Data Quality White Paper

Written by: Kyoungho Ahn[1], Hesham Rakha[2], and David Hill[3]

[1] Senior Research Associate, Center for Sustainable Mobility at the Virginia Tech Transportation Institute
[2] Director, Center for Sustainable Mobility at the Virginia Tech Transportation Institute
[3] Vice President, FreeAhead Inc.

Notice

Quality Assurance Statement

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-HOP-08-038		

4. Title and Subtitle **Data Quality White Paper**	5. Report Date **June, 2008** 6. Performing Organization Code
7. Author(s) **Kyoungho Ahn(1) , Hesham Rakha(1) , and David Hill(2)**	8. Performing Organization Report No.
9. Performing Organization Name and Address **(1) Virginia Tech Transportation Institute, 3500 Transportation Research Plaza, Blacksburg, VA 24061** **(2) FreeAhead Inc., 268 Golden Woods Court, Great Falls, VA 22066**	10. Work Unit No. (TRAIS) 11. Contract or Grant No. **DTFH61-06-D-00004**
12. Sponsoring Agency Name and Address **Federal Highway Administration** **Office of Operations** **1200 New Jersey Avenue, SE** **Washington, DC 20590**	13. Type of Report and Period Covered **Interim Report** **August 2007 - June 2008** 14. Sponsoring Agency Code

15. Supplementary Notes

Performed under subcontract to Cambridge Systematics Inc. (Task Order CA04-029)
Contracting Officer's Technical Representative (COTR) Mr. James Pol

16. Abstract

This paper looks at the issue of data quality within the context of transportation operations and management. The objective of this paper is to investigate data quality measures and how they are applied in existing systems. This paper explores the relevance of the data quality measures that were defined in a report entitled "Traffic Data Quality Measures" and presents an overview of the requirements for the implementation of a real-time information program.

Specifically, this paper focuses on the real-time travel information applications within six primary interfaces (traffic management information, maintenance and construction management, transit management and information, information service provider information, parking information, and emergency management information) and their associated applications as identified in the publication of "Interim Guidance on the Information Sharing Specifications and Data Exchange Formats for the Real-Time System Management Information Program".

The paper examined the quality of traffic data in existing real-time ATIS applications for both the public and private sector. The paper provides recommended data quality measures for three widely utilized traffic-related parameters, travel time, speed, and weather information. These recommendations were defined for each of the six data quality measures, accuracy, completeness, validity, timeliness, coverage, and accessibility.

17. Key Word	18. Distribution Statement
data quality, quality measures, real-time information program, accuracy, completeness, validity, timeliness, coverage, accessibility	**No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161**

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	**Unclassified**	**47**	

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

Table of Contents

List of Tables

List of Acronyms

AADT	Annual Average Daily Traffic
ADMS	Archived Data Management System
ATIS	Advanced Traveler Information Systems
AVL	Automated Vehicle Location
Caltrans	California Department of Transportation
CATT Lab	Center for Advanced Transportation Technology Laboratory
CCTV	Closed Circuit Television
FHWA	Federal Highway Administration
FZRA	Freezing Rain
GPS	Global Positioning System
HAR	Highway Advisory Radio
HOT	High Occupancy Toll
IP	Internet Protocol
ISP	Information Service Provider
ITS	Intelligent Transportation System
MAPE	Mean Absolute Percent Error
PDA	Personal Digital Assistant
PeMS	Performance Measurement System
PI	Precipitation Identifier
RADS	Regional Archived Data Server
RMSE	Root Mean Squared Error
RTSMIP	Real-Time System Management Information Program
RWIS	Roadway Weather Information System
SAE	Society of Automotive Engineers
SAFETEA-LU	Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users
SN	Snow
TMDD	Traffic Management Data Dictionary
TMS	Traffic Management System
TTE	Travel Time Estimation
TxDOT	Texas Department of Transportation
VMS	Variable Message Sign
V-SPOC	Volume, Speed, and Occupancy

Chapter 1 Introduction

Traffic data collection, within the context of transportation operation and management, is becoming an increasingly valuable asset for today's transportation arena. Significant traffic data have been generated from Intelligent Transportation Systems (ITS) technologies in recent years. The data have been widely utilized in managing system operations and providing information on traffic conditions. However, public and private users are finding that the utilization and operation of the data is an increasingly difficult task since the data are collected with different levels of accuracy and resolution, and data formats are incompatible. Furthermore, the problem worsens as the amount of data continues to grow. The quality of data in data collection, operation, and management efforts has resulted in the underutilization of data and increased utilization costs. Various problems were identified in recent research efforts regarding the quality of data for transportation operations, planning, traffic congestion information, transit and emergency vehicle management, and/or commercial truck operations. [1-6].

Data quality has been questioned since the earliest stages of traffic data collection. Since a variety of ITS applications and various travel information systems have unique data requirements, the matter of data quality has become more urgent. Furthermore, in the last few years, this intricacy has been made more complex due the emergence of private services which are providing traffic information services to the public. Turner[6] gave a definition of data quality as "*the fitness of data for all purposes that required it. Measuring data quality requires an understanding of all intended purposes for that data*". Traffic data has different meaning(s) to different consumers and the intended uses of data should be considered and understood when designing, implementing and operating data collection systems and applications.

Traditional data collection systems may not assure the quality of data that satisfy the state-of-the-art transportation applications. There are urgent needs that the specific data quality measures should be considered for each traffic data application. This paper investigates the data quality measures for transportation data and presents an overview of the requirements for the implementation of a real-time information program.

1.1 Background

Section 1201 of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU) requires the Secretary of Transportation to establish a Real-Time System Management Information Program (RTSMIP) to provide, in all states, the capability to monitor, in real-time, the traffic and travel conditions of the major highways of the United States and to share that information to improve the security of the surface transportation system, to address congestion problems, to support improved response to weather events and surface transportation incidents, and to facilitate national and regional highway traveler information. Section 1201 also requires the establishment of data exchange formats to facilitate the exchange of information.

The purposes of the RTSMIP are to (1) establish, in all States, a system of basic real-time information for managing and operating the surface transportation system, (2) identify longer range real-time highway and transit monitoring needs and develop plans and strategies for meeting the needs, and (3) provide the capability and means to share the data with State and local governments and the traveling public. RTSMIP will provide the capability to monitor the real-time traffic and travel conditions of the major U.S. highways. Furthermore, RTSMIP will share that information to improve surface transportation system security, address congestion, improve response to weather events and incidents, and facilitate national and regional highway traveler information.

This proposed program requires establishing minimum parameters and requirements for States to make available and share traffic and travel condition information via real-time information programs and also involves general uniformity among the real-time information programs to ensure consistent service to travelers and other agencies. Information sharing specifications and data exchange formats were developed by the Federal Highway Administration to accelerate the sharing of traffic and travel-condition information. Interim guidance was published to engage the transportation community on an appropriate course of action to simplify the exchange of real-time information program content.

Satisfying data quality and data accuracy requirements is a key step in the implementation of real-time information programs. The data quality should be considered in advance of developing congestion management and traveler information system applications that rely upon data from various sources. Specifically the data quality requirements should be defined by each application. For example, some applications such as High Occupancy Toll (HOT) operations and other congestion and value pricing applications require higher accuracy and more rapid availability of data in comparison to other applications.

1.2 Project Objective and Scope

The objective of this paper is to investigate the data quality measures and how they are applied in existing systems. This paper explores the relevance of the data quality measures that were defined in a report entitled "*Traffic Data Quality Measures*" and presents an overview of the requirements for the implementation of a real-time information program.

Specifically, this paper focuses on the real-time travel information applications within six primary interfaces (traffic management information, maintenance and construction management, transit management and information, information service provider information, parking information, and emergency management information) and their associated applications as identified in the publication of "Interim Guidance on the Information Sharing Specifications and Data Exchange Formats for the Real-Time System Management Information Program".

1.3 Organization of Paper

This paper is organized into six chapters. The second chapter provides a review of previous studies on data quality measures. The third chapter presents an overview of the utilization of data quality measures in public and private sectors. Chapter four discusses the data quality measures for real-time travel information applications. Chapter five investigates the proposed real-time information program. Finally, chapter six provides a summary of the findings, the conclusions of the research effort, and recommendations for further research.

Chapter 2 Literature Review of Previous Efforts

This chapter highlights the existing literature in data quality measures with an emphasis on traffic data. The subject of traffic data quality has been an issue since the earliest days of traffic data collection. The growing deployments of ITS projects across North America and worldwide require extensive evaluation of data quality issues. While new data collection technologies, data collection methods, and their analytical studies are widely used in ITS projects, relatively few studies have been conducted to evaluate the quality of traffic data. The following section discusses recently conducted research efforts regarding traffic data quality.

2.1 "Guidelines for Data Quality for ATIS Applications" (2000)

Recent research and analysis efforts have identified several issues regarding the quality of traffic data available from ITS applications for transportation operations, planning, or other functions. For example, ITS America's Advanced Traveler Information Systems (ATIS) Committee formed a Steering Committee and developed the "ATIS Data Gaps Workshop" in 2000 which identified information accuracy, reliability, and timeliness as critical to ATIS. The key findings of the "Quality Advanced Traveler Information System (ATIS) Data", are the following [7]:

- Guidelines for quality ATIS data are desirable
- Need for further refinement in classifying types of data, quality attributes for each type of data, and quality levels for each attribute
- Guidelines for quality data go beyond ATIS.

The Steering Committee identified five reasons for publishing this document:

- Raise awareness of the need for data collection planning
- Increase the amount of traffic data being collected
- Increase the quality of traffic data being collected
- Increase the recognition of the value of data
- Encourage similar efforts for traffic management, transit management, and transit-related and rural traveler information data collection

One of the earliest efforts for data quality was drafting the guidelines for Quality ATIS data. The report *Closing the Data Gap: Guidelines for Quality Advanced Travel Information System Data* provided useful insight into the required processes for data quality [7]. To establish these guidelines, two separate issues should be considered, namely: data content and data access. The data content defines the data type, coverage, and quality of the data collected while the data access issue covers availability of data to organizations for use in creating ATIS products and services. The report also states that the most frequently cited reason for insufficient data quality is inadequate geographic coverage. The data quality issues were mainly raised from incomplete data collection efforts in metropolitan areas with multiple jurisdictions, particularly with respect to traffic speeds. The report identified inadequate geographic coverage, inaccurate information, insufficient update frequency, lack of data timeliness, and inadequate spatial resolution as the most common complaints. The data guidelines defined four types of real-time traffic data: traffic sensor data, incident/event reports, images, and road/environmental sensor station data. Each data type consists of the attributes and the desired data quality levels. Table 1 illustrates traffic sensor data types, their attributes, and data quality levels.

Table 1. Attributes and Quality Levels of Traffic Sensor Data

Attributes	Quality Levels
Nature:	Limited Access Highways – Aggregated Point Data Principal Arterials – Aggregated Section Data
Accuracy	< 15% error
Confidence	Qualitative measure of suspicious data communicated along with the data
Delay	< 5 minutes
Availability	> 95% availability
Breadth of Coverage	Limited Access Highways – Major Roadways Principal Arterials – Major Roadways
Depth of Coverage	Limited Access Highways – Between Major Interchanges Principal Arterials – Between Major Arterials/Limited Access Highways

The guidelines also developed the quality levels, "good", "better", and "best" to assess the data attributes. A "good" quality level is the minimum level of data collection that should be designed for each attribute and "better" and "best" quality levels provide an improved level of service.

ATIS data guidelines are a useful indicator that offers the opportunity to enhance and improve the available ATIS data and applications. While the guidelines are limited to real-time or dynamic traffic-related information to offer traveler information services in the near-term, the guidelines provide the resources to be used for ATIS data collection. Also the guidelines need further refinement in classifying types of data, quality attributes for each type of data, and quality levels for each attribute.

2.2 "Traffic Data Quality Workshop Proceedings and Action Plan" (2003)

The quality of the traffic data and the information are critical factors since traffic and travel condition information affects the management of transportation resources and is utilized by the traveling public in making travel decisions. In 2003, FHWA designed traffic data quality workshops and developed an action plan to help stakeholders with traffic data quality issues. The workshops were designed to present the findings of three white papers in order to stimulate discussion and obtain input from the participants on how to address the concerns of traffic data quality. The three white papers are as follows:

- Defining and measuring traffic data quality
- State of the practice for traffic data quality
- Advances in traffic data collection and management

The following sections summarize the three white papers and discuss an action plan report to improve traffic data quality [6, 8-10].

2.2.1 Defining and Measuring Traffic Data Quality

The definition of data quality is a relative concept that could have different meaning(s) to different consumers. Even if data are good enough for one user; the same data might not be of acceptable quality for another consumer. Thus it is important to consider and understand all intended uses of data before attempting to measure or prescribe data quality levels. Many researchers [11-13] defined data quality as "fit for use by an information consumer", "fitness for all purposes in the enterprise processes that require it", "phenomenon of fitness for 'my' purpose that is the curse of every enterprise-wide data warehouse project and every data conversion project," and "consistently meeting knowledge worker and end-customer expectations." A white paper *Defining and Measuring Traffic Data Quality* was prepared for the traffic data quality workshops [4]. The white paper defines the data quality as "the fitness of data for all purposes that require it. Measuring data quality requires an understanding of all intended purposes for that data."

The white paper also proposed the following data quality characteristics:

- **Accuracy** – The measure or degree of agreement between a data value or set of values and a source assumed to be correct. It is also defined as a qualitative assessment of freedom from error, with a high assessment corresponding to a small error.
- **Completeness** (also referred to as **availability**) – The degree to which data values are present in the attributes (e.g., volume and speed are attributes of traffic) that require them. Completeness is typically described in terms of percentages or number of data values.
- **Validity** – The degree to which data values satisfy acceptance requirements of the validation criteria or fall within the respective domain of acceptable values. Data validity can be expressed in numerous ways. One common way is to indicate the percentage of data values that either pass or fail data validity checks.
- **Timeliness** – The degree to which data values or a set of values are provided at the time required or specified. Timeliness can be expressed in absolute or relative terms.
- **Coverage** – The degree to which data values in a sample accurately represent the whole of that which is to be measured. As with other measures, coverage can be expressed in absolute or relative units.
- **Accessibility** (also referred to as **usability**) – The relative ease with which data can be retrieved and manipulated by data consumers to meet their needs. Accessibility can be expressed in qualitative or quantitative terms.

While there are several other data quality measures that could be appropriate for specific traffic data applications, the six measures presented above are fundamental measures that should be universally considered for measuring data quality in traffic data applications. The white paper also recommended that goals or target values for these traffic data quality measures be established at the jurisdictional or program level based on a better and more clear understanding of all intended uses of traffic data. It is evident that data consumers' needs and expectations, as well as available resources, vary significantly by the implementation program and by geographic area. The facts preclude the recommendation of a universal goal or standard for these traffic data quality measures. Finally the paper also recommended including metadata in establishing data quality.

2.2.2 State-of-the-practice for Traffic Data Quality

The White Paper *State of the Practice for Traffic Data Quality* examines what operations and planning applications use traffic data and what are the quality requirements for these applications, the causes of poor quality in traffic data, quality issues specific to ITS-generated traffic data, and possible solutions to quality problems [10].

The study highlights the traffic data collection procedures by types and applications. Several types of traffic data are collected by both "traditional" and ITS means. While the basic nature and definitions of the data collected are the same, there are subtle differences in data collection methodologies that may lead to problems with data sharing and quality. For example, for planning purposes traffic volume is typically collected continuously at a limited number of sites statewide; 24-48 hour counts cover most highway segments; data are usually aggregated to hourly averages for reporting purposes. However, for many ITS applications traffic volumes are gathered continuously on every segment (1/2 mile spacing is typical on urban freeways); data often are collected at 20-30 second intervals in the field; data are aggregated and reported for later use anywhere from 20-30 seconds up to 15 minutes. The paper explores various types of data and applications with the comparisons of current (or traditional) data and ITS-generated data.

The characteristics of traffic data quality are explained defining "Bad" data. Bad or inaccurate traffic data are a result of various factors such as type of equipment, interference from environmental conditions, installation, calibration, inadequate maintenance, communication failures, and equipment breakdowns. In order to detect the bad data, a variety of methods are used, including internal range checks, cross-checks, time series patterns, comparison to theory, and historical patterns are used. Once the bad data are found, imputation appears to be most applicable where small intermittent gaps appear in the data instead of editing the measurement values. Various techniques including time series smoothing and historical growth rates have been explored while there is little consensus in the profession on what techniques to be used, or if imputation should be done at all.

The study also pointed out the difference between operational and traditional use of ITS generated traffic data. Several differences are introduced as these points: volumes vs. speeds, data quality control methods, level of accuracy, data collection nuances, data management, level of coverage, vehicle classification definitions, institutional and data sharing issues. Finally, sampling of ITS locations and data streams, shared resources, maintenance, calibration, and performance standards, contractual arrangements, more sophisticated operations applications as a data quality leader, and new technologies are recommended as possible solutions to improve traffic data quality in the study.

2.2.3 Advances in Traffic Data Collection and Management

The white paper *Advances in Traffic Data Collection and Management* identifies innovative approaches for improving data quality through quality control. The study recommends innovative contracting methods, standards, training for data collection, data sharing between agencies and states, and advanced traffic detection techniques [9]. Each methodology to improve data quality is described in this section.

The paper first introduced the innovative contracting methods that can improve data quality. A few agencies around the country have already invested resources in developing new contracting methods as a means of ensuring data quality. The study introduced the examples of Virginia and Ohio as case studies to show the potential data quality improvement though innovative contracting methods such as performance-based lease criteria for payment of data collection services and a task-order-type contract for maintenance.

The development of standards is introduced as an important aspect of traffic data quality. While standards development is still at an early stage in the United States, many European countries such as Germany, the Netherlands, and France have developed national standards for data collection equipment. All equipment purchased for national traffic data collection utilize the same formats and protocols for communication purposes. The standardization in European countries has increased the quality and accuracy of the data collected, decreased the effort needed to transfer data between agencies or offices, and increased the reliability of field equipment. However, the standardization increased the initial cost of the equipment when compared to non-standard equipment.

Training of personnel is an essential part of ensuring data quality since rapid changes and improvements of hardware and software require constant training.

Data sharing between agencies can result in cost savings and provides alternate means to meeting data quality needs. For example, the white paper demonstrated that the states of Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont have cooperated to help each other and share transportation data such as inventory, travel monitoring data, and performance data. By working together for many years, these states have improved data quality in a more efficient and cooperative environment.

Finally advanced traffic detection techniques are discussed to ensure that the data gathered are accurate. The study demonstrates that inductive loop detectors continue to effectively serve their needs. Also most failures originated from improper sealing, pavement deterioration, and foreign material in the saw slot, not because of the loop wire itself. Recent research efforts found that multi-lane detectors are most competitive from a cost and accuracy standpoint. Video imaging systems also provide an image of traffic, which is often useful in spot-checking traffic conditions.

2.2.4 Action Plan Development of Traffic Data Quality Workshops

The report defines action plans to address traffic data quality issues. The action plan presents a blueprint to address the traffic data quality based on the findings in the white papers and input received from the regional workshops [8]. The following ten priority action items were identified.

1. Develop guidelines and standards for calculating traffic data quality measures.
2. Synthesize validation procedures and rules used by various states and other agencies for traffic monitoring devices (or compilation of business rules, data validity checks, and quality control procedures).
3. Develop best practices for installation and maintenance of traffic monitoring devices.
4. Establish a clearinghouse for vehicle detector information.

5. Conduct sensitivity analyses and document the results to illustrate the implications of data quality on user applications.
6. Develop guidelines for data sharing resources for traffic monitoring activities.
7. Develop a methodology for calculating life-cycle costs.
8. Develop guidelines for innovative contracting approaches for traffic data collection.
9. Conduct a case study or a pilot test.
10. Provide guidance on technologies and applications.

The ten action items were categorized into three potential groups of activities to implement the action plan, namely: research studies, workshops, and clearinghouse case. The research studies were related to (1) the development of guidelines and standards for calculating data quality measures, (2) compilation of business rules/data validity checks and quality control procedures, (3) best practices for equipment installation and maintenance, (5) sensitivity studies to demonstrate "value of data", and (10) guidance on technologies and applications are required. Also, action items that require workshops are (6) guidelines for sharing resources, (7) life-cycle costs of detection equipment, and (8) improved contracting approaches. Finally, action items such as (4) clearinghouse for vehicle detector information and (9) case study or pilot tests should be implemented through case studies.

2.3 "Traffic Data Quality Measurement" (2004)

In order to address the most demanding recommendation from the Traffic Data Quality Workshops which is "Developing guidelines and standards for calculating traffic data quality measures", the report *Traffic Data Quality Measurement* develops methods and tools to enable traffic data collectors and users to determine the quality of traffic data that they are providing, sharing, and using. Specifically, the report presents a framework that can provide methodologies to develop and evaluate the data quality measurement for different applications. Also the report provides guidelines for developing and calculating traffic data quality.

The developed framework is based on six data quality measures, namely: accuracy, completeness, validity, timeliness, coverage, and accessibility. The framework is constructed as a sequence of steps in calculating and accessing the data quality. The structure of the framework is as follows;

- Step1. Know your customer
- Step2. Select measures
- Step3. Set acceptable data quality targets
- Step4. Calculate data quality measures for unique data
- Step5. Identify data quality deficiencies
- Step6. Assign responsibility and automatic reporting
- Step7. Complete the feedback cycle

Case studies were introduced to demonstrate how the data quality measures were calculated. Table 2 shows the traffic data quality scorecard for the Austin, TX case study. The results indicate that the quality of traffic detector data in the Austin case reasonably fits to the data quality target while only 13% of freeway sections are covered.

Table 2. Traffic Data Quality for Austin Case Study

Data Quality Measures	Original Source Data	Archive Database	Traveler Information
Accuracy • **MAPE** • **RMSE**	One-minute speeds: 12.0% 11 mph	Hourly volumes: 4.4% 131 vehicles	Travel times: 8.6% 1.56 minutes
Completeness • **Percent Complete**	Volume: 99% Occupancy: 99% Speed: 98%	Volume: 99% Occupancy: 99% Speed: 99%	Website: 100% Phone: 96%
Validity • **Percent Valid**	Volume: 99.9% Occupancy: 99.9% Speed: 99%	Volume: 97% Occupancy: 98% Speed: 99%	Route travel times: 97%
Timeliness • **Percent Timely Data** • **Average Data Delay**	99.8% 28 seconds	90% 3 hours	96% n.a.
Coverage • **Percent Coverage**	Freeways: 13% with 0.4 mile spacing	Freeways: 13% with 0.4 mile spacing	Freeways: 13% with 0.4 mile spacing; Arterials: 0%
Accessibility • **Ave. Access Time**	Archive admin.: 8 minutes; ISP: 10 minutes	Retrieve AADT values: 12 minutes avg. access time	Website: 20 second avg. access time Phone: 60 second avg. access time

* MAPE(Mean Absolute Percent Error), RMSE(Root Mean Squared Error)

The guidelines include acceptable data quality targets, level of effort required for traffic data quality assessment, specification for using metadata, and guidelines for data sharing agreements. Data quality targets are defined for different applications using six data quality measures and prepared for the acceptable quality based on the data user's needs and applications. Table 3 shows the estimated data quality targets for sample applications.

Table 3. Sample Data Quality Requirement

	Traveler information (Travel time)	Air Quality Conformity Analysis (VMT)
Accuracy	10-15% RMSE (Root mean sq. error)	10%
Completeness	95-100% valid data	At a given location 50% - two weeks per month, 24 hr
Validity	Less than 10% failure rate	Up to 15% failure rate -48 hour count, Up to 10% failure rate –permanent stations
Timeliness	Data required closed to real-time	Within 3 years of model validation year
Typical coverage	100% area coverage	75% Freeways 25% Arterials 10% Collectors
Accessibility	5-10 minutes	5-10 minutes

The guidelines also present the data sharing agreement which explains the roles, expectations, and responsibilities among the data providers and users. The data sharing agreement typically does not include the data quality specifications between the data providers and the data users. The report recommends the following three steps to add the data quality provision into data sharing agreements.

- Reporting/documenting the quality of the data
- Specifying what the quality of the data must be
- Structuring payment schedules based on amount of data passing minimum criteria

2.4 "Quality Control Procedures for Traffic Data Collection" (2006)

Quality control procedures which monitor and identify data quality problems are a critical factor in improving traffic data quality as defined in the action plans of the traffic data quality workshops. The report *Quality Control Procedures for Archived Operations Traffic Data: Synthesis of Practice and Recommendations* was prepared for the Federal Highway Administration. The report summarizes the data quality control procedures and provides recommendations for a set of quality control procedures for system-specific data quality issues [5].

Three general categories are typically utilized to indentify the validity of traffic data. The first method is univariate and multivariate range checks which set minimum, maximum, or a range of expected values for a variable or multiple variables. Secondly, spatial and temporal consistency checks are also widely utilized to check for data validity. The method evaluates the consistency of traffic data as compared to nearby locations (either across lanes, or upstream and downstream monitoring locations) or previous time periods. The third method, detailed diagnostic, requires detailed diagnostic data from traffic detectors that are not typically available in archived traffic data. This criterion can be used in diagnosing the cause(s) for poor data quality at specific detector locations.

The study also reviewed the validity checks of the nine data archives: ADMS Virginia, California PeMS, CATT Lab, Central Florida Data Warehouse, FHWA Mobility Monitoring Program, Kentucky ADMS, Phoenix RADS, PORTAL, and WisTransPortal V-SPOC and found the following problems:

- The validity criteria are similar among the nine different data archives
- The validity criteria are less sophisticated and complex than those described in the literature
- Nearly all of the validity criteria are programmed on a simple pass/fail basis
- Most of the validity criteria do not have a specified order or sequence.
- It appears that all validity criteria are applied even if previous criteria indicate invalid data

Finally, the report also provides the following recommendations for quality control procedures:

- Recognize that validity criteria (i.e., quality control) are only one part of a comprehensive quality assurance process
- Provide metadata to document quality control procedures and results
- Provide metadata to document historical traffic sensor status and configuration
- Use database flags or codes to indicate failed validity criteria
- At a minimum, implement basic foundation for data validity criteria
- Further develop other spatial and temporal consistency criteria for ADMS (Archived data management systems)
- Use visual review to supplement the automated validity criteria.

2.5 Summary of Previous Efforts

The above review of the literature has shown the significance of traffic data quality and various contributing factors that can improve the quality of traffic data. One of the foremost recommendations suggested from researchers and workshops is that there is an urgent need to develop guidelines for traffic data quality. While the previous research endeavors have attempted to build these procedures, methodologies, and guidelines, the proposed approaches are too general to satisfy the requirements for a real-time information program. The following chapters will explore the data quality measures with associated applications for a real-time information program.

Chapter 3 Use of Data Quality Measures in Existing Systems

As more traffic data have become available, the issue of data quality has become a greater concern. Furthermore, the advent of private services that provides real-time traffic information made the issue of data quality more complicated. Traditionally, the public and private sectors generally have differing points of views and consequently differing expectations with regard to traveler information systems. The public side installs and operates roadway sensors and provides traffic information, while the private side focuses on offering detailed information to travelers. However, more recently, this relationship has become more complicated in structure. ITS technologies are evolving fast and the cost of installation and operation has increased. Thus, in many ways, the fast-paced private sector shares a role in ITS projects including real-time traveler information applications. In the following section, the issues of data quality for both public and private sectors will be covered.

3.1 Public Sector Use

Much of the interest from the public sector has been to deliver an acceptable level of traveler information. Public agencies increasingly rely upon traveler information products to convey transportation system status to travelers. Data quality thus is seen as an aspect of their services that instills a sense of public confidence.

The National ITS Architecture provides a common structure for the design and implementation of ITS including traveler information applications. The National ITS Architecture defines the functions, the interfaces and information flows, and the communication requirements for the flow of information. In addition, the National ITS Architecture identifies and specifies requirements for standards needed to support national and regional interoperability and product standards. The formal definition of the physical interfaces and information exchange requirements are included in these standards. The National ITS Architecture played a key role in establishing the initial data quality standard for ITS applications [14].

511 is the nationally available telephone number that provides travelers access to traveler information in nearly every state in the country. The traveler information service being provided at the state and metropolitan area levels provides real-time traveler information using telephone and internet websites. Information quality is a major concern for each of the 511 deployments. The *Implementation and Operational Guidelines* developed by the 511 Deployment Coalition employed the accuracy, timeliness, reliability, consistency of presentation, and relevancy of information as important parameters for content quality and consistency across systems [15, 16]. The recommendation of the 511 Deployment Coalition for each attribute is as follows:

Accuracy

– Reports are recommended to contain information that match actual conditions. If the system reports construction events that are not occurring (or worse, does not report a construction event that is occurring) or a road closure is not reported, callers will start to distrust the information provided. If inaccuracies persist, callers will discontinue their use of 511.

Timeliness

– Closely related to accuracy, information provided by 511 is recommended to be timely to the greatest extent possible in accordance with the speed of changing conditions. While it is recognized that non-urban areas may have more difficulty collecting, inserting and updating information quickly, it is recommended that every attempt be made in both urban and non-urban areas to update information as soon as there is a known deviation from the current route segment or service report. Thus, the timeliest reports are based on changing conditions and not on regular interval updates.

Reliability

– Often, transportation management systems are staffed during normal working hours. But travelers use highways 24 hours a day, 7 days a week. In fact, often the most challenging travel conditions are at nighttime and on weekends. Methods must be developed to provide callers with a reliable stream of information 24/7.

Consistency of Presentation

– It is recommended that reports use the same, or similar, terminology to describe conditions. Lack of consistent terminology leads to misunderstanding and confusion amongst callers and consistent terminology will make the system more usable as users move from system to system. The use of existing and evolving standards, such as the TMDD and SAE J2354, for messages enables this consistency.

Relevancy

– The information that is provided needs to be relevant to the caller given their location, modal choice and/or actions they may need to take as a consequence of weather and road conditions or service disruptions.

Good data quality is required to provide quality information to travelers. Data sources for 511 are generally provided from the state DOT, the highway patrol and police departments, transit agencies and sometimes local jurisdictions and private companies. The traffic data are generated from direct measurement and estimation. The report *Roadway Content Quality on 511 Services* provided the recommended quality guideline for 511 as shown in Table 4 [16]. The report also included recommendations developed by Caltrans as part of its TMS Detection Plan stating that for speed the accuracy level should be within +/- 5 mph over 30 second intervals with 99% availability for speed and 95% accuracy between detection points for travel time with 95% availability.

Peirce [17] investigated the end user acceptance of traveler information using 2003 survey data that were collected in the Seattle metropolitan area. The study found that only 10% of the travelers utilize traveler information accessed via the internet websites, TV, radio, VMS and less than 1 % of travelers make a change in response to traffic information. The study also presented six factors affecting the decision to use traveler information as: the broader regional context, awareness levels, trip characteristics, information quality, the presence of delays, and the availability of alternatives. The study concluded that improved data quality and sufficient geographic coverage could increase ATIS usage with user demands.

Table 4. Data Quality Recommendations for 511

Applications	Data Quality Guideline
Traffic Data	• Data from general purpose lanes and special purpose lanes (e.g., high occupancy vehicle lanes) should not be mixed; • No more than 15% mean error in reported data (e.g., a true 60 MPH average speed being reported between 51 and 69 MPH); • No more than a five minute delay in data (e.g., data collected at 6:00 p.m. should be available on the 511 service by 6:05 p.m.); and, • Data should be available for a given road segment at least 90% of the time, on average (e.g., equipment and communications failures should result in no report being available for a road segment for no more than 876 hours throughout the course of a year).
Incident/Event Data	• No more than 10 minutes from the time an incident/event occurs to when it is available in a 511 service. • Incident/event reports are verified in some fashion prior to being included in 511 messages. • Incident/event report information (such as location, nature, severity, duration, etc.) is fully accurate in at least 85% of the reports.
Weather Data	• Conditions (fog, dust, snow, etc.): 95% accuracy and 99% availability.

3.2 Private Sector Use

While many public transportation agencies provide traffic information to travelers, private sectors are also involved in the dissemination of traveler information. Convenient accessibility of traveler information attracts travelers to use traffic information services from private firms. In addition, various public-private partnerships encourage private partners to participate in traveler information system deployment. The major roles of the private sector, as an information service provider (ISP), are to collect basic traveler information from public agencies, supplement it with additional information, process and combine it for presentation in useful ways, and use it in the derivation of information to provide added services [14].

Although the technology for generating and providing sophisticated traveler information services exists, the marketing of these services is relatively new. Internet map service providers such as Google Maps, Yahoo Maps, Mapqest.com, and Microsoft Live Search Maps provide real time traffic conditions with additional traffic information. Also private data provision services such as INRIX and Traffic.com are now being used as the principal data source in a few traffic control systems offering more complete services.

For example, Google Maps shows real-time traffic information across major US cities. Google Maps illustrates a layer that colors the roads in green, yellow, red, or gray. The colors represent how fast the traffic is moving as follows:

- Green: more than 50 mi/h

- Yellow: 25 - 50 mi/h
- Red: less than 25 mi/h
- Gray: no data available

The traffic data that are provided for major highways is aggregated from several sources including road sensors as well as car and taxi fleets. Google Maps is not the only company providing real-time online traffic data. Yahoo Maps and Mapquest.com also provide real-time traffic information services and provide symbols designating specific traffic incidents. Recently Microsoft released its latest software technology called “ClearFlow” through Live Maps in April 2008. ClearFlow which was developed using an artificial intelligence algorithm provides real-time traffic data to help drivers avoid traffic congestion including major arterials. Clearflow predicts traffic patterns, while taking into account traffic congestion, and then reflects the back ups and their consequential spill over onto city streets [18, 19].

INRIX, a private data service provider, aggregates and enhances data from hundreds of sources to provide comprehensive traffic data information including real-time reporting of traffic flow information and improved quality through proprietary error detection and correction of individual road sensors. In addition to real-time traffic information, INRIX provides a dynamic predictive flow service. INRIX traffic speed prediction algorithms includes short-term predictions (next 2-3 hours) using current traffic, weather forecasts and other metadata impacting traffic, medium and long-range predictions (days, weeks and months ahead) using weather forecasts, school, construction and event schedules, error detection and correction of real-time flow data [20].

Traffic.com also provides comprehensive real-time traffic information for major US metropolitan areas including main arterial traffic information. In addition to traffic condition information, Traffic.com offers estimated travel times based on real-time traffic conditions, incidents, construction, events, and mass transit information. The service includes an alternate drive feature which suggests an alternate route based on real-time traffic conditions including current estimated delay information when the major route is congested.

While numerous private companies provide traffic information via a number of media, including the Internet, cell phones, radio, satellite radio, and television; the accuracy of their information has not been systematically verified. Table 5 illustrates and summarizes the data quality attributes of private sector traffic information services.

Table 5. Data Quality for Traffic Information Services from Private Sectors

Data Quality Measures	Traffic Information
Accuracy	• No comparison with ground truth data • Accuracy level: Illustrates 3 to 4 levels of congestion condition (good, mild delay, congestion)
Completeness	• Percent complete 100 %
Validity	• No comparison with ground truth data • Error detection and correction by INRIX, however has not been documented
Timeliness	• Percent timely data: 100% (24 hr/7 days) • Average data delay: Updated less than 5 minutes
Coverage	• Major U.S. cities (Highways and Arterials*)
Accessibility	• Access time: Real time

* Limited service available

In many cases the algorithms used by the private sector are not made public, presumably to promote competitive advantage and brand differentiation. Verification and validation of the information products in this environment is extremely limited. Some method needs to be devised that protects the private sector investment in data quality algorithms while also providing their customers adequate assurance that the data and information products are indeed valid.

Chapter 4 Information Sharing Specifications and Data Exchange Formats

FHWA published an Interim Guidance on Information Sharing Specifications and Data Exchange Formats for the Real-Time System Management Information Program in October 2007 [21]. The real-time information program recognized under Section 1201 of SAFTEA-LU was intended to institute a standard data format for the exchange of travel- and traffic-related data between State and local government agencies and the traveling public.

While ITS standards have been developed over the past decades, non-standardized interfaces and different versions of the standards have been deployed in various ITS applications such as traffic management, transit management, and emergency management systems causing difficulties in traffic data exchange among agencies. In addition, traffic information which is available to use for transportation operators and the traveling public is not always accessible due to the lack of standard interfaces. Thus the standardization of data exchange formats is key to establishing the Real-Time System Management Information Program.

There are a variety of traveler information systems in operation today. A system may cover a single metropolitan area, an entire state, or an even larger area such as a multi-state corridor. The types of information and the modes covered can also vary widely. A system might use data from a single transportation entity, a metropolitan transit operator, multiple agencies and/or private entities.

Data for use in a traveler information system are often collected for other purposes. However, these data have a valuable second use when they are processed and packaged in forms that can be used to influence travelers' trip-making decisions. Currently, both public agencies and private organizations are providing information to travelers in many ways. In addition, technological advances are expanding travelers' options on how information can be obtained: telephone, Internet, radio, TV, variable message signs, PDAs, and more. The following sections identify the data quality specifications for a sample real-time traveler information program.

4.1 Sample Applications for Data Quality Measures

Six primary interfaces and their associated applications were defined in the Interim Guidance on Information Sharing Specifications and Data Exchange Formats. These high-level specifications were mapped from the ATIS01-Broadcast Travel Information Market Package in the National ITS Architecture. Table 6 summarizes the six primary interfaces and the associated applications. These applications are described in more detail in the following sections.

Table 6. The Six Primary Interfaces and Associated Applications

Primary Interface	Application Areas	Sample ITS Applications
Traffic Management Information	• Road network conditions • Road weather information • Traffic information coordination • Road network probe information • Traffic incidents • Air Quality data	511, VMS, traffic.com traffic report, weather warning system, VMS
Maintenance and Construction Management	• Maintenance and construction work plans • Roadway maintenance status • Work zone information	511, traffic.com roadwork report, VMS
Transit Management Information	• Emergency transit schedule information • Road network probe information • Transit and fare schedules • Transit incident information • Transit system data	511, Website Information from Transit Operators
Information Service Provider Information	• Broadcast information • Road network probe information • Traveler information • Emergency Traveler Information	511
Parking Information	• Parking locations • Parking availability	Advance parking management system, ITS truck parking service
Emergency Management Information	• Evacuation information • Disaster information	511, VMS, HAR

4.2 ITS Applications using Real-Time Information

511

511 is a nationally available traveler information service that provides pre-trip and en-route traveler information. 511 programs provide information through cell phone, an internet website, TV and radio programs, traveler information kiosks, in-vehicle devices, radios, and/or other wireless devices at local, metro, and statewide levels. Typical 511 services involve current travel information such as traffic delay/congestion, travel time, weather, roadwork, incident, transit, and/or event information allowing travelers to make better choices - choice of time, choice of mode of transportation, and choice of route of travel.

The sample data needs include:

- Delay/congestion information: travel time and speed
- Travel time information: travel time and speed
- Weather information: air temperature, visibility, and precipitation
- Roadwork/construction information: lane closure, duration, and location

- Incident Information: lane closure and estimated clearance time
- Transit Information: service disruption and service adherence

Traffic.com

Traffic.com is an independent private provider of traffic information and services in major U.S. cities. Traffic.com provides real-time traffic conditions, travel time, incident, construction, event, and mass transit information via a number of media, including the Internet, cell phones, radio, satellite radio, and television. The company utilizes four types of traffic data sources: digital traffic sensors, GPS/probe devices, commercial and government partners, and traffic operations center staff members including their own network of sensors to disseminate real-time traveler information. In addition, Traffic.com uses a variety of means such as police and fire scanners and monitoring traffic cameras to collect information and combines these to provide travelers with traffic information.

Data Needs:

- Delay/congestion information: travel time and speed
- Travel time information: travel time and speed
- Weather information: air temperature, visibility, and precipitation
- Roadwork/construction information: lane closure, duration, and location
- Incident Information: lane closure and estimated clearance time
- Transit Information: service disruption and service adherence

Real-Time Weather Information System

More accurate and accessible weather information could improve road maintenance and decrease fatal crashes. The real-time weather and road information is collected from Doppler weather radar, the national lightening detection network, a road sensor network, the agricultural weather network, satellite data, and reports from state highway personnel and state patrol officers. Wind speed and direction, cloud thickness, precipitation, air temperature, dew point and humidity, and radar depiction are/or provided for weather information and overall road conditions, pavement temperature, pavement condition (dry, wet, icy) , road dew point, road freeze point, and/or road snow depth are provided. In addition to these parameters, some agencies provide high resolution road images and/or video images to help the user to inform the driving public of real-time roadway conditions. Many RWIS systems are built up with various types of a map-based system that allows the user to select weather and pavement condition parameters for a specific area within an entire coverage area. The real-time information is typically updated every a 30 to 60 minutes.

Data Needs:

- Weather information: wind speed and direction, cloud thickness, precipitation type and intensity, air temperature, dew point and humidity, and radar depiction
- Road information: overall roadway condition, visibility or visible distance, pavement temperature, pavement condition (dry, wet, icy) , road dew point, road freeze point, and/or road snow depth

Advanced Parking Management Systems

Real-time parking information enhances mobility by avoiding parking problems and traffic congestion. The search for parking often keeps vehicles on the road needlessly and may cause lengthy queues that block adjacent streets. Slowing and stopped vehicles in travel lanes may create safety hazards. Advanced parking systems can save total travel time, improve safety, shift demand to other modes/destinations, and reduce traveler frustration and anxiety.

There are three major areas where parking information is generally in high demand – downtown areas, airports/terminals, and park-and-ride lots. Real-time en-route parking information systems provide information in the form of real-time parking availability, parking locations, and shuttle/transit service. While parking information can be provided through the Internet before a person embarks on a trip, a more common method of dissemination is through dynamic message signs that allow travelers to make a decision en-route. Entry/exit counters and space occupancy detectors are the most frequently utilized method to check the real-time parking availability [23].

Data Needs:

- Parking lot availability: entry/exit counters and space occupancy detectors
- Parking lot location: parking lot availability, parking lot location information, and transit/shuttle service

Incident Management Systems

Incident management systems can reduce the effects of incident-related congestion by decreasing the time to detect incidents, the time for responding vehicles to arrive, and the time required for traffic to return to normal conditions. A variety of surveillance and detection technologies including inductive loop or acoustic roadway detectors and camera systems providing frequent still images/ full-motion video can help detect incidents quickly. Real-time Information dissemination systems help travelers safely navigate around incidents on the roadway. While most 511 systems include a real-time incident information system, most traffic management centers also share real-time incident information with travelers through technologies deployed as part of incident management programs, such as dynamic message signs, highway advisory radio as well as internet websites.

Data Needs:

- Incident Information: lane closure, estimated clearance time, and detour information

4.3 Data Quality Measures

Real-time travel information applications require various traffic-related parameters. For instance, traffic incident applications involve the location of incidents, estimated clearance time, speed, travel time, and other parameters. Also transit service information includes routes, schedules, schedule adherence, and fare information as well as transit service information during emergency evacuation. Each application requires a unique set of traffic-related parameters, different levels of data flows, and database management. This section investigates the data quality measures of three of the most widely utilized traffic-related parameters, travel time, speed, and weather information.

4.3.1 Travel Time

The travel time data collection handbook (1998) defines travel time as “the time required to traverse a route between any two points of interest”. Accurately estimating travel time or Travel Time Estimation (TTE) is a critical component of a traveler information system. Traveler information users may alter their route of travel, change their mode, or cancel their trip based on the information provided to them. However, there is no perfectly accurate ATIS travel time estimate that computes the duration of any trip option considering different departure times. Furthermore, it is not possible to estimate perfectly accurate travel times given that some form of prediction is required [24].

Travel time is typically estimated using various mathematical models using traffic data collected from a variety of technologies. TTE is a cross-subject that needs various advanced technologies in computer-aided engineering, electronic engineering, automatic control engineering, and telecommunication engineering, all of which are applied in data collection, information transmission, and signal processing. The major technologies currently being used in TTE include loop detectors, probe vehicle technologies, license plate matching, test vehicle technologies, GPS, Automated Vehicle Location (AVL) using transponders or toll tags, cell phones as probes, and aerial surveys.

Accuracy

The accuracy of travel time depends on the data collection and analysis methods. Although each data collection method/technology has some specific advantages, each also has some particular shortcomings, causing the device or technology to work improperly under certain circumstances. For example, loop detectors have trouble measuring low-speed vehicles and only provide point time-mean speeds to estimate link travel times. Probe vehicle technology cannot provide around-the-clock traffic data collection. License plate matching is time consuming and needs large handling efforts that still cannot avoid errors. Test vehicle technology requires massive labor efforts, with slow data processing.

Furthermore, although every travel-time estimation method has its own advantage, none of the methods can provide consistently satisfying outcomes on different common cases. Currently, there is no unanimously satisfying methodology that can estimate travel time with certain accuracy that could be used in traveler information and transportation management applications. Therefore a method’s application accuracy will be questioned until it has yielded consistently from various applications. The accuracy of data collection methods is also a key aspect that has been discussed by many researchers as part of real-time travel time estimation algorithms [25-28]. For example, the data derived from inductance loop detectors are typically screened in order to enhance the quality of the data used in transportation applications. The most commonly used approach is to determine minimum and maximum acceptable values of volume, speed, and/or occupancy. Any data outside these ranges are regarded as invalid. For example, Coifman [29] introduced a method that identifies detector errors using dual loop speed trap data at both the upstream and downstream detectors in the same lane. Hablas [27] attempted to investigate the impact of detector failure frequency and failure duration on the accuracy of loop detector speed, flow, and density measurements using a Monte Carlo simulation approach and developed regression models to relate loop detector accuracy to detector failure data.

Cheevarunothai [26] presented an algorithm and its implementation for identifying and correcting loop sensitivity problems. Loop sensitivity-level discrepancies between two single loops forming a dual-loop detector and unsuitable sensitivity levels of the single loops are two major causes of quality degradation in dual-loop data. The proposed algorithm identifies dual-loop sensitivity problems using individual vehicle data extracted from loop event data and corrects dual-loop sensitivities to enhance the reliability of dual-loop detectors and improves the quality of traffic speed and bin volume data. Hellinga[30] examined the issue of the accuracy of mean travel times as estimated from probe vehicles. The study concluded that the reliability of probe-based average link travel times is highly affected by sampling bias.

While the accuracy of the travel time information is heavily dependent on the data collection technique and travel time estimation methods, Toppen [24] recommended the range of 13-21% travel time error is acceptable for ATIS applications using example applications in Los Angeles. The author concluded that an accuracy drop below a critical point deems relying on experience more efficient. Also at the highest levels of accuracy, little is gained by making further improvements and if the accuracy error is below 5%, it makes little sense to invest in improved accuracy.

Meanwhile, a recent study [31] also recommended that the acceptable accuracy range of travel time is between 10 and 20 percent for travel information applications. The study also found that if the error exceeds 20 percent, the public lose confidence in the information source, undermining the support and usefulness of the system. In addition, the application does not necessarily benefit from an increased accuracy below the specified range. The study also introduced different error ranges for different applications. For example, traffic engineering and traffic management applications requires errors between 5 and 10 percent to travel time systems, while transportation planning applications, including any type of planning or long-range monitoring activity, require a 5 to 15 percent error range for travel time data.

The report *Traffic Data Quality Measurement* proposed that the accuracy of travel times for traveler information systems be maintained within a 10 to 15 percent error range [2]. In addition, the report developed methods and tools to enable traffic data collectors and users to determine the quality of traffic data. While various studies recommend a 5 to 21 percent error range of travel time for real-time traveler information applications, the range of 10 to 17 percent travel time error would be a reasonable target for traveler information applications.

Coverage

Fujito[32] investigated the impact of sensor spacing along freeway corridors on the computation of performance measures using a travel time index. The study evaluated the effectiveness of loop detector spacing of 0.3, 0.6, 1.0, 2.0, 3.0, and 4.0 miles using data from Atlanta and 0.5, 1.0, 2.0, and 3.0 mile spacing using data from Cincinnati. The study found that increasing the sensor spacing led to over- or underestimating the travel time index relative to the baseline condition while no evidence was found that the travel time index measure became “worse” as the sensor spacing increased. However, the results appear to suggest that more sensors are not necessarily “better,” depending on the usage of the data. It does appear that, as a general rule, detector spacing of up to 1.0 mi should provide a reasonable estimate of performance measures for tracking congestion. The actual spacing between two adjacent detectors may be narrower or wider, depending on the local highway geometry (e.g., interchange locations). The study also showed that the actual location of the sensors is important in estimating the travel time index for a corridor. Thus strategically located sensors could significantly improve the performance measures of traffic collection systems.

The report *Traffic Data Quality Measurement* estimated “100 % area coverage” is recommended for the coverage measures. The coverage includes highway sections and major arterials.

Validity

TransGuide is an ATIS application which was designed by the San Antonio District of the Texas Department of Transportation (TxDOT). TransGuide provides the traveling public with real-time traveler information on traffic conditions, travel times, accidents, and construction. Turner [33] investigated the data quality analyzing loop detector data from the TransGuide system in San Antonio. The study utilized three attributes of data quality that are relevant to ITS data archiving: suspect or erroneous data, missing data, and data accuracy. The study found that in the analysis of TransGuide data, missing data accounted for about 25 percent of all data records. Error detection rules were developed to screen for suspect or erroneous data. It was concluded that data quality procedures are essential for ITS data applications.

Hablas[27] investigated the relationships between the failures of loop detector and the accuracy of loop detector measurements using a Monte Carlo simulation approach. The research concluded that the errors of performance measures such as density, flow, and space mean speed increase as the frequency and duration of failures increase. The study also developed regression models that predict the accuracy of loop detector measurements using input parameters such as the failure frequency, the failure duration, and the traffic stream density. The report *Traffic Data Quality Measurement* projected that less than 10 % of the detector failure rate is acceptable for real-time travel time information.

Timeliness, Completeness, and Accessibility

The report *Traffic Data Quality Measurement* proposed that for the timeliness measure “the data is required close to real-time”. Also Tarnoff [34] suggested delay should be less than 1 minute for local implementation and less than 5 minutes for national implementation for the timeliness measure. The reports also suggested that 95 to 100 percent coverage was required for real-time travel time information.

For the accessibility measure, the real-time travel information application can be adequately serviced with access times in the 5 to 10 minute range while predictive traffic flow methods should access the information within 30 seconds. However, the traffic information from some private service providers shows a warning sign if the traffic information is more than 5 minutes old. Also sensor networks typically update real-time speed and volume information every few minutes.

4.3.2 Speed

Traffic stream speeds are typically measured in the field using a variety of spot speed measurement technologies. The most common of these spot speed measurement technologies is a presence-type loop detector, which identifies the presence and passage of vehicles over a short segment of roadway (typically 5 to 20 meters long). When a vehicle enters the detection zone, the sensor is activated and remains activated until the vehicle leaves the detection zone.

The average traffic stream speed can be computed in two different ways: a time-mean speed and a space-mean speed. The difference in speed computations is attributed to the fact that the space-mean speed reflects the average speed over a spatial section of roadway and thus is weighted by the traffic stream density, while the time-mean speed reflects the average speed of the traffic stream passing a specific stationary point. In other words, time-mean speed is the arithmetic mean of the speeds of vehicles passing a point on a highway during an interval of time. Alternatively, the space-mean speed is the harmonic mean of the speeds of vehicles passing a point on a highway during an interval of time.

Specifically, Daganzo [35] demonstrates that the space-mean speed is a density weighted average speed, while the time-mean speed is a flow weighted average speed. The space-mean speed reflects the spatial dimension of speed and thus is utilized in the standard speed-flow-density relationships.

Accuracy

The traditional practice for estimating speeds using single loop detectors is based on the assumption of a constant average effective vehicle length. Studies, however, have shown that this assumption provides speed estimates that are sufficiently inaccurate as to severely limit the usefulness of these speed estimates for real-time traffic management and traveler information systems[36]. In addressing these issues researchers have investigated the use of filtering techniques. For example, Dailey[37] developed a Kalman filter on vehicle length estimates while Hellinga[36] used exponentially smoothed adjacent dual loop detector vehicle length measurements to enhance the speed estimates of single loop detectors. Hellinga demonstrated that the exponential smoothing of 20-s average vehicle length measurements from adjacent dual loop detectors enhanced the accuracy of the speed estimates by approximately 20 percent. Wang and Nihan [38] used screening procedures to remove intervals with long vehicles and space-mean speed estimates were derived from the intervals with passenger cars only. Alternatively, researchers have investigated the use of median as opposed to mean statistics in order to enhance the robustness of the statistics by ensuring that the measures are not influenced by outlier observations. For example, Lin [39] used the median vehicle passage time as opposed to the mean passage time to estimate speeds from single loop detectors. Similarly, Coifman[40] computed the median speed from the median occupancy in order to reduce speed estimate errors when a wide range of vehicle lengths are present in the traffic stream.

While the speed estimation is a key aspect of the accuracy of speed information for real-time traveler information, studies have recommended the speed accuracy requirements for a variety of applications. Tarnoff [34] proposed that for traveler information applications 20% or less error range is adequate for both local and national level implementations while the study recommended a 5 -10 % error range for traffic management applications. A recent study [31] recommended the following thresholds for speed accuracy:

- 1 - 5%, for traffic engineering applications
- 2 - 10%, for transportation planning applications
- 5 - 10%, for traffic management applications, and
- 5 - 20%, for traveler information applications.

In addition, Table 7 [2] summarizes the acceptable speed errors for various applications.

Table 7. Speed Accuracy Requirement for Transportation Applications

Applications	Recommended Accuracy Levels
Free Flow Link Speeds for Planning Applications	15 – 20 %
Congested Link Speeds for Planning Applications	At V/C < 1, 10 mph At V/C >=1, 2.5 mph
Transit Vehicle Speeds for Planning Applications	15 – 20 %
Free Flow Link Speeds for Traffic Simulation Applications	5.0 %
Congested Link Speeds for Traffic Simulation Applications	2.5 %
Corridor-level Vehicle Speeds for Congestion Management Applications	5.0 %

Completeness, Validity, Timeliness, Coverage, and Accessibility

The attributes of speed data is highly correlated with the travel time. Thus, similar to the real-time traveler information applications, the real-time speed information applications require more than 90% valid data, a less than 10% failure rate, real-time data, 100% area coverage, and 5-10 minute access time.

Additionally, the report *Traffic Data Quality Measurement* proposed acceptable levels of data quality for various applications, as summarized in Table 8 [2]. The table demonstrates that most speed applications require a 90 – 100 % validity check and 100 % coverage of the study area excluding transit vehicle speed applications, which require more than 95% completeness and 10 – 15% failure rates. Also the timeliness of each application varies from 6 months to three years.

Table 8. Data Quality Requirement for Transportation Applications

Applications	Completeness	Validity	Timeliness	Coverage
Free Flow and Congested Link Speeds for Planning Applications	90-100% validity for instrumented floating car data collection	90-100% validity for instrumented floating car data collection	Within three years of model validation year	100% freeway 100% major arterial 80-100% collectors 10% local road
Transit Vehicle Speeds for Planning Applications	Less than 5% - one peak and one off-peak route	Up to 15% failure rate for 48 hours counts Up to 10% failure rate for permanent counts	Within three years of model validation year	100% of study area
Free Flow Link Speeds for Traffic Simulation Applications	90-100% validity for instrumented floating car data collection	90-100% validity for instrumented floating car data collection	Within one year of study	100% of study area
Congested Link Speeds for Traffic Simulation Applications	90-100% validity for instrumented floating car data collection	90-100% validity for instrumented floating car data collection	Within one year of study	100% of study area
Corridor-level Vehicle Speeds for Congestion Management Applications	90-100% validity for instrumented floating car data collection	90-100% validity for instrumented floating car data collection	Within six month year of study	100% of study area

4.3.3 Weather Data

Low visibility, precipitation, high winds, and extreme temperature can affect driver capabilities, sight distances, vehicle performance, and infrastructure characteristics. ITS applications allow traffic managers to disseminate advisory and regulatory traveler information to motorists. These systems also facilitate sharing of road weather data among managers in multiple agencies and neighboring jurisdictions. To improve traffic operations under adverse environmental conditions, traveler information may be furnished through roadside warning systems, web-based applications, interactive telephone systems such as 511, and roadway weather information websites. The goals of the systems are to provide decision support information to travelers in a manner that may enhance efficiency and safety. One key feature in this decision support information is to improve access to real-time and forecasted weather conditions [41, 42].

511 typically provides information on current and changing travel conditions and forecasts for upcoming weather events that are likely to impact the ability to travel. Weather information for 511 on a segment-by-segment basis needs to be focused on the travel impact of weather conditions. Segments

need to be defined at a logical length to reflect the possible weather conditions and variation in conditions along segments. 511 weather service offers the following weather related information for a segment or location depending on the system location: temperature, wind speed and direction, precipitation rate, sky condition, visibility in miles and eighths of a mile once visibility is below a mile, accumulation (for snow events), air quality, and pavement temperature and condition(dry, wet, icy) [43].

Many state DOTs also provide textual and graphical road weather information on the internet [44]. The most advanced is the Washington State DOT traffic and weather information website that collects data from a variety of sources, and displays current and forecasted pavement and weather conditions on a color-coded statewide map. The DOT accesses real-time data from meteorological observing networks, a CCTV surveillance system, mountain pass reports, and various satellite and radar images. Also interactive voice response technology to provide route-specific road condition reports and six-hour weather forecasts to drivers on highways is utilized as Weather Information Systems in many other states [41]. The data quality measures of weather information are described in the following sections [43, 44].

Accuracy

Weather Information should contain information that matches actual conditions. Hourly comparison between FAA and National Weather Service weather sensor observations and road–weather sensor observations in close geographical proximity should be considered.

The accuracy of weather information is also dependent on data collection technologies such as weather sensors. Visibility is based on light scattering. The visibility sensor projects a beam of light over a very short distance, and the light that is scattered is detected by a receiver. The amount of light scattered and then received by the sensor is converted into a visibility value. A one-minute average visibility is calculated and the value is stored for the next 10 minutes. A harmonic mean is used rather than an arithmetic mean because it is more responsive to rapidly decreasing visibility conditions. The location of the visibility sensor is critical. The sensor should be located in the area of most concern. For aircraft navigation, most primary visibility sensors are placed near the touchdown zone of the primary instrumented runway. The sensor must be located at least 10 feet above ground level. Visibility is reported in quarter-mile increments up to two miles, then at 2.5 miles, then at every mile to a maximum of 10 miles. Visibilities greater than 10 miles are still reported as 10 miles. Values less than a quarter mile are reported as a quarter mile [27].

The Precipitation Identifier (PI) sensor distinguishes between rain (RA) and snow (SN) while the Freezing Rain (FZRA) sensor detects freezing rain. The Freezing Rain sensor measures accumulation rates as low as 0.01 inches per hour. The PI sensor reports data every minute as a 10-minue moving average and stores the data in memory for 12 hours. If more than three data items are missing, the algorithm reports “missing”, if an equal number of different precipitation types are reported in the last 10 minutes the heavier is reported. After the determination of the precipitation type the algorithm calculates the intensity (light-moderate-heavy) and it is determined from the highest common intensity derived from three or more samples of data [27].

Completeness
100% of weather data (24 hr/7 days) should be prepared to provide a reliable stream of information to travelers.

Timeliness
Weather information is recommended to be timely to the greatest extent possible in accordance to the speed of the weather change anticipated. In many rapidly evolving situations this can imply an hourly update with change notices of weather variations at hourly intervals. While it is recognized that non-urban areas will have more difficulty collecting, inserting and updating information quickly, it is recommended that every attempt be made in both urban and non-urban areas to update information as soon as there is a known deviation from the current route segment report.

Coverage
100% of functional weather/roadway sensors of study area should be covered.

Other attributes
It is also recommended that weather reports use the same, or similar, terminology to describe conditions. Lack of consistent terminology leads to misunderstanding and confusion and consistent terminology will make the system more usable as users move from one system to another.

4.3.4 Summary of Data Quality Measures
The previous sections described data quality measures for sample parameters including: travel time, speed, and weather information. Table 9 presents a summary of recommended data quality measures for each of the sample parameters.

Table 9. Recommended Data Quality Measures for Real-time Travel Information Applications

	Travel Time	Speed	Weather Information
Accuracy	10-17% error range for data collection and travel time estimation	5-20% error range for speed measurement or estimation	Recommended to contain information that matches actual conditions
Completeness	95-100% temporal coverage	95-100% temporal coverage	100% (24 hr/7 days)
Validity	90-100% validity for sensor or instrumented car data collection	90-100% validity for sensor or instrumented car data collection	90-100% validity for sensor
Timeliness	Less than 1 minute for local implementation and less than 5 minutes for national implementation	Real-time	An hourly update
Coverage	Sensor spacing of 1 mile and 100% area coverage	100% area coverage	100% area coverage
Accessibility	Less than 5 minutes and warning system if traffic information is more than 5 minutes old	Less than 5 minutes and warning system if traffic information is more than 5 minutes old	5-10 minutes

Chapter 5 Proposed Rule for Real-Time Information Program

On May 4, 2006, the FHWA published a notice in the Federal Register (71 FR 26399) outlining some proposed preliminary program parameters and seeking public comments on the proposed description of the Real-time System Management Information Program, including its outcome goals, definitions for various program parameters, and the current status of related activities in the States. Presented in this chapter is some of the material contained in the notice along with a summary of the responses received and the current direction the FHWA is heading with regards to implementing the requirements established in Section 1201 of the SAFETEA-LU.

5.1 May 2006 Request for Comments

The material presented in this section provides the major points that were contained in the May 2006 Request for Comments and the FHWA's proposed approach for implementing the requirements for the Real-time System Management Information Program.

5.1.1 Program Purpose

The purpose of the Real-time System Management Information Program is to provide the capability to monitor, in real-time, the traffic and travel conditions of the major highways of the United States and to share that information to improve surface transportation system security, address congestion, improve response to weather events and surface transportation incidents, and to facilitate national and regional highway traveler information.

5.1.2 Program Funding

A State may use its National Highway System, Congestion Mitigation and Air Quality Improvement program, and Surface Transportation Federal-aid program apportionments for activities related to the planning and deployment of real-time monitoring elements that advance the goals of the Real-time System Management Information Program. The FHWA has issued policy guidance, available at *http://www.ops.fhwa.dot.gov/travelinfo/resources/ops_memo.htm*, indicating that transportation system operations activities, such as real-time monitoring, are eligible under the major Federal-aid programs noted previously, within the requirements of the specific programs. State planning and research funds may also be used for activities relating to the planning of real-time monitoring elements.

5.1.3 Program Goals

By September 30, 2009, the Real-Time System Management Information Program shall:

(1) Establish, in all States, a system of basic real-time information for managing and operating the surface transportation system;

(2) Identify longer range real-time highway and transit monitoring needs and develop plans and strategies for meeting those needs; and

(3) Provide the capability and means to share the data with State and local governments and the traveling public.

Section 1201 does not specify a time frame for implementing the Real-time System Management Information Program. The FHWA proposed the implementation date of September 30, 2009, since it coincides with the expiration of the SAFETEA–LU authorization.

5.1.4 Program Outcomes

The Real-Time System Management Information Program shall result in:

(1) Publicly available traveler information Web site(s) providing access to information that is derived from the real-time information collected by the system established under the program;

(2) 511 Travel Information telephone service(s) providing to callers information that is derived from the real-time information collected by the system established under the program;

(3) Regional Intelligent Transportation System (ITS) Architectures updated to reflect the systems established under the program; and

(4) Access to the data collected by the system established under the program in an established data exchange format through standard Internet protocol (IP) communications links.

Outcomes (1) and (2) relate to commonly available methods used by public sector agencies to disseminate traffic and traveler information. Outcome (3) relates directly to a requirement in section 1201(c)(1) regarding regional ITS architectures. Outcome (4) relates to the use of common data exchange formats required by section 1201(c)(2).

5.1.5 Program Parameters

As part of describing the Real-time System Management Information Program, it is necessary to establish definitions for various parameters under the program. These parameters will define the content and context for systems developed and implemented under the program. As noted above under the program purpose, traffic and travel conditions of major highways are to be monitored in real-time. This notice proposed definitions for three principal terms used in describing the program's purpose—major highways, traffic and travel conditions, and real-time.

Major Highways

The FHWA proposed that, as a minimum, major highways to be monitored by the systems implemented under the Real-time System Management Information Program include all National Highway System (NHS) routes and other limited access roadways. In metropolitan areas, major arterials with congested travel should be included in the coverage areas of systems implemented under the Real-time System Management Information Program.

The NHS includes the Interstate Highway System as well as other roads important to the nation's economy, defense, and mobility. The NHS was developed by the DOT in cooperation with the States, local officials, and metropolitan planning organizations. More detailed information about the NHS is available from the FHWA at http://www.fhwa.dot.gov/hep10/nhs/. Because of the criteria under which the NHS was developed, it provides a sound foundation for the highways to be monitored under the program. Adding major arterials in metropolitan areas helps the program address congestion as noted in the purpose of the program.

Traffic and Travel Conditions

The FHWA proposed that the basic traffic and travel conditions to be monitored by systems implemented under the Real-time System Management Information Program include:

—Road or lane closures because of construction, traffic incidents, or roadway weather conditions;

—Roadway weather or other environmental conditions restricting or adversely affecting travel;

—Extent and degree of congested conditions, i.e., length of roadway experiencing stop-and-go or very slow (e.g., prevailing speed of traffic less than half of speed limit) traffic;

—In metropolitan areas that experience recurring traffic congestion, travel times or speeds on limited access roadways; and

—In metropolitan areas that experience recurring traffic congestion, disruptions to public transportation services and facilities.

These basic traffic and travel conditions are based on work conducted by the National 511 Deployment Coalition (Coalition) in developing its guidelines for implementing 511 travel information telephone services. The Coalition guidelines are available from the 511 Deployment Coalition at http://www.deploy511.org. In general, the minimum conditions are intended to capture events and occurrences that reduce the capacity of highways (lane closures and adverse weather conditions) or present unsafe travel conditions (congestion). In congested metropolitan areas, the minimum conditions are enhanced through the addition of travel times and transit service disruptions as a way of capturing system performance.

Real-Time

Systems implemented under the Real-time System Management Information Program will monitor and reflect current traffic and travel conditions according to the following minimum criteria:

—Construction activities affecting travel conditions, such as implementing or removing lane closures, will be available as program information within 30 minutes of the change, with changes to be available within 15 minutes in metropolitan areas with frequent or recurring traffic congestion;

—Roadway or lane blocking traffic incident information will be available as program information within 15 minutes of the incident being detected or reported and verified;

—Roadway weather conditions are updated as program information no less frequently than 30 minutes;

—Traffic congestion information will be updated as program information no less frequently than 15 minutes; and

—Travel time information, when reported and available as program information, will reflect travel conditions occurring no older than 10 minutes.

—Public transportation service disruptions, when reported, will be updated as program information no less frequently than 30 minutes.

Since the Real-time System Management Information Program applies to all States, these minimum criteria reflect systems that employ manual entry of information. Systems that use more automated or integrated information entry processes may be able to reflect changes in conditions virtually immediately. These criteria are intended to present aggressive but realistic time frames for reporting and entering information including manual entry, remotely polled sensor stations, or calculation of values. The proposed criteria also consider the usefulness of the information to travelers, hence the decreased amount of time for recording construction activities in congested metropolitan areas.

5.1.6 Information Quality

The quality of the real-time system management information depends on the techniques and technologies used to record the information. The Real-time System Management Information Program will not specify technologies or methods to be used to collect information; however, levels of quality for general attributes may be provided.

The following proposed levels of quality for two attributes are based on the report ‘‘Closing the Data Gap: Guidelines for Quality Advanced Traveler Information System (ATIS) Data’’ that is available from the DOT at http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPT_MIS/13580.html (Intelligent Transportation Society of America, ATIS Committee; September 2000).

Accuracy

Accuracy indicates how closely the recorded information matches the actual conditions. All sensors and data collection systems are subject to inaccuracies from situations such as physical obstructions, weather conditions, and radio frequency interference. The more accurate the data are, the higher the quality of information recorded by the system. This attribute is typically characterized using percentages, either as a percentage of accuracy or as an error percentage. For example, a system may be characterized as being 90 percent accurate or having a 10 percent error rate. This attribute is used to describe the average performance of the sensors or data collection system. The FHWA is considering proposing that systems implemented under the Real-time System Management Information Program are to be 85 percent accurate at a minimum, or have a maximum error rate of 15 percent.

Availability

Availability indicates how much of the data designed to be collected is made available. While sensors and data collection systems are usually designed to operate continuously, inevitably a user of the data will lose access from time to time. This attribute describes the average probability that a given data element will be available for use from a particular sensor or data collection system. For example, if a sensor records average speeds at a specific point over five minute intervals, 12 data points are generated each hour. Over the course of a year, 105,120 data points should be recorded; however, if

2,100 data points were not available for use over the course of the year, the availability would be 98 percent. This attribute essentially combines factors such as sensor or system reliability, maintenance responsiveness, and fault tolerance into a single measure related to data output. The better the traffic sensor data collection system is designed, operated and maintained, the higher the availability. The FHWA is considering proposing that systems implemented under the Real-time System Management Information Program are to have 90 percent availability at a minimum.

5.1.7 Data Exchange Formats

Section 1201(b) requires that within two years of the date of enactment of SAFETEA–LU, the Secretary of Transportation is to establish data exchange formats to ensure that the data provided by highway and transit monitoring systems, including statewide incident reporting systems, can be readily exchanged to facilitate nationwide availability of information. States shall also incorporate these data exchange formats in the systems they implement to support the Real-time System Management Information Program. If after development, the data exchange formats are officially adopted through rulemaking by the DOT, part 940 of title 23, Code of Federal Regulations, requires in section 940.11(f) that all ITS projects funded with highway trust funds shall use the applicable DOT-adopted ITS standards.

On October 15, 2007, the FHWA published a notice in the Federal Register (72 FR 58347) Interim Guidance on the Information Sharing Specifications and Data Exchange Formats for the Real-Time System Management Information Program[21].

5.2 Responses to the Request for Information

The comments submitted in response to the May 2006 Request for Comments presented in the previous section were used to develop a proposed rulemaking regarding the RTSMIP that is anticipated to be released sometime in 2008. The FHWA received a total of 44 comments to the docket, of which 22 of the submissions were from State Departments of Transportation (DOT's). Responses also were received from representatives of the private sector and national associations.

Many of the State DOT's that responded identified that they were capable of achieving many of the goals outlined in the notice by 2009, provided that there would be a phased approach for achieving key milestones. The public sector responses often cited funding limitations, budget and planning cycles, and the lack of data collection infrastructure as obstacles to fully achieving all of the program goals by a 2009 date. All of the private sector responses indicated that all of the stated objectives could be achieved by 2009 and perhaps sooner.

The private sector respondents generally felt that having the information on nearly every road, at least in urban areas, was a reasonable goal. Many State and local public sector respondents did support reporting of conditions along arterial highways, but preferred to define which ones locally. Respondents generally noted that rural and urban areas might have different needs for coverage. Several rural States noted that monitoring the National Highway System plus other limited access roadways would overwhelm their strained resources and would not necessarily improve the quality of the traffic and travel conditions reporting. One private sector respondent suggested using the same definition of "major highway" as the mapping industry.

There was general support for including travel times and speeds, as well as extent and degree of congested conditions in urban areas. Several rural States objected to the congestion requirement. Several States suggested adding expected duration for incidents, scheduled events, Homeland Security emergency notifications, maintenance work zones as well as construction work zones, hurricane evacuation, and terrorist acts. There was strong and articulate opposition from States about including information on public transportation disruptions.

There was general support for the proposed definition of "real-time" for congestion, travel time, and lane blockage information. There was no consensus among the respondents concerning the proposed thresholds for timeliness and accuracy: private sector respondents commonly suggested more stringent thresholds, some State agencies suggested weaker thresholds; some overall respondents agreed with the thresholds identified in the notice. Several respondents, including State DOTs, noted that a more stringent timeliness threshold (5 minutes or less) would be more useful to the public. A few State agencies and private sector organizations noted that they were already meeting and exceeding these proposed threshold requirements. A few States objected to the timeliness threshold requirements as inappropriate for rural areas. Several respondents noted that the timeliness threshold requirements imply either a fully automated system or a 24/7 staff, which is likely not available immediately in all areas of the country.

Overall the responses reflected reasonable support for the proposed scope of the program, with the acknowledgement that there were dissenting opinions on some details. Nearly all the respondents anticipated that the FHWA would propose a rule to establish a program to advance the level of traffic and travel conditions reporting available today. The FHWA determined to propose a rule to exercise the authority established by Congress to provide for congestion relief and to support the Department's Congestion Relief Initiative. It is expected that this proposed rule will enable various methods for mitigating the effects of recurring and non-recurring congestion by assisting agencies in providing 511 traveler information; enhancing traffic incident management; improving work zone mobility; updating and coordinating traffic signal timing; and providing localized bottleneck relief.[1]

The comments that were received in the docket that were of significant concern and are expected to be addressed by the proposed rule are in the areas: program phasing and content requirements. There was a clear preference for a phased approach in achieving the program implementation milestones. The FHWA is considering two distinct dates for establishing a real-time information program: one deployment for all Interstates within a specified date after the final rule is published in the Federal Register, and the other for non-Interstate highways in metropolitan areas by a later date from when the final rule is published in the Federal Register. The FHWA noted the interest of many public sector respondents about their preference to select the routes for traffic and travel conditions reporting.

There was wide variability in the content requirements for traffic and travel conditions reporting, especially in selecting a threshold for disseminating information after it has been collected. The FHWA

---

[1] Additional information about FHWA's focus on congestion is available at the following URL: http://www.fhwa.dot.gov/congestion/toolbox/index.htm.

considered the responses in parallel with the types of information that are needed to provide congestion relief. Based on the comments, the focus of the information to be reported centered on non-recurrent events like construction/maintenance; road closures and major delays; major special events; and, weather and road surface conditions.[2]

5.3 Transportation System Operations Enhancements Enabled by the Proposed Rule

A critical factor in the ability of transportation managers to respond effectively to a wide variety of events and situations is the availability of information that conveys the operating status of transportation facilities in real-time. Through the availability of information that improves upon today's geographic coverage, data accessibility, accuracy, and availability, transportation system operators will have the tools necessary to reduce congestion, facilitate incident management, and improve management of transportation systems assets.

Real-time information programs can be established so that States easily can exchange information on the real-time operational status of the transportation network with other States and with the private sector, value-added information market.[3] This cooperation and sharing of information could stimulate the dissemination of traffic and travel conditions that include Web or wireless access to route-specific travel time and toll information; route planning assistance using historical records of congestion by time of day; and communications technologies that gather traffic and incident-related data from a sample of vehicles traveling on a roadway and then publishing that information to travelers via mobile phones, personal digital assistants (PDAs), in-car units, or dynamic message signs.

The establishment of real-time information programs could enable the exchange of commonly applied information among public and private partners, which will stimulate national availability of travel conditions information. Real-time information programs could increase the available quantity of data for conditions prediction, expand commercial markets that broker information, provide validated and accurate data for performance measure development and reporting, and stimulate new information products that could not be achieved with present day methods.

The Real-Time System Management Information Program is focused upon making data available for a range of applications that benefit States and travelers. The proposed rule would provide a substantial foundation for the collection and gathering of data in a manner that would provide coherent use for other applications. The 511 Implementation and Operational Guidelines Version 3.0[4] (2005)

---

[2] These types of content are consistent with those documented in Implementation and Operational Guidelines for 511 Services, v.3.0 2005, available at the following URL:http://www.deploy511.org/implementationguide.htm. The guidelines were prepared by the 511 Deployment Coalition of the American Association of State Highway and Transportation Officials (AASHTO), ITS America, the American Public Transportation Association (APTA), and the USDOT to promote service consistency to help achieve a nationwide 511 system.

[3] The value-added information market creates products intended for commercial use, for sale to a customer base, or for other commercial enterprise purposes. The market may rely on information gathered by State, from other sources, or from the market's own capabilities to create the information.

[4] Available at the following URL:http://www.deploy511.org/implementationguide.htm

illustrate what detailed information from a real-time information program could be provided for other applications:

- Location – The location or portion of route segment where a reported item is occurring, related to mileposts, interchange(s) and / or common landmark(s).
- Direction of Travel – The direction of travel where a reported item is occurring.
- General Description and Impact – A brief account and impact of the reported item.
- Days / Hours and / or Duration – The period in which the reported item is "active" and possibly affecting travel.
- Travel Time or Delay – The duration of traveling from point A to point B, a segment or a trip expressed in time (or delay a traveler will experience).
- Detours / Restrictions / Routing Advice – As appropriate, summaries of required detours, suggested alternate routes or modes and restrictions associated with a reported item.
- Forecasted Weather and Road Surface Conditions – Near-term forecasted weather and pavement conditions along the route segment.
- Current Observed Weather and Road Surface Conditions – Conditions known to be in existence that impact travel along the route segment.

The extent of the proposed rule would be solely the provision of real-time information, yet the outcomes possible through this program would also reach the business of the private sector and the public sector. It is expected that the proposed rule will not be centered on a particular technology nor on a technology-dependent application. States establishing a real-time information program would be able to employ any solution chosen to make the information available. States and public agencies can enter into collaborative agreements with the private sector for establishing the program and gathering the data. States and public agencies can purchase value added information products from value added information providers. States and public agencies can apply combinations of these, and other, approaches to establish a successful real-time information program.

Chapter 6 Conclusions and Recommendations

Satisfying data quality and data accuracy requirements is a key step in the implementation of real-time information programs. Data quality is a key factor in the effectiveness of congestion management and traveler information system applications that rely on data from various sources. This white paper identified the data quality measures that should be considered within ITS applications and how they can be applied to existing systems. The study initially provided an overview of previous studies followed by a discussion of data quality issues associated with public and private sectors, providing specifications for data quality measures for sample real-time travel information applications, and finally identifying the needs of a real-time ITS program.

The paper examined the quality of traffic data in existing real-time ATIS applications for both the public and private sector. For instance, the 511 Deployment Coalition employed the accuracy, timeliness, reliability, consistency of presentation, and relevancy of information as important parameters to enhance data quality and consistency across various systems. However, in many cases the algorithms used by the private sector are not made public, presumably to promote a competitive advantage, and thus verification and validation of these systems is extremely limited.

Six primary interfaces and their associated applications were defined in the *Interim Guidance on Information Sharing Specifications and Data Exchange Formats* report. The study summarized sample applications related to real-time ATISs and their data requirements. Real-time ATIS applications require various traffic-related parameters. Each application requires a unique set of traffic-related parameters, different levels of data flow, and database management. This paper provides recommended data quality measures for three widely utilized traffic-related parameters, travel time, speed, and weather information. These recommendations were defined for each of the six data quality measures, accuracy, completeness, validity, timeliness, coverage, and accessibility.

ITS technologies are evolving fast and new data collection technologies are being deployed. In addition, data collection methods are frequently updated within ITS applications. The growing deployments of ITS projects require extensive evaluation of data quality and more extensive validation of these applications is recommended. Moreover, the development of data quality standards for different applications as well as data collection equipment will become an important aspect of real-time ATIS applications. The proposed standardization could increase the quality and accuracy of the data collected, decrease the effort needed to transfer data, and increase the reliability of field equipment.

References

1. Battelle, *Traffic Data Quality Workshop Proceedings and Action Plan*. 2003, Prepared for Federal Highway Administration: Washington D.C.
2. Battelle, *Traffic Data Quality Measurement*. 2004, Prepared for Federal Highway Administration: Washington D.C.
3. Quiroga, C., K. Hamad, R. Brydia, R. Rajbhandari, R. Benz, and S. Sunkari, *Transportation Operations Data Needs and Recommendations for Implementation*. 2007, Prepared for Texas Department of Transportation and the Federal Highway Administration: Austin, Texas.
4. Turner, S., *Defining and Measuring Traffic Data Quality: White Paper*. 2002, Prepared for Federal Highway Administration: Washington D.C.
5. Turner, S., *Quality Control Procedures for Archived Operations Traffic Data: Synthesis of Practice and Recommendations*. 2007, Prepared for Federal Highway Administration: Washington D.C.
6. Turner, S., *Defining and Measuring Traffic Data Quality: White Paper on Recommended Approaches.* Transportation Research Record, 2004(1870): p. 8.
7. ITS America and U.S. Department of Transportation, *Closing the Data Gap: Guidelines for Quality Advanced Traveler Information System (ATIS) Data*. 2000: Washington, D.C.
8. Fekpe, E. and D. Gopalakrishna, *Traffic Data Quality Workshop Proceedings and Action Plan*. 2003, Prepared for Federal Highway Administration: Washington, D.C.
9. Middleton, D., D. Gopalakrishna, and M. Raman, *Advances in Traffic Data Collection and Management: White Paper*. 2003, Prepared for Federal Highway Administration: Washington D.C.
10. Margiotta, R., *State of the Practice for Traffic Data Quality: White Paper*. 2002, Prepared for Federal Highway Administration: Washington D.C.
11. Strong, D.M., Y.W. Lee, and R.Y. Wang, *10 Potholes in the Road to Information Quality.* Computer, 1997(Aug.): p. 38-46.
12. English, L.P., *7 Deadly Misconceptions About Information Quality*. 1999, Brentwood, Tenn.: Information Impact International, Inc.
13. English, L.P., *Improving Data Warehouse and Business Information Quality*. 1999, New York: John Wiley and Sons.
14. Mitretek, *Developing Traveler Information Systems Using the National ITS Architecture*. 1998, Prepared for Federal Highway Administration: Washington D.C.
15. 511 Deployment Coalition, *America's Travel Information Number: Implementation and Operational Guidelines for 511 Services*. 2003.
16. 511 Deployment Coalition, *Deployment Assistance Report #7: Roadway Content Quality on 511 Services*. 2003.
17. Peirce, S. and J. Lappin. *Why don't more people use advanced traveler information? Evidence from the Seattle area*. *Presented at 83rd Annual Meeting of the Transportation Research Board*. 2004. Washington, D.C.
18. Wikipedia. *Live Search Maps*. 2008 [cited April 14, 2008]; Available from: http://en.wikipedia.org/wiki/Live_Search_Maps.
19. Markoff, J., *Microsoft Introduces Tool for Avoiding Traffic Jams*, in *The New York Times*. 2008: New York.
20. INRIX. *INRIX: The Leading Provider of Traffic Information*. 2008 [cited April 15, 2008]; Available from: http://en.wikipedia.org/wiki/Live_Search_Maps.

21. Federal Highway Administration, *Publication of Interim Guidance on the Information Sharing Specifications and Data Exchange Formats for the Real-Time System Management Information Program*, in *Federal Register*. 2007: Washington, D.C. p. 58347 - 58379.
22. Battelle, *Final Report of the Evaluation of the FORETELL Consortium Operational Test: Weather Information for Surface Transportation*. 2003, Prepared for U.S. Department of Transportation: Washington D.C.
23. Federal Highway Administration, *Advanced Parking Management Systems: A Cross-Cutting Study*. 2007: Washington, D.C.
24. Toppen, A. and K. Wunderlich, *Travel Time Data Collection for Measurement of Advanced Traveler Information Systems Accuracy*. 2003, Prepared for Federal Highway Administration: Washington D.C.
25. Zhang, W., *Freeway Travel Time Estimation Based on Spot Speed Measurements*, in *Civil Engineering*. 2006, Virginia Tech: Blacksburg, VA.
26. Cheevarunothai, P., Y. Wang, and N.L. Nihan, *Identification and Correction of Dual-Loop Sensitivity Problems.* Transportation Research Record, 2006(1945): p. 9.
27. Hablas, H.E., *A Study of Inclement Weather Impacts on Freeway Free-Flow Speed*, in *Civil Engineering*. 2007, Virginia Tech: Blacksburg, VA.
28. Weijermars, W.A.M. and E.C.V. Berkum, *Detection of Invalid Loop Detector Data in Urban Areas.* Transportation Research Record, 2006(1945): p. 7.
29. Coifman, B., *Using Dual Loop Speed Traps To Identify Detector Errors.* Transportation Research Record, 1999(1683): p. 12.
30. Hellinga, B. and L. Fu, *Assessing Expected Accuracy of Probe Vehicle Travel Time Reports.* Journal of Transportation Engineering, 1999. **125**(6): p. 7.
31. Tarnoff, P.J., S.E. Young, J. Crunkleton, and N. Nezamuddin, *Guide to benchmarking operations performance measures: Preliminary draft, Final report*. 2008, University of Maryland, Center for Advanced Transportation Technology: College Park, Maryland.
32. Fujito, I., R. Margiotta, W. Huang, and W.A. Perez, *Effect of Sensor Spacing on Performance Measure Calculations.* Transportation Research Record, 2006. **2006**(1945): p. 11.
33. Turner, S., L. Albert, B. Gajewski, and W. Eisele, *Archived Intelligent Transportation System Data Quality: Preliminary Analyses of San Antonio TransGuide Data.* Transportation Research Record, 2000(1719): p. 8.
34. Tarnoff, P.J., *Getting to the INFOstructure*. 2002, Prepared for TRB Roadway INFOstructure Conference.
35. Daganzo, C.F., *Fundamentals of Transportation and Traffic Operations*. 3rd ed. 2003, Oxford, UK: Elsevier Science.
36. Hellinga, B.R., *Improving Freeway Speed Estimates from Single-Loop Detectors.* Journal of Transportation Engineering, 2002. **128**(1): p. 10.
37. Dailey, D.J., *A Statistical Algorithm for Estimating Speed from Single Loop Volume and Occupancy Measurements.* Transportation Research. Part B: Methodological, 1999. **33**(5): p. 313.
38. Wang, Y. and N.L. Nihan, *Freeway Traffic Speed Estimation with Single-Loop Outputs.* Transportation Research Record, 2000(1727): p. 7.
39. Lin, W.-H., J. Dahlgren, and H. Huo, *Enhancement of Vehicle Speed Estimation with Single Loop Detectors.* Transportation Research Record, 2004(1870): p. 6.
40. Coifman, B., S. Dhoorjaty, and Z.-H. Lee, *Estimating Median Velocity Instead of Mean Velocity At Single Loop Detectors.* Transportation Research. Part C, Emerging technologies, 2003. **11**(3): p. 12.
41. Pisano, P.A. and L.C. Goodwin. *Surface Transportation Weather Applications*. in *the ITE Annual Meeting and Exhibit*. 2002. Philadelphia, PA, USA.

42. Osborne, L.F. and M.S. Owens, *Evaluation Of The Operation And Demonstration Test of Short-Range Weather Forecasting Decision Support within an Advanced Rural Traveler Information System*. 2000, Prepared for Federal Highway Administration: Washington D.C.
43. 511 Deployment Coalition, *America's Travel Information Number: Deployment Assistance Report #6, Weather and Environmental Content on 511 Services*. 2003.
44. Fayish, A.C. and P.P. Jovanis, *Usability of Statewide Web-Based Roadway Weather Information System.* Transportation Research Record, 2004(1899): p. 11.

Federal Highway Administration
U. S. Department of Transportation
1200 New Jersey Avenue, SE
Washington, DC 20590
Toll Free HelpLine 866-367-7487
www.ops.fhwa.dot.gov
Publication Number: FHWA-HOP-08-038