Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX-99/1752-4		

4. Title and Subtitle	5. Report Date
PROCEEDINGS FROM THE TEXAS ITS DATA USES AND ARCHIVING WORKSHOP	February 1999
	6. Performing Organization Code
7. Author(s) Shawn M. Turner and William L. Eisele	8. Performing Organization Report No. Report 1752-4
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. Project No. 0-1752
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Office P.O. Box 5080 Austin, Texas 78763-5080	13. Type of Report and Period Covered Research: August 1998-September 1999 14. Sponsoring Agency Code

15. Supplementary Notes

Research performed in cooperation with the Texas Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.
Research Project Title: TransLink® Research Program

16. Abstract

The "Texas ITS Data Uses and Archiving Workshop" was held November 10, 1998, in Austin, Texas, to discuss issues and opportunities related to archiving data from intelligent transportation systems (ITS). The workshop participants represented several of Texas' urban areas (Austin, Dallas, Ft. Worth, Houston, and San Antonio) as well as many different disciplines within transportation, including planning, system operations and management, system integration, research and evaluation, emergency management, and air quality analysis.

The purpose of the workshop was threefold: 1) share information about ITS data archiving activities and plans at the national level as well as within Texas' major urban areas; 2) discuss common issues and areas of concern related to ITS data archiving and data needs; and 3) identify common themes, best practices, and issues that need to be addressed. Technical presentations at the workshop focused on the following topics: 1) national activities related to the archived data user service (ADUS), traffic management data dictionary (TMDD), and the national data registry; 2) state activities related to ITS architecture, statewide system integration, and the information sharing policy; 3) data archiving and management approaches at several TxDOT districts; and 4) planning data collection practices and needs. Break-out discussion groups covered the topics of data user groups and needs, data warehousing technology, and data standards.

17. Key Words	18. Distribution Statement
ITS Data, Data Archiving, Data Warehousing, Data Standards, Transportation Management Centers	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)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	136	

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

Proceedings From the Texas ITS Data Uses and Archiving Workshop

FHWA/TX-99/1752-4

TransLink®

Innovative Surface Transportation Management

Texas Transportation Institute
The Texas A&M University System
College Station, Texas

Texas Department of Transportation

in cooperation with the
U.S. Department of Transportation
and the Federal Highway Administration

PROCEEDINGS FROM THE TEXAS ITS DATA USES AND ARCHIVING WORKSHOP

by

Shawn M. Turner, P.E., Editor
Assistant Research Engineer
Texas Transportation Institute

and

William L. Eisele, P.E., Co-Editor
Assistant Research Scientist
Texas Transportation Institute

Report 1752-4
Project Number 0-1752
Research Project Title: TransLink® Research Program

Sponsored by the
Texas Department of Transportation
In Cooperation with the
U.S. Department of Transportation
Federal Highway Administration

February 1999

TEXAS TRANSPORTATION INSTITUTE
The Texas A&M University System
College Station, Texas 77843-3135

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 Texas Department of Transportation (TxDOT) or the Federal Highway Administration (FHWA). This report does not constitute a standard, specification, or regulation, nor is it intended for construction, bidding, or permit purposes. This report was edited by Shawn Turner (Texas certification number 82781) and William Eisele (Texas certification number 85445).

ACKNOWLEDGMENT

The editors acknowledge the support and assistance of the project director, Mr. Al Kosik, Traffic Operations Division, Texas Department of Transportation.

This research is being performed in cooperation with the TransLink® Research Center. The TransLink® Partnership includes the Texas Department of Transportation (TxDOT), Metropolitan Transit Authority of Harris County (METRO), United States Department of Transportation (U.S. DOT), Rockwell International, Southwestern Bell-Technology Resources, and the Texas Transportation Institute (TTI).

The editors acknowledge the support and assistance of Mr. Christopher Poe, who provides overall leadership and direction for the TransLink® Research Program at TTI. The authors also wish to thank the following students who assisted in transcribing and editing the workshop proceedings:

- Luke Albert
- Kyle Dooley
- Pete Ferrier
- Emily Stubbs

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ACRONYMS

AASHTO	American Association of State Highway and Transportation Officials
ADUS	archived data user service
ATIS	advanced traveler information system
ATMS	advanced traffic management system
ATR	automatic traffic recorder
AVI	automatic vehicle identification
CMAQ	congestion mitigation/air quality
CVO	commercial vehicle operations
DBA	database administrator
DMS	dynamic message sign
FHWA	Federal Highway Administration
FOT	field operational test
FTP	file transfer protocol
FY	fiscal year
GB	gigabyte
GIS	geographic information system
HAZMAT	hazardous materials
HGAC	Houston-Galveston Area Council
HSM	hierarchical storage management
IEEE	Institute of Electrical and Electronics Engineers
ISP	information service provider
ITE	Institute of Transportation Engineers
ITS	intelligent transportation system
LCS	lane control signal
LCU	local controller unit
MIS	major investment study
MMDI	Metropolitan Model Deployment Initiative
MPO	metropolitan planning organization
NHS	National Highway System
NTCIP	National Transportation Communication for ITS Protocol
OLAP	on-line analytical processing
OLTP	on-line transaction processing
SDIWG	statewide development and integration working group
SRI	Southwest Research Institute
TMC	transportation management center
TMDD	traffic management data dictionary
TMG	Traffic Monitoring Guide
TNRCC	Texas Natural Resource Conservation Commission
TP&P	Transportation Planning and Programming
TRB	Transportation Research Board
TTI	Texas Transportation Institute
TxDOT	Texas Department of Transportation

WORKSHOP SUMMARY

The "Texas ITS Data Uses and Archiving Workshop" was held November 10, 1998, in Austin, Texas, to discuss issues and opportunities related to archiving data from intelligent transportation systems (ITS). The workshop participants represented several of Texas' urban areas (Austin, Dallas, Ft. Worth, Houston, San Antonio) as well as many different disciplines within transportation, including planning, system operations and management, system integration, research and evaluation, emergency management, and air quality analysis. The Texas Department of Transportation (TxDOT), the TransLink® ITS Research Center, and ITS Texas sponsored the workshop.

The purpose of the workshop was threefold:

1) share information about ITS data archiving activities and plans at the national level as well as within Texas' major urban areas;
2) discuss common issues and areas of concern related to ITS data archiving and data needs; and
3) identify common themes, best practices, and issues that need to be addressed.

The following paragraphs summarize the major findings and conclusions from the workshop.

Strong Interest from TxDOT. The Traffic Operations Division of TxDOT has expressed strong interest in archiving and sharing data from ITS applications. The primary motivation for archiving ITS data for this group has been the documentation of ITS benefits. However, this group also recognizes that other data users, such as planners and researchers, have a stake in archived ITS data. TxDOT Policy Statement 2-98, "ITS Information Sharing," explicitly underlines this commitment by requiring that all TxDOT districts "shall make all transportation related information available, including . . . archived historical data."

Activities at the District Level. As noted in the technical presentations, three of Texas' traffic management centers (TMCs) have plans for or are already archiving data being collected or generated by ITS applications. Staff at Houston's TranStar are currently working with the Houston-Galveston Area Council (HGAC) on a pilot data warehousing project, which they hope will complement a larger, more ambitious effort at developing an enterprise management system, of which a comprehensive ITS data warehouse will be a major component. At Fort Worth's TransVISION, an on-line analytical processing function is included in the ultimate system design, although lack of funding may delay its implementation for several years. At San Antonio's TransGuide, data from several ITS applications are archived daily as standard operating procedure. In addition, TransGuide staff have supported several data sharing and warehousing efforts with the Texas Transportation Institute (TTI), including the development of ITS DataLink, a prototype ITS data management system. These activities indicate a strong interest as well as a commitment at TxDOT's district level to archiving and sharing data being collected at Texas' TMCs.

Involvement of Statewide Systems Integrator. The Southwest Research Institute was selected by TxDOT as the statewide systems integrator, and has been tasked to support TxDOT in ITS standardization efforts in Texas. Southwest Research Institute is currently providing technical assistance to Houston TranStar on their pilot data warehousing project. It appears that, given the different data archiving approaches in several of TxDOT's districts, the statewide integrator could play a role in any standardization issues related to archiving ITS data. The statewide integrator could also play a role, along with other data stakeholders, in implementing the components of the newly formed archived data user service (ADUS).

Dialogue with Data Users. As plans move forward for archiving and warehousing ITS data in Texas, it will be important to establish and maintain a dialogue with potential data users (e.g., planners, researchers, etc.). This workshop provided a forum for exchanging information about ITS data being collected by TMCs as well as understanding the data requirements and needs of planners. Based upon discussions at the workshop, there was a strong interest in continuing to exchange information among these and other data user groups. Additionally, workshop participants identified many data needs that currently deployed ITS applications may not be able to provide in the near future, including arterial street coverage, intermodal connectivity, and system expandability. These concerns point to the possibility that archived ITS data, even from ITS applications deployed in the near future, may not be able to address every single data need from the many data user groups.

Cost of Implementation. There were several discussions at the workshop about the cost of implementing ITS data archiving or warehousing. Cost was an issue for several reasons: 1) scarce resources at all government levels require agencies to do more with less; 2) funding for ITS projects may be scrutinized more closely than traditional roadway funding; and 3) the operating agency that typically collects ITS data is not seen as the primary beneficiary for the archived ITS data. Agencies interested in archiving or warehousing data should seek innovative solutions for implementation, such as cost sharing among potential data users, or replacing/supplementing traditional manual data collection program with data collected by ITS applications.

Data Warehouse Design Considerations. Workshop participants identified a wide range of data needs, user groups, and user applications. For example, user groups and applications ranged through all aspects of transportation, including planning, design, construction, operations, maintenance, safety, and evaluations. Group discussions centered on the difficulty in meeting data needs from myriad user groups and applications with a single data warehouse design. It appears that the majority of users can be best served by a data management system with two functions: 1) an on-line, easily accessible data warehouse with a core set of data applications, such as providing data summaries in user-prescribed formats with graphical capabilities; and 2) an off-line data archive that contains all data elements of interest in the most disaggregate form. With this approach, the data management system can provide ease of use and accessibility for typical data users, while retaining flexibility of analysis for power data users.

The following workshop proceedings contain edited transcripts from presentations and group discussions. All presenters were given the opportunity to review and/or clarify their edited transcripts. Where visual aids were used in presentations, the editors have included these visual aids within the transcript itself. The workshop was divided into the following four sessions (Figure S-1):

ITS as a Data Resource
Creation of an Archived Data User Service - Rich Margiotta, Cambridge Systematics
Statewide Architecture and System Integration - Doug Lowe, TxDOT
TxDOT Information Sharing Policy - Al Kosik, TxDOT

Approaches to Data Storage, Archiving, and Management
Houston TranStar - Sally Wegmann and Cindy Gloyna, TxDOT
Fort Worth TransVISION - Abed Abukar, TxDOT
San Antonio TransGuide - Pat Irwin, TxDOT
DataLink: Development of an ITS Data Management System - Shawn Turner, TTI

Panel Discussion on Data Needs
TxDOT Planning Perspective - Dayton Grumbles and Mark Hodges, TxDOT
Development of a National Data Registry - Ed Seymour, TTI
Data Needs: Traffic Management Data Dictionary - Steve Dellenback, Southwest Research Institute

Break-Out Discussion Groups
Data Needs and Uses
Data Warehousing Technology and Data Standards

TEXAS ITS DATA USES AND ARCHIVING WORKSHOP

Overview

Intelligent transportation systems (ITS) are maturing in urban areas across the U.S. The deployment of ITS components is generating, and will continue to generate, vast amounts of data. There are primary uses of this ITS data in 'real-time' applications to operate and manage the transportation system. These applications range from incident management to the presentation of real-time traffic conditions for travelers. There are also secondary uses of this data beyond operational uses including planning, designing, and evaluating the transportation system. These secondary uses, however, require *management* of the data to ensure utility of the data for additional users. ITS data management is a comprehensive approach consisting of data collection, storage, aggregation, access, and quality control.

The following one-day workshop is planned to bring together representatives from many agencies that have a stake in ITS data. The goals of the workshop are:

1) Present an overview of national and state ITS data archiving activities
2) Present current approaches to ITS data collection, storage, and management
3) Discuss data needs for secondary users
4) Identify action items to facilitate implementation and research.

FINAL TECHNICAL PROGRAM

Tuesday, November 10th

10:00 am **Welcome, Al Kosik from Traffic Operations Division, Texas Department of Transportation**

10:05 am **Workshop Overview, Christopher Poe, Director, TransLink® Research Center, Texas Transportation Institute**

10:15 am **ITS as a Data Resource**

- National Perspectives, Rich Margiotta, Cambridge Systematics
- Statewide Architecture and System Integration, Doug Lowe, TxDOT - TRF
- TxDOT Information Sharing Policy, Al Kosik, TxDOT - TRF

11:00 am **Approaches to Data Storage, Archiving, and Management**

- Houston TranStar, Sally Wegmann and Cindy Gloyna, TxDOT-Houston
- Fort Worth TransVISION, Abed Abukar, TxDOT-Fort Worth
- San Antonio TransGuide, Pat Irwin, TxDOT - San Antonio
- DataLink: Development of an ITS Data Management System - Shawn Turner, TTI

12:00 pmLunch

1:00 pm **Panel Discussion on Data Needs**

- TxDOT TP&P Perspective, Mark Hodges, TxDOT - TP&P
- National Traffic Management Data Dictionaries Effort, Steve Dellenback, Southwest Research Institute
- Development of a National Data Registry, Ed Seymour, Texas Transportation Institute

2:00 pm **Break-out Groups**

- Break-out Group A: Data Needs and Uses (data uses, analysis requirements, data quality)
- Break-out Group B: Data Warehousing Technology (current data storage and warehousing technology, privacy issues, security issues)
- Break-out Group C: Data Standards (ITS data standards, data dictionaries, meta data, system architecture)

3:30 pm **Break**

3:45 pm **Presentations from the Break-out Groups**

4:30 pm **Workshop wrap-up (recap of major issues and future efforts)**

5:00 pm **Workshop Adjourns**

Workshop Outcome

The Texas Transportation Institute will record presentations and discussions at the workshop. This material will be synthesized into a workshop proceedings. The proceedings will be used to help guide future efforts in information and data sharing, ITS deployment, and ITS research.

Figure S-1. Texas ITS Data Uses and Archiving Workshop Agenda

OPENING REMARKS

INTRODUCTION

Al Kosik
Texas Department of Transportation

I would like to welcome you to the intelligent transportation system (ITS) data workshop this morning, particularly on behalf of Tom Newbern, the director of the Texas Department of Transportation's (TxDOT's) Traffic Operations Division, who is unable to be here today because of previous travel commitments. He told me to tell you that he appreciates your participation in this workshop today. We hope to have a good workshop as well as good discussion and results. In the traffic operations area, we have been concerned about the data that is being collected by traffic management systems for quite some time. We have been using the data for various operational analyses, as well as some planning applications. It really started becoming critical for us as we talked about how to document the benefits of ITS. We are questioning whether we have been collecting the right data. We also are questioning whether we are saving the right data. What we need to do now is to collect and store the correct data so that two years from now we can perform ITS benefits and other operational analyses, as well as having the ability to do better planning. This is a very key objective for us, and we hope to get good results from this workshop.

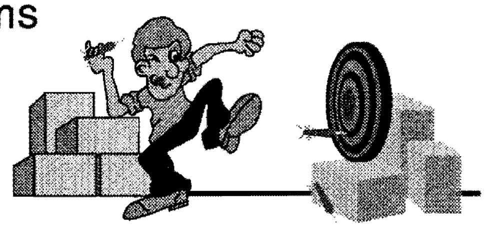
WORKSHOP OVERVIEW

Christopher Poe
Texas Transportation Institute

I appreciate everyone being here today for the workshop. The idea for a Texas ITS data workshop grew from some of the ITS data archiving discussions that were happening statewide. There is a significant amount of work going on nationally related to ITS data archiving and the archived data user service (ADUS). We have some presenters here today to talk about these national activities. The individual TxDOT districts and their partners in the urban areas are starting to move forward with assessing data needs, and there is some research that is occurring there as well. Because of all these efforts, we thought it was an appropriate time to bring these groups within Texas together and share information and ideas. I would like to recognize that this workshop is sponsored by TxDOT through the TransLink® ITS Research Program at the Texas Transportation Institute (TTI). As director of the TransLink® Center, I, along with other TTI researchers, will be helping to facilitate the workshop today. ITS Texas is also helping to sponsor the workshop, as they will be providing the catered lunch today. We hope to distribute the workshop proceedings to the participants as well as the general membership of ITS Texas.

Workshop Goals

- Present an overview of national and state ITS data archiving activities
- Present current approaches to ITS data collection, storage, and management
- Discuss data needs
- Identify action items

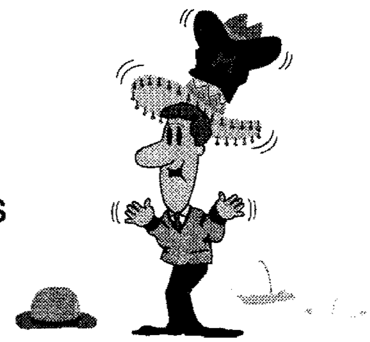


TransLink® Research Center Texas Transportation Institute

First I would like to discuss what we want to accomplish today. We want to present an overview of ITS data archiving activities that are happening at the national and state level. One of the things that everyone should have received with their invitation packet is the Federal Highway Administration (FHWA) report titled "ITS as a Data Resource," which was authored by Rich Margiotta. We are glad to have Rich here today to give some background on this report and other national activities related to ADUS. We have other workshop participants here that have been leading related national efforts, and we hope that they can contribute to the discussions today. We also have several representatives from the urban TxDOT districts to talk about current ITS data archiving activities in these urban areas. After these scheduled presentations, we would like to start more informal discussions about several topics, such as the data needs of various stakeholders. I think that one of the points that Rich will convey this morning is that the ITS data area is complicated because there are so many stakeholders, more so than many of the other user services in the National ITS Architecture. We hope to emerge at the end of today with some action items, in terms of where to go from here, who are the stakeholders, and what we can do to advance the state-of-the-practice.

Agenda for the Workshop

- Overview of national and state ITS data efforts
- Current ITS data efforts from TxDOT districts
- ITS data uses
- Break-out groups
  - Data needs and uses
  - Data warehousing
  - Data standards



TransLink® Research Center | Texas Transportation Institute

The workshop agenda includes technical presentations this morning and early afternoon, with some break-out groups scheduled for later in the afternoon. The first group of presentations will provide an overview of activities at the national level as well as within TxDOT's Traffic Operations Division. The second group of presentations will detail specific efforts in several Texas urban areas. After lunch, we will have one last group of presentations that will discuss data needs, the traffic management data dictionary (TMDD), and the Institute of Electrical and Electronics Engineers (IEEE) national data registry concepts. After these presentations, we have three break-out groups scheduled to discuss data needs, standards, and technologies. We will let you select your break-out group, and the presentations this morning and afternoon should help you in making this decision. We do hope to have an even distribution, so we encourage you to spread yourselves out among these three groups. Break-out group one will talk about data needs and data uses. Break-out group two will talk about data archiving and warehousing technology, and more specifically the tools to store and archive ITS data. The third break-out group will focus on data standards. To conclude, we will have a final session that summarizes the break-out groups as well as identifies future activities.

Output from the Workshop

- Raise awareness of ITS data issues in Texas
- Workshop proceedings
- Provide input to the TxDOT Statewide System Integrator effort
- Interim information to local partners working on ITS data initiatives

TransLink® Research Center — Texas Transportation Institute

With this workshop, we hope to have heightened awareness among the Texas participants about what is going on nationally, as well as identifying the issues related to archiving and warehousing ITS data. We are planning to produce proceedings from this workshop that will be made available to workshop participants. Shawn Turner of TTI will be developing these proceedings and has a number of TTI participants who will be helping to record today's discussions. We hope that today's discussions will provide some input to Doug Lowe of TxDOT, who is managing the statewide systems integrator contract, in examining ITS data archiving and warehousing issues. We also hope that today's discussions will benefit those from the urban areas, who are currently trying to address data archiving, data warehousing, and data storage needs. The participants from the urban areas need immediate guidance on some issues, and our hope is that some of the discussions here can help as they move forward with implementation. In conclusion, those are the goals for this workshop. I would like to add that this is quite a bit of material to address in one day. We might not be able to accomplish everything today, but we hope that this workshop has brought together the right stakeholders, so that we can identify further efforts, meetings, or working groups that are needed.

ITS AS A DATA RESOURCE

CREATION OF AN ARCHIVED DATA USER SERVICE

Rich Margiotta
Cambridge Systematics

I would like to talk to you today about some of the national activities that are related to the archived data user service (ADUS). In addition to the information in the FHWA report ("ITS as a Data Resource"), I would like to tell you about more recent and upcoming activities that FHWA is leading in this area. These current activities can be categorized into three main areas: 1) develop revisions to the National ITS Architecture, 2) input to the standards setting and TMDD efforts, and 3) conduct focused research on relevant topics.

You may be asking "what is a user service?" I am a transportation engineer by training and vocation, and I have been swimming in the Architecture for about the last six months. Needless to say, it is a daunting experience. I will attempt to give you an "outsider's" view, or a non-systems perspective, on the National ITS Architecture, including experiences that I have had. A user service is basically the definition of a primary function that ITS is supposed to perform. Incident management, traffic management, and highway-rail intersection operations are all examples of user services within the National ITS Architecture. There are currently 31 user services contained in the Architecture, which are intended to encompass a full range of transportation management activities.

Why Are We Doing This?

- **ITS collects large amounts of continuous and (usually) real-time data to implement control strategies**
  - **Freeway surveillance data for ramp metering control**
  - **Electronic fare payments**
- **ITS-generated data have a huge potential for secondary (nonreal-time) uses**
- **ITS National Architecture does not currently include a specification for archiving data**
- **FHWA wants to promote the use of ITS-generated data as a supplement to existing data programs**

CAMBRIDGE SYSTEMATICS

You may ask "why do we need an archived data user service?" At this point, I think that we can all agree on the first two bullets of this slide. Many ITS applications do collect large amounts of data that can be used not only for ITS control strategies, but for secondary uses such as planning and evaluation. Given this potential, there has been strong federal interest in revising the Architecture and putting ITS data archival into practice. When these discussions first began, those involved felt that the best way to put this into widespread practice was to revise the Architecture and develop a new user service.

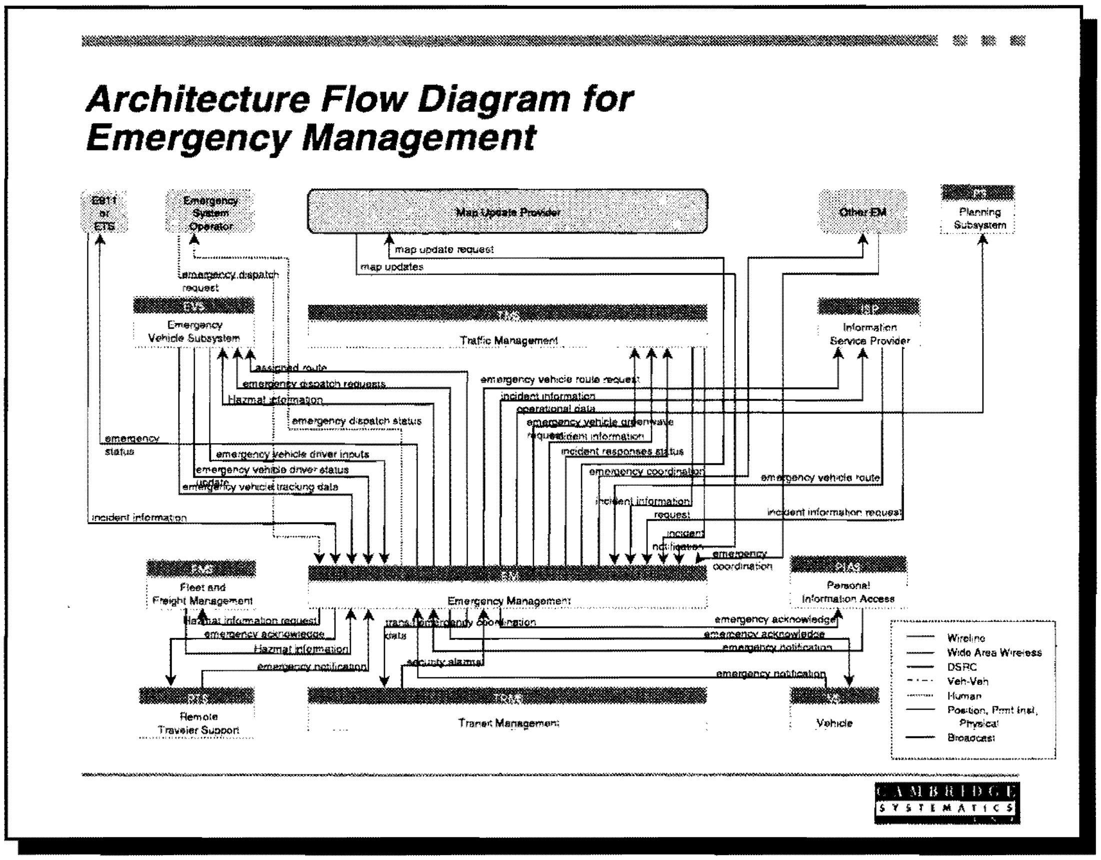
Stakeholders for ITS-Generated Data

- MPO and state transportation planners
- Traffic management operators
- Transit operators
- Air quality analysts
- Freight and intermodal planners
- Safety planners and administrators
- Transportation system monitoring
- Design, construction, and maintenance personnel
- Commercial vehicle enforcement personnel
- Emergency management personnel
- Transportation researchers
- Private sector users
- Land use regulation and growth management

By far the largest number of stakeholders for a National Architecture User Service

CAMBRIDGE SYSTEMATICS

As Chris mentioned earlier, there are a variety of stakeholders that have an interest in ITS data. Most of the other user services in the National ITS Architecture are very narrowly focused and the stakeholders that were defined were very limited (most were between one and three). With ADUS, however, there are 13 stakeholder groups. The stakeholders include transportation planners, which are the obvious ones, as well as many other groups that have an interest in ITS data, such as emergency management personnel planning for system deployment, air quality analysts, etc. It appears that most participants here today are traffic management operators, those who operate ITS, or local/regional traffic operations personnel. The workshop participants should be aware that there are a variety of other stakeholders who have a vested interest in the data that you may be collecting and archiving.



As I mentioned earlier, I have been "swimming" in the Architecture for approximately the past six months. I would like to talk about the Architecture from a transportation planner/engineer perspective. At the most basic level, the National Architecture is a flow diagram. In general terms, this flow diagram includes boxes, which represent the functions that ITS can provide, and lines that connect boxes, which represent data flow between ITS functions. As you can see from this slide, there are many boxes. This is an example architecture flow diagram from the emergency management user service. Each box on the diagram has its own function to perform. To perform a given function, it has to interconnect with other functions. This connection with other functions is then accomplished by data flows throughout the Architecture. In essence, this is a very quick overview of the National ITS Architecture. There are many levels in the Architecture. There is a top level, and there is a very detailed level where these data flows are defined, perhaps not necessarily in data dictionary terms but at least at the conceptual level. The hope is that the National ITS Architecture can be used to influence regional ITS architectures that are being designed throughout the country. You may ask "why do we need a national architecture?" The concept is that to ensure true interoperability, we need to have building standards and a basic game plan. We need systems design guidance for people who are designing regional architectures. That is the reason for a National ITS Architecture.

What ITS-Generated Data Are Available?

1. Freeway traffic flow surveillance
2. Ramp meter and traffic signal preemptions
3. Ramp meter and traffic signal cycle lengths
4. Visual and video surveillance
5. Vehicle counts from electronic toll collection
6. TMC-generated traffic flow metrics
7. Arterial traffic flow surveillance
8. Traffic signal phasing and offsets
9. Parking management
10. Transit usage (boardings)
11. Transit route deviations and advisories
12. Rideshare requests
13. Incident logs
14. Train arrivals at HRIs
15. Emergency vehicle dispatch records

CAMBRIDGE SYSTEMATICS

There are numerous data elements available within the National ITS Architecture. As I mentioned, the Architecture is an all-encompassing blueprint for ITS, and it covers many areas. Also, it is extremely detailed and is basically a common denominator. The "worst case" scenario is accounted for in the National ITS Architecture. With regional architectures, you may not want to go to that level, but at least it is there for guidance. If you get into the National Architecture, you will find particular types of data that could be of use for secondary applications. There are 30 different types of data in the Architecture that have been identified in the FHWA report ("ITS as a Data Resource"), and this slide shows the first 15 types. I think that everyone here is familiar with ITS surveillance data (e.g., loop detectors, probe vehicles, etc.) and system performance data from systems monitoring devices.

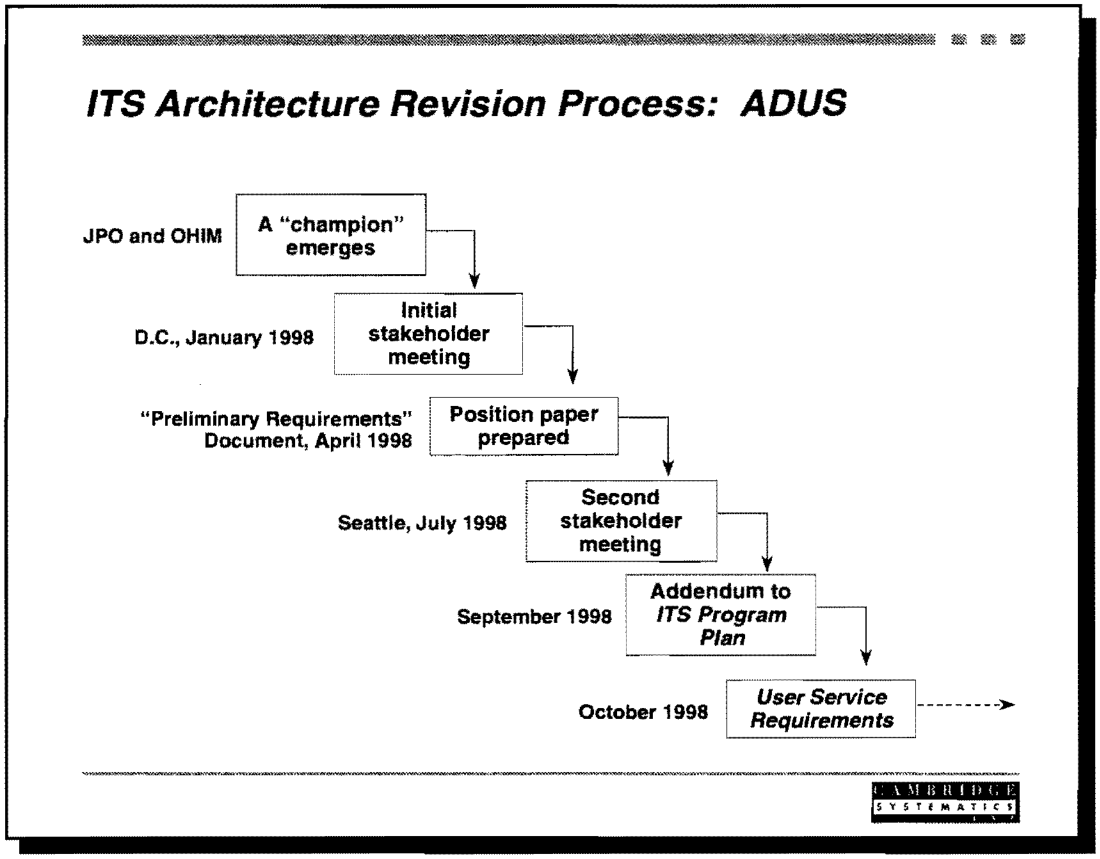
What ITS-Generated Data Are Available? (continued)

16. Emergency vehicle locations
17. Construction and work zone ID
18. HazMat cargo identifiers
19. Fleet activity reports
20. Cargo identification
21. Border crossings
22. On-board safety data
23. Emissions management
24. Weather data
25. Location referencing
26. Probe data
27. VMS messages
28. Vehicle trajectories
29. Route guidance advisories
30. Parking roadway pricing changes

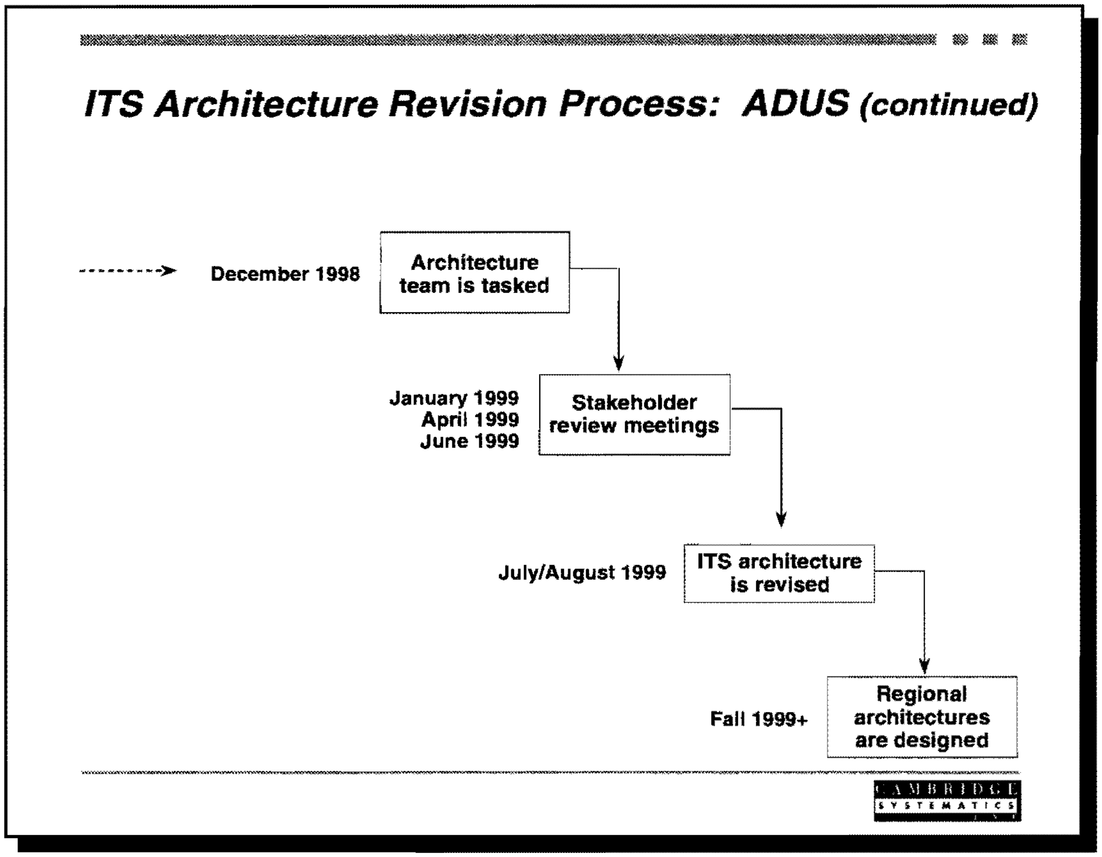
ITS-generated data have basically the same nature as "traditional" data, but are usually collected continuously and in real- or near real-time

CAMBRIDGE SYSTEMATICS

In addition to the surveillance data, which probably have the most widespread use and are certainly the most ubiquitous, there are many types of non-traffic surveillance data that could be of potential use to stakeholders. These include data items such as transit boardings, commercial vehicle operations (CVO) data, cargo data, hazardous materials (HAZMAT) data, environmental data, weather data, etc. So the main point of this slide is to remind you that when you are thinking about archiving data, surveillance data is probably the place to focus your efforts. However, please be aware that there are other data types that could be of use to stakeholders as these applications are developed.



I would like to provide an overview of the Architecture revision process in general, and include information about the most recent revision for ADUS. The first step is that a "champion" for a proposed user service must emerge and officially request incorporation of the proposed user service into the National Architecture. In this case, the ITS Joint Program Office and FHWA's Office of Highway Information Management officially requested the inclusion of ADUS in the Architecture. An initial stakeholders' meeting was held just prior to the Transportation Research Board's (TRB) Annual Meeting in January 1998, and an attempt was made to invite participants from all stakeholder groups. The result of that meeting was largely the FHWA report ("ITS as a Data Resource"), which defined the functions of which ADUS should be capable. A second stakeholder meeting was held in Seattle in July 1998 to discuss the results of the previous workshop and discuss the issues in more depth. The result of this second stakeholder meeting was the development of an addendum to the National ITS Program Plan, which is the guidance document for the National ITS Architecture. In addition, from the Program Plan addendum, a very specific set of user requirements was developed for ADUS. These two documents have recently been approved, and that is currently where the revision process is to date.



At this point in the Architecture revision process for ADUS, we are going from conceptual thinking to more focused activities. The next step in the process is to have the Architecture team get involved and let them work more closely with stakeholders in designing the ADUS architecture. We are nearly ready to start this process of involving the Architecture team. As far as this step of the process is concerned, there will be additional stakeholder meetings where the Architecture team will be presenting the results of their initial work for review. The next stakeholder meeting is scheduled again right before the TRB meeting in January 1999, and some of the workshop participants here may get an invitation. In addition, the Architecture team, in their normal course of business, will meet with stakeholders and try to observe how they do their jobs, so that they can better understand the user needs and requirements. The time frame for these activities to conclude are approximately late summer 1999, and by then the revisions to the Architecture should be made. After that, a miracle occurs and the regional architectures are designed and deployed based upon the National ITS Architecture.

ADUS
General Principles

- Base on *existing* data flows within the National Architecture; new data flows may require additional user services
- Accommodate both centralized and decentralized structures
- Develop detailed *metadata* that document data collection, quality control, and editing procedures in addition to basic attributes
- Data should be stored on-line in *field-reported form* for a length of time
- Data should be aggregated to levels useful for stakeholders, depending on the type of data
- Compatible with other ITS standards and principles: location referencing, NCTIP/TCIP, TMG, privacy

CAMBRIDGE SYSTEMATICS

There are several basic principles associated with ADUS. The first is that ADUS will largely be based upon existing data flows within the Architecture. However, we also will be considering the expansion for certain data flows to help meet stakeholder needs wherever possible. This possible expansion of new data flows is something we will be thinking about as ADUS is developed, but certainly the main effort is to consider the data flows currently available in the National Architecture. The desire is to keep the system design as flexible as possible. We want to be able to accommodate both centralized and decentralized structures when the user service is actually deployed. The third general principle is extremely important, and it is the concept of metadata, or data about data. In addition to knowing what the data definitions are from a data dictionary standpoint, we want to expand it to include other information about the conditions under which the data have been collected. The metadata attributes would also include information about what has been done to that data since data collection. Basically we would like an audit trail of information associated with the data element. Those participants familiar with traffic monitoring practices probably know this best as the "truth in data" concept, where you always know what has happened to data since it has been collected.

To accommodate all of the stakeholders, there should be provisions for saving the data as it is received from the field collection equipment for some period of time. We are not going to specify how long to keep it or what format to store it in, but it is very important to keep the actual raw data and make it available for some period of time (or at least have the capability of storing it). We are not going to specify storing it on magnetic tape or CD-ROM, but at least think about these options when designing your systems. Again, you should have the capability of saving the information in the most basic form as it is received from field equipment. For the data to be usable for most stakeholders, some level of aggregation may be necessary for the data. This possibility of aggregation is particularly true for ITS surveillance data. Typically, 20-second loop detector data is not very useful to transportation planners, but it might be useful to transportation researchers. Again, that is the reason for keeping it at the most disaggregate level. But in most cases, however, some level of aggregation will be necessary for the data to be most useful to the primary stakeholder groups. The last general principle for ADUS is that any existing data standards and principles should be applied to data that is stored in the archives. With ADUS, we are not going to be defining new data standards, but we are going to adopt existing data standards. By this, I mean definitions from some of the other federal data systems, as well as some of the ongoing data standards being developed by the Institute of Transportation Engineers (ITE) or IEEE, which we will hear more about later today.

Technical and Institutional Issues

Revising the National Architecture is necessary but insufficient for achieving successful implementation. Other issues include:

1. **Development, operation, and maintenance costs**
2. **Ownership**
3. **System access**
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 in the National Architecture**
15. **Metric conversion**
16. **Training and outreach**

CAMBRIDGE SYSTEMATICS

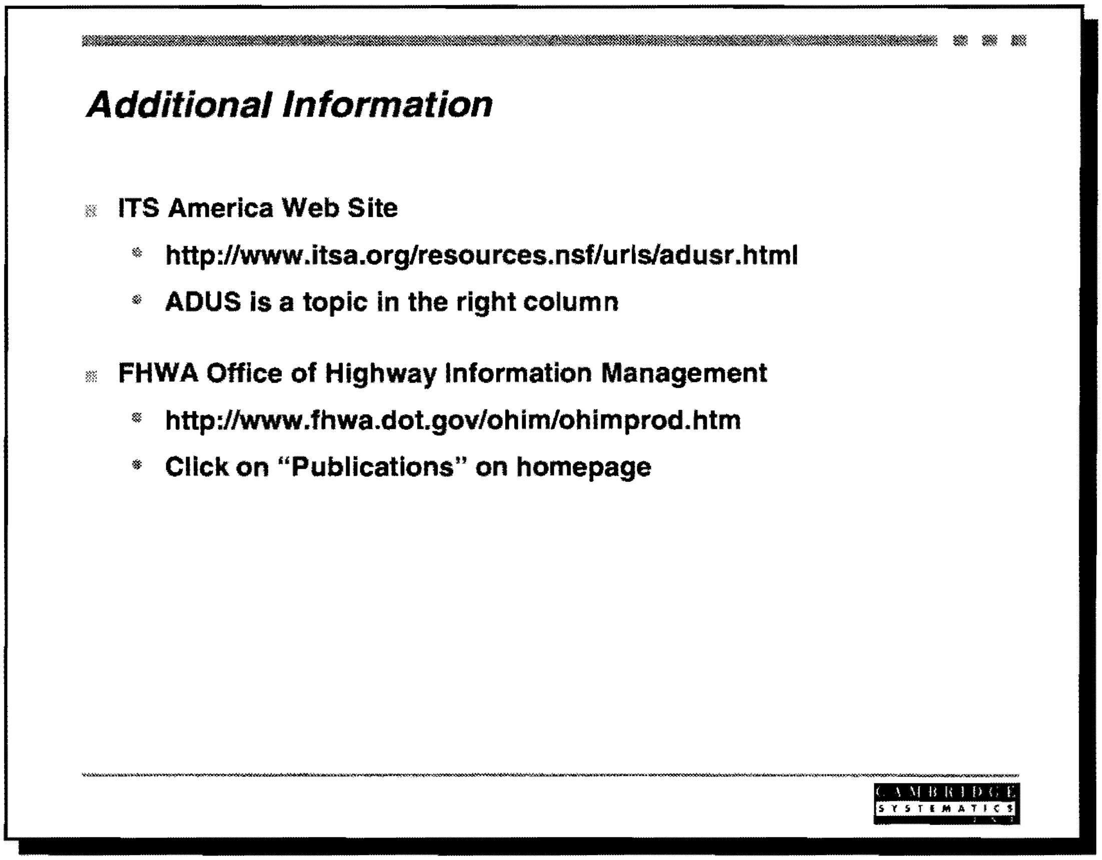
I think I know what most of you are probably thinking. The National Architecture has been revised to include ADUS, now "so what?" because the Architecture is merely guidance. You are partially correct, because there are many technical and institutional issues that need to be addressed before ADUS can move into implementation. The development, maintenance, and operations costs will be a big issue. Who pays for it? Who benefits the most? How to get funding to develop this extra feature within existing ITS infrastructure? Who owns it? Who's in charge? Who is responsible for the system? Who makes the big decisions about the actual operational level? There are privacy concerns associated with archiving data. Most traffic management centers (TMCs) throughout the country, including here in Texas, do not save or archive video images from surveillance cameras mainly for liability reasons (or to avoid being approached by lawyers). However, there are stakeholder groups (e.g., safety researchers) who are excited about the potential of having this video data available for their safety research. This type of data might help them get out the thorny issue of accident causation some day. You should be aware that there are privacy concerns that are associated with saving certain data (i.e., toll tag reads from individual vehicles). Indeed, there are some large issues associated with data archiving that do not exist when you use the data in real-time and let it "evaporate." The retrofitting of existing systems is another issue to consider. It is much easier to build a data archiving function from scratch then it is to retrofit an existing system. In conclusion, that is an overview of the Architecture revision process and the associated issues.

Recommendations for Implementation

- **"Best Practice" procedures for quality control/editing and data analysis should be developed**
- **Coordination with ongoing data dictionary efforts should be established**
- **Stakeholders identified here should be involved in ITS standard setting whenever relevant data are involved**
- **Integration with "traditional" data programs should be promoted**
- **Field demonstration projects highlighting archived ITS-generated data should be undertaken (e.g., FOTs, next MMDIs)**

CAMBRIDGE SYSTEMATICS

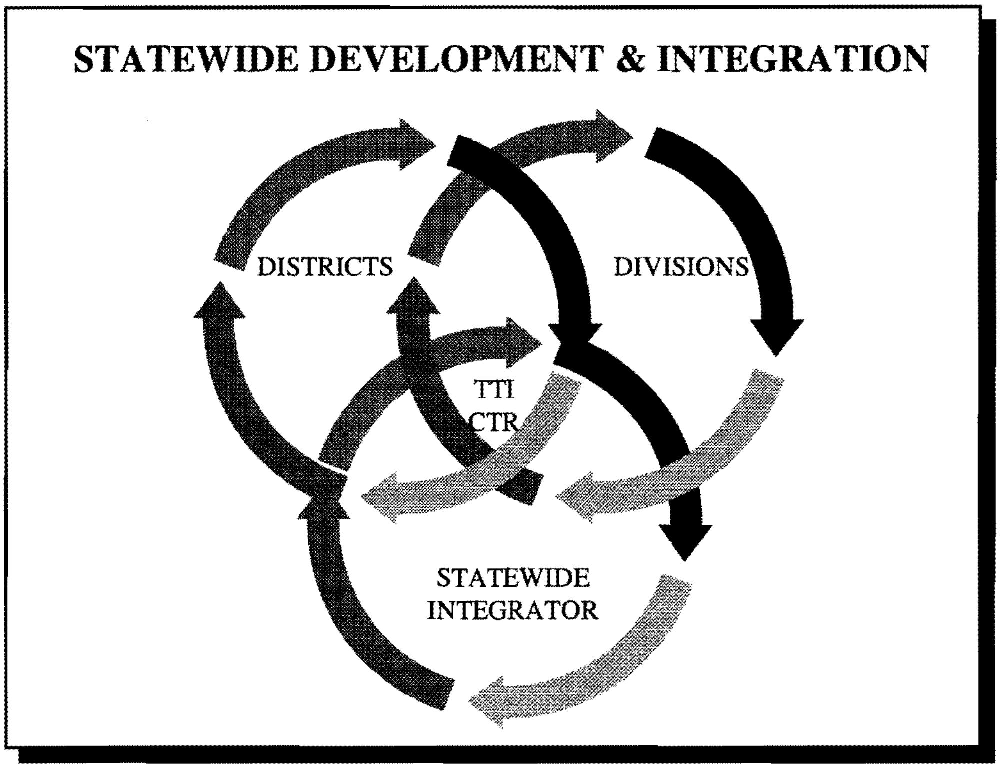
In addition, there are several activities being considered or conducted by FHWA that have an impact on archived data. The first activity is the development of best practices for performing quality control, edits, and analyses of archived ITS data. There are currently no standards or standardized methods for performing quality control checks or editing ITS-related data. The second and third items on the slide are activities that we are definitely involved with, and that is providing input to the data dictionary and standard setting efforts. This is a useful activity, because this has brought representation to these committees that did not exist before. The integration of archived ITS data with traditional data programs should be promoted. There is a research project currently being funded by FHWA through Oak Ridge National Labs to look at integrating archived data into traditional data programs. The idea of a field operational test (FOT) is also being considered as an option to advance the state-of-the-practice in ADUS. An FOT could also serve as a model for how to implement data archiving, as well as identifying what the real issues are in terms of implementing ADUS.



My last slide shows where you can find some additional information on ADUS and current national activities. The two main web sites are located at ITS America and FHWA's Office of Highway Information Management.

STATEWIDE ARCHITECTURE AND SYSTEM INTEGRATION

Doug Lowe
Texas Department of Transportation



The presentation that I will give today is very similar to the presentation I did for a similar group several weeks ago. I do see many new faces in the group, so I was considering how much to talk about the statewide systems integrator and how much to really focus on archived data. Hopefully I can strike a balance today. Rich's comment about "swimming" in the National ITS Architecture was appropriate, as I am also "swimming" in the Architecture (or maybe drowning might be a better choice of words).

STATEWIDE INTEGRATOR CONTRACT

- Concept approved by ITS Committee
- Request for offer distributed to qualified information system vendors
- Southwest Research Institute selected
- Initial period of service 24 months with estimated total cost of $2,500,000
- ITS statewide integration started 1/15/98
- Option to renew for additional 24 months

TxDOT now has an ITS integrator under contract. We went through all the processes of getting the ITS integration contract approved and advertised. After reviewing proposals about a year ago, we selected the Southwest Research Institute. We started working January 15, 1998, and the initial period of service was 24 months, with a contract value of about $2.5 million. The contract does allow for a two-year extension if this is so desired.

STATEWIDE INTEGRATOR QUALIFICATIONS

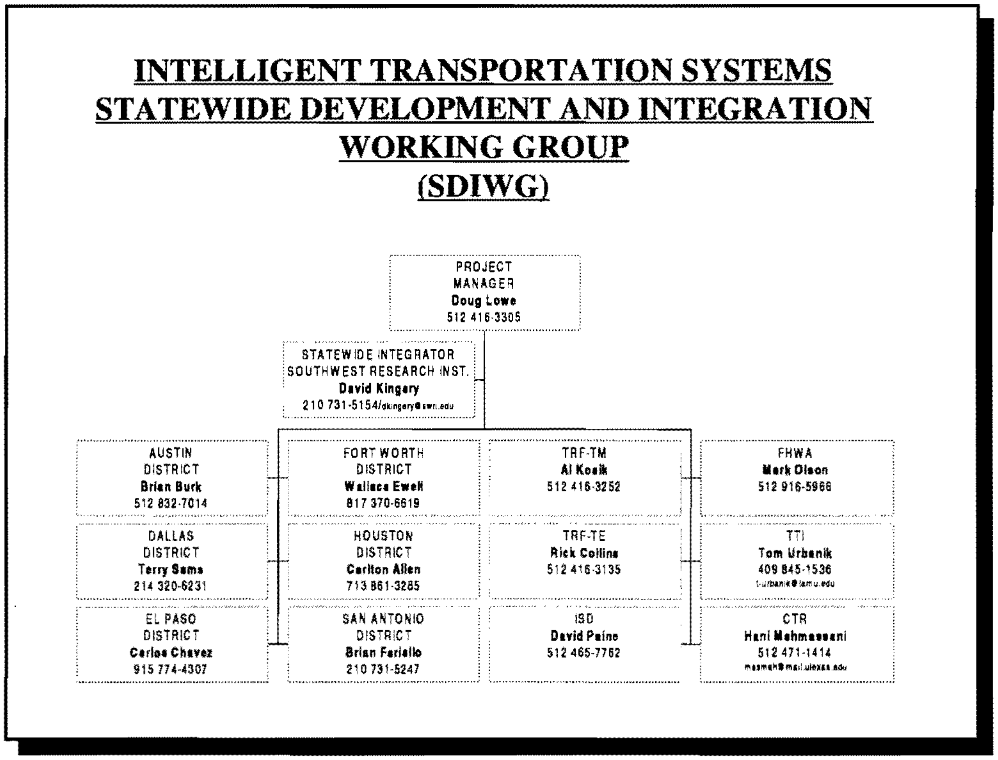
- System and software engineering expertise
  - development and integration processes
  - design and documentation standards
- Software maintenance of TransGuide
- Development and integration of San Antonio Model Deployment Initiatives

In terms of statewide integrator qualifications, Southwest Research Institute has a wide range of experience in systems and software engineering, particularly ITS work in Texas and several other states. They have been handling the TransGuide software maintenance contract in San Antonio for about two years. They managed the systems development and integration for the Metropolitan Model Deployment Initiative (MMDI), which included some very instrumental and complex projects. Southwest Research Institute completed MMDI on time with very good results, so they have a good background and TxDOT is lucky to have them as the statewide systems integrator.

STATEMENT OF WORK

- Define statewide standards
- Develop standard, reusable systems
- Develop standard, reusable database
- Support life-cycle maintenance
- Support transition of existing systems to ITS standards

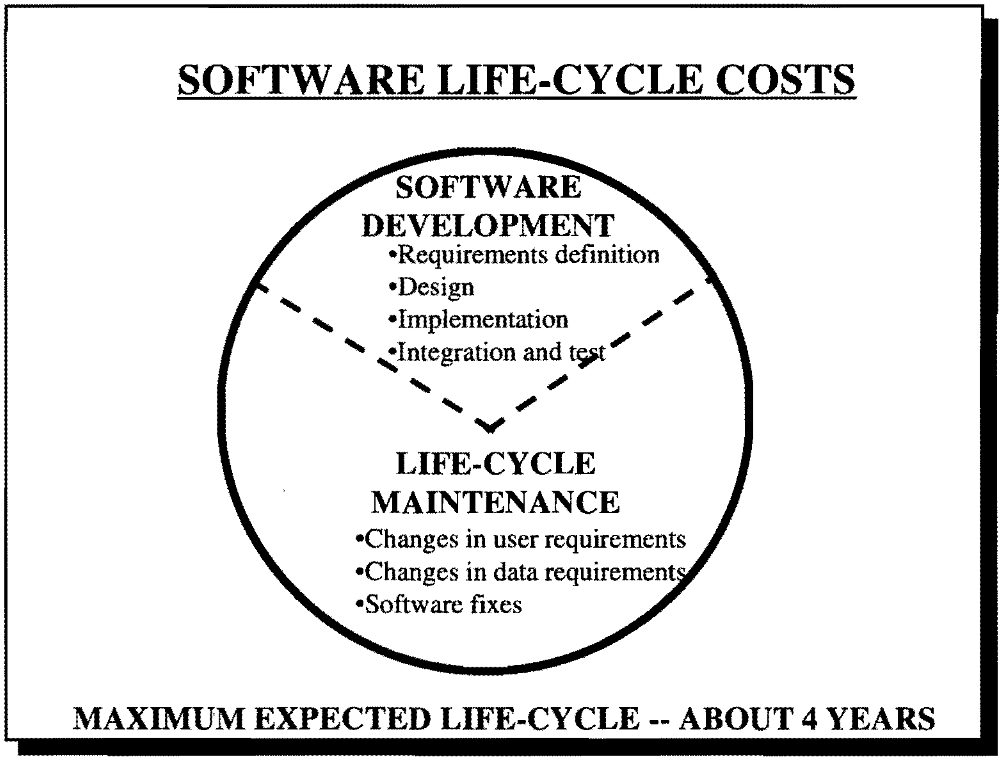
The statewide systems integration contract covers anything related to ITS software development, maintenance, etc. All of these activities are encompassed in the contract, so it is broad enough that we can make a lot of things happen, including the transition of our existing standards to be consistent with national ITS standards.



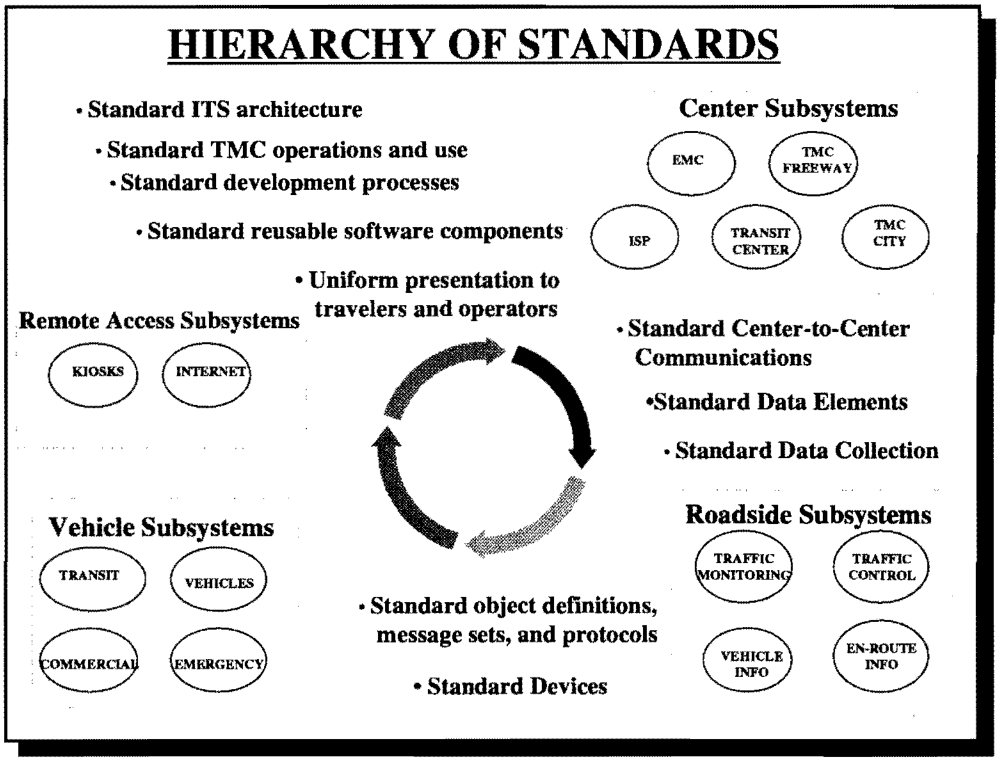
The statewide development and integration working group (SDIWG) was created to maintain a strong partnership with TxDOT's districts and divisions, the statewide integrator, and our research institutes, such as the University of Texas and TTI. The diagram on the slide shows the members of this working group, including a representative from each of the six major urban TxDOT districts as well as Mark Olson from FHWA. The working group meets about every three months to discuss current issues, then attempts to resolve the work activities for the next three months.

The objectives of the statewide integration effort are to:

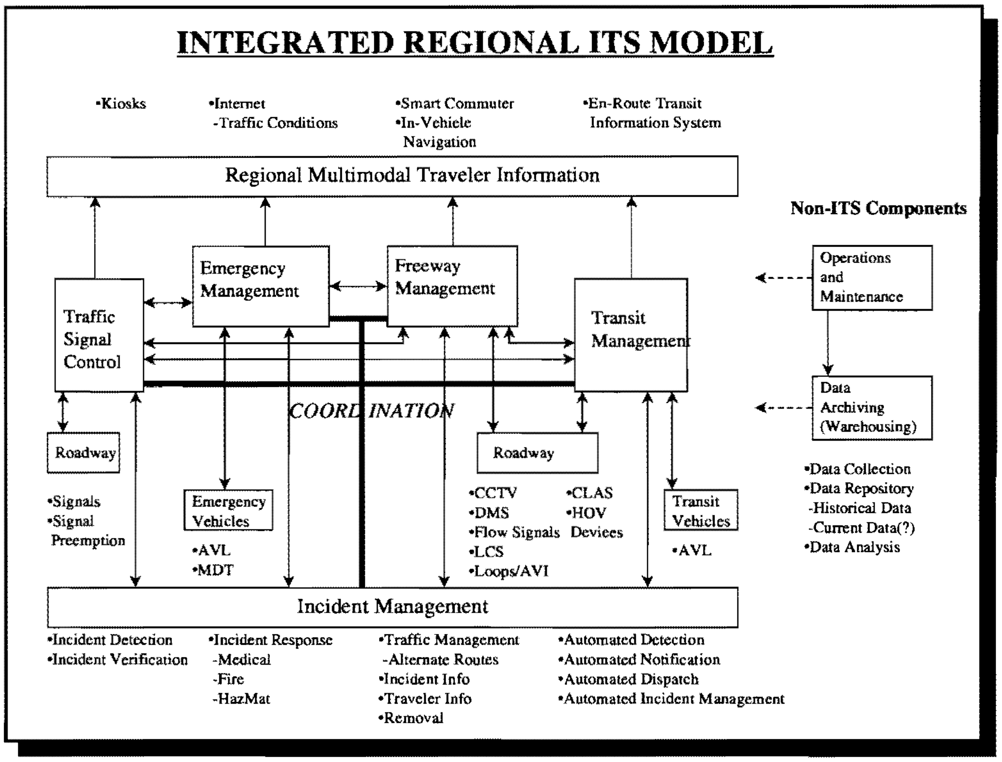
- set the course for ITS in Texas;
- coordinate ITS resources of districts, divisions, and statewide integrator;
- improve traffic management and incident management operations;
- focus on standard operations of devices visible to the public;
- develop a realistic approach to standardization; and
- provide affordable ITS benefits to the travelers of Texas.



As this slide shows, software development costs only account for about one-third of the total software life cycle costs. The other two-thirds of your costs are in life-cycle maintenance, which includes software fixes as well as software updates to accommodate new user or data requirements. The typical maximum expected life cycle for software is about four years, so these costs are likely to be repeated every four to five years.

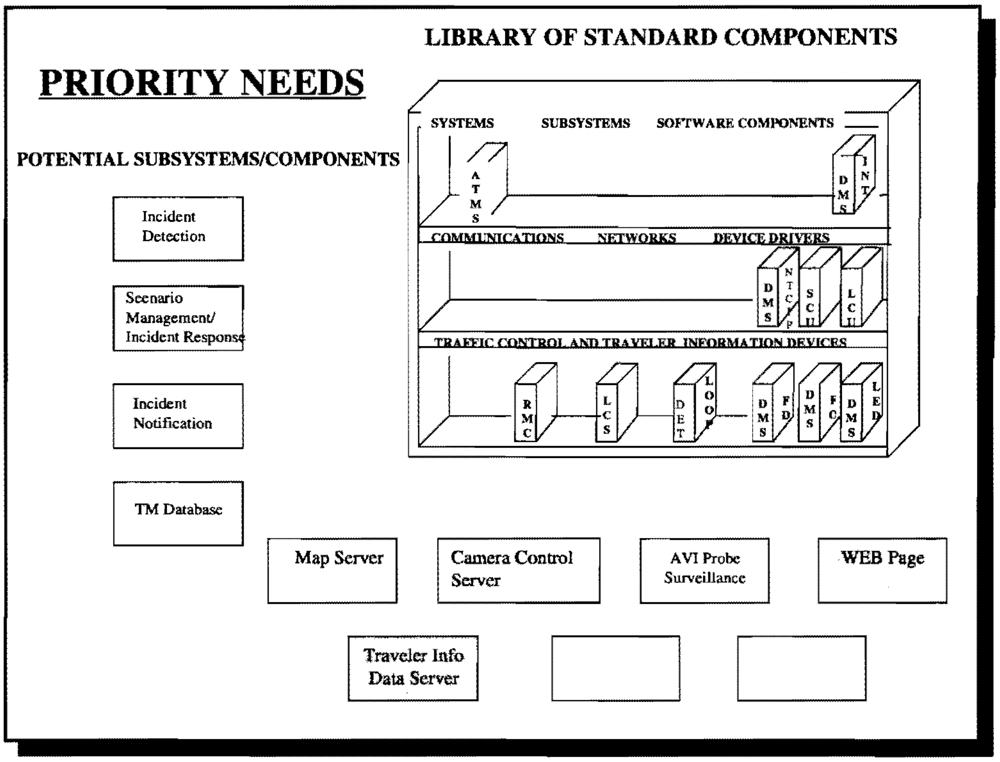


There is a hierarchy of standards we are trying to consider, and I would like to address some of these today. It starts from standard ITS architecture (top left corner of slide), and encompasses everything in between, down to standard devices (bottom of the slide). The first to consider are the four basic types of subsystems in the National ITS Architecture, including roadside, vehicle, remote access, and center subsystems. There are more subsystems than the slide shows, but I was trying to tailor the diagram to the Texas situation. All of the subsystems should be designed to work together and not in isolation. In fact, you almost have to start with standard data elements throughout the system if you hope to have a standardized system. It is unimaginable to me to have a system that can communicate center-to-center (which is another one of the requirements here) without having standard data elements. In addition, I have not listed everything here related to data. Basically the point is that if you don't have standard data elements then the National Architecture may be a waste of time.

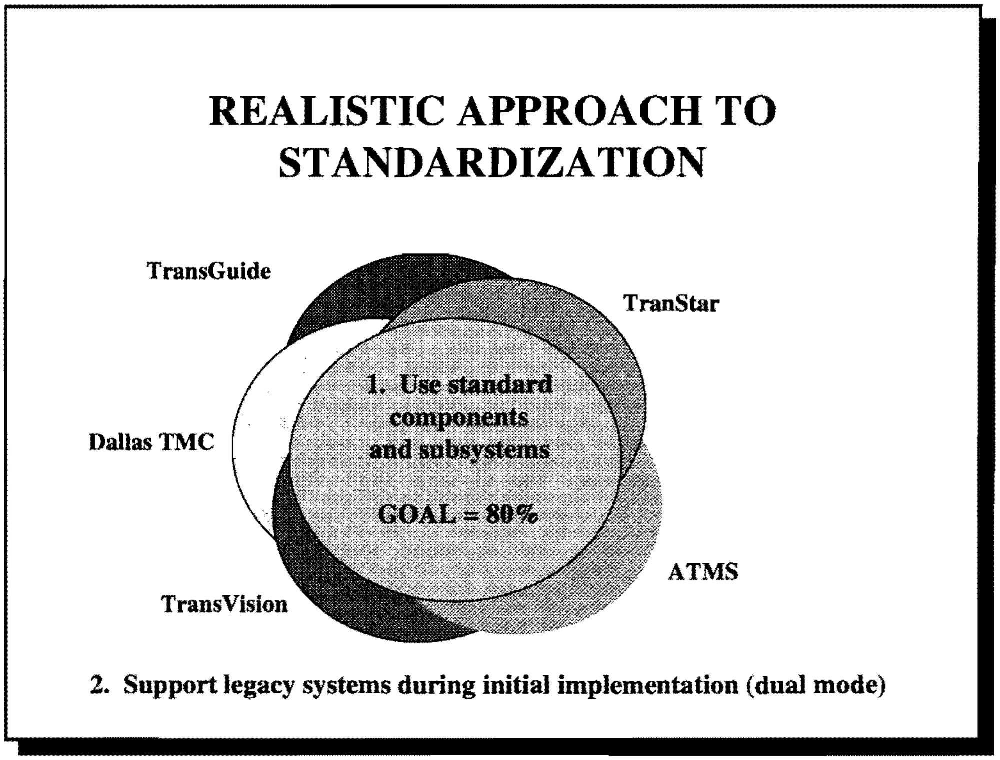


This slide shows an example of an integrated regional ITS model. I won't cover everything on this figure in detail, but the real key is to focus on the most important activities. I think that everybody accepts that incident management is one of the main ITS activities. People from different urban areas or states might have a slightly different opinion, but basically incident management is one of our major activities. In addition, traffic management is also another focus activity for TxDOT.

The real key to integrated ITS is that you can not perform any of these activities in isolation (for example, coordinating all of the incident management activities among the different subsystems). That was the point of the previous slide, that you need standard data elements, message sets, etc. to communicate between systems. Of course, there are many more functions than incident management, and traveler information is closely related to it as well as being a priority for many people. We have placed the data archiving function there as a place holder, because we will need data archiving. We do have a traffic management database shown on this slide as well. There are many items listed on the last line of the slide that still have to be done on nearly all centers that we currently operate.



We have been trying to identify some of the critical functions for these centers, work at standardizing them, and move them to the common "shelf." The common "shelf" reflects one of our goals, which is to reuse software components as much as possible, whether it is a subsystem or a component of a subsystem. We would like to start reusing software components throughout Texas, and that is where standardization really pays off. Data flows within these subsystems have also been a priority issue to standardize, and we have talked about having some common data elements. As I mentioned, I have listed a few priority needs here, but we could easily add 15 or 20 more. One of the things we have been working on recently is integrating dynamic message signs (DMSs) and developing standard device drivers and software for the different centers. Even though I show this component as being on the common "shelf," it is not quite there yet.



When we first started this statewide integration effort and considered standardization, we realized that all of Texas' TMCs are not the same and they have different operating parameters. We are striving to get a common core of at least 80 percent of the center subsystems being standardized. We would like it to be at this level so that we do not have to develop and maintain the software components for every TMC in Texas as they come on-line. For example, we can develop video camera control system software once and then use it statewide. There will be unique cases or situations where system components may be different from area to area, and that is why our goal is not 100 percent standardization. Meanwhile, we will have to support many of the legacy systems currently in operation.

FY 1999 TASKS

- Development and installation support to TRF for ATMS in Austin and El Paso
- Deployment support for national standards
- Design integrated Dynamic Message Signs system software
- Houston integration support:
  - Integrated High Occupancy Vehicle (HOV) requirements
  - Review of development, integration, and deployment plans
  - DMS NTCIP driver development and integration
- Austin ATMS network security and effectiveness evaluation

The statewide integrator worked on the following tasks in fiscal year (FY) 1998:

- site visits to TMCs in Texas;
- technology transfer of TranStar system software;
- documentation of user requirements for DMS;
- development support to Traffic Operations Division for advanced traffic management systems (ATMS); and
- deployment support for national standards.

The work tasks for FY 1999 are shown on this slide. Under the integration support for Houston, we are trying to work with them to identify the data requirements for a data warehouse. In other words, we are asking questions such as "what types of data do we need to archive?" and "how are we going to use the data?" These data and user requirements will then lead into defining the data warehouse system architecture.

SUMMARY

Affordable ITS benefits for travelers in Texas depend on:

- Strong partnership
- Shared resources
- Standard operations and use
- Development of reusable components
- Conformance with national standards initiatives

In summary, this last slide summarizes the main goals of the statewide integration efforts. These are all focused on providing affordable ITS benefits for travelers in Texas.

TXDOT INFORMATION SHARING POLICY

Al Kosik
Texas Department of Transportation

My name is Al Kosik and I work in the Traffics Operations Division, and we have been in the ITS "business" for quite some time. We collect a lot of data on our systems, and perhaps some of that data could be useful for your applications. We tried to get a number of different workshop participants here today (not just TxDOT personnel) to discuss uses for the data that we collect. We have been working for about the past two years on a TxDOT information sharing policy. We developed this policy to provide guidance to our districts on transportation information sharing and to ensure that the information sharing is conducted in a uniform and consistent manner. The information sharing policy also addresses the issues of standard presentation formats (e.g., web site layouts and designs, etc.). We are now attempting to develop standards for implementing this information sharing policy, as well as developing guidelines for distributing traffic management center (TMC) data that could be helpful to other people.

We developed guiding principles for information sharing about a year ago, and after much discussion and review, we developed this policy from those principles. Also, last year we had TTI do a state-of-the-practice and synthesis on other states' and cities' information sharing processes. Very few of the TMCs across the country have access to traffic-related information contained in other systems. Everybody would like to have that type of information, but we do not distribute the information very effectively. The invasion of privacy was really not an issue, or it was considered to be a minor issue. We thought at first that privacy was going to be a problem and we would have to really worry about it, but it turned out not to be a major issue. Many of the other TMCs across the country require the media to pay a connection or access fee. We also found from this survey that if traffic data were archived, if at all, it was archived primarily for planning purposes (and some of this archived data wasn't even valuable to planners).

The information sharing policy that I have just passed out (Figure 1) was developed because we believe that sharing transportation information with both the public and private sectors will increase the mobility, safety, and efficiency of the transportation network. That is the driving force behind the development of this policy. We believe that we need to share as much information as we have in our TMCs to help the overall transportation network. We will attempt to make all transportation-related information available. We have the capability to assess fees for access to this transportation information if that is desired. Any formal sharing of transportation data will be through an agreement or contract. For any of the data that is collected using TxDOT funds, we will be retaining intellectual property rights to that data. We will be encouraging information providers to add value to the data that we have shared. We do not have the resources or the capability to contribute a lot of value-added services, as we are mainly interested in the operation of our transportation system. There could be niche markets, such as trucking companies, taxis, etc., where they obtain our data and add value to it themselves. We do not think there is much of a market now, but there certainly could be in the future.

We want to make as much transportation information available over the Internet as possible. Currently, I know that many of our TMCs are doing work in this area. We would like that Internet-based information to have the same "look and feel" for all of our TMCs. Texas is blessed (or cursed) with six major urban areas, with three of the areas being in the top 10 population-wise in the country. We will be working with our existing TMCs in the next six months to try and develop some of these Internet presentation standards.

Again, this policy document contains broad principles and guidelines and there is not much detail. We even say in the policy that the standards are in the early stages of development and that we will be developing them as we move forward. We think that this policy will give us background, guidelines, and general principles that we can reference when issues arise in sharing data. This information sharing policy has been discussed at our district engineers' level as well as with our general counsel. Everyone in TxDOT has provided input to this policy and we have made numerous changes, but we have held steadfast to the principle that we want to share our transportation data as widely as we possibly can. So with that as an introduction, we will be talking in depth today about the data that we are collecting in our TMCs, as well as how that data can be useful, the data you would like to see us collect, and whether or not we can actually collect that data.

Tm

TxDOT Policy Statement 2-98

Texas Department of Transportation

Subject: ITS Information Sharing

☒ New ☐ Supersedes *Effective: October 5, 1998*

Introduction

Current Intelligent Transportation System (ITS) technology makes it possible for TxDOT to capture transportation related information through the department's transportation management centers.

Policy

TxDOT will share this transportation information with both the public and private sectors in order to increase the mobility, safety, and efficiency of the transportation system.

Responsibilities

Districts shall make all transportation related information available, including video images from closed circuit television (CCTV) cameras, current transportation data existing in on-line data bases, and archived historical data.

Principles

In the implementation of this information sharing policy, the following principles shall be emphasized and employed:

- TxDOT shall retain intellectual property rights to all data gathered at TxDOT's expense.
- TxDOT shall reserve the right to assess fees or negotiate for other benefits for data and information made available.
- TxDOT shall continuously examine practices and policies with regard to information sharing in other cities, states, and countries. TxDOT shall work to improve compatibility in information sharing in accordance with other states' policies.
- TxDOT shall share transportation information with other government agencies and the private sector, and shall share the information to the greatest extent possible for the increased benefit of all.
- TxDOT shall provide specific information on a partnership basis, reserving the right to negotiate for benefits, either monetary or in-kind, in return for the information.

(continued...)

Page 1 of 2

Figure 1. TxDOT Information Sharing Policy

TxDOT Policy Statement *Page 2 of 2*

Principles *(continued)*

- All formal means of sharing information with both public and private sectors shall be through binding written agreements or contracts (not, for example, a memorandum of understanding).
- Private sector information systems providers shall be encouraged to add value to TxDOT information resources.
- Selected transportation management center information shall be available on the Internet when practical. Typically this would include traveler information such as traffic flow data maps, highway incidents, road closures, and video images from CCTV.
- All information viewed by the public shall be presented in a similar statewide standardized presentation format. The standard format will be developed by the districts with support and oversight from the Traffic Operations Division and the Information Systems Division.

Compliance Requirements

District Engineers will be responsible for compliance, and the Traffic Operations Division (TRF) will monitor for compliance.

TxDOT information sharing standards are in the early stages of development. However, this policy shall be followed unless an exception has been approved by TRF.

Manual

This policy will be incorporated into the *Traffic Management Volume* of the *Traffic Operations Manual.*

Reference/Authority

Transportation Code, §201.101.

Why Policy is Needed

This policy is needed to ensure that districts have guidance for the release of transportation information, there is uniformity in the manner in which information sharing is conducted at TxDOT transportation management centers, traveler information is made available to the public, and information is presented in a standard format.

Approved: Charles W. Heald	Date: 10-5-98

Figure 1. TxDOT Information Sharing Policy (Continued)

APPROACHES TO DATA STORAGE, ARCHIVING, AND MANAGEMENT

HOUSTON TRANSTAR

Sally Wegmann
Texas Department of Transportation

When I became the director of traffic operations in the Houston district about one and one-half years ago, the issue of data warehousing was certainly a topic of discussion, both its advantages and disadvantages. Initially, a data warehousing component had been considered in the planning for TranStar, but in some cases good plans can not be implemented because of limited resources. When resources for the TranStar center became scarce, data warehousing was one component that was dropped because of the necessity of completing core functions for the TMC. At the time, there were discussions about whether data warehousing was a necessary component to ITS or whether it was simply a researcher's luxury.

In the last few months, we have been asked to quantify the benefits of ITS applications and the TranStar center, including actual measures of effectiveness and the corresponding level of benefits. TTI has done two of these ITS benefits studies for us, and they have really struggled with obtaining the data to develop quantifiable, measurable benefits. Certain transportation commissioners want to know how many cubic yards of concrete do not have to be poured because of ITS, and we have a difficult time responding to questions such as this. Because of these experiences, we began to realize that not only is data warehousing a researcher's dream of data availability, but it is also a necessary component for TxDOT to constantly evaluate ourselves and to measure the benefits for an expensive system. We can no longer simply say "my gut feeling is that we are doing a really good job." Our partners at TranStar, which consist of Harris County, the city of Houston, Houston-Galveston Area Council (HGAC), and Metro, are also facing the same issue of scarce resources as they get new commissioners and executive directors. They are having to demonstrate that, in fact, TranStar is providing measurable benefits. From that point of view, there is a realization that some form of data warehousing (that includes the ability to calculate measures of effectiveness and benefits) is necessary. We need to provide the appropriate data and analysis tools for these benefit analyses, because we are only hurting ourselves if we do not. Intuitively, we know we are providing more benefits than what we can document from the limited historical data sets.

The Houston district of TxDOT and our TranStar partners have agreed in principle that, within the next couple of years, we will have some form of data warehousing. There are a number of related activities that are occurring in Houston. First, HGAC has an ITS Priority Corridor project to identify planning data requirements, the objective being to identify the data being collected by TranStar that should be saved. This project is currently in the preliminary engineering stage. By utilizing data currently being collected by TranStar, HGAC can eliminate redundant data collection as well as obtain data in a shorter time frame than traditional data collection. HGAC might also be able to use TranStar data for evaluating congestion mitigation/air quality (CMAQ) projects and related compliance requirements. There is definitely some information that we are collecting that HGAC needs, and we are very fortunate to have a good relationship with HGAC. We exchange data freely and have a symbiotic relationship between our district's planning department and HGAC.

The second effort in Houston that relates to data warehousing is the involvement of the statewide ITS integrator, Southwest Research Institute, to help us with data issues. They are reviewing the data elements currently being collected at TranStar and determining what data should be stored for a number of possible end users. Doug Lowe is the contract manager for the statewide integration contract, and he and Southwest Research Institute have been working with us to define end user data requirements. We also have been working with our information systems staff to address year 2000 (Y2K) issues as well as trying to address the hardware and software needs for a data warehouse. We have done some preliminary estimates to determine the costs for a data warehouse, as well as the data warehousing software that we could retrofit into our existing TranStar computer system. For our proposed data warehouse, we have not only TxDOT as an end user, but also our partner agencies. We will likely have to store not only traffic operation information, but also incident and emergency management data. TranStar is unique in that it has an emergency management function, and we are finding that emergency/disaster management and traffic management go "hand-in-hand." I do not know if traffic is a disaster, or if disasters are a result of the traffic. On every single occasion between flooding, ice, etc., the traffic and the disaster response has had to be interlaced together to meet the needs of our city.

With that short overview, let me introduce Cindy Gloyna, who can explain the technical details of our activities. She will explain the various upgrades that we are considering, as well as the actions that we would have to take to make our system compatible for a scalable data warehouse as we identify our user needs.

Cindy Gloyna
Texas Department of Transportation

I am the information resource administrator for the Houston district and the TranStar center. We are very fortunate to have an Oracle database administrator (DBA), David Yuan, who has worked for TxDOT for the past seven years. Basically, I provide support to TxDOT and the partner agencies at the TranStar center. David Yuan and I have been working on a hardware and software upgrade that would fit within our existing system.

The cost of this hardware and software upgrade is about $400,000. That will take care of many of the Y2K issues we are facing. It will also bring us to a level where we believe we can begin looking at implementing new technologies, which in this case is data warehousing. Once we complete that upgrade in the next eight months, we would like to start using Oracle 8. However, there are many things we have to do to get there. We need Oracle 8 because of the proposed $1.2 million Houston will be spending in hardware and software for their data warehouse. Part of this hardware and software upgrade is not simply data warehousing. It is not just a server that we are procuring, and that we are going to load or connect to our Oracle database, and then build a data warehouse. It is an enterprise management system for all of our systems, regardless of the platform. It may be Hewlett Packard-Unix, which is the primary network operating system for our traffic management functions. We also have an administrative network that consists of applications such as Microsoft Office, desktop applications, as well as our emergency management functions. We have bridged these two networks together using an intelligent router. This intelligent router essentially enables communications between the two networks, so that applications such as electronic mail or others can "talk" back and forth. In the enterprise management system that we are proposing, we have estimated that we may need about 1,000 gigabytes (GB) of storage space. The enterprise management system with this amount of storage has been estimated to cost $1.2 million, but at this point these estimates do not directly reflect our user requirements.

I would like to briefly discuss the user requirements now. I was in a meeting yesterday morning with Rita Brohman, a Priority Corridor manager who works for Sally. We were meeting as requested by Traffic Engineers, Inc., who is a contractor for HGAC's Priority Corridor Work Order 16. In that meeting, they were asking typical questions of me, such as: "What does the TranStar system look like?" "What is the data diagram, the Oracle database diagram, the conceptual, logical database sketches?" "Where are they?" and "We need those and all the details of the system." The answers to their questions can be found in many binders of documentation. We were talking about the possibility of this preliminary work being a pilot project to the data warehouse upgrades that we are currently planning. Everyone at the meeting thought that would be a great idea. We do not make those decisions, but that would seem to be a great way to approach this. A pilot project, in my experience, has always been a very viable way to approach a major hardware/software upgrade. Since we believe that data needs and the end user requirement are the most critical, this meeting went to our white board. Basically, we drew a chart to communicate to the group what David Yuan (our DBA) and I need to able to design an effective data warehouse. We believe that without adequately defining user requirements, the data warehousing project will fail. We will spend a lot of our time and effort defining these user requirements. In the places where data warehouses have failed, they failed because system designers did not provide the end user what they needed. When I refer to end users, I am also including all levels of management. Not just me as an end user, or Doug Lowe as an end user, but also Mr. Heald, our executive director, as an end user. He would also be an end user and would need to be consulted as to what he would need to see in the system.

We would like to develop (hopefully we will be getting this from Southwest Research Institute) a matrix of end user requirements. Of course, it would probably be in the form of a report, but perhaps also a table (Table 1) that might have listed the user's data requirements in the first column. In another column might be the functions, such as planning, construction, operation, maintenance, etc. For Houston, another column might be the agency that has the requirement. Is it the city, county, TxDOT, METRO, HGAC, or another agency? This table could be multiple pages, or a large plotter sheet, or it could be in rows in a computer spreadsheet or database. Another column could describe the type of data, such as whether it exists. Basically what I am describing is a data inventory.

Table 1. Example of a User Requirements Matrix

User Requirements	Functions	Agency	Type of Data
• List of data elements based on user requirements	• Planning • Construction • Operation • Maintenance	• City • County • TxDOT • METRO • HGAC	• Planned or Proposed

This user requirements document should be at a high level and provide an overview of what the entire packet would consist of, which would include all of the details. In these end user requirements, we are hoping that Southwest Research Institute will go through the process of identifying all of the requirements. For example, we would have to identify all users, and not just in the Houston area, but also the traffic operations staff at TxDOT headquarters in Austin. They are very important as well. We would need to identify the type of data they need, what their requirements are, and what type of data that they have. We would also need to identify the existing data formats as well as storage mechanisms. We would also need to identify the most efficient way to get it into a relational, on-line, real-time database. In Houston, Oracle is our database of choice. Other users may prefer Sybase or some other relational database software. Specific database software is not really an issue for us at this point in defining user requirements.

What we need is for someone to go out and really hammer this out for us, get those details, and give them to us so that the TranStar Information System (IS) folks (i.e., Ray Lickey, our database administrator) can then work with another consulting firm or whomever we get on contract. I think it is a large project, and it is not something that one person can do. The point is to get those user requirements met. As we go through this process, we would envision that the contractor would group the requirements by area. For example, are they planning, construction, operations, or maintenance requirements. We could then identify where there are duplications and whether these duplications cross agency boundaries, or if they are within an agency.

In addition to defining these basic user requirements, we would like to see another document that gets to the end user directly. With the automated systems that are in existence, this should not be too difficult. I think that maybe even Excel might be able to do it. If you have created a database or a spreadsheet of the type discussed above, another important aspect of that would be whether the data already exist. If the data do already exist, then we would like to know its location, if it is being used, and if it is really needed. Therefore, another table might need to be created that would not focus just on the type of data that is out there and which agency has it, but it would then go to the end user level. For example, consider Joe Schneider, our traffic controller at TranStar. We would need to interview Joe and find out if the data is existing, if he is using it, if it really meets his needs, and if he still needs it. We need to identify if we really need to continue using it because such a system really is garbage in, garbage out. Joe is the user. Joe is the one who is going to be filling out the system, or working with the system that you are hoping to feed your data warehouse. If Joe is not using that system, does not know how to use that system, or did not know it existed, then that is a problem. Maybe there would be columns that would show the end-user perspective and then tie it back by identifying who is the end user, where are they, what is it that they need, and how are they using it (Table 2). If we could construct a table like this (Table 2) and link it back to the previous table (Table 1), it would indicate where we really are and where we need to go. This would be in terms of what we have that is existing, what our users think we have, what our users are currently using, and how that can help us in our data warehousing initiative.

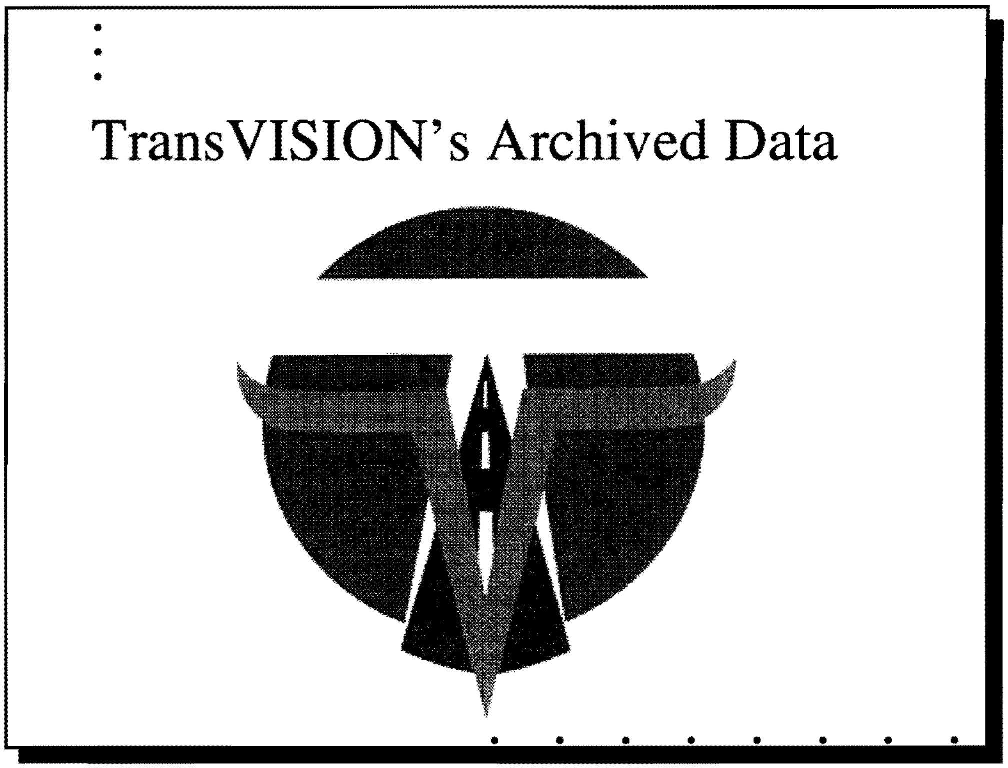
Table 2. Example of Incorporating End User Perspectives

Who is the end user?	Where are they?	What is it that they need?	How are they using it?	Does it meet their needs?	Is it still needed?

I hope that I was able to adequately describe ongoing and planning data warehousing activities in the Houston area. Basically, we are upgrading our computer system infrastructure, and everyone in this room understands rapid changes in technology and the need to upgrade capabilities. Otherwise, our planning data warehousing system may not fit. We may not be able to implement a data warehouse in Houston if we do not spend the next eight months getting our computer infrastructure upgraded. Therefore, that effort is ongoing. Parallel to this effort, we hope to be able to procure required hardware and software. In the meantime, we believe that Southwest Research Institute, working with the Traffic Operations Division in Austin, will be defining these user requirements, which we desperately need for a successful data warehouse.

FORT WORTH TRANSVISION

Abed Abukar
Texas Department of Transportation



We have been considering the potential of ITS data archiving and sharing just like everyone else in Texas. We have been facing the same types of problems as Houston. We are keeping a close eye on Houston and San Antonio since we believe they are ahead of us in terms of ITS implementation. This slide shows our logo for Forth Worth's TransVISION. Our system will be completed by September of 1999. By then, hopefully, we should be able to invite everyone to come up and visit the system that we have built. TransVISION will be a basic system that is similar in design to Houston TranStar. We differ in terms of database software, as we use Sybase instead of Oracle to adhere to TxDOT software standards.

Logic TransVision Uses

1) What type of data is available?

2) Would it add value?

3) In what form should the data be saved or logged?

4) How do we share this data with others?

This slide illustrates the basic questions we are asking at this time. One question is "what type of data is available?" We have a lot of data coming in and "do we want to save it all?" " Do we need it all?" "Would 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 cities of Fort Worth and Arlington, the Dallas District of TxDOT, and others in our area?" These are some of the questions that we are currently trying to address as we design the TransVISION database.

Data Anticipated in TransVISION's Future

- Traffic Incident Information
- Flow Signal Timing
- DMS stored messages
- LCS patterns
- GIS data
- Traffic Signal Timing
- CCTV Control Functions
- Electronic Work Order
- Detection System Data
- Road Construction, Maintenance, and Special Events Data

This slide shows selected data that we should be able to collect and put in our database. Incident information requires significant computer disk space. We will also collect flow signal data; however, the flow signal data will be somewhat independent from TransVISION because the vendor was unable to meet Windows NT software requirements in time to include it as an integrated TransVISION function. Another problem is the dynamic message sign (DMS) system. This data comes in from different areas and different vendors with proprietary systems, and is outside of the overall database right now because of functionality problems. Lane control patterns are technically within our system. Geographic information systems (GIS) base maps and data are within the system as well. We are working with the cities on traffic signal timing, and are trying to get more details resolved.

Control functions for closed circuit television (CCTV) cameras are another potential source of data that we would like to consider. Because of a good relationship between operations and maintenance in our district, we anticipate using electronic work orders to make our maintenance request process more efficient. Right now we have a very primitive electronic work order system in place. The majority of our data will come from the inductance loop detection system, and that will provide the maximum load on our system. Roadway information, including construction, maintenance and special events, is information that can be saved from one year to the next and could prove to be beneficial as well.

Data Logging Currently Designed for TransVISION

Incident Data

Closure Data

Operator Data

Detector Data

We have identified four types of data from the TransVISION system that should be logged. The first is incident data. We have an incident identifier, which is a predetermined code that identifies the specific accident or specific incident. We would also like an incident count that provides a cumulative total. For example, during a single day, week, or month, we want to know many accidents or incidents that TransVISION personnel have helped verify, detect, or manage. We could then do trend analyses and comparisons, say to another month in the same year, or to the same month last year, etc. For the location of incidents, we would like to have the roadway designation and location, cross-street information (street name), direction of travel, and the lane that has been blocked (e.g., left lane, central lane, right lane, inside lane, or outside lane, etc.), and the number of vehicles involved in an accident. These are the types of data that we would like to save within the incident data. We would also like to know the incident status. Is it an active or inactive accident? What is happening on it? What is the operator identification and console identification that initiated the work order for the incident? Which operator has confirmed the incident? Who is the last operator to modify anything on the incident? What is the last modification and what is the accident time span? What was the time the incident was detected and by what agency or person was it detected? What time was the incident confirmed?

Once the operator confirms that the suspected incident is a true incident, then we need to have that information embedded within our system as well. For example, what time was the incident moved to the shoulder of the freeway? When was it cleared? What were the operator's comments? When did the response team arrive on the site, and when did they depart from the site? We would also like to know the roadway conditions at the time of incidents. Was it raining? What is happening on the roadway? Is it under construction or maintenance? This gives you some examples of the type of data we would like to keep in our database. In our existing system, we capture images of major incidents from CCTV cameras and store these in our database. We would like to continue this image capture for major incidents and store these images in our proposed TransVISION system. In some cases, though, our operators have to decide when an incident is considered "major." If we are able to save these images of incidents, then we will be able to retrieve the images later and capture all of the data I just mentioned if we would like to review certain circumstances.

For lane closures, we would like to save data about the roadway names, direction of travel, the beginning and ending location of the closure, current status of the closure, lanes that are affected by the closure, duration, and reason for the closure. We would like to know the contact person and the source of the closure as well. For example, is it maintenance, construction, the city, the state? Who is in charge of the closure?

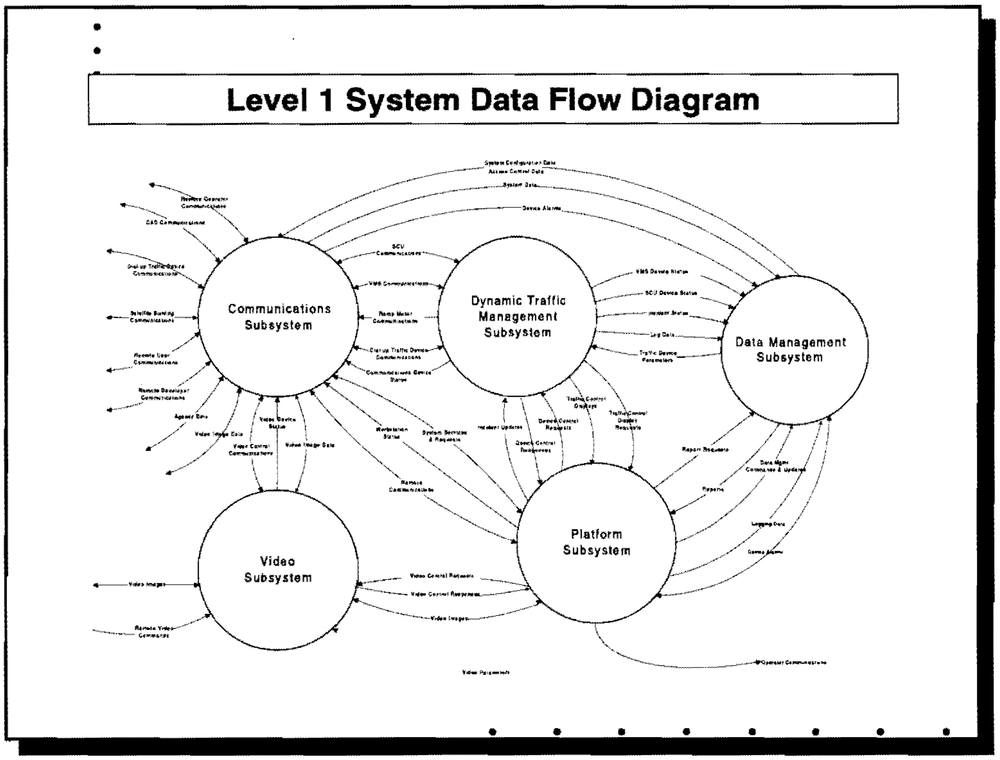
With respect to operator data, we want to know the date and type of action that is taking place because of an operator's action. We want to know the console identification number and where action originated. We want to know the operator identification number and who initiated the action. We also would like to know the action time and details on it. We also want to know the incident identification number.

Detector data are also very important to us. Those are very important because we want to know the measurements of date and time, sample duration, and location of the accident. We want to know the device identification and time the incident was detected. We want to know the volume and occupancy within that general area. We want to know the average speed, travel time, and volume count. All that needs to be kept within our database so we will be able to determine what is happening in the duration of that accident. We also like to know what error rate we are getting from the system. Is the system giving us a lot of errors, and how do we determine that by looking at it at a different time? Maybe we have a maintenance problem there? These are the four basic data elements that we would like our proposed TransVISION system to save (i.e., incident data, closure data, operator data, and detector data).

System Architecture

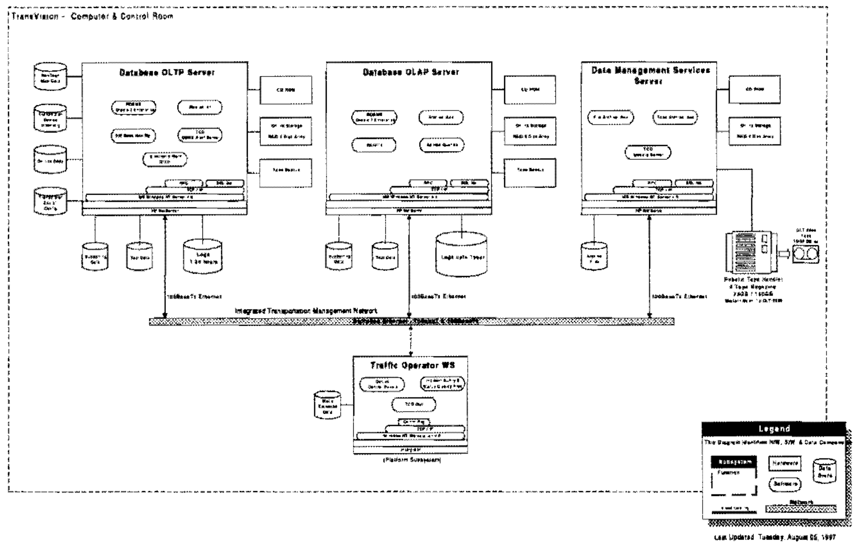
- **Overall System Divided into Five Major Subsystems**
  - Communications
  - Video
  - Dynamic Traffic Management
  - Data Management
  - Platform

As I mentioned earlier, our TransVISION system has a design very similar to the Houston TranStar system. We have the five major subsystems, but I am only going to talk about the data management.



This logical architecture diagram shows how the system is linked together. Data comes in from the field and then goes into the dynamic traffic management system. From there it goes into the database. Our operator on the platform subsystem can call for specific types of information from the database and get it through reporting. Then the operator can get information on what is happening as well as receiving regular updates.

Data Management Subsystem Architecture



This slide shows the physical architecture of the proposed TransVISION system. Lockheed Martin designed the system, and the design is very similar to Houston TranStar. We asked them to do a complete system build-out design rather than just a "stopgap" design that would end at the on-line transaction processing (OLTP) server. We were trying to think ahead about the possibilities of a data warehouse server. We wanted to see how this data warehouse server, or an on-line analytical processing (OLAP) database server, would fit into our system, as well as examine the capabilities for storing a full year of data. Our current design for the OLTP server includes the ability to save the most recent 30 days of data. At the end of (and also potentially during) the 30-day period, there will be some replication taking place between the OLTP and OLAP servers. After a one-year period in the OLAP database, the data will be archived to some permanent storage format. Our current system specifications stop at the OLTP server, so at this time we will only be able to store the past 30 days of data. Eventually we do not want this 30-day limitation. Because we have already considered a full build-out design that includes an OLAP database and data archiving / management functions, it should not be too difficult to add additional capabilities as resources become available.

Data Management Subsystem Architecture

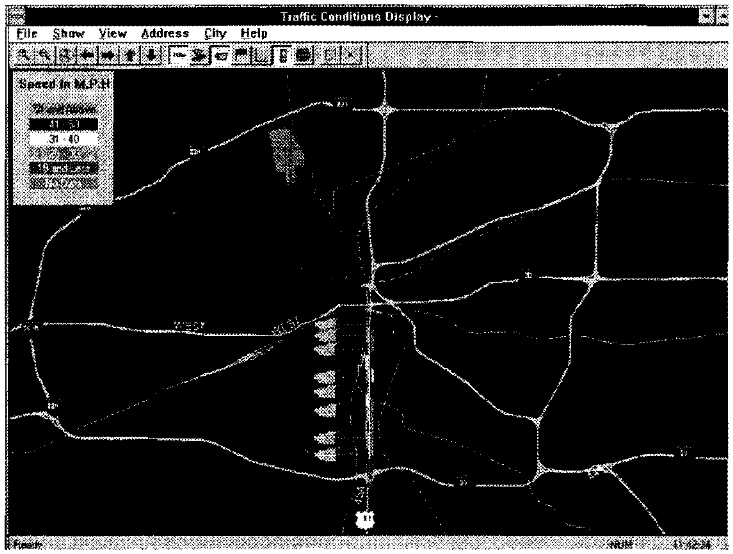
- On-Line Transaction Processing (OLTP)
- On-Line Analytical Processing (OLAP)
- Data Management Services
- Reports
- Data Management
- Hardware Configuration Items
- Roadway Traffic Conditions Display (TCD)

This slide shows the different parts of the data management subsystem architecture. Each one of these parts has a specific function and relationship to one another. The OLTP part deals mostly with the relational database, the management of the database, with the data model for real-time operation. The OLTP database contains the most recent information about what is happening on the network and updates that information on a regular, real-time basis. It has an alert function that has the ability to send out an alarm if these parameters have gone beyond certain limits. This alarm goes to an operator and describes the possible incident and its location. The OLTP database also maintains a relationship with the GIS base map because it interacts with the map itself. It also maintains all of our operational manuals, our operational requirements, and the operation procedures in the system. Also, it provides replication from OLTP to OLAP so information can be automatically transferred from one database to the other.

The OLAP function is basically a relational database, but its function focuses mostly on the complex analyses and relationships that are recorded by the system. For example, if we want to know what happened on the network last month or six months ago, we can go back and pull that information. We can also pull the information if we want to know how we dealt with a certain problem on the system. The OLAP database uses "off-the-shelf" tools to allow the query and a generation of those reports. It was designed to work automatically with the OLTP database in receiving and duplicating data. The OLAP database can also send that data to be saved in a tape or CD-ROM library.

For our data management services, we will have a server with the main function of maintaining and updating all of the data, monitors, and consoles equally. It does provide that service to all the consoles within the control room. It has a robotic tape handler that can obtain up to 150 GB of data to support the OLAP function. For the types of reports that we have been getting out of the system, we would use "off-the-shelf" tools for TxDOT to configure their own reports. We will be able to take whatever report we want from the system and create that report. We will be using a system report writer and Microsoft Access to be able to extract those reports and analyze them to see what is happening in the system itself. It can also report maintenance, equipment, and incident information, as well as reports on operator activity and system activity. Also, it will tell you the types of errors you have had within the system. The data management portion of it allows the data administrator to define new items within the database; change, remove, and view existing items within the database; query records; record lobbying; archive historical data; set up schedules; and do data replications. It can differentiate between what is mandated, what is optional, and what is read-only as far as fields within the database. The hardware that we are planning for the system is basically what Lockheed has proposed (Dell Power Edge 6100s, 200 MHz Pentium dual processor, 256 MB RAM and six, two GB hard disks). This hardware should provide for the necessary OLTP functions that I have discussed.

Roadway Traffic Conditions Display



For the roadway traffic conditions display, we will have dynamic, real-time, color-coded maps that will provide current traffic conditions (in terms of speed, occupancy, and level of service) on the network. If there is an incident based on pre-defined traffic conditions, the system will send out an alarm function, consisting of an icon that tells the operator to look for a potential incident or traffic problem.

Traffic Flow Data

-Traffic Flow Data Measurement Interval

-System supports two different measurement intervals

-First not less than one minute

-Second not less than five minutes

We are currently discussing the aggregation levels that we would like to use for our traffic data. At this time, we are considering one- and five-minute intervals, but this will be a decision made by the operators and the assistant administrator. It will really be a question of how much detailed information do I need? Do I need it in one minute, or do I need it in five-minute intervals, or 15-minute intervals? The TransVISION system is capable of providing data in not less than one-minute intervals. We know that the local controller unit (LCU) provides the information in 20-second intervals, but it gets accumulated before being sent into our central database.

Data Load

-Logging up to a month of data

100 bytes / loop-record *2,000 loops *1.2 records / min.
* 60 min/hr *24 hr/day *30 day/month =
10.3 G bytes/month

-Lockheed Martin recommended Sybase Version 11.8 to gain row-level locking.

This slide shows some of the data load issues that we are currently considering. This is a rough estimation of the data load that our OLTP server will handle if we feed it loop detector data once every minute. You are looking at about 100 bytes per loop per record, about 1.2 records per minute, with more than 2,000 loops. Given this data load, we are looking at about 10.3 GB of data generated per month. That is a significant data load, and once Lockheed did these calculations they indicated their preference for Sybase version 11.8 (instead of 11.5). With this newer version of Sybase, we gain the ability to lock only one line for a specified table (as opposed to the entire page), which makes it easier to make adjustments to the database. So, this is the data load that we are faced with, and it only includes relatively new data. It does not include the three other basic types of data that I described earlier. As you can see, we have already accumulated quite a bit of information within our system within a one-month period. So, when Cindy talks about a 1,000 GB system, that is a significant amount of disk space. Our estimate is only for 2,000 loops for a single month, and I am sure that Houston has many more loops than we do in Fort Worth.

Unresolved Issues

- Is the current design direction adequate?
- Does new technology cause a change in direction for archive storage?
- How much do we squeeze 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 (security / system load / etc.)?

We still have several unresolved issues. We have discussed some of these with Lockheed, but others we have not discussed yet. Sholeh Karimi, who works for the city of Arlington, has been very helpful. She has been very helpful with Steve Connell, our operations supervisor, in terms of defining end user requirements for archived ITS data. One of our questions is whether the direction of our current design is adequate. Do we need to revisit that design again in a year or two and be able to update it based on any problems that we may have? Does the technology that we will use one or two years from now fit? How is that going to help us in terms of warehousing? How much can we squeeze out of the data? Do we need one-minute data or can we go to 15 minutes, or an hour? Which level is necessary that would give us the maximum ability of our existing system without having to go to the next level within the next six-month period? And what data is needed by other agencies? We understand our responsibility in the Dallas-Fort Worth area to share the data that we collect through our TransVISION system. How important is the need for sharing this information? How important is it to TxDOT? We have that responsibility. How is that going to affect us? How is that going to affect our system? What type of security and system abilities do we need for our system?

SAN ANTONIO TRANSGUIDE

Pat Irwin
Texas Department of Transportation

I would like to applaud Sally's opening statements. She did an excellent job identifying many of the overall needs that we have in managing ITS data. I believe that we have a responsibility to the public to get this information to them and let them make some decisions on their own. We do not need to tell them everything that they need to do. We need to get this information to the media as well. We also need to get this information to transportation planners and researchers, who have supported us in the past. We have to get this information out to them or we are not doing our job.

I have been given 10 minutes and I assure you that I will comply with that requirement. That is nine minutes more than I need to tell you what I know about this topic. I do have a "cheat sheet" here, so forgive me if I refer to this too much. I will try to answer any questions that you may have, but if I can't we have some good folks right here who will help me in answering those more detailed questions.

Data archiving within TransGuide consists of:

- identifying the data sets,
- developing a storage hierarchy that identifies migration paths for the data and the appropriate storage media,
- determining the factors that affect storage and distribution of data,
- developing backup strategies,
- acquiring the necessary resources to sustain these strategies, and
- building and maintaining the infrastructure necessary to fulfill the data sharing and research requirements set out for TxDOT and its ITS facilities.

Data Sets

- Operational Data
- Procedural Data
- Performance Data
- System Data
- Administrative Data

Texas Department of Transportation

Data gathered, processed, and stored within TransGuide® can be classified by these five general categories.

- Operational data is that set of data derived from direct operation of the advanced traffic management system (ATMS) and consists of observations, readings, and responses from the automated traffic detection sub-systems.
- Procedural data is that set of data derived from accessing and using system resources and is a function of the number of users on the system, the amount of activity, and the current state of the hardware and software.
- Performance data is that set of data derived from monitoring the system for errors, for demands on resources, system responses, and system loads. This set includes performance tuning parameters for the database management system.
- System data is that set of data comprised of the software components that make up the advanced traffic management systems, as well as design documentation, and network configuration data.
- Administrative data is that set of data comprised of administrative and maintenance records that consist of reporting on the operational state of equipment, acquisition and repair records, network schematics, system documentation, and the like.

Operational Data

- Real-Time Traffic Data
  - 20-Second Data from Inductive/Acoustical Loop Detection Subsystem
  - 15-Minute Data Aggregated from 20-Second Data
- Scenario Execution Logs
- AVI Tag Reader Data
- Theoretical Database

Texas Department of Transportation

Operational data is comprised of those observations, readings, and responses generated by automated vehicle detection subsystems. Presently these consist of inductive loop detectors, acoustical detectors, and the antenna arrays of the automated vehicle identification (AVI) system. This data is collected in various intervals. The most frequent interval is 20-second readings from the inductive loop/acoustical detection systems that feed the alarm incident handling subsystem of the ATMS. Operational data also consist of operator-generated responses to alarm incidents, and static, theoretical data that is updated periodically through statistical sampling of traffic data on major roadways and routes in the San Antonio metropolitan area.

Procedural Data

- User Access Logs
- System Accounting Records
- Resource Utilization Logs

Texas Department of Transportation

Procedural data consist of access logs to the system by users and processes, of system accounting records, and resource utilization logs. This information is collected continuously and is used to administer computing and communications resources and to determine access violations and excessive or inappropriate use of resources.

Performance Data

- System Error Logs
- System Access Logs
- Database Performance Monitoring Logs

Texas Department of Transportation

The performance data set consists of log files that record errors in both applications and operating systems for the various platforms in TransGuide®. This data set also includes logs that record the use of system resources and monitor the performance of the database management system. This data is used to monitor the operational state of TransGuide® systems, to manage corrective maintenance, and to direct performance tuning.

System Data

- Source Code
- Executables
- Design Documents
- Equipment Configurations
- Network Management Databases

Texas Department of Transportation

System data consist of critical software components and configuration records that determine how TransGuide® systems will function. This data set receives the highest priority in backup strategies and in the amount of resources allotted to its maintenance and preservation. This data set consists of the custom source code and executables written for TransGuide® that make up the advanced traffic management system and the Model Deployment Initiative projects. It also consists of design documentation, field equipment and network device configurations, and critical network management databases for the telecommunications network.

Administrative Data

- Operator Logs
- Equipment Maintenance Records
- System Configuration Data

Texas Department of Transportation

The administrative data set consists of electronic operator logs, equipment maintenance records, system configuration data that derives from the actual building of the network and computerized systems, and various reporting requirements, like acquisition records, budget development, equipment inventories, network management records, etc.

Storage Hierarchy

- Migration Path
  - On-Line
  - Near-Line
  - Off-Line
- Media
  - Hard Disks
  - Optical
  - Tape

Texas Department of Transportation

TransGuide® uses a conventional approach widely adopted by hierarchical storage management (HSM) systems to determine the migration paths for data sets and to identify the appropriate storage media. Like sophisticated HSM systems, TransGuide® has automated this approach to a large degree through the use of commercial software and custom scripts. Secondary storage devices execute backup routines through a complex set of instructions tailored to fit maintenance windows, system resources, and backup strategies. Unlike most HSM systems, however, TransGuide's implementation is modest and requires a fair amount of operator intervention because we do not employ tape library devices or auto-loaders. Our array of secondary storage devices now consists of 4-mm DAT and 1/2-inch DLT tape units.

Also, TransGuide® has future plans for the use of optical disks to bring data sets to the "near-line" state. Presently, all data is either stored on hard disk or tape, and distribution is handled electronically. In the near future, large historical data sets will be available via CD-ROM.

Factors Affecting Storage

- System Performance
- Storage Capacities
- Cost
- Life Expectancy
- Data Sharing Requirements

Texas Department of Transportation

There are several factors that affect data storage requirements at TransGuide®. Whenever and wherever system performance is adversely impacted by the collection of data, immediate action is taken to move data off systems to other platforms, or to even move it off line to other media. Recently, we dramatically increased the amount of on-line storage in TransGuide®, which should result in more and better access to historical and performance records.

Storage capacities for all media type determine how much data is available and in what media form. Tape media is relatively inexhaustible compared to hard drive space, but also makes access more difficult. It is our objective to keep as much relative and pertinent data as can be feasibly accommodated on-line. Cost has been an influential factor when purchasing storage devices and media. It has not yet been a limiting factor. Life expectancy of data is largely dependent on our constituency of researchers and entrepreneurs. Real-time traffic data quickly grows stale from an operational perspective. Its value beyond the first moment or two lies in its statistical significance.

And finally, data sharing requirements for those data sets TransGuide® will make available will be very influential in this process. TransGuide® will make accommodations to this requirement consistent with TxDOT Policy Statement 2-98.

Backup Strategies

- Complete
- Differential
- Incremental

Texas Department of Transportation

Backup strategies for TransGuide® translate into how best to create and manage the various data sets. In contrast to the factors that influence storage, the backup strategies are more attuned to the daily operation of facilities. We are now maintaining a 21-hour work day, which leaves a small three-hour maintenance window from 1:00 AM to 4:00 AM. As secondary storage requirements grow and the data sets become larger and more complex, backup will have to be conducted around the clock.

The methods that we employ include differential backups (backing up all files that were changed since the last complete backup), incremental backups (backing up all files that were changed since the last backup, regardless of what kind of backup it was), and complete backups (typically incorporate capturing disk images and require complete system down time).

These strategies will affect the timing of data from the viewpoint of data consumers. Whether consumers can influence these strategies is not yet well understood since their needs are lacking precision and clarity.

Factors Affecting Backups

- Maintenance Window
- Backup Intervals
- Rotation Schemes

Texas Department of Transportation

I have already mentioned the small three-hour daily maintenance window and its effect on backup strategies. There are also other influencing factors at work in TransGuide®. These include backup intervals, such as daily, weekly, monthly, and quarterly, as well as rotation schemes, such as "grandfather - father - son" generations.

Historical traffic data is typically stored in a complete backup, after which it is purged from on-line systems. In contrast, system data, as described earlier, is backed up at frequent intervals.

Data sharing will not likely have any effect on these factors. For the most part, TransGuide® conducts backups in concert with internal requirements and remains relatively unaffected by external interests.

Data Sharing Strategies

- Real-Time Data Feeds
- On-Line Resources
- Off-Line Resources
- Future Near-Line

Texas Department of Transportation

Data generation, storage, and maintenance in TransGuide® is servicing the needs of a growing constituency that is both internal and external in nature. Those needs span the entire spectrum of possibilities with respect to timing and levels of abstraction. Some constituents require real-time data in its most raw form, and others require historical data at the highest levels of aggregation. In this regard, the term data archiving is very limiting and does not fairly describe the gamut of activities managed by TransGuide®, since the term is generally understood in the trades to describe historical data used to feed decision support systems.

TransGuide® collects a wealth of data on all facets of the operations. Thus far, a small portion of that data has been the object of interest to researchers. Data sharing, as described in TxDOT policy, has been readily applied to TransGuide®, and all indicators would suggest that this will become a larger part of daily administration.

This slide enumerates the sources of data in TransGuide® from the consumer viewpoint. They vary from real-time data feeds to permanent storage on CD-ROM.

Data Set Availability

- Operational Data
- Limited System Data

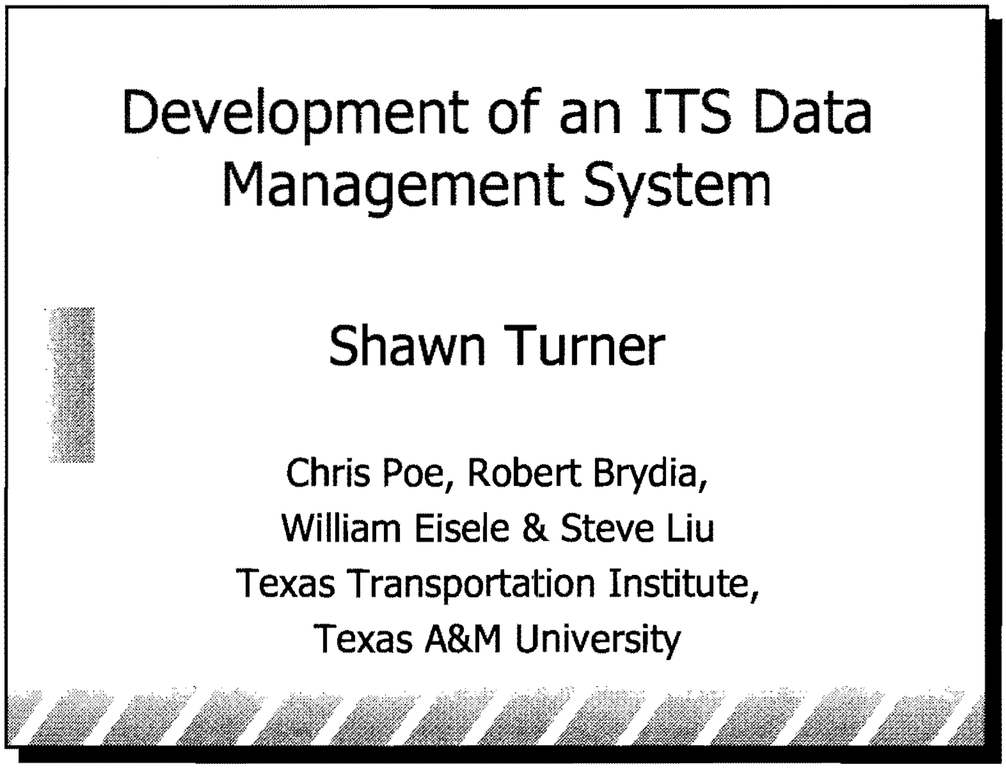
Texas Department of Transportation

The preceding slides illustrated the variety of data generated by TransGuide® systems as well as the host of issues we deal with in administering and managing the backup and archiving tasks. Presently, the data sets available for consumption are limited to those described earlier as operational, along with a very limited portion of system data as it pertains to design.

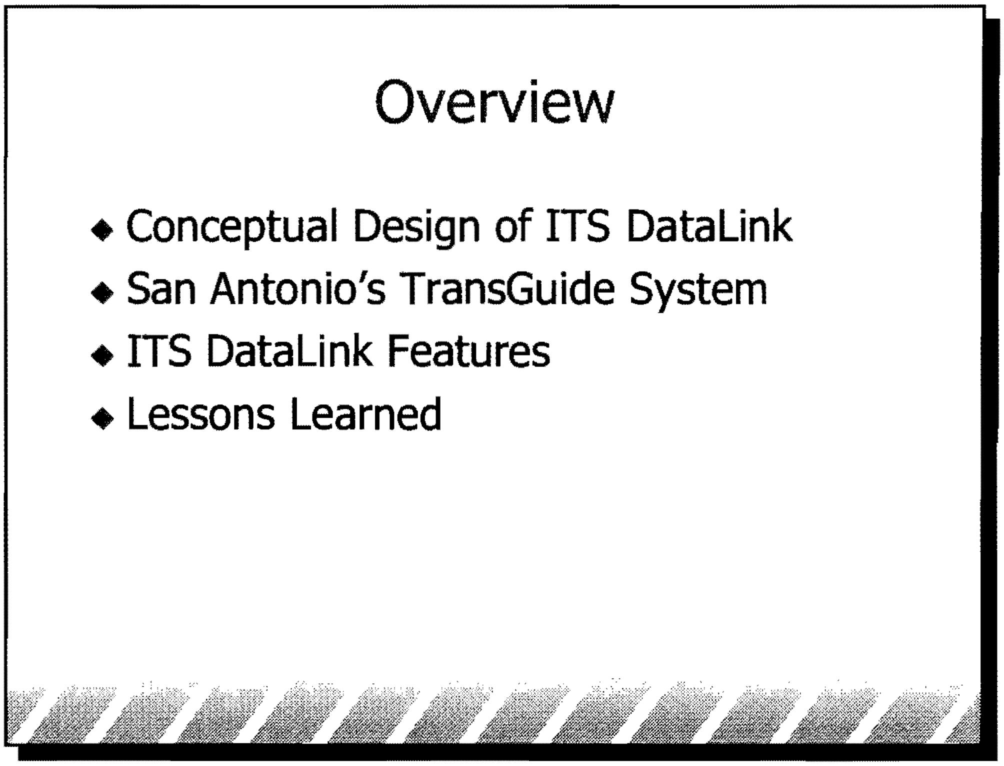
Systems based on the analysis of historical data have been constructed by researchers at a fair number of organizations. In contrast, systems outside of TransGuide® itself that operate on real-time (or near real-time) data are far fewer. Both perspectives have been examined in depth and TransGuide® is accommodating nearly all interested parties to the greatest extent possible. A number of issues pertaining to the integrity and security of the TransGuide® network and facilities have played a large role in defining the extent of that accommodation. However, we continue to explore all opportunities to facilitate the need for information and we have developed management strategies that give us flexibility on these issues.

DATALINK: DEVELOPMENT OF AN ITS DATA MANAGEMENT SYSTEM

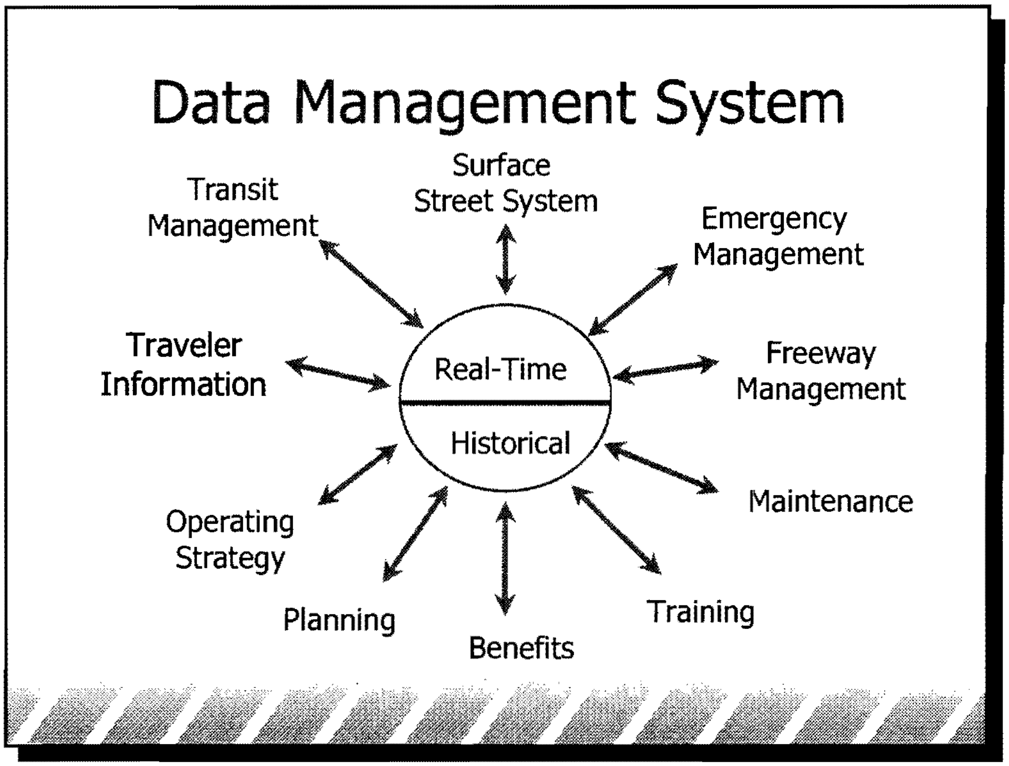
Shawn Turner
Texas Transportation Institute



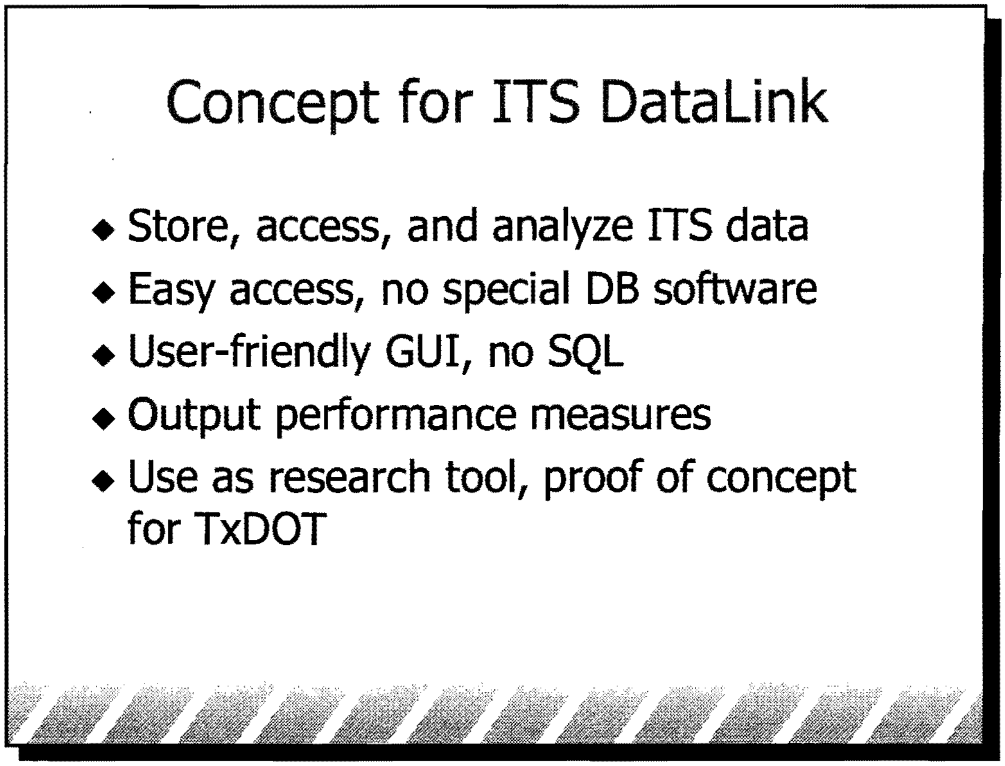
There are several people that have been involved and played a key role in developing what I will refer to as DataLink, or TTI's ITS data management system. Several of these people are here in the room today, including Chris Poe, Bill Eisele, and Steve Liu (who has been the brains behind the development of our system). Steve helped us get a prototype system developed in a short period of time. I think that I am up here giving the presentation because I am the only one who has Powerpoint on my laptop.



This afternoon, I would like to talk about several things. I will briefly discuss the conceptual design of the ITS data management system that we have developed. I will also talk briefly about the TransGuide system and its loop detector data used to develop our system. I will also talk about some of the features that are in the DataLink system. Most importantly, though, I want to convey some of the lessons that we have learned in the past year in dealing with the vast amounts of data and developing this data management system.



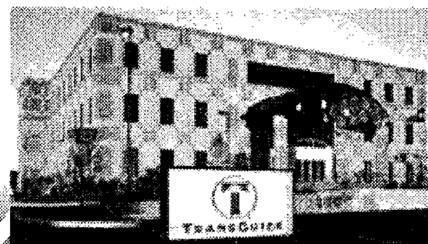
I think we all understand the concept that there is a lot of data that is being collected and used in real-time for numerous ITS applications. The part of the figure that we need to add, however, is the use of historical databases for a variety of other applications, including planning, benefits analyses, evaluation, training, maintenance, etc.



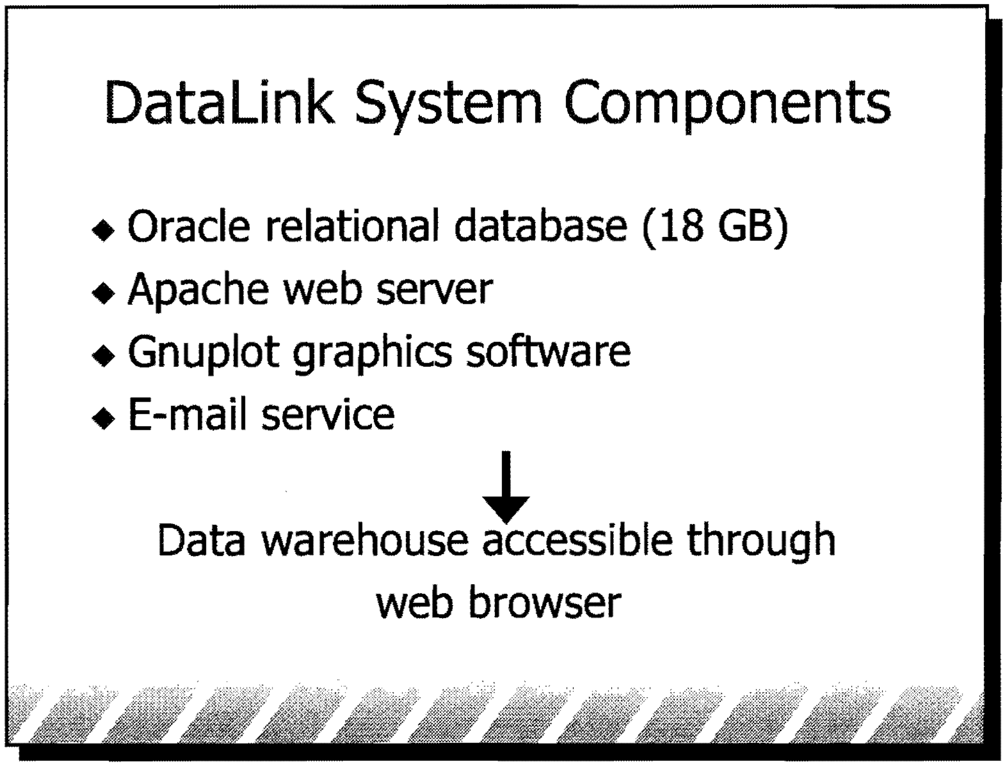
The concept for the ITS DataLink system originated about two years ago. We were trying to analyze loop detector data from San Antonio's TransGuide and we thought "wouldn't this large amount of data be easier to work with if we had a central repository for it and we could get to it easily? Wouldn't it be nice if we didn't have to know a special database language to query the system?" We also wanted to be able to output different types of performance measures, such as travel time, vehicle miles traveled, and other measures that are commonly used in planning applications. At this time, there was a need for such a system because of several ongoing or recently started research projects that were to use this data in developing incident detection algorithms or documenting system performance. Because of these needs, we intended to develop the DataLink system as a research tool for numerous studies. We also wanted to develop the system as a "proof of concept" for TxDOT, as a way to demonstrate one approach to storing ITS data and making it accessible to end users. That was the motivation behind the development of DataLink about two years ago.

TransGuide Phase One, San Antonio

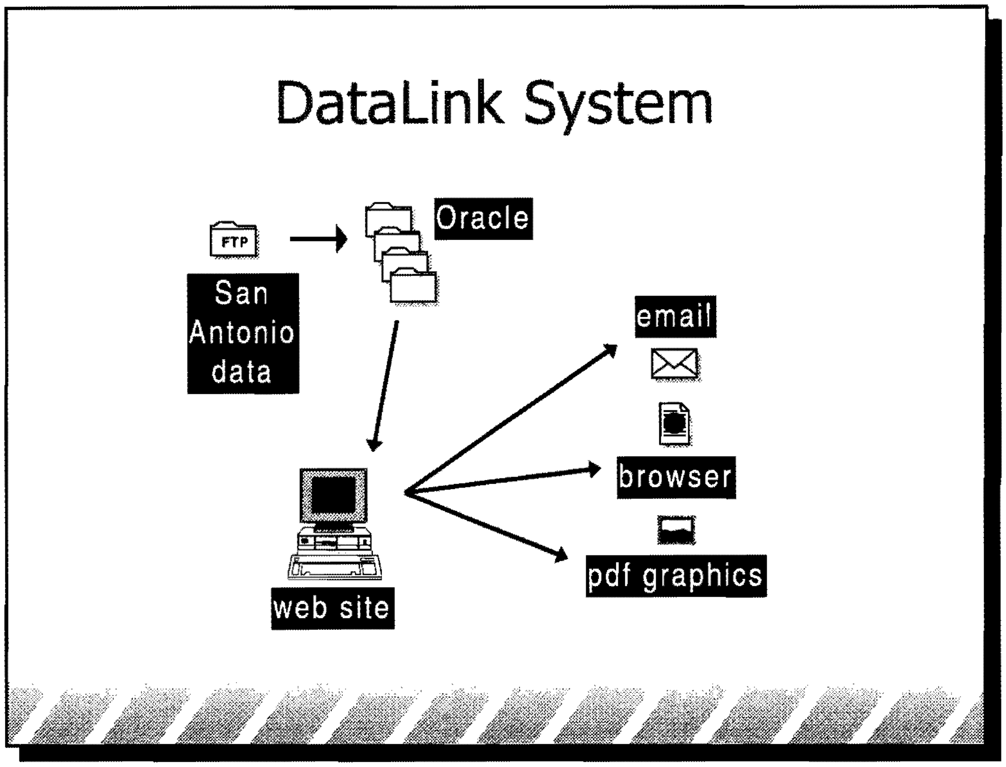
- Over 530 loop detector stations on mainlanes and ramps
- 20-second polling pattern
  - volume, speed, loop occupancy
- ±120 megabytes per day



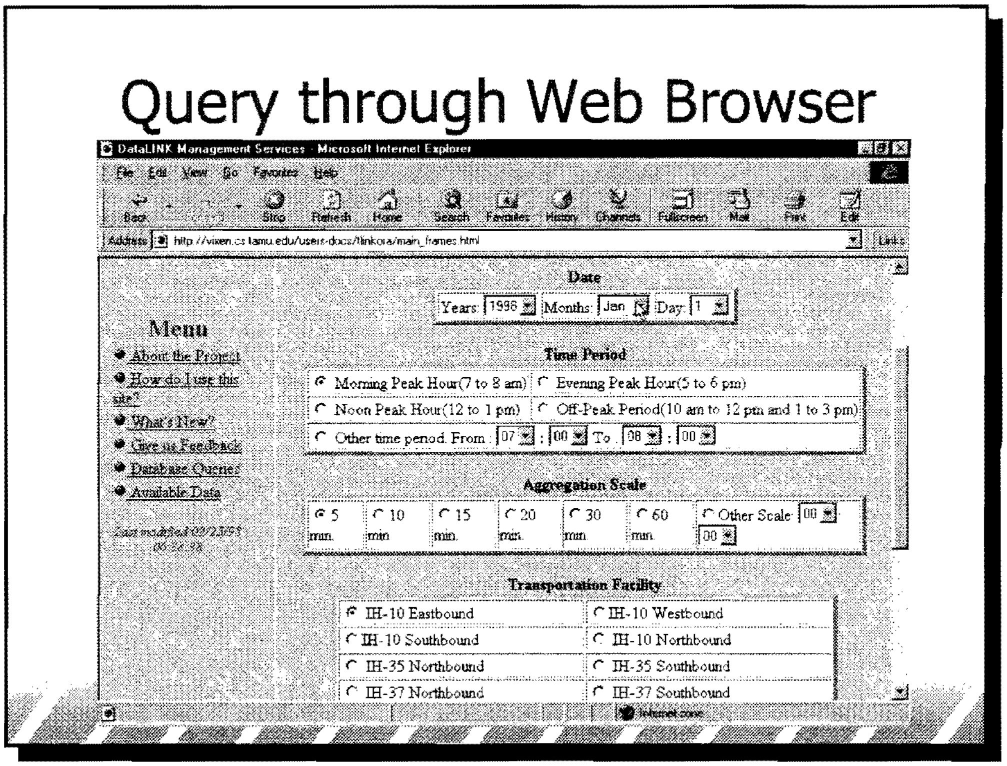
The primary data that is in our data management system is from Phase One of San Antonio's TransGuide system, which includes over 530 loop detector stations on the mainlane and exit/entrance ramps. This loop detector data is collected in a 20-second polling pattern. Vehicle volume and occupancy data are available from all loop detector stations, and vehicle speeds are available only from the mainlane detectors, which are in a double-loop configuration. Approximately 120 megabytes of data are being archived (in a "flat" ASCII-text file) per day from Phase One, which is only about 26 miles of the freeway system around downtown San Antonio. From this daily estimate of file storage requirements, you can see that the file sizes for an entire year can certainly add up.



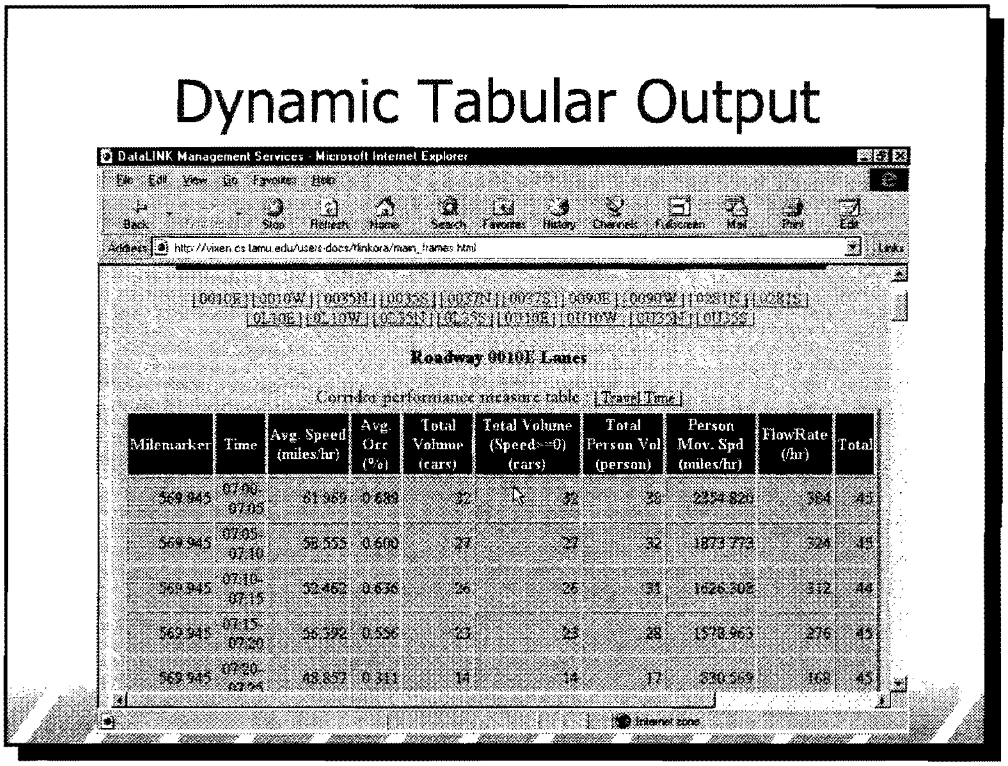
The ITS DataLink system that we developed consists of several components. The main component of the system is an Oracle relational database. We started out with 18 GB of storage capacity, and that quickly grew to 40 GB. We currently have about a full year of data on-line with this 40 GB capacity. Because we wanted the DataLink system to be accessible through a web browser, it was necessary to add a web server to the relational database. We also added a few components to the system that we hoped would make the data more accessible to users as well as provide a user-friendly interface. For example, we used free graphics software that generates two- and three-dimensional graphics. Another component we added was e-mail service, which provides the capability to have query results automatically sent to a specified e-mail address. These query results are sent as comma-separated values that could easily be imported into most spreadsheet applications. The integration of these components basically provided a scaled-back data warehouse that is accessible with a web browser. You do not need to know any special database languages to make data queries. You really just need to know how to get on the Internet and point-and-click. Once you have entered the system, you can define your query and the desired outputs. The next several slides will show some screen shots from the system. Again, the DataLink system is accessible via the web, so if you are interested we can establish a user account and password for you to examine the system.



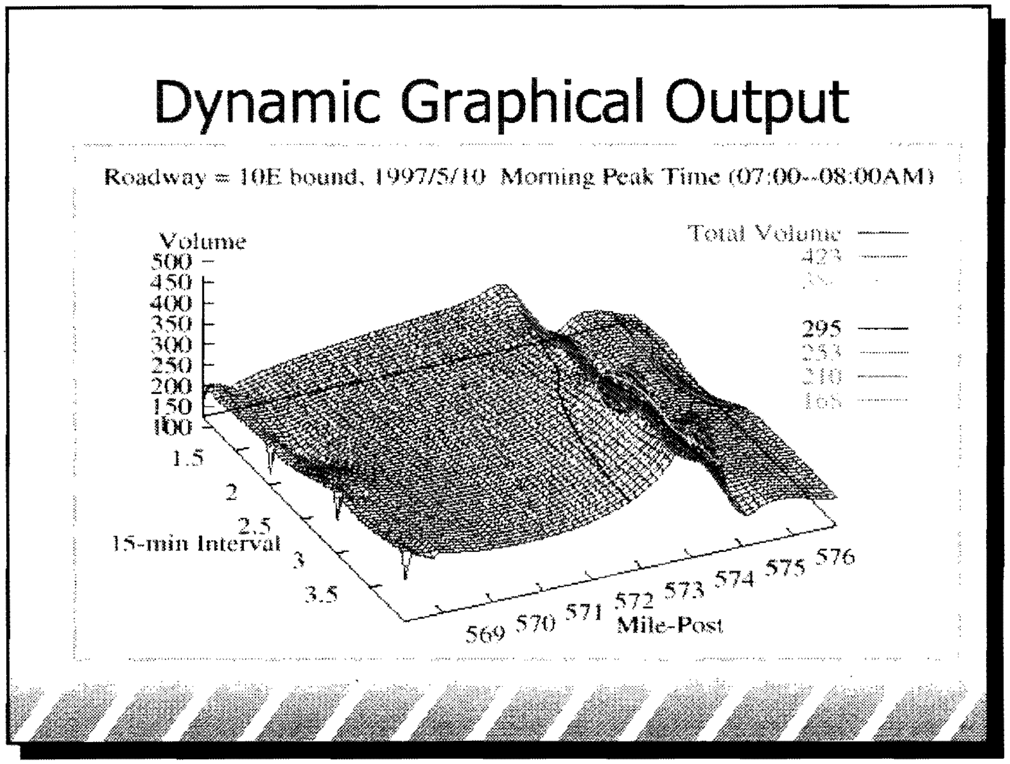
This slide shows a schematic of the DataLink system. The TransGuide system is currently archiving the loop detector data to a hard drive, which is accessible to us via file transfer protocol (FTP). On a daily basis, we download the data from the TransGuide FTP site and load it into the Oracle database. In loading the database, we do summarize the 20-second raw data into five-minute time periods. At the outset, we made a design decision to warehouse loop data at the five-minute level because: a) we probably did not have the computer resources to handle 20-second data in a large database for any amount of time, and b) we felt, at the time, that the predominant number of uses would require five-minute data (at the time we were talking about mostly planning applications). Since then, however, we have wanted to have access to the raw 20-second data for other research purposes. Because this raw data is being archived at TransGuide for at least the most current year, all we have to do is download the data from the FTP site for the day(s) that we are interested. At this point, we have the five-minute data that is available through DataLink, as well as the raw data archives that are available for download through TransGuide. The main point of entry for most people is through the DataLink web site. We have a number of different outputs that we can get from DataLink through ad hoc or structured database queries. The DataLink system can provide query results through e-mail, or you can have data summary tables returned in the web browser itself. Two- or three-dimensional graphics can also be generated depending upon the query.



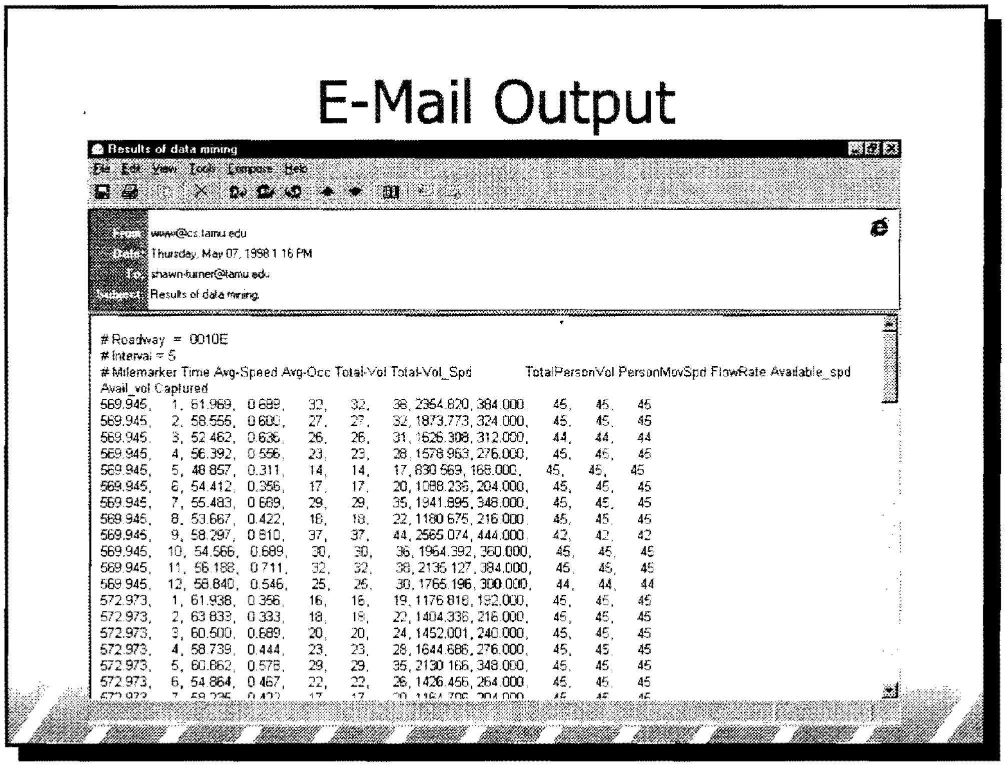
Again, the idea behind the DataLink system is to make large ITS databases available to someone like me that does not know database query or programming languages. I can basically go into the DataLink system and perform the queries by point-and-click. We have some pre-defined time periods, such as the peak hour or peak period, as well as the ability to define whatever time period you would like to have. You can even define a daily summary. The system also provides the ability to aggregate the data. The data is stored as five-minute data, but has the capability to summarize to 10, 15, 30, 60 minutes, or you can set the aggregation level to anything under 24 hours. If you need six-hour averages, you can get six-hour averages. We have tried to provide the flexibility in the system so that we have fairly detailed data (five-minute level) available, but we also have the ability to accommodate a wide variety of users and uses in terms of being able to aggregate up to different levels of detail. If you would like to select a specific freeway corridor to analyze, you can also select these by direction.



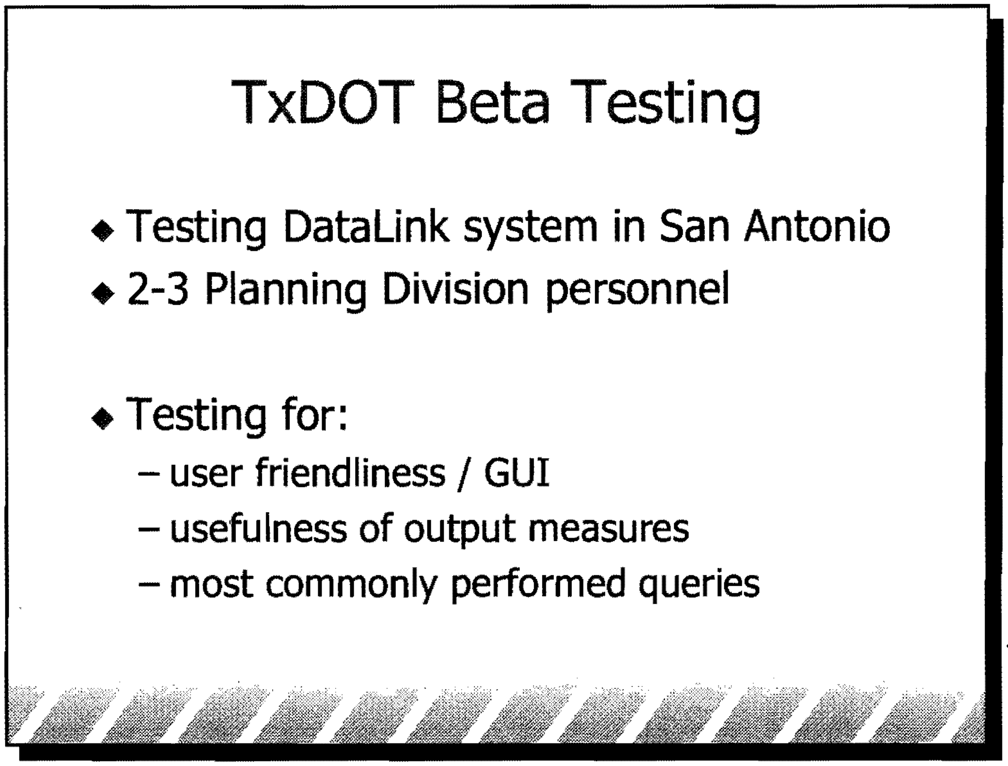
This slide shows an example of the tabular outputs that are returned to the web browser window. The DataLink system has gone through some evolution. We started out with a fairly complicated, detailed output of performance measures. For this example slide, I believe, we are showing numerous performance measures. We have tried to simplify the outputs to some basic measures, but again the idea is to be able to output different types of data and performance measures to a wide variety of users to make it easy and accessible to those same users.



This slide shows an example of the three-dimensional graphics available through DataLink. We use free graphics software that has certain limitations, but we have also provided the ability (through the e-mail service) to import the data into more sophisticated graphing or analysis packages. With the e-mail service, comma-separated values are included in the body of the e-mail, which can be directly imported into most spreadsheet and database applications.



This next slide shows an example of the e-mail output. It is a fairly simple process to import these comma-separated values into most spreadsheet or database applications.

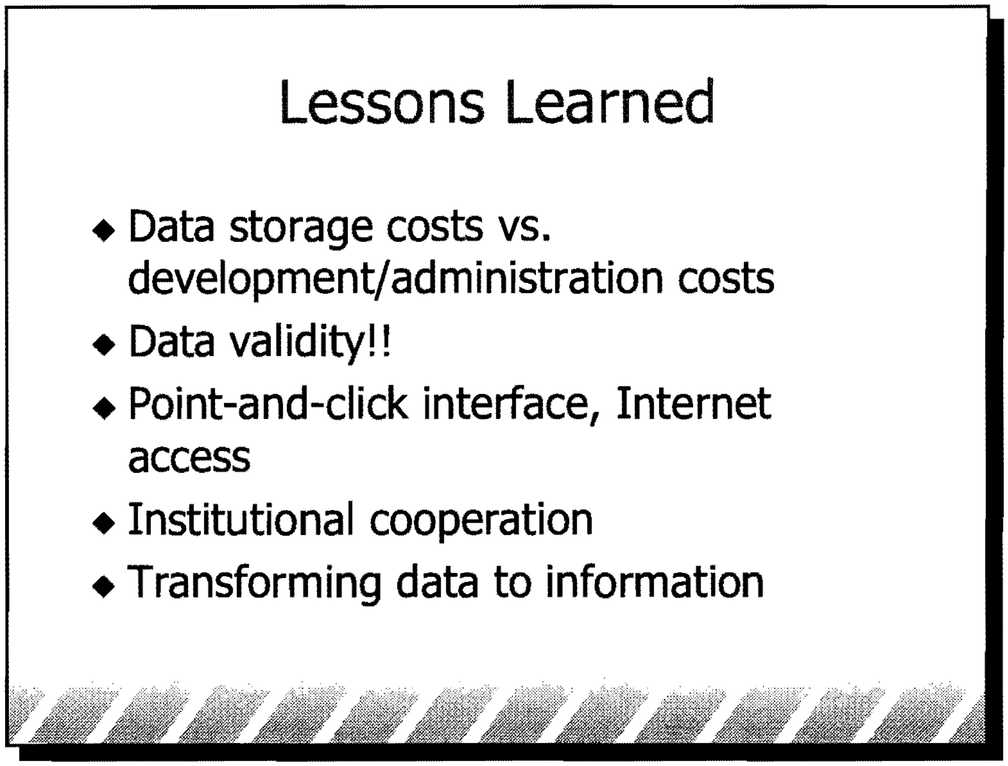


As I mentioned earlier, the DataLink system has been evolving over the past 9 to 12 months. We have had several transportation planners from TxDOT's San Antonio district doing some beta testing for us. They were looking at aspects of the system, such as user-friendliness, graphical user interface, and the usefulness of the output measures. We were also trying to determine the most commonly performed queries so that we can develop pre-defined queries for regular system performance reports. For these pre-defined queries, you could log onto the system, punch one button, and get a daily or weekly performance summary. We would like to take this feedback from the beta testing and improve the DataLink web site.

TxDOT Data Needs Survey

- Surveyed 9 urban districts (operations and planning groups)
- Commonly specified data needs:
  - congestion management systems
  - level of service analyses
  - project/pavement design
  - demand forecasting model calibration
  - MIS/corridor studies

We have also performed a data needs survey. If you go to the data uses and applications break-out group, Russell Henk should be distributing a copy of the results there. TTI sent out data needs surveys to the operation and planning engineers in nine of the urban TxDOT districts. We asked them "we have certain data elements that are or will be available through ITS, so can you tell us what you think you might be able to use these data elements for and what types of format, aggregation level, etc., that you would desire." We were really trying to identify some of the user needs and requirements. Some of the most commonly specified applications that could use ITS data include congestion management systems, level of service analyses, project and/or pavement design, demand forecasting model calibration, and major investment studies (MIS)/corridor studies.

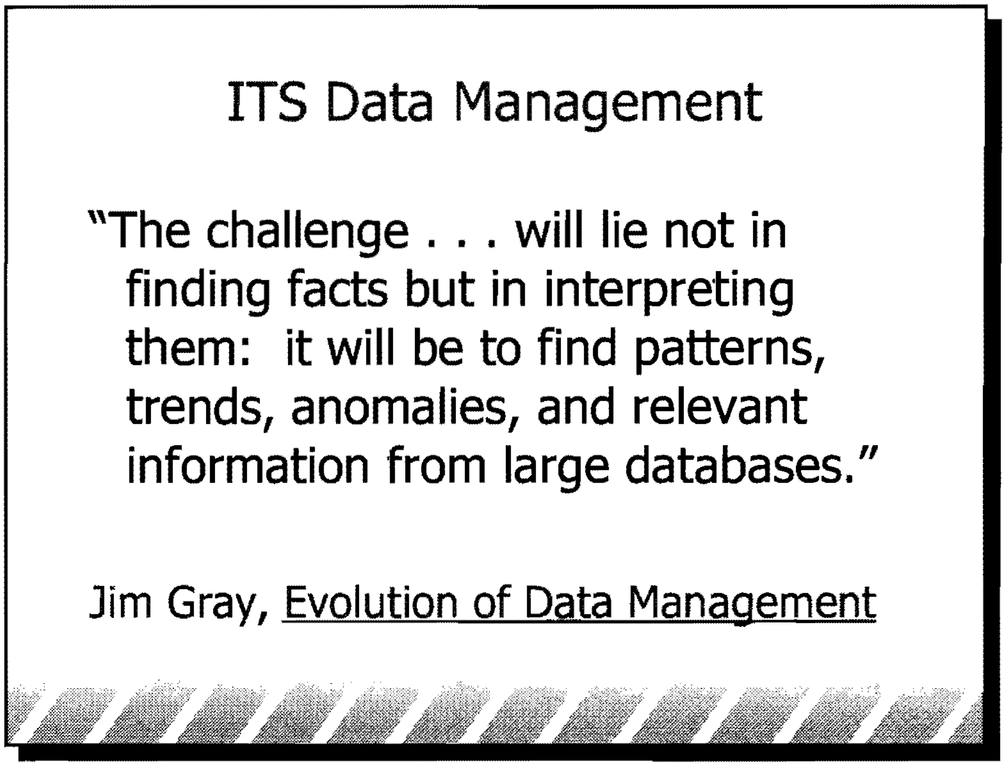


Over the past year or two we have made some mistakes, but I think we have also managed to do several things right. At the start, one of the big issues (at least from my perspective) was the costs for data storage. Instead, we found that the database development and administration costs were much more significant than data storage costs. The costs of computer storage is dramatically decreasing, whereas the costs for a database administrator for developing applications and maintaining a data management system can be upwards of $60,000 per year. Data validity was an issue that came up several times in developing and testing the DataLink system. Before you start loading a database, you should: a) understand the data that you are collecting and b) have quality control procedures in place so that you can identify bad or suspicious data. If you don't have good quality control procedures, the old adage of "garbage in, garbage out" certainly applies. We thought that the most beneficial part of the DataLink system was the intuitive, point-and-click interface provided via a web browser interface. You can be anywhere and get to the database, as long as you have Internet access. Institutional cooperation is a key element in data archiving and warehousing projects. I have heard some very positive things here this morning about the sharing of information resources. I think that will be necessary because there are many stakeholders that are interested in archived ITS data. The last point relates to the ability to transform data into useful information. We could very easily get overwhelmed with the vast amounts of data that are being collected by ITS control centers. We need to ensure that we have a mechanism (i.e., analysis and reporting tools) that can turn the data into information people can use.

Current Research Efforts

- Performance measures
- Error detection
- Aggregation/storage guidance

This slide lists some of the current research efforts that we have going on either through the TransLink ITS research center or other research projects at TTI. We have an effort aimed at developing performance measures. We are also trying to develop error detection algorithms that can be used to identify bad or suspicious data at the controller level. Ideally, we would like to be able to detect errors as far "upstream" in the data collection process as possible. Additionally, we are developing guidance for ITS data storage and aggregation levels based, in part, on the data that is currently being collecting in San Antonio and Houston. This effort is examining the benefits of saving data at a very detailed level versus an aggregated level, and the ability to accommodate a number of different uses.



With that said, I will put up the following quote, which I thought is particularly appropriate for the ITS data archiving and warehousing activities that we are talking about today.

PANEL DISCUSSION ON DATA NEEDS AND STANDARDS

TXDOT PLANNING PERSPECTIVE

Dayton Grumbles
Texas Department of Transportation

I would like to talk about a few things this afternoon from a data collection standpoint. I work in the technical services group of TxDOT's Transportation Planning and Programming (TP&P) Division. Right now, we are making about 85,000 counts per year using rubber tubes (40,000 in urban areas), as well as about 8,500 ramp counts per year. In some cases, we are placing our rubber tube counters very close to inductance loop detectors that are also collecting data. If we could get data from ITS, we feel that we could save a significant amount of money. We are currently spending about $100,000 to $140,000 per year through contracts to collect ramp traffic volume data. This contract amount does not include additional traffic counts that TxDOT has to perform for special studies.

There is also a safety issue when we have personnel installing road tubes. Even if we are doing traffic counting through contracts, there are still people who are vulnerable and could be injured. We have been lucky so far to not have many accidents. Another issue with manual data collection is the scheduling and data processing activities. There is a significant effort in simply scheduling data collection by contractors. Because of these reasons, we feel there can be a tremendous amount of savings if we could collect traffic data from ITS centers. So it could be a money-saving activity as well as addressing safety issues of manual data collection, if we could work together in obtaining and sharing data from ITS applications. I believe that this is the direction that we should head.

Mark Hodges
Texas Department of Transportation

My name is Mark Hodges, and I am the Traffic Monitoring System (TMS) manager in the TP&P division. As Dayton mentioned, we make about 85,000 traffic volume counts per year. There are several things that are important for our data collection programs within TP&P. The quality of the data is certainly one thing about which we are concerned. Another concern is data transfer from the field, as we currently have about 160 ATR stations across the state that do continuous traffic volume counts. We have been working with Shawn Turner of TTI to do some control checks of this ATR data versus equivalent loop detector data from an ITS center. In these control checks, Shawn is comparing ATR data from two stations in San Antonio to traffic volumes collected by nearby TransGuide loop detectors. These control checks should give us an indication of whether these devices are providing comparable traffic volume reports.

There are some standard guides that we use when we perform data collection. The two guides that we use most often are FHWA's *Traffic Monitoring Guide* (TMG) and the *AASHTO Guidelines for Traffic Data Programs*. Both of these guides are in the process of being revised and rewritten. Mark Hallenbeck with the University of Washington has been contracted by FHWA to rewrite the TMG guidelines. An AASHTO subcommittee is working to revise the AASHTO guidelines, which should reflect the additions that Mark is doing for the TMG guidelines. In these revision efforts, we are hoping to provide some insight and standards for using data from ITS. We are also considering the best approach to incorporating this ITS data into existing data formats that we have been using for planning data. For example, vehicle classification data has typically been subdivided into 13 categories or vehicle classes. This vehicle classification data is formatted into Records 2 and 4, which are defined in FHWA's TMG. Another example is continuous traffic volume count data, which is formatted into Records 1 and 3. These TMG guidelines and reporting formats are necessary because we are required to report data to FHWA on an annual basis.

This ITS area is quite new to us in TxDOT's TP&P. We did meet with Al Kosik of the Traffic Operations Division about 10 years ago, and there were some discussions about what types of data were needed in TP&P. At the time, we thought that they were crazy thinking they could give us the data that we needed. But it now appears that the data is available and could potentially be used by us.

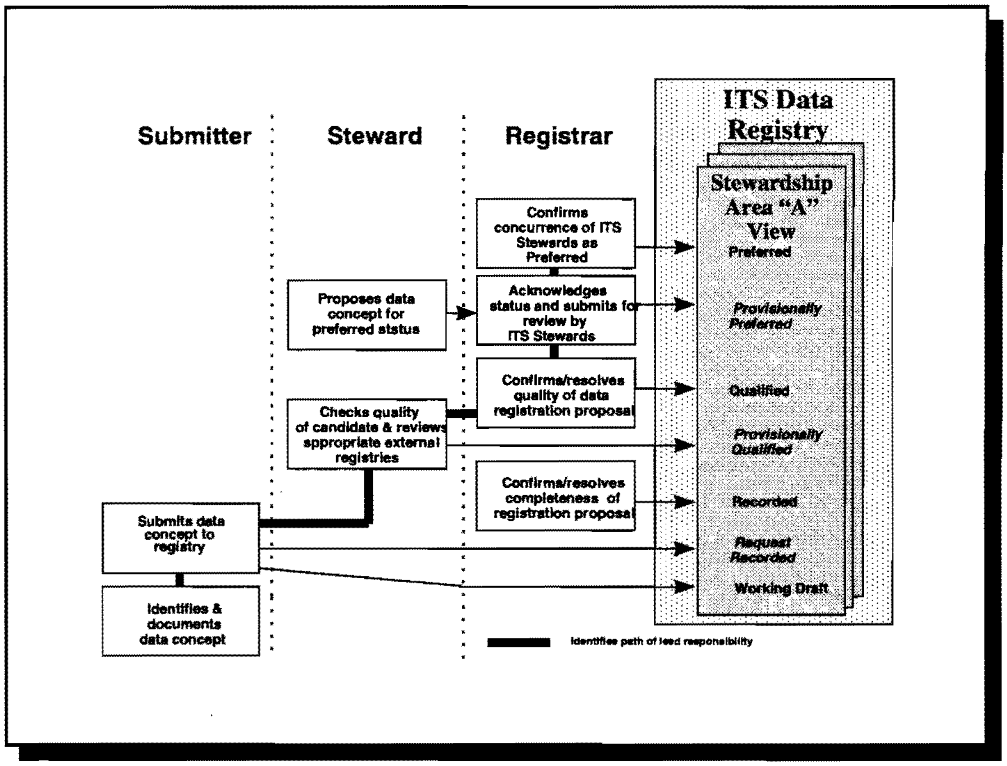
DEVELOPMENT OF A NATIONAL DATA REGISTRY

Ed Seymour
Texas Transportation Institute

I would like to spend some time this afternoon talking about the concept of the National Data Registry. Standards efforts are really an outgrowth of the National ITS Architecture, and they are user-driven in the sense that (especially for many roadside devices) there has been a desire to develop standards for the ITS applications for many years. Many of these standards efforts are interrelated or are applicable for numerous ITS applications. For example, Steve Dellenback will be talking about one of them–the traffic management data dictionary (TMDD). And there are others as well. There are groups that are developing advanced traveler information system (ATIS) standards. There also is the National Transportation Communication for ITS Protocol (NTCIP), which is developing many roadside device standards.

Because of the many standards efforts, there will be many data elements defined. These data elements will be used in control applications and many other processes associated with these applications. You can imagine that, with all of these different standards efforts, there are thousands of data elements that eventually map back up to the National Architecture. For example, the TMDD deals with the traffic management center communications. However, the way it is built, it leaves out all of those things that are being devised or defined somewhere else. It leaves out all the data elements that are being collected at the roadside from traffic control devices because they are being covered somewhere else. Pretty soon, it becomes difficult to figure out where all the data elements are defined. Different standard development organizations are working on these (ITE, IEEE, etc.), and the data element definitions are maintained at different locations.

The National Data Registry was conceived as a repository for the numerous data element descriptions and the message set descriptions that are being developed in standards efforts. IEEE has taken the lead on this data registry, and I have passed out a handout (see pages 88 to 94) that has some key points from this particular effort. I do want to say that this effort has not yet officially started. At this point, a limited effort has been initiated to define how the data registry will work. It has not been prototyped nor has the prototyping been funded. However, I think the funding of the data registry is eminent, and they will develop a small-scale deployment of the data registry in the near future.



One of the main concepts of the data registry is that data moves through different quality levels, where the highest quality level is analogous to a standard. This high quality level means that the ITS community has bought into it, and that the data element has been standardized by a standards development organization. At the lowest quality level are the working group and recorded levels, which contain draft data element that have been or are being proposed in standards efforts. The concept behind the data registry, then, is that you take all the data elements that have been defined in standards efforts and put them in the data registry at some quality level. In addition, you may encourage people in their own deployment to take a data element from their system and post them to the data registry. As you build a system, for example, you can put in a data element at a working group draft level or recorded level on behalf of the state or city. You do not necessarily need a sponsor, you just submit it to the registry. In effect, you are sharing your data element description with other people.

At the provisionally qualified level (the quality level above the recorded level), you have to get data stewards involved. Data stewards are basically sponsors of the data elements. To have data elements registered at this quality level, then, you essentially have to work with some standards development organization in sponsoring the data element and moving it up the chain of quality levels. At the lowest quality level, you can just submit a data element with a minimal form, and you have not necessarily tried to coordinate with others to see if it is a good data element. As you move the data element up in quality level and get a steward to sponsor it, you get more buy-in from the ITS community. As this occurs, you must verify that all the data registration forms are filled out correctly, that all the data meta-attributes and other considerations are properly documented, and you must coordinate with everyone else. Eventually, data elements may move up the quality level to a preferred status. With the archived data that we are talking about today, these data elements could follow the same path. The data element descriptions could be posted at the beginning at the working draft level or the recorded level, and then a standards development organization could help sponsor this data element to move it up in quality level.

IEEE will be developing the data registry with the goal of making it a self-sustaining effort, so there will likely be fees or charges for access to the data registry documents. I have no idea how much the cost will be, as I have no indication from IEEE. I do not think that IEEE will project access fees until after they do the pilot project and assess the likely costs of maintaining the registry. The pricing concept is that a city, county, or state organization would pay a fee and then have access to the data registry. Standards development organizations sell their paper standards documents for $50 to $70 each, and you might have to go to 10 different groups to get the applicable standards. With the data registry, you would pay a single fee and go to a single source to get the data element descriptions.

DATA REGISTRY

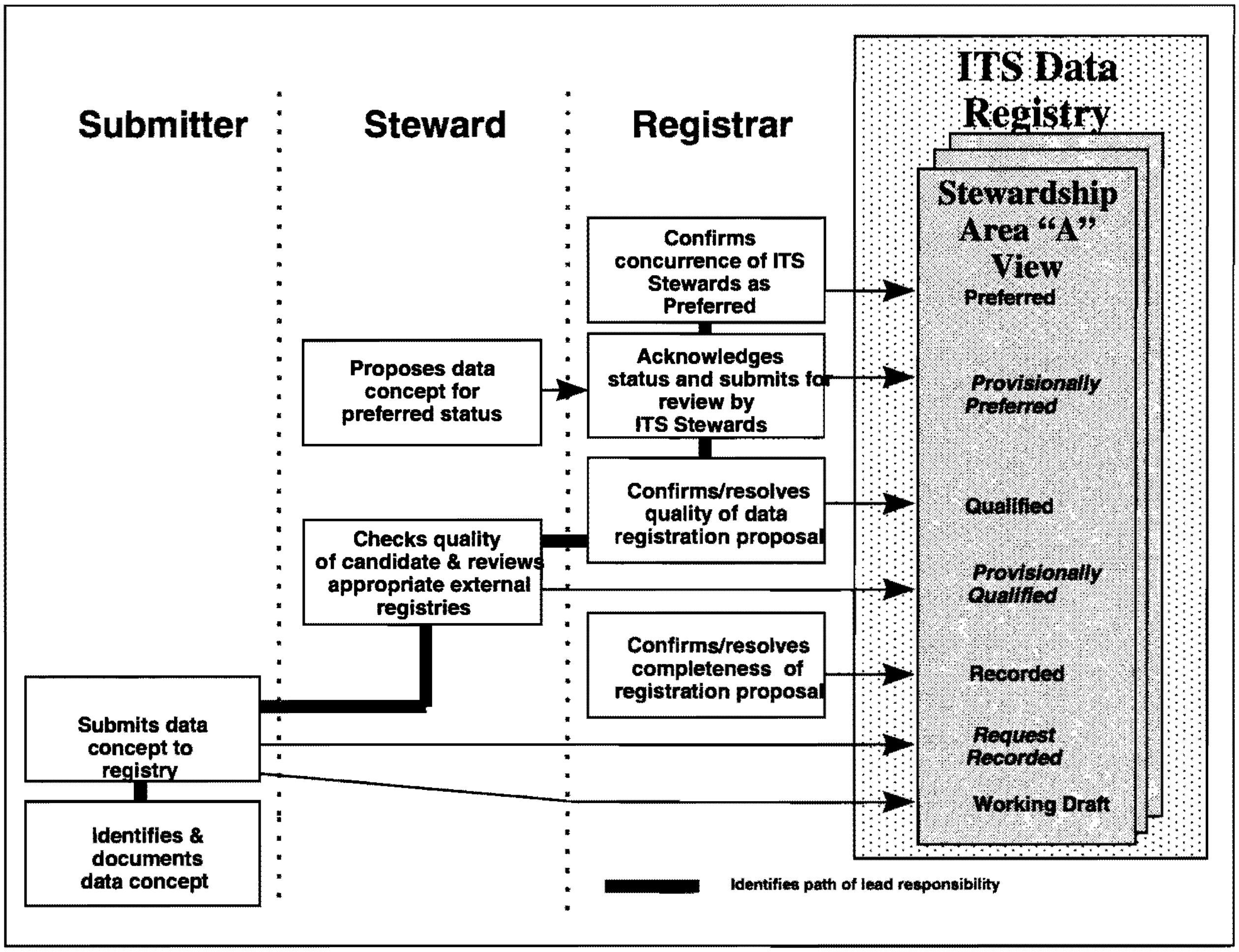
By Ed Seymour
TTI
eseymour@tamu.edu
Tel: 972-994-0433
Fax: 972-994-0522

Discussion Outline

- Serves as a central repository for ITS data element definitions.
- Sponsoring Standards Development Organization (SDO) is IEEE.
- Plan is to structure the effort as fee supported at some point in the future.
- Classifies data elements through "quality status" designations.
- Does not replace SDO standardization efforts.
- Any ITS data elements (sponsored by a public agency, private sector, or SDO) could be "registered."
- See excerpted "Functional Activity Overview."
- See excerpted "ITS Data Registry Interim Participants."
- Web site for this effort is http://standards.ieee.org/regauth/itsdr/index.html .

All info is excerpted from a 9/28/98 draft of the "Data Registry Functional Operating Procedures" document, posted at http://grouper.ieee.org/groups/scc32/datareg/DRFOP8.DOC.

Functional Activity Overview



Registry functional activities

The process will result in data concepts registered in one of seven status categories:

a) Working Draft: Data concepts that a Submitter submits to the Registry but for which meta attributes may not yet be complete or reviewed by a Steward. Working draft data concepts are not maintained under version control; that is, updates to Working Draft data concept are over-written. Changes or other updates to data concepts in Working Draft status are by replacement (i.e., the changed entry entirely over-writes the previous entry). The previous entry is then not retrievable. Note: The Submitter may retire a data concept in the registration status of Working Draft at any time, without warning.

b) Recorded: Working Draft data concepts for which the Submitter has requested Recorded status and the Registry system has verified entries in all mandatory meta-attributes (including Relevant Groups and presence of the ASN.1 Name). Note that these

All info is excerpted from a 9/28/98 draft of the "Data Registry Functional Operating Procedures" document, posted at http://grouper.ieee.org/groups/scc32/datareg/DRFOP8.DOC.

mandatory meta attributes may not be in conformance with quality requirements for such meta attributes and the ASN.1 Name may not be unique. The Submitter may retire a data concept in the registration status of Recorded at any time, without warning.

c) Provisionally Qualified: Recorded data concepts for which a Steward has confirmed that the meta attributes are complete and conform to applicable Registry meta attribute quality requirements. Steward Organization Name is mandatory and the ASN.1 Name must be unique in the registry for data concepts at Provisionally Qualified or higher status.

d) Qualified: Provisionally Qualified data concepts for which the CCC has confirmed that the meta attributes are complete and conform to applicable quality requirements (Qualified corresponds to Certified in ISO/IEC 11179.).

e) Provisionally Preferred: Qualified data concepts that a Steward proposes as Preferred for use in the ITS community. However, certification of Preferred status of the data concept by the CCC is not yet complete.

f) Preferred: Provisionally Preferred data concepts that the CCC confirms as a Preferred data concept for use in the ITS community. (Preferred corresponds to Standardized in ISO/IEC 11179.)

g) Retired: Data concepts in the registration status of Qualified or higher that have been approved by the CCC as no longer recommended for use in the ITS community. Also, data concepts in the registration status of Working Draft or Recorded that the Submitter has retired. Such data concepts are retained in the Registry or its archival storage facility for historic reference purposes.

All info is excerpted from a 9/28/98 draft of the "Data Registry Functional Operating Procedures" document, posted at http://grouper.ieee.org/groups/scc32/datareg/DRFOP8.DOC.

ITS Data Registry Interim Participants

SDO Sponsor	Functional Area(s)	Contacts
Institute of Transportation Engineers (ITE)	TMDD TCIPDD	Overall POC Lyle Saxton [540.347.9512; lsaxton@erols.com] Board of Directors Mark Norman [202.554.8050x126, mnorman@vax.ite.org] James Checks [202.8050x131, jchecks@vax.ite.org] Configuration Control Committee Lyle Saxton (TMDD) Eva Lerner-Lam (TCIPDD) [800.756.9542;[lernerlam@palisadesgroup.com] Steward(s) Lyle Saxton (TMDD), Eva Lerner-Lam (TCIPDD)
Society of Automotive Engineers (SAE)	ATISDD	Overall POC Marcy Lucas [724.772.8557; lucas@sae.org] Board of Directors Marcia Lucas Configuration Control Committee Joel Markowitz [jmarko@mtc.dst.ca.us] Cecil Goodwin [???] Steward(s) Joel Markowitz, Cecil Goodwin
American Association of State Highway and Transportation Officials (AASHTO) National Electrical Manufacturers Association (NEMA)	NTCIP	Overall POC Bo Strickland [703.281.6510; strickbo@aol.com] Board of Directors Dave Hensing (AASHTO)[202.624.5812; dhensing@aashto.org] Frank Kitzantides (NEMA) [703.841.3258; fra_kitzantides@nema.org Configuration Control Committee Ed Seymour [972.994.0433; e-seymour@ttimail.tamu.edu] Bruce Schopp [703.841.3231; bru_schopp@nema.org] Steward(s) Ken Vaughn [703.471.0838; kvaughn@mail.viggen.com] Bob DeRoche [850.562.2253; rderoche@transyt.peek_traffic.com]
Institute of Electrical and Electronic Engineers (IEEE)		Overall POC Tom Kurihara [703.516.9650; t.kurihara@ieee.org] Board of Directors (Chair) Jerry Walker [732.562.3823; j.t.walker@ieee.com]

All info is excerpted from a 9/28/98 draft of the "Data Registry Functional Operating Procedures" document, posted at http://grouper.ieee.org/groups/scc32/datareg/DRFOP8.DOC.

(IEEE)		Configuration Control Committee (Chair) Tom Kurihara Registration Authority point of contact and BOD/CCC Secretary: Anita Ricketts [732.562.3847, aricketts@ieee.org] Registrar: Burt Parker [703.979.9499; parkerbg@idsonline.com] Steward(s) Tom Kurihara
Federal Highway Administration (FHWA)	JPO	Overall POC Mike Schagrin [202.366.2180; mike.schagrin@fhwa.dot.gov] Board of Directors Mike Schagrin Configuration Control Committee (advisor) National Architecture Bruce Eisenhart [703.367.1671; bruce.eisenhart@lmco.com] Steward(s) [Not applicable]
Commercial Vehicle Operations (CVO)	CVO Data Dictionary	Overall POC Ray Yuan [240.228.6356; raymond-yuan@jhuapl.edu] [Note: Participation unknown as of second DR Design Group meeting]
American Society of Testing and Materials (ASTM)		Overall POC Dan Smith [Note: Participation unknown as of second DR Design Group meeting]

All info is excerpted from a 9/28/98 draft of the "Data Registry Functional Operating Procedures" document, posted at http://grouper.ieee.org/groups/scc32/datareg/DRFOP8.DOC.

Annex B

ITS data registration form

<table>
<tr><th colspan="5">DATA REGISTRATION PROPOSAL</th></tr>
<tr><td colspan="5">Proposal type: New [ ] Change [ ] Retire [ ]
Expected Registration Level: Recorded [ ] Qualified [ ] Preferred [ ]</td></tr>
<tr><th colspan="5">ADMINISTRATIVE INFORMATION</th></tr>
<tr><td colspan="3">Submitter Organization Name:</td><td colspan="2">Submitter Phone Number:</td></tr>
<tr><td colspan="3">Steward Organization Name:</td><td colspan="2">Steward Phone Number:</td></tr>
<tr><td colspan="3">Registrar: (If an external registrar)</td><td colspan="2">Registrar Phone Number:</td></tr>
<tr><th colspan="5">REGISTRATION INFORMATION</th></tr>
<tr><td colspan="2">Descriptive Name:</td><td colspan="3">Descriptive Name Context:</td></tr>
<tr><td colspan="5">Definition:</td></tr>
<tr><td>Class Name:</td><td colspan="3">Classification Scheme Name:</td><td>Classification Scheme Version:</td></tr>
<tr><td>Value Domain:</td><td colspan="3">Representation Class Term:</td><td>Data Type:</td></tr>
<tr><td>Valid Value List, Range, or Rule:</td><td colspan="3">Keyword:</td><td>Formula: (if applicable)</td></tr>
<tr><td>Data Concept Type:</td><td colspan="3">Security Class:</td><td>Source:</td></tr>
<tr><td colspan="3">Related Data Concept:</td><td colspan="2">Relationship Type: (One for each related data concept)</td></tr>
<tr><td colspan="3">Relevant Group:</td><td colspan="2">User:</td></tr>
</table>

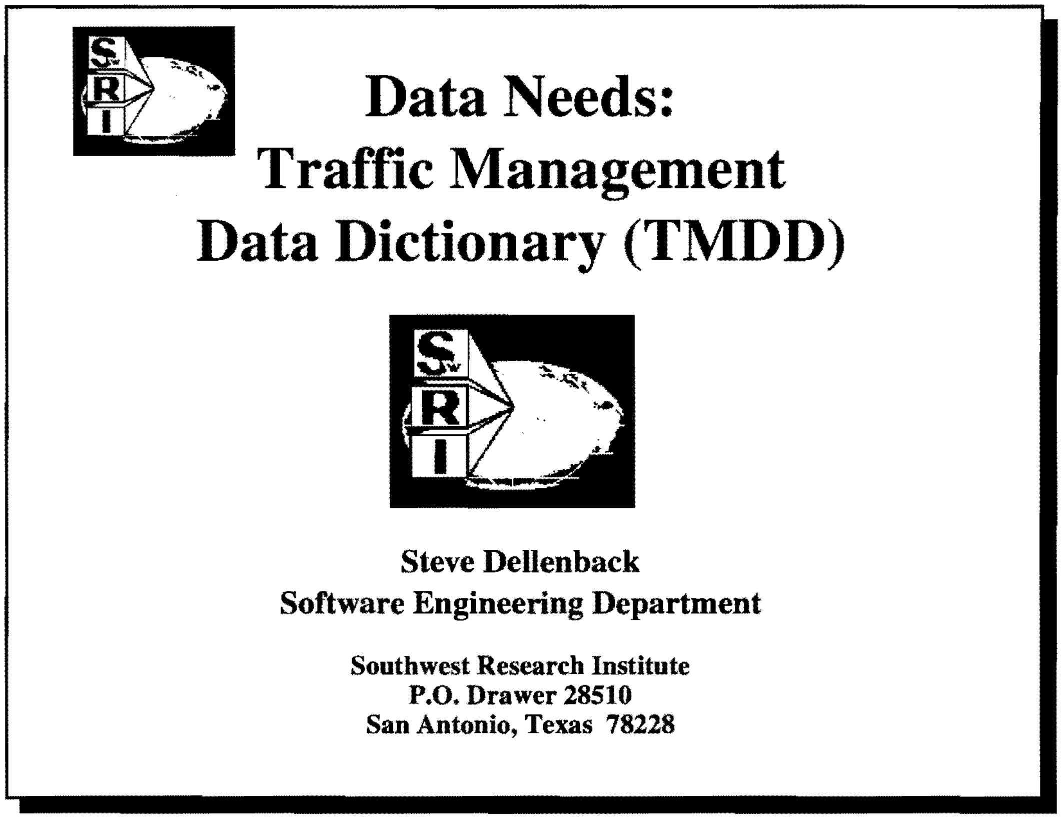
All info is excerpted from a 9/28/98 draft of the "Data Registry Functional Operating Procedures" document, posted at http://grouper.ieee.org/groups/scc32/datareg/DRFOP8.DOC.

<table>
<tr><td colspan="2">View:</td><td colspan="3">Configuration Baseline:</td></tr>
<tr><td colspan="2">Synonymous Descriptive Name:</td><td colspan="3">Synonymous Descriptive Name Context:
(One for each synonymous name)</td></tr>
<tr><td colspan="2">Symbolic Name:</td><td colspan="3">Symbolic Name Usage: (One for each name)</td></tr>
<tr><td colspan="5">Representation Layout:</td></tr>
<tr><td colspan="5">Constraints: (One for each internal name)</td></tr>
<tr><td colspan="5">ASN.1 Name:</td></tr>
<tr><td colspan="5">Remarks:</td></tr>
<tr><td colspan="5">(Five attributes below are mandatory for change proposals; otherwise reserved for Registrar use)</td></tr>
<tr><td>Data Concept Identifier:</td><td>Data Concept Version:</td><td>Registration Status:</td><td>Date Registered:</td><td>Last Change Date:</td></tr>
</table>

All info is excerpted from a 9/28/98 draft of the "Data Registry Functional Operating Procedures" document, posted at http://grouper.ieee.org/groups/scc32/datareg/DRFOP8.DOC.

DATA NEEDS: TRAFFIC MANAGEMENT DATA DICTIONARY (TMDD)

Steve Dellenback
Southwest Research Institute



The traffic management data dictionary (TMDD) has been around for about two years, and I would like to give you an overview this afternoon. Since I am a software engineer, I will probably give you a different perspective than everyone else that has talked today. First, TMDD standards are an outgrowth of the National ITS Architecture, and what we are trying to focus on is center-to-center data communications. So why are we talking about that? I think one of the most obvious uses is for corridor-level traveler information. For example, if a commuter lives in Fort Worth and works in Dallas, they need the capability to get traveler information from both cities as well as any other cities along the travel corridor. At a national level, we may be concerned about travelers driving from Houston to San Antonio to El Paso, and eventually to Los Angeles. We need to be able to share information between TMCs to be able to provide different levels of traveler information.

What Is TMDD?

- **Standard: High priority standard which is critical to successful deployment of Intelligent Transportation Infrastructure (ITI).**
- **Data dictionary: Provides a unique definition and description of the data elements used in the communication of messages and information between systems and subsystems.**
- **Method: Steering committee exists. The steering committee forwards recommendations to ITE and AASHTO for approval.**
- **Overlap: The data dictionary has overlap with other standards efforts (e.g., NTCIP, ATIS). These are being resolved at the committee level.**
- **Schedule: To be completed in early 1999.**

A data dictionary provides a unique definition and description of basic data elements. One concern at the national level is that there are several groups developing standard data dictionaries. For example, there are the ATIS and NTCIP efforts, as well as many others. I am guessing that, in the last year, half of our time in the TMDD committee was spent addressing the issue of duplicate data elements between different data dictionaries. In fact, we have trimmed out duplicate data elements and reduced the TMDD size by about 30 to 40 percent by using NTCIP data element definitions. We are still struggling to address overlap with the ATIS data dictionary committee. The issue of duplicate data elements is something that we are very concerned about and we hope that the data registry concept helps to address these problems. The TMDD effort is run by a steering committee, who then forwards recommendations to ITE and AASHTO for approval. The TMDD effort should be complete in early 1999.

Who is Supporting TMDD?

- **TMDD is sponsored by:**
  - **ITE: Institute of Transportation Engineers**
  - **FHWA: Federal Highway Administration**
  - **AASHTO: American Association of State Highway and Transportation Officials**

  GOAL: To develop a standard data dictionary.

- **MS/ETMCC is the companion program:**
  - **Message sets for external traffic management center communications**

  GOAL: To develop message sets for TMC to "ITS" centers.

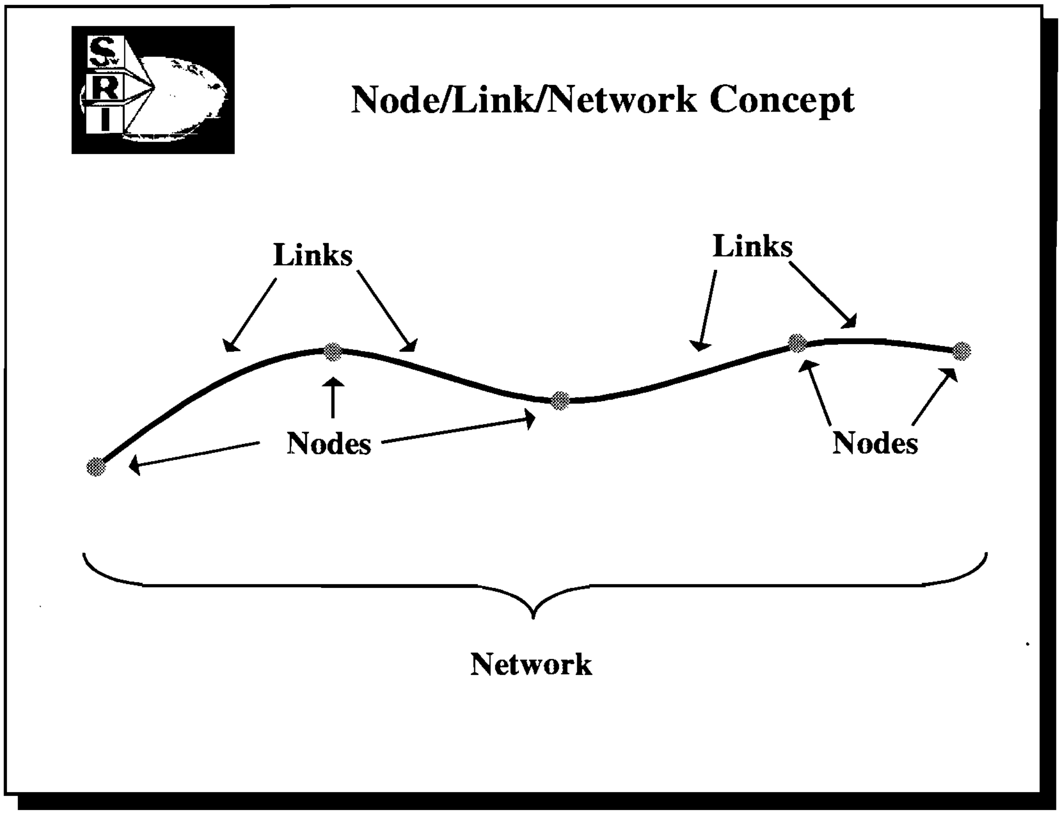
- **Effort is being carefully coordinated with NTCIP, ATIS, and IEEE efforts (many "co-committee members").**

The TMDD effort is co-sponsored by ITE, FHWA, and AASHTO, and the formal balloting is through ITE and AASHTO. An important component of the data dictionary is the message set, which logically groups data elements according to a necessary ITS function. When we refer to the TMDD, we are actually talking about the data dictionary as well as the standard message sets. Again, we are trying to coordinate with other standards efforts and working group committees.

TMDD Scope of Work

- **Section 1: Traffic Data**
  - **Links**
  - **Nodes**
  - **Traffic Data**
- **Section 2: Incidents**
  - **Incident Management**
  - **Construction**
  - **Alarm**
- **Section 3: Traffic Control:**
  - **Traffic Control**
  - **Actuated Signal Control**
  - **Traffic Detectors**
  - **Vehicle Probes**
  - **Ramp Metering**
  - **Traffic Modeling**
- **Section 4: DMS/Video/etc.:**
  - **Dynamic Message Signs**
  - **CCTV**
  - **Environmental Sensors**
  - **Gate controllers**
  - **Highway Advisory Radio**
  - **Weather Stations**
- **Message Sets**
  - Allow traffic management systems to exchange near real-time data with other transportation center based systems.

The TMDD is divided into four sections. The four main sections include 1) traffic data which consists of link, node, and network information; 2) incident information; 3) traffic control information; and 4) DMS, video, and other types of information. One of the things you will not see in TMDD, for example, is "how do I pan a camera?" That is not part of the TMDD. We are focusing more on the basic traffic information, status information, and some device information, although we will rely on NTCIP to provide a lot of that data. The other component that I have mentioned is the message set, and I will give an example of this later in the presentation.



I would like to briefly discuss and define links, nodes, and networks. The nodes represent specific points on the traffic network, links are sections of the transportation network that connect nodes, and the network is simply an interconnected series of links and nodes. If you use TranStar for an example, this diagram might represent the I-10 corridor as it crosses the city of Houston. We can use San Antonio for the same example, but in this case the nodes could represent different geographical locations.

Example TMDD Elements

- LINK_IdNumber_number
- LINK_BeginNodeId_number
- LINK_EndNodeId_number
- LINK_BeginNodeLatitude_location
- LINK_BeginNodeLongitude_location
- LINK_EndNodeLatitude_location
- LINK_EndNodeLongitude_location
- LINK_Direction_code
- LINK_Length_quantity
- LINK_Capacity_quantity
- LINK_SpeedLimit_quantity
- LINK_SpeedLimitTruck_quantity
- LINK_LanesMinimumNumber_quantity
- LINK_ShoulderWidthRight_quantity
- LINK_ShoulderWidthLeft_quantity
- LINK_MedianType_code

- LINK_PavementType_text
- LINK_RestrictionAxleCount_quantity
- LINK_RestrictionHeight_quantity
- LINK_RestrictionWidth_quantity
- LINK_RestrictionLength_quantity
- LINK_RestrictionWeight_quantity
- LINK_OversaturatedThreshold_percent

- NETWORK_IdNumber_number
- NETWORK_Name_text
- NETWORK_Jurisdiction_text
- NETWORK_LinkSetSize_number
- NETWORK_NodeSetSize_number
- NETWORK_LinkSetList_number
- NETWORK_NodeSetList_number

TMDD is sharing MANY data elements with other standards efforts!!!

This slide shows examples of several TMDD data elements. The data element names are somewhat wordy because of the naming convention. This slide was made about three months ago (before the most recent iteration/revision), and some of these data elements have since changed.

TMDD Data Element Format: Example Data Element: P1489

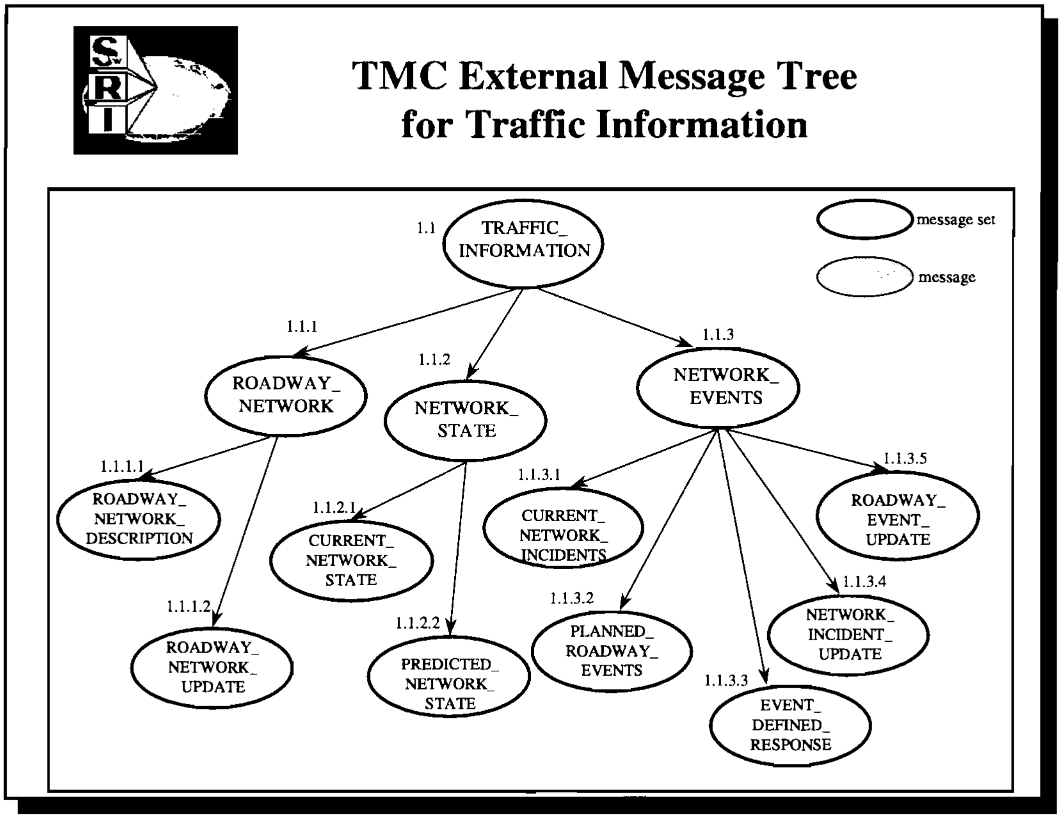
Descriptive Name	PHASE_MaximumGreen_Quantity
Descriptive Name Context	Manage Traffic
Definition	The maximum length of time that the respective phase may be held green in the presence of an opposing serviceable call.
Class Name	Actuated Signal Controller
Classification Scheme Name	IEEE P1489, Annex B
Classification Scheme Version	19971009, V0.0.7
Keywords	Phase Maximum Green
Related Data Concept	PHASE_MinimumGreen_Quantity
Relationship Type	Not Applicable
ASN1 Name	Phase-maximum-green
ASN1 Data Type	Integer
Representation Class Term	Quantity
Value Domain	ANSI NCITS.310; seconds
Valid Value Range	1 to 255
Valid Value List	Not Applicable
Valid Value Rule	Not Applicable
Internal Representation Layout	999
Internal Layout Maximum Size	SIZE(8)
Internal Layout Minimum Size	SIZE(8)
Remarks	
Data Concept Identifier	3245
Data Concept Version	V1.1
Submitter Organization Name	TMDD
Last Change Date	19880220

This slide shows an example of the defined data attributes for a single data element. Similar attributes are defined for all other data elements in TMDD. The data attributes are in the P1489 format, which is an IEEE standard for defining data elements. It has changed since then, and I have not updated my slide, so I apologize for the minor differences. For the data element shown in this example, we are looking at the maximum green phase for a traffic signal, and the attributes for this data element are defined as shown on the slide.

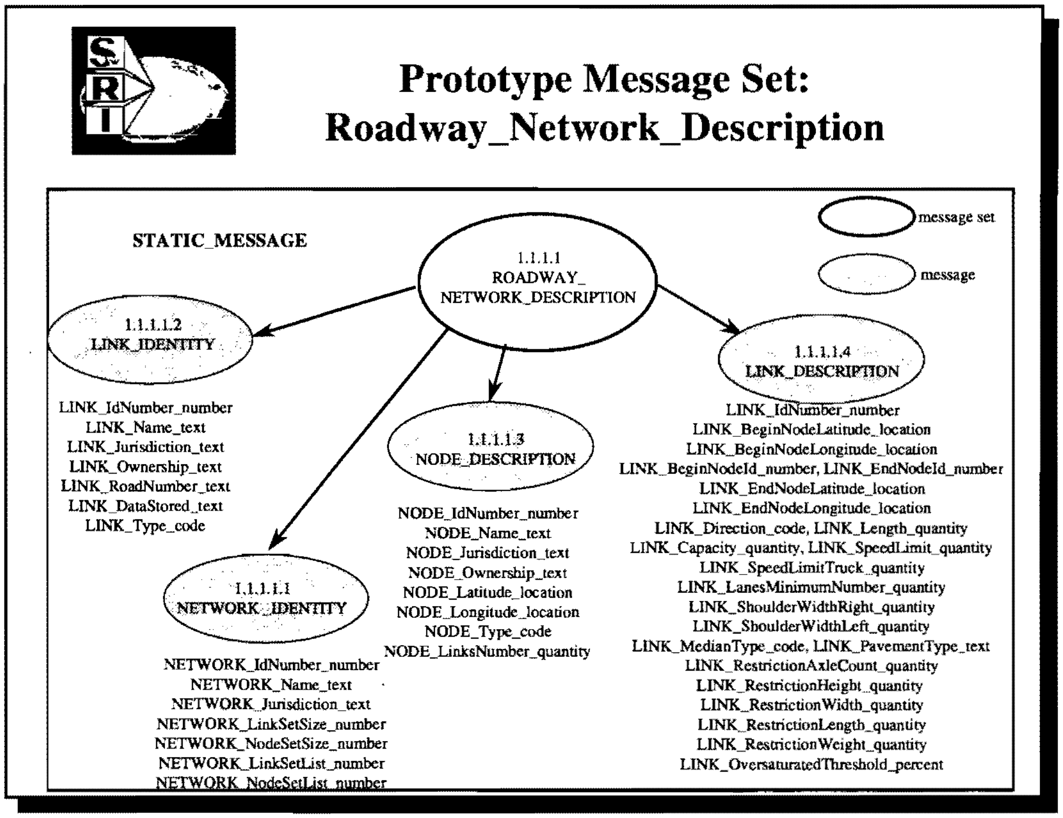
TMDD Message Sets

- Message sets are needed to communicate TMC data
- Example Messages:
  - Roadway_Network_Description
  - Roadway_Network_Update
  - Current_Network_State
  - Predicted_Network_State
  - Current_Network_Incidents
  - Planned_Roadway_Events
  - Event_Defined_Response
  - Network_Incident_Update
  - Roadway_Event_Update

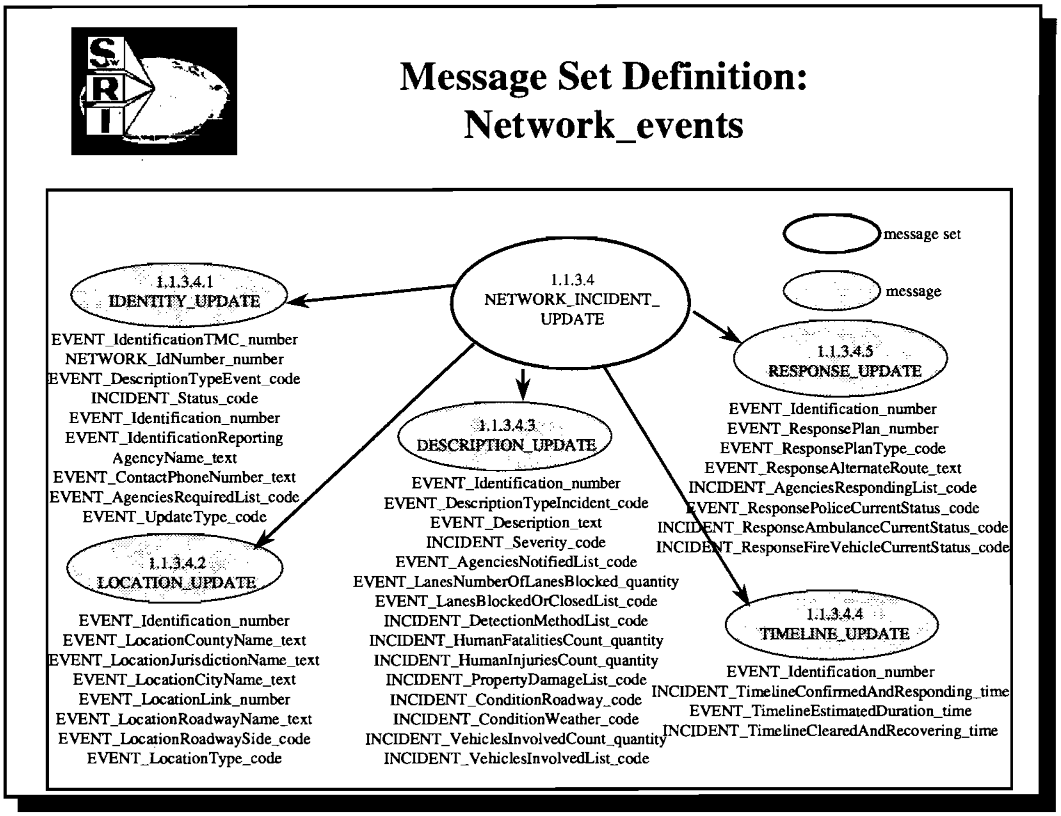
The TMDD message sets are also important. Once you have all the basic data elements and attributes defined, we take the data and combine it into message sets for transmission to the outside world.



This slide shows an example of a traffic information message tree that has been generated by a TMC. The diagram shows message sets that are being passed from the TMC. There are a ton of messages illustrated here, so we will need to go to the next lower level of detail to see actual message sets. These two diagrams show that messages are comprised of message sets, and each message set is comprised of data elements that are defined in the data dictionary. It certainly is not a week long or weekend effort by a programmer to put together these data elements and message sets, as this is a significant effort.



This slide shows an example of a roadway network description. For example, if I am sitting in San Antonio and I want a description of the Houston roadway network, I would request a roadway network description message set. This message set would provide me with a description of the network and its many links and nodes. From this message set, then, I could build a roadway network in the San Antonio system that describes the Houston roadway network. Continuing this example, if I am subscribing to their message service, I could get a constant stream of data coming into San Antonio from Houston that describes updates or changes to the network. One of the reasons for putting this complex slide up here is that I do somewhat want to overwhelm you with the level of detail that is in the data dictionary and message sets. Message sets are not small, and they are not something that you can transmit over a 1,200 baud serial modem, and you can not do it with a weekend's worth of software programming.



This slide shows an incident update for the same traffic network. This would occur every time you have an update on an incident. For example, let’s say an incident went from being reported to having emergency response on-site. They are now doing something. You would have to change the appropriate data elements and they would have to be transmitted to subscribing TMCs. The problem is, how do you get back to what goes there? Each one of these is a data element, and they are not necessarily in the TMDD. So when we put together the message set, under the TMDD umbrella, only some of these may be within the TMDD. That is why the data registry is so important.

Status

- **Section 1: Traffic Data** - **Balloted, final editing**
- **Section 2: Incidents** - **Balloted, final editing**
- **Section 3: Traffic Control** - **Balloting process initiated, ballots due December 18th**
- **Section 4: DMS/Video/etc.** - **Balloting process initiated, ballots due December 18th**
- **Message Sets** - **TMDD balloted; going out for ITE/AASHTO ballot**

This slide shows information about the status of TMDD. Sections one and two have been balloted by ITE and AASHTO, and they are in final editing and will be released soon. Sections three and four ballots are due in mid-December so we are progressing well on that. The message sets appear that they should be out in January 1999.

How Should Data Needs Be Established?

- INPUT NEEDED FROM USERS!!!!
- Requirements of end users are seriously lacking.
- TMDD is based on "experience" and "best guesses."
- The TMDD will change once it is implemented.
- Do NOT allow contractors to define data requirements!

What we really need from the data needs and data archiving perspective is input from users. You do not want your developers and contractors telling you what you need. Most of the people in this room with backgrounds in transportation areas need to tell us what you require from the data. What I have seen going around to different organizations is the requirements of the true end users (for example, sitting down with the metropolitan planning organization (MPO) and finding out how they would like to use the data).

I do not mean to offend anyone who has worked on these committees, but TMDD is based on experience and best guesses. We have designed TMDD from committees. We have never built one. We have not actually deployed it yet; therefore, we know it is going to change. Let us say that Cindy requires a contractor in Houston to build a TMDD, and then a year later we change. All of a sudden, she is no longer in compliance. When you buy into standards, you also buy into the evolution of your systems. You cannot just go put something in and say I am not going to change this for 10 years. Also, do not let me define your data requirements. I need to keep reiterating that because I do not see enough coming back from the other direction.

Data Provided by TMDD

- **Monitoring**
  - **Links, nodes, networks:**
    - **Physical description**
    - **Speed**
    - **Volume**
    - **Occupancy**
    - **Nominal**
  - **Incidents**
    - **Causes**
    - **Effects**
    - **Reactions**
  - **Equipment information (values and status)**
- **Controlling**
  - **Incidents (input to incident status)**
  - **Equipment (requests being submitted)**

When you talk about ITS data users, this is the contractor's perspective of what you need to store. This is what Cindy was talking about with her stuff in Houston. From a link perspective, you have the physical description, speed, and traffic volume. You have information about incidents and what caused the incidents. I think that is important information. Was it an accident, flat tire, or mattress in the road? What type of blockage occurred? What type of speed reduction occurred? Something that is important from my perspective, and something that we, often times, do not ever address are the reactions. In other words, what DMSs and lane control signals (LCSs) did you change? For example, in TransGuide, Shawn talked about the loop data. We also capture any time changes that we make. If we have a particular accident, we can give you a time-based historical perspective on how we changed the DMSs and LCSs to cope with that incident. That is important information to put together, especially for people trying to plan a better way to an intersection. The TMDD does not include much information about controller equipment. We talk a lot about getting information about an incident. Because what you may find is that you have a lot of different jurisdictions and different control centers, each knowing something about an incident. How do you get that information together and merged so that there is one view of the overall incident?

Final Thoughts

- **Get something on the table...**
- **Storing data is expensive (not a "traditional" transportation cost).**
- **Lack of "testing" standards will "complicate" implementation.**
- **Standards will "cost" (both dollars and schedule).**
- **Low-bid may not be the optimal procurement method for systems that are to be based on "moving" standards...**
- **A number of good standards/tools are available; time to stop "talking" and start "walking"...**

This slide shows my final thoughts. We need to get something on the table. There have been many discussions about data archiving and standards. Someone has got to go out and do it. A model deployment program does that. In San Antonio's MMDI program, we were about a year too early to effectively use the National Architecture. We do need to see another round of model deployment initiatives funded.

Storing data is expensive and it is not a traditional transportation cost. It seems every six months or every year or so there is another version of the software. Every time I get a new commercial release, it changes my commercial code. The changing software industry is very expensive, and we have to look at that. One of the big issues that is out there, besides the data registry, is the "testing" the standards. Some may interpret that standards may be based upon how much money I have. I simply throw that thought out. You know you have low-bid, you have a standard, you want to be in compliance, you want it robust, and you want an operative system. This is sometimes in direct conflict with procurement.

You have no way to validate. How do you know when you hire a contractor, that they will build the system to that standard? If you go buy some commercial package off the street, how do you know it will meet standards? What you do, is you buy two of them and see if they will "talk." Well, then, all of a sudden, it is "well, whose fault is it?" It does not matter whose fault it is, guess who is paying–you. Standards will cost. It is the bottom line. It is sometimes difficult to do long-term planning when we have short-term requirements, and we get funded for two years out. I know you have plans for what you need in 10 years, but sometimes you have to short-circuit long-term vision with the short-term reality of funds. In Austin, in the long-term, I feel that standards will save you money. In particular, at the statewide level, I know they will. But sometimes when you are the local jurisdiction trying to get something squeaked in, and then find out that standards will cost you 10 percent more, it is a tough issue.

I will also throw this thought out, that if you really want good data archiving, you cannot put ITS into low-bid construction contracts. Low-bid software gets you . . . Well, I think we all know. Finally, this is one of my comments, and one of the things we talk about with TMDD, and that is why we are glad we got something out on the streets. How long are we going to talk about the National Architecture? It seems like an eternity to me. I am really happy to see the enthusiasm in doing standardization because I think it really can help out.

SUMMARY OF BREAK-OUT DISCUSSION GROUPS

After providing the workshop participants with an overview of the different national and regional ITS data activities, discussion groups were moderated in the afternoon to focus on specific topics and issues. Two break-out discussion groups met to discuss these general topics:

- data needs and uses; and
- data warehousing technology and data standards.

A summary of the discussions and major conclusions from these two groups are presented below.

DATA NEEDS AND USES

Facilitator
Russell Henk
Texas Transportation Institute

Recorders
Bill Eisele
Texas Transportation Institute

Pete Ferrier
Texas Transportation Institute

The data needs and uses break-out group included representatives from several agencies. Participants were represented from the following groups:

- Public sector: research, metropolitan planning organization (MPO), department of transportation (DOT), and city personnel.

- Private sector: transportation consultants.

Workshop participants were supplied a copy of the FHWA report entitled, *"ITS As a Data Resource: Preliminary Requirements for a User Service."* A handout of the survey results of selected TxDOT district personnel regarding secondary uses of data gathered from ITS components was also handed out after the break-out session. It was the intent of the session facilitators to obtain as much unbiased feedback from the break-out session participants as possible. Therefore, the FHWA report was introduced and briefly discussed at the beginning of the break-out group, but the session was then quickly turned over to the audience for their feedback. The feedback from the participants can be categorized as follows:

- data needs;
- user groups;
- user applications; and
- implementation considerations.

Data Needs

The participants in the break-out session identified several data needs. These included the following:

- **Arterial street data needs.** The majority of systems that have been developed have been based upon instrumentation of the freeway system. There is a need to fill the data void for arterial streets. The arterial data must be coordinated with the freeway data. It was suggested that perhaps the National Highway System (NHS) could serve as the basis for this coordinated system.

- **Intermodal connectivity.** The point was repeatedly made that these ITS data issues are not just for roadway considerations. Rather, ITS data should be shared among all modes including roadways, airports, train/rail, and transit. For example, parking availability information at an airport that informs motorists which lots are full.

- **Crash/safety data.** Fundamental problems with the timeliness, quality/accuracy, and accessibility of crash/safety-related data were discussed. Can ITS data alleviate these concerns?

- **Rural data needs.** Data needs and issues are often considered for urban locations. However, a need was expressed for ITS data in rural locations as well.

- **Data quality.** One of the pressing questions is the quality of the data. One need is a better estimation of how good the data are that are currently being used.

- **Need for detailed and summarized elements.** Not only are there evident research needs for detailed (disaggregate) data, there is also a need for data summaries at specified levels of aggregation. For example, the suggestion was provided to summarize information for a network or city and give a "grade" for the area performance for the day. Although this would "dilute" the detail of some factors, this suggests that there is a range of uses of the data from very detailed data users to users of aggregated data.

- **Map interface needs.** The point was made that a map is necessary for visual presentation of the real-time ITS data. An Internet web page with links to weather (environmental) and video images is also desirable. The Internet provides an excellent tool by allowing a user to click on a particular section of the map and obtain the information for the desired roadway section or region.

- **System expandability.** There was a consensus that an important element of the system design include consideration of system expandability. Expandability of the system is necessary as the needs and expectations of the system will likely grow in the future.

User Groups

Many public and private user groups were identified in the break-out session. These included those shown in Table 3. It is important to note that the consideration of other users beyond the transportation field are considered in Table 3. These potential end users include other private sector entities such as consultants, research specialists, insurance groups, and other independent service providers (ISPs) that may also benefit from the ITS data produced. These potential markets should be considered because they may be willing to fund portions of an ITS database management system.

Table 3. Public and Private User Groups Identified for ITS Data Uses

Public Sector	Private Sector
• TxDOT • MPOs • Texas Natural Resources Conservation Commission (TNRCC) • Cities • Customs • Research Institutions	• Information Service Providers (ISP) • Commercial Vehicle Operations (CVO) • Insurance groups • Research

User Applications

Given the data needs and users, different types of user applications were then discussed. These included examples in several different areas including planning, construction, operations, and maintenance of transportation facilities.

Planning

A significant amount of discussion was focused on planning applications. It was noted that point source data have historically been collected for planning purposes (e.g., counts from automatic traffic recorders, tubes, or inductance loop detectors) and what may really be of interest are the data that are available from probe vehicle data collection methods (e.g., utilizing distance measuring instruments, the global positioning system, or automatic vehicle identification techniques). It was further discussed that trip data (e.g., travel time, origin-destination) are desired for planning applications, and probe vehicle data collection allow these data to be collected more easily. The comment was also made that detailed and disaggregate data are valuable in the planning profession for use in prediction equations and forecasting of air quality and emissions as well as travel characteristics.

It was also noted that for planning applications, aggregate data have historically been used. In addition, these data are often collected on a limited number of days. The group felt that a larger data source would be beneficial to provide more accurate estimates of common traffic characteristics (e.g., speed, volume, occupancy, classification, k-factors, d-factors) and performance measures (e.g., vehicle-miles of travel).

Construction

Several comments for the use of ITS data were also mentioned as they relate to roadway construction. It was mentioned that it is necessary to keep data on weekends since construction often is performed on weekends. Further, it was noted that the data that are kept must be consistently collected before, during, and after the construction to provide accurate quantitative assessment of the benefits of the project. Finally, coordination of construction must be communicated to adjacent area engineers to ensure mobility during construction phases.

Maintenance

The benefit of ITS data during a lane closure situation was discussed. Participants thought highly of the ability to have up-to-the-minute data to "catch" problems occurring at freeway sections and to take immediate and appropriate actions.

More detailed data collection was also noted as being of benefit to pavement management professionals. With more detailed data about pavement conditions, better decisions can be made about reconstruction, rehabilitation, and maintenance of pavements.

Operations

The comment was made that ITS would allow for real-time management of the transportation system. It was noted that the real-time ITS data could be used for adjusting signal timings along arterial streets and for dynamic lane control applications. It was also noted that detailed and real-time ITS data would allow for improved incident detection.

Safety

The problems associated with accurate accident data also surfaced. Concerns with accident data include timeliness, quality/accuracy, and accessibility to the data. The question was raised about how ITS data will be able to alleviate these concerns and/or automate the accident reporting procedures.

Evaluation

The evaluation of transportation improvements is often a difficult process. Further, some project impacts are relatively difficult to quantify. The need for consistent and reliable data for project evaluations was also recognized by break-out session participants.

Implementation Considerations

After discussing these user applications, the group began to discuss the implementation of an ITS data warehousing system to address these needs. The questions brought up by the audience were important for the fundamental development of an ITS data stream. The questions included:

- Where do we go from here?
- Can we do this?
- Can data warehousing help everyone?
- Will new users seek out and discover a system that is useful?
- What do we want, what do we really need, and what will it cost?

The session participants seemed hesitant that a data warehousing system that contains only summarized and/or partially aggregated data could provide for the needs of all users. There was a strong desire within the group to identify all users prior to system development. Further, it was noted that the raw data should be kept. Users could then download the data and manipulate it with their algorithms and software packages for their purposes.

DATA WAREHOUSING TECHNOLOGY AND DATA STANDARDS

Facilitator
Shawn Turner
Texas Transportation Institute

Recorders

Cesar Quiroga	Luke Albert
Texas Transportation Institute	*Texas Transportation Institute*

Participants were represented from the following groups:

- Public sector: research, metropolitan planning organization (MPO), department of transportation (DOT), and city personnel.

- Private sector: transportation consultants.

The break-out discussion was intended to be a free-ranging discussion about topics of interest of the participants. In order to enhance the discussion, the moderator would periodically propose a question to the group. The discussion topics could be grouped into the following categories:

- data storage methods
- data storage responsibility
- data access
- data standards

Data Storage Methods

One approach to storing data is archiving. Archiving involves only long-term storing of the data that have been collected, with extremely limited access. Since the cost of information storage has decreased so dramatically, the cost of archiving data is minimal.

Data warehousing is the other approach to storing data. This involves storing the data and having the ability of ready access. This is much more expensive than data archiving because in addition to the costs of storage, a database administrator needs to manage these data. It was estimated that a database administrator with between two and five years of experience would demand approximately $70,000 per year.

Data Storage Responsibility

Since data warehousing is more expensive than data archiving, an important issue is who uses the data that are warehoused. Most current ITS applications are real-time, which use only current data. Transportation planners and researchers are the primary users of warehoused data. The group seemed to believe that the state DOT should only be responsible for archiving data and that a higher level organization, such as an MPO, should warehouse data. Most ITS agencies do not have the funds to warehouse data, so warehousing is not a priority. Due to the high cost of data warehousing and the lack of funds of smaller districts, the majority opinion of the group was that there should be some type of statewide data depository for the warehoused data. Another possibility mentioned was to allow the private sector to commercialize and operate the ITS data warehouse.

Data Access

Prior to attending the workshop, most of the group were familiar with the DataLink web site, which allows users to execute queries on warehoused data. One consensus of the group was that in a data warehouse system, there should be a simple user-friendly interface to make queries easier. It was agreed that the interface on DataLink was an effective interface for users with limited computer knowledge. The consensus of the group also was that any access mechanism must be web browser based, with easy to use query tools.

There is an issue of what information needs to be available on the public side of the firewall. The data that are made available to the public should be organized inside the firewall, so that the end users only need to view the data, usually through making a query. The public should not be able to manipulate the data.

An important feature of an efficient data warehouse is allowing more than one user to access the data concurrently. It was mentioned that the end user's computer should provide the power with which to access the data when data are being accessed concurrently. If the supplier of the data also provided that power, their computers would have to be too powerful.

Another topic of discussion was the level of security of the ITS data. Due to the costs associated with making a web site secure, the DataLink site is password protected but not secure. There were also privacy issues relating to the data that were discussed. The group seemed to agree that different types of users should be allowed different levels of data access. Each user would be given a password, and the computer would allow the user certain access rights based on that password.

Another question brought up was how recent does the data being accessed need to be? The data on the DataLink site is downloaded every night, and for applications other than real-time this time of delay seemed to be sufficient.

There is also a question as to what level the data needs to be aggregated to. The data from the loop detectors are 20-second data, and queries on the DataLink site can be run for a range of levels, with the shortest being five-minute aggregation. The question of desired levels of aggregation was posed to the group, and only the researchers were interested in warehoused 20-second data. Researchers use the 20-second data to develop and calibrate models and algorithms. For the other participants, five-minute data seemed to be sufficient.

Data Standards

A final topic of discussion was the standards that are associated with ITS data. Since most people need the same types of data, it should not be too difficult to come to an agreement on standards. The initial cost of standardizing is rather expensive, but it should end up costing less in the long run. The best solution for the standards seemed to be keeping the standards rather flexible and open, hoping to include the needs of as many groups as possible. The goal for standardization proposed by Southwest Research Institute is 80 percent standardization between the groups involved.

APPENDIX: LIST OF PARTICIPANTS

Mr. Abed Abukar
Texas Department of Transportation
P.O. Box 6868
Fort Worth, TX 76115-0868
T: (817) 370-6621 ; F: (817) 370-6707
E-mail: aabukar@mailgw.dot.state.tx.us

Elidia Banda
City of San Antonio
P.O. Box 839966
San Antonio, TX 78283
T: (210) 207-6906 ; F: (210) 207-4418

Mr. Jerry Bobo
Houston-Galveston Area Council
P.O. Box 22777
Houston, TX 77227-2777
T: (713) 993-4571 ; F: (713) 993-4508
E-mail: jbobo@hgac.cog.tex.us

Mr. Darrell Borchardt
TTI-Houston
701 N. Post Oak, Suite 430
Houston, TX 77024
T: (713) 686-2971 ; F: (713) 686-5396
E-mail: d-borchardt@tamu.edu

Mr. Brian Burk
Texas Department of Transportation
125 E. 11th Street
Austin, Texas 78701-2843
T: (512) 832-7014 ; F: (512) 832-7148
E-mail: bburk@mailgw.dot.tx.us

Mr. W.R. Cairns
Mitretek Systems
600 Maryland Ave SW, Suite 755
Washington, DC 20024
T: (202) 863-2986 ; F: (202) 863-2988
E-mail: bcairns@mitretek.com

Tung-Lung Cheng
Houston-Galveston Area Council
P.O. Box 22777
Houston, TX 77227-2777
T: (713) 993-2443 ; F: (713) 993-4508
E-mail: tcheng@hgac.cog.tx.us

Steve Connell
Texas Department of Transportation
P.O. Box 6868
Fort Worth, TX 76115-0868
T: (817) 370-6784 ; F: (817) 370-6707
E-mail: sconnel@mailgw.dot.state.tx.us

Mr. Steve Dellenback
Southwest Research Institute
P.O. Drawer 28510
San Antonio, TX 78228-0510
T: (210) 737-2983 ; F: (210) 737-9310
E-mail: sdellenback@swri.org

Mr. William Eisele
Texas Transportation Institute
Texas A&M University System
CE/TTI 404F
T: (409) 845-8550 ; F: (409) 845-6008
E-mail: bill-eisele@tamu.edu

Mr. Brian Fariello
Texas Department of Transportation
P.O. Box 29928
San Antonio, TX 78284-3601
T: (210) 731-5247 ; F: (210) 731-5310
E-mail: bfariel@mailgw.dot.state.tx.us

Mr. Richard Foster
Peek Traffic Inc.
1500 N. Washington Blvd.
Sarasota, FL 34236
T: (800) 245-7660 ; F: (941) 365-0837
E-mail: fozzy@dircon.co.uk

Mr. Byron J. Gajewski
Texas Transportation Institute
Texas A&M University
CE/TTI Building Room 405D
T:(409) 845-9925 ; F: (409) 845-6008
E-mail:byrong@stat.tamu.edu

Mr. Kenneth Gathright
TNRCC
P.O. Box 13087, MC 164
Austin, TX 78711-3087
T: (512) 239-0599 ; F: (512) 239-1514
E-mail: kgathrig@tnrcc.state.tx.us

Ms. Cindy Gloyna
Texas Department of Transportation
P.O. Box 1386
Houston, TX 77251-1386
T: (713) 802-5147 ; F: (713) 802-5140
E-mail: cgloyna@mailgw.dot.state.tx.us

Mr. Harry Grothues
Southwest Research Institute
P.O. Drawer 28510
San Antonio, TX 78228-0510
T: (210) 522-2949 ; F: (210) 522-5499
E-mail: hgrothues@swri.org

Mr. Dayton Grumbles
Texas Department of Transportation - TP&P
4305 Bull Creek Rd
Austin, TX 78731
T: (512) 465-7318 ; F: (512) 465-7585
E-mail: dgrumbl@mailgw.dot.state.tx.us

Mr. Russell Henk
Texas Department of Transportation
3500 N.W. Loop 410, Suite 315
San Antonio, TX 78229
T: (210) 731-9938 ; F: (210) 731-8904
E-mail: rhenk@tamu.edu

Mr. Mark Hodges
Texas Department of Transportation - TP&P
125 E. 11th Street
Austin, TX 78701-2483
T: (512) 486-5122 ; F: (512) 486-5153
E-mail: mhodges@mailgw.dot.state.tx.us

Mr. Pat Irwin
Texas Department of Transportation
3500 N.W. Loop 410
San Antonio, TX 78229
T: (210) 731-5249 ; F: (210) 731-5310
E-mail: pirwin@mailgw.dot.state.tx.us

Ms. Sholeh Karimi
City of Arlington
P.O. Box 231, MS 01-0340
Arlington, TX 76004-0231
T: (817) 459-6361 ; F: (817) 459-6379

Mr. Al Kosik
Texas Department of Transportation
125 E. 11th Street
Austin, TX 78701-2483
T: (512) 416-3252 ; F: (512) 416-3299
E-mail: akosik@mailgw.dot.state.tx.us

Dr. Steve Liu
Texas Transportation Institute
Texas A&M University System
Bright Building, 502B
T: (409) 845-8739 ; F: (409) 847-8578
E-mail: liu@cs.tamu.edu

Dr. Tim Lomax
Texas Transportation Institute
Texas A&M University System
CE/TTI 404E
T: (409) 845-9960 ; F: (409) 845-6008
E-mail: t-lomax@tamu.edu

Mr. Douglas T. Lowe
Texas Department of Transportation
125 E. 11th Street
Austin, TX 78701-2483
T: (512) 416-3305 ; F: (512) 416-2854
E-mail: dlowe@mailgw.dot.state.tx.us

Dr. Richard Margiotta
Cambridge Systematics
1651 Ila Perdue Drive
Knoxville, TN 37931
T:(423) 670-8516 ; F: (423) 670-8517
E-mail: ram@camsys.com

Mr. Darren McDaniel
Texas Department of Transportation-TRF
125 E. 11th
Austin, TX 78701-2483
T: (512) 416-3331 ; F: (512) 416-3161
E-mail: dmcdanie@mailgw.dot.state.tx.us

Mr. Bubba Needham
Texas Department of Transportation
7901 N. Interregional Highway
P.O. Drawer 15426
Austin, TX 78761-5426
T: (512) 832-7053 ; F: (512) 832-7250
E-mail: bneedha@mail.dot.state.tx.us

Mr. Carol W. Nixon
Texas Department of Transportation
P.O. Box 1386
Houston, TX 77251-1386
T: (713) 802-5308 ; F: (713) 802-5894
E-mail: cnixon@mailgw.dot.state.tx.us

Mr. Andy Oberlander
Texas Department of Transportation
P.O. Box 3067
Dallas, TX 75221-3067
T: (214) 320-4438 ; F: (214) 320-4492
E-mail: aoberla@mailgw.dot.state.tx.us

Mr. Mark Olson
FHWA-Austin
300 E. 8th Street, Room 826
Austin, TX 78701
T: (512) 916-5966 ; F: (512) 916-5881
E-mail: mark.olson@fhwa.dot.gov

Mr. Christopher Poe
Texas Transportation Institute
Texas A&M University System
CE/TTI 411C
T: (409) 862-3558 ; F: (409) 845-6001
E-mail: cpoe@tamu.edu

Mr. Cesar A. Quiroga
Texas Transportation Institute
3500 N.W. Loop 410, Suite 315
San Antonio, TX 78229
T: (210) 731-9938 ; F: (210) 731-8904
E-mail: c-quiroga@ttimail.tamu.edu

Mr. Jeff Reding
Texas Department of Transportation - TP&P
125 E. 11th Street
Austin, TX 78701-2483
T: (512) 465-7949 ; F: (512) 465-7585
E-mail: jreding@mailgw.dot.state.tx.us

Mr. William A. Reichert, Jr.
Texas Department of Transportation
125 E. 11th Street
Austin, TX 78701-2483
T: (512) 416-3172 ; F: (512) 416-3349
E-mail: wreicher@mailgw.dot.state.tx.us

Dr. Laurence R. Rilett
Texas Transportation Institute
Texas A&M University System
CE/TTI Building Room 303D
T: (409) 845-9880 ; F: (409) 845-6481
E-mail: rilett@tamu.edu

Dr. Thomas W. Rioux
Dept. of Civil Engineering
University of Texas at Austin
Austin, TX 78712-1076
T: (512) 471-0513 ; F: (512) 475-8744
E-mail: rioux@mail.utexas.edu

Dr. Edward J. Seymour
Texas Transportation Institute
9441 LBJ Freeway, Suite 103
Dallas, TX 75243
T: (972) 994-0433 ; F: (972) 994-0522
E-mail: eseymour@tamu.edu

Mr. Rick Stalowski
Peek Traffic, Inc.
1500 N. Washington Blvd.
Sarasota, FL 34236
T: (800) 245-7660 ; F: (941) 365-0837
E-mail: pksrqeng@kudos.net

Ms. Cathy Stephens
TNRCC
P.O. Box 13087, MC 164
Austin, TX 78711-3087
T: (512) 239-1749 ; F: (512) 239-1514
E-mail: cstephen@tnrcc.state.tx.us

Mr. David Stone
City of Austin - EMS
P.O. Box 1088
Austin, TX 78767
T: (512) 469-5041 ; F: (512) 482-9407
E-mail: stone@ci.austin.tx.us

Mr. Shawn Turner
Texas Transportation Institute
Texas A&M University System
CE/TTI 405C
T: (409) 845-8829 ; F: (409) 845-6008
E-mail: shawn-turner@tamu.edu

Ms. Sally G. Wegmann
Texas Department of Transportation
P.O. Box 1386
Houston, TX 77251-1386
T: (713) 802-5171 ; F: (713) 802-5186
E-mail: swegman@mailgw.dot.state.tx.us

Mr. Joe Wilkinson
Chaparral Systems Corporation
649 Harkle Road, #E
Santa Fe, NM 87505
T: (505) 984-6255 ; F: (505) 988-5009
E-mail: joe@chapsys.com

Mr. Kenneth Zigrang
Texas Department of Transportation
P.O. Box 29928
San Antonio, TX 78284-3601
T: (210) 615-5923 ; F: (210) 615-6925
E-mail: kzigran@mailgw.dot.state.tx.us