidents, weather conditions, detours, closures/alternate routes, transit scheduling	Lane and ramp closures, congestion, travel time information in the Milwaukee area, web site statewide,
Madison Metro ITS Initiatives	Transit scheduling	Madison
Fleet Online	Road surface conditions, weight restrictions, travel times, congestion levels, incidents, weather conditions, detours, closures, in-vehicle road guidance, route specific and point-to-point travel information	GCM corridor
Lane closures	Road surface conditions, detours, closures/alternate routes, delay at work zones	Milwaukee (work zone delay information), Statewide (all other work zone information)

Wisconsin Statewide ATIS Plan

To better serve the purpose of providing traveler information via ATIS tools to the users, WisDOT is working on the Statewide Advanced Traveler Information System (ATIS) Plan. The Statewide ATIS Plan develops the statewide framework for how traveler information is to be shared between transportation districts and the methods for disseminating information to travelers in Wisconsin.

The Advanced Traveler Information Systems (ATIS) and ITS Architecture Advisory Group was established by the WisDOT in early 2000. The Advisory Group was charged with the responsibility for leading the development of two very important planning projects: The Statewide ATIS Plan and the Statewide ITS Architecture.

Figure 2.2 provides a graphical representation of the process used to develop the ATIS Plan and Statewide ITS Architecture. Wisconsin's long-term vision for ATIS is that a mix of public- and private-sector entities will deliver accurate, consistent and reliable information to travelers in the state. And this vision identifies three priority goals: (1) to make travel safer, (2) to make travel more efficient, and (3) to increase customer satisfaction.

The plan defines the existing information gaps, the ATIS data quality standards, database design, ATIS public/private sector policy (roles and responsibilities), ATIS development guidelines and implementation plan.

Four business models have been developed: (1) Public-Centered Operations, (2) Contracted Operations, (3) Franchise Operations, and (4) Private, Competitive Model (National Weather Service Model). Also there are four basic market forms identified: (1) Perfect competition, (2) Monopolistic competition, (3) Oligopoly, and (4) Monopoly.

Figure 2.2. Process to Develop the Statewide ATIS Plan and ITS Architecture

Several factors must be balanced in choosing a business model for ATIS deployment, and no single model will be ideal for accomplishing all of the objectives for all projects. Some of the criteria that should be considered are cost to the public agency, level of competition encouraged in the market, the diversity of products offered to the consumer and whether policy objectives are likely to be achieved. The appropriate business model for ATIS in Wisconsin should be determined in two to three years once the critical components of the statewide ATIS data collection system and database have been implemented.

2.4.2 Wisconsin Statewide ITS Architecture

The statewide ITS Architecture defines the relationships and information sharing needs between ITS systems in the region. The main goal of the Statewide ITS Architecture for the Wisconsin Department of Transportation is to provide a framework for the development of ITS systems in Wisconsin that will allow for the integration and interoperability of disparate systems. The secondary goal of development is to conform to the National ITS Architecture, developed by the United States Department of Transportation in 1998.

The process used to develop the Wisconsin statewide ITS Architecture is a three-tiered architecture (Figure 2.3). This architecture approach allows for flexibility in its development. One can start from the project level, the regional level, or the statewide level and move toward the other tiers to fill the three-tiered architecture.

Figure 2.3. Three-Tiered Architecture Design Approach

The State was first divided into regions, which are defined by the Wisconsin Department of Transportation District boundaries. From these regions, metropolitan and rural regions were identified. Two methodologies were then applied: Metropolitan Regional Architectures and Rural Regional Architectures.

Metropolitan regional architectures refer to the regions that have a significant effort in place to develop a regional architecture and the population density is great enough to qualify as a metropolitan region. Districts 1, 2 and 3 fall into this category and have already taken steps toward creation of their district's architectures.

Rural regional architectures refer to the regions that do not currently have a district wide effort ongoing to develop an architecture, yet have relative transportation user needs that could be served by various ITS deployments. These include Districts 4, 5, 6, 7, and 8.

The Statewide ITS Architecture covers the entire state of Wisconsin, but it was developed using the five regional architectures from Districts 4, 5, 6, 7, and 8. Districts 1, 2, and 3 are currently in the process of development or require modifications to each of their respective architectures. That being said, the Statewide ITS Architecture is merely an extension of the regional architectures in conglomerate.

The Wisconsin Department of Transportation Statewide ITS Architecture is an open, modular ITS planning tool. Just due to this open, modular nature, the architecture will be able to receive the other district architectures upon completion, as well as, any additions that may be necessary in the future.

Table 2.5 Wisconsin Statewide ITS Architecture Market Packages

Category	Market Packages	Deployable Market Packages
ATMS	Network Surveillance Probe Surveillance Surface Street Control Freeway Control Traffic Information Dissemination Regional Traffic Control Incident Management System Virtual TMC and Smart Probes Road Weather Information System	Network Surveillance Surface Street Control Freeway Control Traffic Information Dissemination Regional Traffic Control Incident Management System Road Weather Information System
APTS	Transit Vehicle Tracking Transit Fixed-Route Operations Demand Response Transit Operations Transit Traveler Information	Transit Vehicle Tracking
ATIS	Broadcast Traveler Information Interactive Traveler Information Yellow Pages and Reservation Dynamic Ridesharing	Broadcast Traveler Information
CVO	Electronic Clearance CV Administrative Processes Weigh-In Motion HAZMAT Management	
EMS	Emergency Response Emergency Routing Mayday Support	Emergency Response Emergency Routing
ADS	ITS Data Warehouse	ITS Data Warehouse

The market packages and the deployable ones for the statewide architecture are summarized in Table 2.5. Note this summary is simply a collection of the market packages that were applied to the District Regional Architectures. And not every market package necessarily has a "statewide" deployment option.

2.5 Summary of Findings

The state-of-the-practice in ITS data archiving can be summarized as follows.

(a) Easy 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. 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.

(b) 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.

(c) Similar Concerns Among Practitioners

Nearly all of current ITS data archiving practices 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).

(d) Improved Interdisciplinary Coordination But Ongoing Dialogue Needed

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 is improving.

But 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.

(e) Concerns about Data Quality and Geographic Location Referencing

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.

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.

(f) Concerns about Different Data Aggregation Levels

There does not appear to be a least common denominator of data aggregation that is significantly favored, other than simply saving raw, disaggregate data. Many of the applications provided throughout this report cover many stakeholders including planning, operations, and research, and they use many different levels of aggregation. In fact, several researchers expressed their interest in performing similar analyses at different aggregation levels. TTI researchers suggest that basic aggregated data are saved for current needs, and yet innovative archiving capabilities exist that can provide advanced data users with access to raw, disaggregate data.

3. ITS Data Management and Archiving Business and Organizational Issues

3.1 Strategic Issues - How Data Management Fits Into Corporate Business

- Performance Measure Is the Key

Traffic performance measure is the key to the decision-making for both the users and providers of transportation, and at both operation and planning levels. Performance measure involves collecting, processing, archiving, and disseminating of traffic data. Figure 3.1 shows the components of a traffic performance measure system.

Figure 3.1. The Components of a Traffic Performance Measure System

- Performance Measure Supports the Investment Decision Process

Given that the priorities for investment should relate to the benefits, the primary criteria for setting priorities should be from the measure of transportation system performance. Figure 3.2 illustrates how archived traffic data support investment decision making.

Figure 3.2. Archived Traffic Data Support Investment Decision Making

3.2 Business Issues

Currently, few transportation management centers have any mechanism for sharing data resources among other transportation groups or agencies within the same jurisdiction. Meanwhile, transportation analysts and researchers often struggle to obtain accurate, reliable data about existing transportation performance and patterns. It is necessary to develop strategies of developing public/private partnerships for managing archived ITS data.

As a key element of an integrated ITS, ATIS data collection efforts must be coordinated with the existing data collection of other areas and systems, such as Advanced Traffic Management Systems (ATMS) or Advanced Public Transportation Systems (APTS). It is important for those developing a strategy for data collection to consider the needs of the various systems at work in their respective areas. This will help to foster a more coordinated regional effort toward data collection as well as a more cohesive plan for use and dissemination of the data from the various systems.

While who has access to data and under what terms is an extraordinarily complex issue for data collectors, the recommended quality level is straightforward: licensed or open access. Good data with limited or no access or inadequate data with sufficient access both result in less than optimal data being available to support ITS products and services.

3.3 Functional and Decision Issues

- Plan for data retention and management

Having a long-term vision for the ITS goals and system will allow stakeholders to evaluate data needs and priorities, as well as develop an incremental plan for ITS data collection deployment. This vision will serve as a regional roadmap for ITS development, and aid in long-range planning for ITS as well as other systems that could contribute to ITS data needs. This vision should also articulate the methods for data sharing.

Identification of data needs from different stakeholders needs an iterative process that establishes a dialogue or partnership between data collectors / providers and data users, thereby affecting future ITS detector / sensor designs.

- Identify appropriate levels of detail

Identifying what aggregation levels are necessary for different applications is indispensable. Different uses of ITS data require different levels of details. Design and operational applications commonly require detailed data for shorter sections and time interval. Planning applications commonly require historical data over extended sections of roadway and period of time. Evaluations require a range of detail levels for both time and space.

- Storage capacity and management

ITS data potentially requires large amount of data storage capacity. With modern computing technology this can be supplied at a moderate cost, however, the data become difficult to manage without an efficient data management strategy. Innovative storage and/or aggregation strategies are necessary to keep costs to a reasonable level and within the reach of a typical agency.

- Data Quality

Quality control procedures are especially critical with ITS data for several reasons: 1) the large potential volume of ITS data makes it difficult to detect errors using traditional manual techniques; 2) the continuous monitoring nature of ITS data implies that equipment errors and malfunctions are more likely during operation than periodic data collection efforts; and 3) archived data users may have different (and potentially more stringent) data quality requirements than real-time users.

- Regional data repository

There is often a desire by transportation agencies for ITS data in a given region. A regional data repository could allow data management to occur at a central location at a regional level.

- Relation to ITS national architecture

ITS national architecture does include some provisions for the storage and management of data in planning subsystems. However, TMCs need additional guidance with these "planning subsystems" and related data dictionaries.

3.4 Data Management Issues

- Standards and modern information technologies that support integration and optimal use of information resources

- Definition and selection of transportation performance measures

After defining goals and objectives of the systems, the appropriate performance measures in terms of effectiveness and efficiency can be selected, such as to reduce congestion and increase economic productivity, to minimize the noise impacts along the corridor, to increase the safety of commuters, etc. To adequately capture all desirable goals of a transportation system, performance measures should characterize more than just traditional notions of mobility or efficiency. Safety, accessibility and equity are non-traditional measures that should be considered in selection of performance measures.

- Online testing and refinement

4. Toward A Wisconsin ITS Data Management and Archiving Model

4.1 Goals and Objectives

The goals are to develop a multimedia state-of-the-art information system to preserve and facilitate utilization of ITS data. Specifically, the objectives are to define the needs and potential uses of ITS data, to define performance measures of applications, to choose the computer hardware and software for ITS data management system, and to resolve the issues of storage, aggregation and manipulation of ITS data.

4.2 System Requirements - What Customization To Be Done for Wisconsin Applications?

- Reliable

Transportation professionals must have reliable data for their planning and operations applications. The most frequently cited reason for insufficient data quality is inadequate geographic coverage. This arises mainly from incomplete data collection in metropolitan areas with multiple jurisdictions, particularly with respect to traffic speeds. Of the problems listed, geographic coverage and spatial resolution are the only ones for which it is possible to develop a national picture based on the public sector responses. To some extent the issues of update frequency and temporal coverage can be explored indirectly. The most common complaints, in order of frequency of citation are: inadequate geographic coverage, inaccurate information, insufficient update frequency, not timely enough, and inadequate spatial resolution.

On average, how closely does the data collected match actual conditions? All traffic sensor data collection systems are subject to inaccuracies. For example, loop detectors in adjacent lanes could both count a vehicle in the process of changing lanes, resulting in double counting the vehicle. Other technologies are subject to weather conditions, radio frequency interference, and occlusion. Clearly, the more accurate the data, the higher quality.

- Effectively archived

The abilities to store and manage large amount of data and to aggregate and summarize data in point, section, and corridor/facility formats are required.

Public sector doesn't want to give out data that: (1) are of unknown or questionable quality; (2) may be easily misinterpreted or misused by naive ISPs who don't understand the transportation data; or (3) are likely to mislead or misinform the public.

- Manageable and affordable

It is important to archive ITS data into a manageable form. The ability to access the database from remote locations without burdensome or costly software requirements is necessary. No user costs are associated with the system except for an Internet connection.

How much of the data designed to be collected is made available? While traffic sensor data collection systems are usually designed to operate continuously, inevitably a user of the data will lose access from time to time. This attribute describes the average probability that a given data element will be available for use from a given traffic sensor. The better the traffic sensor data collection system is designed, operated and maintained, the higher the data availability.

- Clear presentation

Data outputs are both tabular and graphical. The ability to calculate and summarize a given set of performance measures is required. A user-friendly query interface that does not require knowledge of special programming languages or relational database applications.

Sometimes, images give a quick impression of traffic conditions that can be easily assessed. However, this data type is not conducive to deriving detailed information such as that can be provided by traffic sensors. Images can be disseminated through multiple outlets including web pages, TV stations, kiosks, etc. Images can also be used by ISPs to verify or identify information that can then be manually inserted into a traveler information service.

- Professional support and maintenance

There are many benefits to having data from TMCs available for future transportation analyses. To ensure that data are provided for these purposes, cross-disciplinary coordination and communication is essential. Professional support and maintenance are the warrant for successful and efficient utilization and sharing of ITS data. Related professional training is encouraged.

4.3 System Architecture

The system includes 4 major components:

- Data storage

For efficiently manage huge ITS database, successful business software, such as ORACLE, must be employed. Other innovative storage and/or aggregation strategies are necessary to maintain enormous amount of ITS data.

- Database construction

The potentially large ITS database cannot be built with traditional desktop computer-based spreadsheet or database applications because of enormous amount of ITS data and diverse data manipulation.

- Access to data

The archived ITS database stored in web server should be accessed by internet, email service and specialized applications or query languages to interact with the database engine. Public sector wants at least a basic level of services available to all taxpayers/citizens, not just for high-income gadget-buyers.

- Easy-to-use query interface

1-D, 2-D and 3-D graphics software should be employed for map query interface.

4.4 System Implementation

Before system implementation, one has to first identify potential users of archived ITS data, such as university research, planners, traffic management centers, etc. and their specific data requirement. ITS market segmentation based on attitudes and values related to control, time, travel, and technology successfully identifies much of the current ITS customer market, differentiating ITS customers from others with similar demographic characteristics.

Control-seekers dominate the ITS customer market. These customers consult ITS to save time, to use their time efficiently, to stay on schedule, and to stay informed. Control-seekers use information more intensively than the general public.

Technology has an important and complex role in ITS. Web Heads comprise the second largest group of ITS customers. However, their allegiance appears linked to the Internet media, and may or may not migrate to other information platforms as the web becomes more mobile.

Individuals in the low-tech pre-trip information seekers market segment had a low acceptance and comfort level with the Internet and web-based information. Nevertheless, this customer segment represents a large portion of the current ATIS customer pool, and can be expected to continue to demand good information services on low-tech media in the future.

Concurrently, consumers' expectations for advanced information services generally are very high. They've been conditioned by the Internet and a computing environment in which information services and electronic devices get faster, better, and cheaper very quickly. In the research and evaluation to date, we see a progression in the expectations and requirements of drivers as they become more experienced ITS consumers.

Another important issue during system implementation is privacy. The conflicts of having ITS data that records the location or travel pattern of an individual with some privacy concerns must be solved. Public sector doesn't want confidential or sensitive information released. And it doesn't want to reduce system security (increased hacking attacks) by opening up systems. After documenting existing regulations, policies, and guidelines addressing the issues of privacy considerations and rights, it is required to recommend public and private roles and responsibilities in the treatment of data and opportunities for partnership.

4.5 Data Mining Model

Statistical clustering and classification techniques can be utilized for data mining. Specifically, hierarchical cluster analysis approach, method of classification and regression tree would be powerful for treating large database.

Hierarchical cluster analysis is a statistical method for finding relatively homogeneous clusters of cases based on measured characteristics. It starts with each case in a separate cluster and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left. When there are N cases, this involves N-1 clustering steps, or fusions. This hierarchical clustering process can be represented as a tree, or dendrogram, where each step in the clustering process is illustrated by a join of the tree. The horizontal scale corresponds to the fusion values obtained from the hierarchical cluster analysis. Methods of hierarchical cluster analysis include single linkage, complete linkage, average linkage, weighted average linkage, mean proximity, centroid, median, increase in sum of squares, sum of squares, flexible, density, etc.

When involved in data mining or analyzing large surveys, k-means analysis and hierarchical cluster analysis can handle different types of variables, such as those in survey questionnaires and database records.

Acknowledgments

This study was sponsored by the Wisconsin Department of Transportation. The authors gratefully acknowledge the valuable assistance and contribution of WisDOT Project Manager John Corbin and administrative contact Linda Keegan. We also want to thank members of the Technical Oversight Committee, who provided many helpful suggestions on project scope and direction.

References

1. ADUS Program, "ITS Data Archiving: Five-Year Program Description", March 2000.
2. Archived Data User Service Resource Page, http://www.itsa.org.
3. Archived Data User Service (ADUS): An Addendum to the ITS Program Plan, Final Version 3.0 (1998). http://www.itsdocs.fhwa.dot.gov/jpodocs/repts_pr/41401!.htm, EDL# 5224.
4. Archived ITS Data: Data Aggregation and Quality Issues. Texas Transportation Institute, 1999.
5. Briggs, V., and Walton, C.M., 'The Implications of Privacy Issues for Intelligent Transportation Systems (ITS) Data", Texas Transportation Institute, SWUTC/00/472840-00075-1, May 2000.
6. California Freeway Performance Measurement (PeMS) Project, 2001.
7. Cambridge Systematics, Inc. "National Transportation System Performance Measures". Report DOT-T-97-04. U.S. Department of Transportation, Washington, DC, April 1996.
8. Cary, N.C., "The SAS Data Warehouse: Information Systems Management and the Data Warehouse Concept." White Paper, SAS Institute, 1996.
9. Chaudhuri, S., and Dayal, U. *An Overview of data warehousing and OLAP technology*. SIGMOD Record 26, 1997, pp 65-74.
10. Chin, S.M., D.L. Greene, J. Hopson, H.L. Hwang, and B. Thompson. "Towards National Indicators of VMT and Congestion Based on Real-Time Traffic Data." Preprint. 78th Annual Meeting of Transportation Research Board, Washington, DC, 1999.
11. Corbin, J.M., Hocking, R.J., Beck, J. "The Southeastern Wisconsin Freeway System Operational Assessment Process: Integrating Freeway System Management and Performance-based Decision Support", ITE 2000 Annual Meeting and Exhibit.
12. Corey, M.J., Abbey M., Abramson I., and Taub B. *Oracle 8 Data Warehousing: A Practical Guide to successful Data Warehouse Analysis, Build and Roll-Out*. Osborne/McGraw-Hill Companies, 1998.
13. Devlin, B. *Data Warehouse from Architecture to Implementatio*n. Addison-Wesley Developers press, 1997.
14. Edelstein, H. *Introduction to data mining and knowledge discovery*, 2nd Edition by Two Crows Corporation, 1998, "http://www.twocrows.com".
15. Flinner, M. "Traffic Data Quality Procedures." The Pooled Fund Project Newsletter, Issue 1, Minnesota Department of Transportation, St. Paul, Minnesota, December 1997.

16. Gifford, J.L., Stalebrink, O.J. “Transportation Asset Management: A Vehicle for Mainstreaming ITS?”, Proceedings of 11th ITS America Annual Meeting, Miami Beach, Florida, 2001.
17. Gray, J. “Evolution of Data Management,” In Computer, October 1996, pp. 38-46.
18. Hackney, D. *Understanding and Implementing Successful Data Marts*. Addison-Wesley Developers press, 1997.
19. Hallenbeck M.E., J.M. Ishimaru, and J. Nee, “Central Puget Sound Freeway Network Usage and Performance”, May 2001.
20. http:\\www.sas.com\software\olap
21. ITS America, “Proceedings from ATIS Data Collection Guidelines Workshop”, ATIS Data Collection Guidelines Workshop, Scottsdale, Arizona, 2000.
22. ITS as a Data Resource: Preliminary Requirements for a User Service, 1998, http://www.itsa.org.
23. ITS Data Archiving Resources CD-ROM (2000), Contact Shawn Turner, Texas Transportation Institute, (979) 845-8829, Shawn-Turner@tamu.edu.
24. Loop Client: Real-Time Freeway Sensor Information Over the Internet, http://www.its.washington.edu/software/loop_cli.html, updated April 8, 1997.
25. Manor, D. “Smart Solution: ATIS Data Collection Made Easy”, Traffic Technology International, 2001.
26. McGurrin, M., Glassco, R., Mertig, A., and Roberts, D., "Applicability of XML to ITS", MITRETEK presentation, March 2000.
27. Mergel, Joseph. “The Capability of ITS to Provide Systems Performance Data: The Use of ITS Data in Planning in the San Antonio Area. Draft Report”. Volpe National Transportation Systems Center, Cambridge, Massachusetts, August 1998.
28. Miller, K.T., Novotny, L.R., Sultana, S. “Development of An Asset Management System for ITS Maintenance”; Proceedings of 8th World Congress on Intelligent Transport Systems, Sydney, Australia, 2001.
29. Oak Ridge National Laboratory. “Proof of Concept of ITS as an Alternative Data Resource: A Demonstration Project of Florida and New York Data”. Statement of Work. Oak Ridge, Tennessee, September 28, 1998.
30. R.E. Brydia, S.M. Turner, W.L. Eisele, and J.C. Liu. “Development of Intelligent Transportation System Data Management”. Transportation Research Record, 1625. Transportation Research Board, Washington, DC, 1998.
31. Scherer, W.T., and Smith, B.L. *The Development of Integrated Intelligent Transportation Systems (IITS)*, Proceedings of Annual Meetings for TRB, Washington, D.C., 1999.
32. TDAD: Traffic Data Acquisition and Distribution. University of Washington, http://www.its.washington.edu/tdad/, March 18, 1999.

33. Turner, S., "Archived ITS Data: Data Aggregation and Quality Issues", TTI Presentation, 1999.

34. Turner, S, "ITS Data Archiving: Case Study Analysis of San Antonio TransGuide Data", Texas Transportation Institute, August 1999.

35. Turner S.M., R.E. Brydia, J.C. Liu, and W.L. Eisele, "ITS Data Management System: Year One Activities", FHWA/TX-98/1752-2, August 1997.

36. Turner S.M., Gajewski, B. J., Widenthal, M., and Dresser, G. B., "Using Archived ITS Data for Sensitivity Analyses in the Estimation of Mobile Source Emissions", SWUTC/00/167406-1, December 2000.

37. Walton, J. “Data Collection Through ITS/CVO Initiatives: The Advantage CVO Program.” Presented at the National American Travel Monitoring Exhibition and Conference (NATMEC). Charlotte, North Carolina, May 1998.

38. Wolfson, D. “Scalable AZTech Data Server Enhancements for Planning and Operations”. Draft Scope of Work. Maricopa County DOT, Arizona, March 1999.

39. USDOT, "ITS As A Data Source", 2000.

40. USDOT, "ITS Data Registry Functional Operating Procedures", Sept., 1998.

41. USDOT. *The National Architecture for Intelligent Transportation Systems*. Washington, D.C., 1996.

42. What Have We Learned About Intelligent Transportation Systems? Chapter 7: What Have We Learned About Cross-Cutting Technical and Programmatic Issues? 2000. http://www.itsdocs.fhwa.dot.gov/jpodocs/EDLBrow/@@301!.pdf, EDL# 13323.

43. Youngworth, P. Data Visualization and the Intranet, Turning Mounds of Data into Shared Corporate Insights Dimensional Insight, 1998. (http://www.datawarehouse.com/whitepaper/dv1.htm)