ITS Archived Data User Service (ADUS) Safety Applications

Review of Related Activities

DTFH61-03-X-30180

Submitted to
Larry J. Brown
FHWA, Safety CBU

Prepared by
Oscar Franzese
Oak Ridge National Laboratory

September 2005

OAK RIDGE NATIONAL LABORATORY

Table of Contents

List of Tables

List of Figures

Acknowledgements

The author gratefully acknowledges the support of the Safety Core Business Unit of the Federal Highway Administration, U.S. Department of Transportation. He also thanks Mr. Larry Brown for his guidance in the course of this work and review of early drafts of this report. The author, however, is solely responsible for any remaining deficiencies or errors.

Executive Summary

The Federal Highway Administration (FHWA) is currently sponsoring projects aimed at determining how data collected by the deployed Intelligent Transportation Systems (ITS) infrastructure can be used to improve safety on the nation's roadways. These research and development efforts include assessing the potential usage of ITS technologies to support innovative real-time ITS safety measures, developing "Safety Level" analysis methodologies that can quantify the roadway/facility hazardous conditions in real-time, and developing Archived Data User Service (ADUS) elements for safety oriented applications. The main objective of the present report, which is part of the last group, was to provide to FHWA an overall evaluation of the progress being made on the development of the ITS ADUS standards, as well as to report on ITS data archiving related projects and their impacts on the safety applications.

In an attempt to understand how evolving standards and guidelines may potentially affect current and future safety applications, this report presents a review of the current state-of-the-art safety methodologies, focusing on the type of data required to develop such applications. In general, these studies have as an objective the real-time assessment of the likelihood of a crash occurring on a given segment of a freeway[1]. Their methodologies take as input, real-time traffic information collected through roadside sensors, as well as other types of information to produce traffic-related hazardous condition assessments. However, the development and testing of these applications rely on archived data, particularly traffic condition information, crash data, weather information, and roadway geometric characteristics. Therefore, the applications reviewed in this report provide guidelines on what type of data, and at which spatial and temporal resolution, is needed to develop real-time safety models and applications. In the most demanding case, the literature shows that data needs to be archived with a resolution of at least 20 seconds, for at least 30 minutes previous to a crash occurrence, not only for the day in which the crash occurs, but also for the same day of the week when no incident occurs, and for sensors upstream and downstream of the crash site.

The National ITS Architecture provides general guidelines regarding Archived Data and other user services. The implementation of these services, however, requires normative procedures and specific guiding principles to assure compatibility and interoperability, which are usually achieved through the use of standards. In the case of ADUS, the development of the standards was assigned to the ASTM ADUS Subcommittee (ASTM E17.54) under the ASTM Committee on Vehicle - Pavement Systems (ASTM E17). As of July 2005, the ADUS Subcommittee has published one guideline –the *Standard Guide for Archiving and Retrieving Intelligent Transportation System-Generated Data* (ASTM E2259-03a)– and is working on two other standards –the *Standard Practice for Metadata to Support Archived Data Management Systems* and the *Standard Specifications for Archiving ITS-Generated Traffic Monitoring Data*. The ASTM E2259-03a document (which is not a "true" standard) encourages best practices in the development of information archiving systems, as well as providing guidelines for data maintenance and quality control. The metadata standard, a "true" standard which is almost ready for publication, describes archived data characteristics such as its content, quality, organization, availability, lineage, and other attributes; but not the data itself. The Traffic Monitoring Standard (also a "true" standard) describes not only data elements but also the structure for an archived data management system, including conventionally gathered traffic data, data collected directly from ITS systems, and travel time data obtained from probe vehicles.

---

[1] At the present time, there is no published study dealing with real-time crash prediction at intersections, even though these types of incidents account for about 30% of all crashes. To remedy this research gap, the Office of Safety at FHWA is currently sponsoring a project aimed at investigating this type of problem.

Although ADUS standards are currently being developed, many traffic operation centers and other organizations have a real and practical need to archive data, and are developing their own systems. Some of these on-going data archiving projects will undoubtedly play a significant role in the development of the ADUS standards. Appendix A of this report presents a detailed update on several on-going data archiving projects, including descriptions of the Virginia ADMS (Archive Data Management Systems), the iFlorida model deployment project, the Cincinnati-Northern Kentucky ARTIMIS (Advanced Regional Traffic Interactive Management and Information System) system, and the California PeMS (Performance Measurement System). The review of these systems indicates that typical traffic information (i.e., volume, occupancy, speed) is archived at very fine temporal resolutions –20 to 30 seconds–, although some systems do this only for snapshots. The aggregation levels range from 1 minute (VA ADMS), to 5 minutes (PeMS), to 15 minutes (ARTIMIS). Spatially, the information is aggregated at the detector and station levels.

As a complement to the information offered by these ADMS systems, this report includes in Appendix B a summary and analysis of the data archiving practices of freeway and arterial management centers across the 78 major metropolitan areas of the nation that participated in the 2002 ITS Deployment Tracking Survey. That survey indicated that 41.4% of the freeway management agencies interviewed, and 12.6% of the arterial management agencies had in place an ADMS system as described in the National ITS Architecture. Information about the type of data archived by those agencies, the level of aggregation, and the quality assurance procedures is also included in Appendix B.

Based of the lessons learned from the review of state-of-the art safety applications, and taking into account the current status of the ADUS standards, particularly the traffic monitoring standard, this report includes several recommendations aimed at making this standard compatible with the development of future safety methodologies and applications.

Traffic monitoring data should to be archived at the finest granularity provided by the sensors to make possible the calculation of the variability, and averages, of traffic parameters used as crash precursors during time intervals centered on the time of the crash. This fine resolution also permits in some cases (e.g., freeways) to determine the time of crash through the use of traffic flow theory methodologies.

The raw sensor data should be archived for a minimal time allowing the collection of statistically significant safety data information. The development of safety applications requires the knowledge of traffic conditions at the location of a crash not only around the time of its occurrence, but also on the same day of the week, and at the same time of day when there were no accidents at that location. This requires a minimum archiving time that could extend for several months, or even longer in areas where there are no incident detection systems deployed (i.e., almost everywhere except on some urban freeways) if the police accident reports do not become immediately available.

The spatial distribution of traffic information should be maintained in the archive. As in the case of temporal distribution of the traffic information, its spatial distribution also plays a key role in the development of safety applications, especially those geared towards real-time hazardous condition identification. Spatial distribution of the traffic information along the direction of travel allows the determination of this space-related variability of traffic parameters (e.g., speed differentials between consecutive detector stations) which are used by some safety applications as crash predictors. The archiving of disaggregated traffic information is also important for the development of new safety technologies that provide collision countermeasure systems (CCS), and may rely on detectors located at different points upstream of surface street intersections. Aggregating the information provided by these spatially distributed detectors would hamper the development of these applications.

The standards should be able to archive new traffic parameters, such as travel time generated from pseudo probe-vehicles. The latest development of real-time safety applications was only possible due to the recent availability of high resolution (temporally and spatially) traffic parameters such as speed, volume, and density. At the present time, this “traditional” traffic information is the only type of data that the ITS infrastructure collects. However, new technologies (e.g., license-plate readers, new generation of inductive detectors that can provide very detailed vehicle signatures, and the US DOT Vehicle-Infrastructure Integration initiative, VII) are being, or will be, deployed that supply other types of traffic information (e.g., travel time), which in the past were either not available or not very accurate. This type of information, once it becomes readily available, would play a very important role in the development of new safety applications. In the present version of the traffic monitoring standard, detector information and probe-vehicle measurements are not linked, simply because the latter do not rely on any roadside sensors. However, in the case of pseudo probe-vehicles, the standard should be expanded to add a relationship between the probe-vehicles and the infrastructure sensors (license plate readers, inductive sensors, VII beacons) that make the traffic parameters measurements to, among other considerations, determine the quality of the information obtained.

Shortly, the ADUS Subcommittee will start developing new standards for transit and incident management. The latter, which has critical implications for the development of safety applications, may be based on the NHTSA Model Minimum Uniform Crash Criteria (MMUCC) Guideline (2nd Edition, 2003). From the view point of the development of safety applications, the exact location and time of the crash are perhaps the most important pieces of information (together with type of crash and number and type of vehicles involved). The other key requirement is the ability to link the data to other sources of information such as traffic monitoring data, roadway geometry, and weather information.

ADUS Safety Applications

Review of Related Activities

Introduction

The Federal Highway Administration (FHWA) is currently sponsoring projects aimed at determining how the wealth of data collected by the deployed Intelligent Transportation Systems (ITS) infrastructure can be used to improve safety on the nation's roadways. These research and development efforts include assessing the potential usage of ITS technologies to support innovative real-time ITS safety measures, developing "Safety Level" analysis methodologies that can quantify the roadway/facility hazardous conditions in real time, and developing Archived Data User Service (ADUS) elements for safety oriented applications.

The main objective of this report is to provide to FHWA an overall evaluation of the progress being made on the development of the ITS ADUS standards, as well as to report ITS data archiving related projects and their impacts on safety applications. ADUS, which was incorporated into the National ITS Architecture in 1999, is perhaps the user service with the highest number of stakeholders. The *ITS Data Archiving Five-Year Program Description* (FHWA, 2000) identifies fourteen stakeholder groups having an interest in the use of the data generated by ITS devices and infrastructure; with safety planners and administrators being one of these groups. As a result, the development of ADUS standards is influenced by many parties that may or may not share a common interest with the needs and requirements of the transportation safety community. Therefore, to understand how these evolving standards and guidelines may potentially affect current and future safety applications, it is necessary to analyze what type of data (including granularity, spatial distribution, and temporal distribution) the current state-of-the-art safety methodologies are using. The next section of this report presents an overview of recently published research that focuses on the application of real-time data in determining safety conditions on roadways. The data required by these methodologies provides a benchmark against which it is possible to assess whether or not the ADUS standards will be able to satisfy the data requirements of safety applications.

The third section reports on the status of ADUS standards development. Although no ADUS standards have been published yet[2], there is a real and practical need to archive data within the transportation community, and hence many traffic and freeway management centers are developing their own archiving systems. Some of

[2] The ASTM sub-committee E17.54 recently published the *Standard Guide for Archiving and Retrieving Intelligent Transportation Systems-Generated Data*, which is not a standard, but a guideline.

these efforts are sponsored by the FHWA and will have a significant impact on the development of the ADUS standards. Appendix A of this report describes several of these systems, while Appendix B includes an analysis of the information contained in the 2002 ITS Deployment Tracking database. This project gathers detailed traffic related information for the largest 78 metropolitan areas, including information related to data archiving for freeway and arterial systems. The analysis, therefore, gives an overview of the current data archiving practices across the nation. Appendix C complements the information in section three by describing other activities that the ASTM Committee E17.54 (i.e., the ADUS Committee) is currently pursuing.

Finally, the last section of this report presents a summary of the findings and recommendations on data archiving and the development of safety applications.

Review of Safety Applications Research

One of the important goals of this report is to provide FHWA with an assessment of how well the ADUS standards can accommodate the data needs of current and future safety applications, particularly those that could provide safety condition evaluations in real-time. In order to make this assessment it is first necessary to determine what these data needs are. While it is difficult to foresee the type of future methods that researchers may develop, it is possible to analyze current state-of-the-art safety applications to determine how far the data requirement envelop is pushed by those applications. This “data requirement envelop” will provide a benchmark against which it is possible to assess whether or not ADUS is able to supply the needed data for safety applications.

A review of the latest published research in this area is presented below. In general, these studies have as an objective the real-time assessment of the likelihood of a crash occurring on a given segment of freeway. Their methodologies take as input, real-time traffic information collected through loop detectors and other sensors, as well as other types of information to produce traffic-related hazardous condition assessments. However, the development and testing of these applications relies on archived data, particularly traffic condition information, crash data, weather information, and roadway geometric characteristics.

In 2002 Lee et al. studied traffic flow characteristics that lead to crashes (i.e., crash precursors) on urban freeways to explore factors contributing to changes in crash rate for individual vehicles and to develop a probabilistic model relating significant crash precursors to changes in crash potential. The study used incident logs and traffic flow data extracted from 38 loop detector stations along a 10-km segment of the Gardiner Expressway in Toronto, Canada. The traffic flow data from loop detectors included information on speed, volume and occupancy, averaged over 20-second time slices, and was collected for weekdays over a 13-month period from January 1998 to January 1999.

The authors analyzed 234 crashes that were reported to the traffic control center during the 13-month period. For 45 of these crashes they obtained more in-depth information concerning the causes and nature of these incidents through police reports. Also, to account for weather effects, hourly variation of weather conditions at a weather station near the study area were obtained and incorporated into the model.

The crash precursors used by the authors consisted of the degree of speed and density variation (low, moderate, and high) during a five-minute period prior to the crash. For the statistical model proposed in this study, the 24-hour speed data were divided into 288 five-minute time periods, during which the variability of the speed and density was determined using the 20-second loop detector information. Loop detector data was also used to determine the actual time of the crash (a critical piece of information), by the analysis of changes in detector speed profiles. This manual speed profile analysis was performed for all 234 crashes in the study.

After controlling for road geometry, weather, and time-of-day, Lee et al. found that the variations of speed and traffic density were statistically significant predictors of crash frequency. They also concluded that by using real-time traffic data, crash potential can be established for any instrumented freeway segment.

The same authors further refined their model (Lee et al., 2003a, 2003b) by better determining the time of crash using the information from the loop detectors upstream of the crash site (i.e., they used the characteristics of the shockwave produced by the crash to determine its time of occurrence). They also added a third crash precursor to the model: the average difference in speed between measurements made at stations upstream and downstream of a specific location (the other two crash precursors were speed and density variation). Another modification was that instead of pre-assigning an arbitrary interval of time in which to determine the values of the crash precursors (e.g., five minutes prior to the incident as in the previous paper); the optimal time offset for that interval was investigated in the study. To do so, the authors looked at twenty different time intervals (relative to the arrival time of the shockwave) ranging from one minute to 20 minutes prior to the crash. Precursor values were obtained for the three variables during all these intervals, identifying which interval size produced the best predictions. Although it is uncertain whether the maximum value identified within the range of one to 20 minutes actually represents a global maximum, it is difficult to imagine that traffic conditions more than 20 minutes prior to the crash occurrence have a significant impact on crashes.

The time offsets were determined to be eight, three, and two minutes prior to the crash, for speed variation, density variation, and speed differential, respectively. This finding implies that the variation of speed has a relatively longer-term effect on crash potential (i.e., it shows potential symptoms of an impending crash earlier) than the other two factors.

Abdel-Aty et al. (2004) used speed variation and average occupancy, both in five minute intervals, as crash precursor indicators. To develop and test their real-time application, these researchers looked at a 13.25-mile stretch of I-4 in Orlando, FL, instrumented with 28 dual loop detectors on each direction, during the period April-November, 1999. These detectors, spaced approximately half a mile from one another, provided information such as average speed, volume, and average occupancy (percent of time the loop is occupied by a vehicle) for every lane at 30-second intervals.

The Orlando Police Department supplied information about the 670 crashes that were recorded during the analysis period (only 375 crashes were used in the study since the rest of them did not have matching traffic information in the detail needed by the proposed methodology). For every crash, Abdel-Aty and his colleagues initially determined its time and location using police reports. In a similar fashion as Lee et al. (2002, 2203a, 2003b), they made necessary adjustments using simple heuristic rules based on space- and time-volume, occupancy and speed diagrams. However, as opposed to Lee et al, these researchers incorporated the spatial distribution of the information provided by the loop detectors surrounding the crash site into their model. That is, for each crash, the nearest loop detector was labeled the station of the crash, and traffic data was extracted for a period of 30 minutes just before the time of the crash. This 30-minute data was also extracted from other loop detectors located over a stretch of three miles, two-and-a-half miles upstream from the location of the crash, and half-a-mile downstream.

In order to predict conditions that may be conductive to a highway crash, it is necessary to look not only at traffic conditions before the event, but also at normal traffic conditions that do not lead to a crash. To do so, Abdel-Aty et al. extracted traffic data on all corresponding days of the week and times to the day of every crash (e.g., if the crash had happened on I-4 eastbound on Tuesday June 1st, 1999 at 4:00 PM at station 40, then data was extracted for all Tuesdays of the same season from 3:30 PM to 4:00 PM, for stations 35 to 41). Also, and to account for seasonal variation in traffic characteristics, two different seasons were defined: summer and non-summer.

The authors of this study found that the likelihood of a crash occurrence at a given location increases when higher occupancy rates upstream, together with high variation in speed downstream of the location are observed during a 5-10 minute interval. The results presented show that using the proposed methodology (i.e., a stratified case-control analysis), the logarithmic odds of crash occurrence may be obtained for real-time observed values of traffic flow parameters. Therefore, the potential "crash location" created due to ambient traffic conditions may be identified to warn the motorists about the impending hazard.

In another project, using the same site and data as in the previous study, Abdel-Aty and Abdalla (2004) developed a model to predict daytime crashes on a freeway when real-time traffic flow information and roadway geometric characteristics are known. The results showed that bad pavement conditions, horizontal curvatures, and the presence of on-ramps all increased the likelihood of a crash. Regarding traffic parameters, the authors found that a high variability in speed during a period of 15 minutes at a certain location increased the likelihood of a crash happening half-a-mile downstream. On the other hand, a low variability in volume over 15 minutes increased the likelihood of a crash occurring within a mile downstream.

Geometric information was also merged with crash data and traffic flow characteristics in a study conducted in California by Ma et al. (2004). These researchers initially looked at 11,045 crashes that occurred in 1998 at different freeway locations in Southern California. However, after merging HSIS (Highway Safety Information System) crash data with local 30-second loop detector data (to obtain traffic conditions preceding the time of the crash); and roadway design data, complete records for only 317 crashes remained. While the main result of this study showed that information on current traffic conditions was very valuable for injury severity prediction, it is worth noticing the need of not only archiving traffic data for safety application development, but also to be able to merge that information with other databases.

In another study, Lord et al. (2004) analyzed 5 years of crash data (1994-98) along one 40.5 km rural segment, and a 6 km urban stretch of Highway A-40 in the Ontario area in Canada. As was the case with the other studies discussed previously, these researchers also used traffic flow characteristics obtained from permanent (urban area) and temporary (rural area) count stations. On a 24-hour basis, the permanent stations utilized loop detectors to record vehicle speed, traffic volume, and vehicle occupancy for each lane at five-minute intervals; later aggregated to one-hour periods. Traffic flow data was sparser for the rural areas.

Lord et al. found that separate predictive models for single- and multi-vehicle crashes should be developed instead of just one common model for all type of crashes. Their results also showed that models that incorporate density and V/C (volume to capacity) ratio offer a more accurate description of crashes occurring on freeway facilities, whether they are located in a rural or urban environment. Predictive models that use traffic volume as the only explanatory variable may not adequately characterize the accident process.

Pande et al. (2005) used four years of traffic data —1999 through 2002, with 2,046 crashes reported in that period— collected from loop detectors along a 36-mile of I-4 corridor in the Orlando, FL metropolitan area. The time of occurrence of the crashes, which is always a critical piece of information needed to identify before and after crash traffic conditions, was estimated by using a combined shockwave and rule-based methodology that employed data from upstream loop detectors on the lane in which the crash occurred. Specifically, for each crash site, the methodology used data corresponding to the crash (case) and five matched non-crashes (controls). The traffic information was aggregated at three- and five-minute levels (the loop detectors collected and archived data at 30-second intervals) using the data from the lane in which the crash occurred and averages across all lanes belonging to the same station on the segment of freeway. That is, for each crash in the database the authors studied four different levels of spatial and temporal aggregation of the data (two for the temporal aggregation and two for the spatial aggregation), using patterns of hazard ratios (i.e., logarithm of the coefficient of variation in speed, the parameter used as a measure of crash risk) to identify freeway "hazardous spots" in real time.

Results of this analysis showed that the coefficients of variation in speed measured at the five-minute aggregation level were slightly better in predicting crash occurrence than those measured at the three-minute interval. Regarding spatial aggregation of the information, the results of the study indicated that combining observations from all lanes was better than using only data from the lane where the crash occurred. The main reason for this was that the aggregation captures across-lane speed variations (or lack of it) and also provides a larger dataset for the analysis with the advantage of producing results even in the case of a loop failure on a certain lane (i.e., since loop detectors tend to have a high rate of failure, the loss of information can be partially compensated by using traffic data from other adjacent loop detectors).

Using the same data as in the previous study, Abdel-Aty et al. (2005) analyzed the effects of high- and low-speed traffic regimes in predicting multi-vehicle freeway crashes in real time. As before, the 30-second raw data were aggregated into five-minute levels to obtain averages and standard deviations. Traffic information was gathered for half an hour previous to the crash, which at the five-minute aggregation level translated into six time slices (slice one was defined as corresponding to the period immediately prior to the incident). For each crash site, the traffic information was gathered from four loop detector stations upstream of the crash site (fifth station) and two downstream of the incident location, for a total of seven stations. The stations were identified as one (the furthest upstream station) to seven (the furthest downstream station).

The authors found that although the procedures used in building the model for low- and high-speed models were similar, different parameters entered in the two models. This was attributed to the fact that crashes occurring under a high-speed traffic regime differ from their low-speed counterparts not only in terms of severity but also in terms of the traffic conditions. Low-speed crashes occur, in general, in persisting congested conditions where queues form and dissipate quite frequently. On the other hand, high-speed crashes tend to happen under "smooth" freeway operation at the crash location and are due to disruptive conditions originating downstream and propagating backwards, causing drivers to make errors, which in turn increases crash potential. The results of the analysis showed, as one may expect, that the accuracy of the high-speed model is not as good as the low-speed model (i.e., in a high speed regime there are less disruptions of the traffic stream that can be used in detecting increases in the crash potential), and that the former model does not include the coefficient of variation in speed as one of the main predictive factors. For the low-speed model, the variation in the speed coefficient measured during time slice two (five-to-ten minutes prior to the crash) from the station closest to the crash location, was the most significant prediction variable.

Abdel-Aty and Pemmanaboina (2005) developed a model to identify significant factors contributing to traffic crashes on freeways based on the frequency of crash occurrence on a 36-mile stretch of I-4 in Orlando, FL, as well as traffic (collected from 139 loop detectors) and geometric and roadway characteristics. For each one of the 139 loop detector stations, information was collected which included the radius of the freeway section, number of lanes, median type, median width, pavement index, pavement surface type, pavement roughness index, and the presence of on- or off-ramps within the influence area of each crash station. Regarding traffic information, the authors used static/aggregated measures (such as the section AADT and VMT values) and disaggregated measures (such as peak 15-minute volumes and 5-to-20 minute average volumes just before the crash occurrence). The results of the model provided strong evidence that detailed measurements of traffic volumes immediately prior to the crash was significant in predicting crashes, with other important factors being road curvature, number of lanes, median type, pavement surface type and presence of on- or off-ramps.

All the studies discussed have as their main goal the development of applications that can take real-time traffic information, combined with weather, and in some cases, roadway geometric characteristics, and assess the likelihood of crashes on a given segment of a freeway. The other side of the equation is how to act on this information; that is, how to affect the behavior of the motorists who are traveling within the hazardous area so that the likelihood of a crash in that area is reduced. Two factors are important to consider: the warning time available and the type of countermeasure strategy to be deployed. Systems that give two or three minutes advance notice (e.g., Lee et al., 2003a, 2003b) do not seem appropriate since they do not give sufficient warning time to deploy any countermeasure strategy. Regarding those strategies, and given the current available technology (e.g., very few, if any, in-vehicle devices), the most sensible approach may be to have some kind of variable speed limit control deployment by which the speed limit on the hazardous segment could be reduced during intervals in which traffic conditions show a high likelihood of crashes. On the other hand, reductions in speed limits also increase system travel time and therefore there is a quantifiable trade-off between reduction in risk (crash potential) and an increase in system travel time that would need to be considered when evaluating this strategy.

Lee et al. (2004) studied the effects of a variable speed limit strategy. They found that, on average, variable speed limits could reduce total crash potential by approximately 25% if speed limits were temporarily reduced during risky traffic conditions. They also found that reduction in crash potential was greatest at locations with high traffic turbulence (e.g., downstream of on-ramps).

One problem with this study, however, is that the analysis did not use real traffic data, but data generated by a traffic simulation model (PARAMICS). Evidently, assumptions in the PARAMICS simulation model need to be empirically validated to have any confidence in the results obtained in evaluating a variable speed limit control strategy using that model. To avoid this type of problem, a soon-to-begin project sponsored by FHWA will study the effects of variable speed limits in work zones in the state of Maryland. The demonstration project will perform before and after evaluations, as well as studying speed profiles and conducting driver performance analysis. The project should help evaluate the impacts of variable speed-limit control in areas under normal traffic operations (no lanes closed for construction or maintenance).

The applications discussed here provide guidelines on what type of data, and at which spatial and temporal resolution, are needed to develop models that permit the real-time assessment of the likelihood of crashes on freeways, which is one of the most data-demanding safety applications. Table 1 below summarizes these guidelines, showing that in the most demanding case, data should be archived with a resolution of 20 seconds, for at least 30 minutes previous to a crash occurrence not only for the day in which the crash happens, but also on the same day of the week when no incident occurs; and for sensors upstream and downstream of the crash site.

Notice that all the papers discussed focused on freeways and not on surface streets. This is because, a literature review of the this area resulted in the identification of no published studies dealing with real-time crash prediction at intersections; even though this type of incident accounts for about 30% of all crashes. To remedy this research gap, the Office of Safety at FHWA is currently sponsoring a project aimed at investigating this type of problem. SAIC, the principal investigator of the project, has proposed to use advanced sensors to automatically measure a more comprehensive set of driving and traffic parameters than the ones currently provided by twin-loop systems. The study is aimed at creating a database that will make possible the investigation of a new generation of safety measures of effectiveness (SAIC, 2003a, 2003b). Specifically, the initial proof-of-concept effort will focus on rear-end crashes at one approach of each of two signalized intersections. This will later be expanded to encompass multiple legs and approaches. One of the intersections to be selected will

Table 1. Data Requirements and Characteristics for Real-Time Freeway Safety Applications.

Information			Study							
			Lee et al. (2002)	Lee et al. (2003a) (2003b)	Abdel-Aty et al. (2004)	Abdel-Aty & Abdalla (2004)	Ma et al. (2004)	Lord et al. (2004)	Pande et al. (2005)	Abdel-Aty et al. (2005)
Crash Precursor	Speed (Sp)	Average					✓		✓	✓
		Temp. Var.	✓	✓	✓	✓			✓	✓
		Spatial Var.		✓						
	Volume	Average						✓perm & temp LD	✓	✓
		Variation				✓			✓	✓
	Density	Average					✓	✓perm & temp LD		
		Variation	✓	✓						
	Occupancy	Average			✓				✓	✓
		Variation							✓	✓
	V/C Ratio	Average					✓	✓		
	Road Condition & Geometry					✓				
Traffic Information	Speed		✓	✓	✓(30 min ptc)	✓(30 min ptc)	✓	✓	✓	✓
	Volume		✓	✓	✓(30 min ptc)	✓(30 min ptc)	✓	✓	✓	✓
	Occupancy		✓	✓	✓(30 min ptc)	✓(30 min ptc)	✓	✓	✓ (by lane)	✓
Weather Information			✓	✓						
Geometric Information			✓	✓		✓	✓			
Temporal Granularity			20 sec	20 sec	30 sec	30 sec	30 sec	5 min	30 sec	30 sec
Temporal Distribution[1]			5 min ptc[4]	Sp Temp Var (8 min ptc) Sp Spatl Var (2 min ptc) Dens. Var. (3 min ptc)	5-15 min ptc (30 min ptc for model development)	15 min ptc (30 min ptc for model development)	10 min ptc	60 min	3-5 min ptc (60 min ptc and 30 min ac[5] for model dev.)	5-15 min ptc (30 min ptc for model development)
Spatial Distribution[2]			NA	Upstream Detectors	4 upstream & 2 downstream stations	4 upstream & 2 downstream stations	Upstream Detectors	NA	4 upstream & 2 downstream stations	4 upstream & 2 downstream stations
Crash Information			234 crashes Traffic Cont Center (45 police reports)	Crashes reported to Traffic Control Center	670 crashes from Orlando Police reports	375 crashes from FL DOT database	11,045 crashes reduced to 317 when merged with traffic info	5 years of crash data	2046 crashes from FL DOT database	1999-2002 FL DOT crash database
Time of Crash	Reported ToC[3]		✓	✓	✓	✓	✓	✓	✓	✓
	Incident Detection		✓							
	From Traffic Flow Theory			✓	✓	✓			✓	✓

[1] All the studies require traffic conditions information at the site of the crash for same day of the week and time of day under normal (i.e., incident free) traffic conditions to develop the methodology.
[2] The spatial distribution indicates how many information-gathering source points (e.g., loop detectors) upstream and downstream of the crash site were used in the methodology.
[3] ToC: time of crash. [4] ptc: prior to crash. [5] ac: after crash.

be characterized by a low-to-moderate crash history, while the other will have a relatively higher crash rate history. At each of these intersections, one approach will be instrumented with advanced detectors —complementing the existing loop detectors and cameras— to capture traffic parameters such as vehicle approach speed, vehicle headways, time-to-collision, critical incidents and conflicts, and others. Similarly to the freeway studies discussed earlier in this report, this urban safety project will also correlate the traffic parameters with crash records.

ADUS Standards

The *Strategic Plan for the Development of ADUS Standards* (Cambridge Systematics, Inc., 2000) defined a sequence of standards development for ADUS around four categories of data –and their associated "market package bundles" as defined in the National ITS Architecture– which included traffic data (Advanced Traveler Information Systems and Advanced Traffic Management Systems), transit data (Advanced Public Transit Systems), incident data (Emergency Management and Advanced Traffic Management Systems), and freight data (Commercial Vehicle Operations).

For each one of these four categories of data, it was established that two types of standards would be considered. The first one included guidelines for data processing, storage, and retrieval. These guidelines would follow general principles (including the avoidance of over-specifying standards, accommodating the widest variety of field sources, assuring the privacy of individuals and firms, storing the original/raw data for a period of time, and documenting data quality, among others) that were established in workshops conducted in 2000. The second one was the development of data dictionaries. It was left to the discretion of the ASTM (American Society of Testing and Materials) ADUS committee and the ADUS Standards Program Manager to determine which one of these two standards should be pursued in each one of the four data categories.

ASTM E2259-03a, Standard Guide for Archiving and Retrieving Intelligent Transportation System-Generated Data (ASTM, 2003)

The 2000 ADUS strategic plan also established an aggressive schedule for the process that was conductive to the development of the ADUS standards. The first guideline (Guidelines for Processing, Storing, and Retrieving Archived ITS Travel Monitoring Data) was originally expected to be approved in June 2002. After some delays, the ASTM Committee E17.54 published the first ADUS standard in 2003: the *ASTM E2259-03a, Standard Guide for Archiving and Retrieving Intelligent Transportation System-Generated Data* (ASTM, 2003). (See Appendix D.)

While the ASTM E2259-03 document promotes a structured methodology for developing the archived data management systems that would host the ADUS functions described in the National ITS Architecture, it is not a standard. In fact the document indicates that

> *"This guide follows the intent of all guides prepared within the ASTM International framework. In particular, it suggests approaches, offers an organized collection of information, or proposes series of options or instructions that give direction without recommending a specific course of action."*

and also

> *"The word Standard in the title of this document means only that the document has been approved through the ASTM International consensus process."*

Because this is not a standard in the narrow sense of the word, the document does not strictly specify formats and processes; its main objective is to provide a degree of consistency in ADMS (Archive Data Management Systems) development for ITS data archiving. It consistently addresses in a general form, issues and challenges expected in the development of the ADMSs; including storage requirements, selection of an operating system environment and database management system, coordination with existing ITS data systems and sources, and the system development process.

The guide also deals with the type of data and information that should be considered for archiving and retrieving. For example, in section 10.15.1 the guide indicates that a management oversight committee should decide what particular data elements are to be archived, but gives particular consideration to user needs. With regard to the preservation of the original source data, the guide considers that the retention capability of an ADMS is one of its most important features and establishes that *"incoming flows should be maintained at its most detailed spatial and temporal resolution for a period of time specified by individual system designs"*. The guide strongly advices against purging the original source data, suggesting instead to permanently store it either online or offline, at the spatial and temporal resolution at which it was collected. If the data is aggregated, it is suggested that provisions should be made to enable interested users to save the detailed data to their own system before the aggregation is performed. Similarly, if an agency decides not to store the original data, the guide advises that provisions be made to enable interested users to save the original source data to their own system before such data is removed from the ADMS. These are all measures that can greatly help in the development of future safety applications by assuring the availability of the necessary archived data.

Regarding data quality control, the guide suggests the creation and documentation of indicators of data quality, collection conditions, and the type of data source. It also recommends providing to the ADMS the capability of accepting user-specified data quality control and editing procedures that would allow flagging and/or removal of suspect, erroneous, or duplicate data from the data archive. The ADMS should also have the capability of alerting the users when they are accessing data that is not original (i.e., imputations).

Guidelines regarding physical and legal security are also included in the ASTM E2259-03a document. Among the former, the guide advises the provision of multi-level hardware and media backup, authorized accessibility, and the designation and control of secure components that are distinguished from insecure or public components. Regarding the latter, it suggests compliance with specific legal and policy restrictions, such as data confidentiality, ITS data privacy, data copyright, data accessibility, and data retention (and destruction) requirements. It also suggests that *"an ADMS for ITS-generated data should avoid and exclude legally restricted data to the maximum possible extent"* and that *"an archived data administrator carefully consider the implications of admitting any legally restricted data into the archive"*.

In summary, the ASTM E2259-03a promotes sound practices for information systems development, data maintenance, and data quality. However, its scope is very broad and it does not rigorously specify formats and/or procedures.

Standard Practice for Metadata to Support Archived Data Management Systems (Currently under development)

The subcommittee is currently in the last stages of the development of a metadata standard which will be used to describe archived data characteristics such as its content, quality, organization, availability, lineage, and other attributes; but not the data itself. The metadata standard, *Standard Practice for Metadata to Support Archived Data Management Systems*, is based on the FGDC's existing *Content Standard for Digital Geospatial Metadata* (FGDC-STD-001-1998) and it is composed of seven main sections. Those sections are: 1) basic information about the data set (i.e., identification information); 2) data quality information; 3) description of the mechanisms to represent spatial information in the data set; 4) spatial reference information (i.e., description of the data set coordinates reference frame and the means to encode these coordinates); 5) information about entity types, their attributes, and their value domains; 6) information about the distributor and how to obtain the data set; and 7) information on how up-to-date the metadata information is. Sections 1, 2, 5, and 7 are mandatory in the standard, while the remaining three are mandatory if applicable. Compound and metadata elements, within each of the seven basic sections, are also classified as: a) mandatory, b) mandatory if applicable, or c) optional (mandatory elements are essential for metadata functionality, while optional elements are desirable but not essential). Besides the seven main sections, the standard also includes three other sections that serve as reference –i.e., citation, time period, and contact information.

The metadata standard was balloted in June 2005 (there were 10 affirmatives and one abstention with comment) and is currently being reviewed by experts outside the ASTM subcommittee.

Standard Specifications for Archiving ITS-Generated Traffic Monitoring Data (Currently under development; July 18th, 2005 version)

The subcommittee is currently developing the *Standard Specifications for Archiving ITS-Generated Traffic Monitoring Data*, a standard practice that describes data elements and the structure for an archived data management system for traffic monitoring data; including conventionally gathered traffic data, data collected directly from ITS systems, and travel time data from probe vehicles. The document defines the names of the data elements, their interrelationships, and their procedural definitions (including data collection instrumentation and methodology as well as recommended procedures for calculating traffic statistics).

Figure 1 illustrates the entities defined in the Standard for traffic data collection using conventional and ITS data-gathering technologies. All the entities shown in that figure are considered punctual (i.e., defined at a given geographical point) and the Standard assumes that there exists a location-referencing system which is adopted for the data archive. The "Roadway" entity provides information about a given road (at a point on that road) and may contain one or two directions of travel. Each direction of travel may have "Station" entities that in turn contain "Lane" entities. The

“Detector” entities are located on the “Lane” entities under a “hierarchical description”, which is what the Standard recommends.

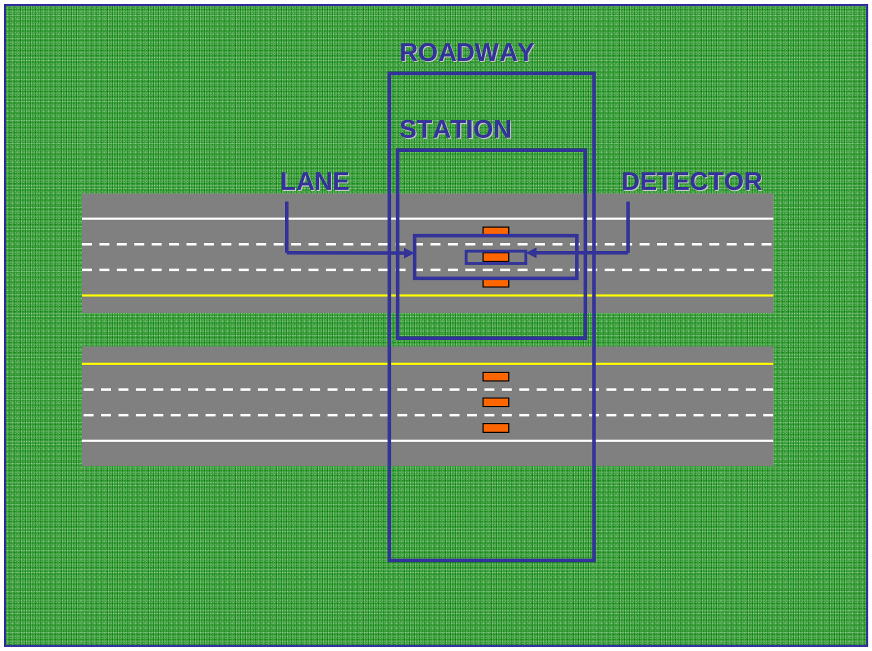

Figure 1. Traffic Monitoring Data Archiving Entities

The Standard also presents specifications for the archiving of traffic information gathered with probe-vehicles. As opposed to the Traffic Information data discussed above in which the sensors are assumed to be located at a point on the roadway, the measurement of travel time using probe-vehicles requires the definition of a segment of road on which these measurements are made. The Standard defines an entity, named “Composite Link”, which provides that framework (again, the Standard assumes that there exists a location referencing system that is used, where necessary, to add the spatial dimension to the archive). The combination of this entity with the entity that archives the position of the probe-vehicle allows computing travel times on that defined link. Further aggregations are possible in the Standard through entities that take into account measurements made on the “Link” entity by multiple probe-vehicles during a given interval of time.

Summary and Recommendations

At the present time, FHWA is sponsoring several research and development efforts aimed at developing “Safety Level” analysis methodologies that can quantify roadway/facility hazardous conditions in real time, assessing the potential usage of ITS technologies to support innovative real time ITS safety measures, and developing Archived Data User Service elements for safety oriented applications.

This report, which is part of the last group of projects, has as a main objective to provide FHWA with an overall assessment of the progress being made on the development of the ITS ADUS standards and how those standards may potentially impact the development of future safety applications. That is, the main research question for this project was to determine how well the ADUS standards can accommodate the data needs of current and future safety applications, particularly those that could provide safety condition evaluations in real time. Clearly, to assess the potential impacts of these standards, it is first necessary to determine the data needs for the applications under consideration.

While it is difficult to forecast what type of future methods researchers may develop, it is possible to analyze current state-of the-art safety applications to determine their data requirements. The methodology used in this report followed precisely this approach. Specifically, all the available state-of-the-art safety applications that required real-time information to assess safety conditions on freeways were analyzed and the data used by the researchers were identified. The information, obtained from published work (see the References section of this report) and also from discussions with researchers that are at the forefront of this field, were used to build a "data requirement envelop" (see Table 1), which in turn provided a benchmark to assess if the archive standards were able to supply the needed data for the development of safety applications.

Section 2 of this report (i.e., Review of Safety Applications Research) presented an overview of current state-of-the-art safety applications that use real-time traffic data and other relevant information (e.g., weather, road geometric characteristics) to assess the likelihood of imminent crashes on freeways. Table 1 summarized these requirements by providing information about the granularity as well as the spatial and temporal distribution of the data needed to develop (and run) these real-time safety applications. In the most demanding cases, the applications required data to be archived with a resolution of 20 seconds, for at least 30 minutes previous to a crash occurrence. This was the case not only for the day in which the crash happened, but also on the same day of the week when no incident occurred, and for sensors upstream and downstream of the crash site.

The Archived Data user service, as described in the National ITS Architecture, is not intended for real-time applications, but rather for the archiving of ITS-generated operational data for historical and secondary uses. A characterization of a "data envelope" based on real-time applications may seem inappropriate for the assessment proposed in this report's methodology. However, this is not the case since the safety applications used for that characterization were developed by means of archived (historical) data, and only require real-time data, obtained directly from the sensors and not from the archive, when they are deployed. Nevertheless, this points to a critical requirement in the archiving of the ITS operational data for the development of safety applications: that of archiving the raw data (i.e., non-aggregated data) produced by the sensors. The National ITS Architecture anticipated this requirement for ADUS and specifies that a data storage function, aimed at preserving original and unaltered data in the master archive for some minimum amount of time should be implemented.

The National ITS Architecture also provides general guidelines regarding this and other user services. However, the implementation of these user services requires normative procedures and specific guiding principles to assure compatibility and interoperability. These objectives are usually achieved through the use of standards. In the case of the Archived Data user services, the development of the standards was assigned to the ASTM ADUS Subcommittee (ASTM E17.54) under the ASTM Committee on Vehicle - Pavement Systems (ASTM E17). The ADUS Subcommittee has published one standard (as of July 2005), the *Standard Guide for Archiving and Retrieving Intelligent Transportation System-Generated Data* (ASTM E2259-03a), which provides guidelines for and encourages best practices in the development of information archiving systems, maintenance of the data (including strong advice to store raw data for at least some period of time), and control of data quality. However, because this is not a "true" standard (i.e., it is a "guideline"), it has a large scope and does not strictly specify formats and/or procedures for the development of these archiving systems.

Two "true" standards are currently being developed by the subcommittee. The first one, which is almost ready for publication as of July 2005, focuses on the development of the metadata standard for ADUS. The *Standard Practice for Metadata to Support Archived Data Management Systems*, which is based on the Federal Geographic Data Committee's *Content Standard for Digital Geospatial Metadata* (FGDC-STD-001-1998), describes archived data characteristics such as its content, quality, organization, availability, lineage, and other attributes, but not the data itself. Although this is an important standard, because of this deficiency its impacts on the development of safety applications are minimal.

The second standard currently being developed by the ASTM ADUS Subcommittee is the *Standard Specifications for Archiving ITS-Generated Traffic Monitoring Data*. This is a standard practice for traffic monitoring data that describes not only data elements but also the structure for an archived data management system. The document defines the names of the data elements, their interrelationships, and their procedural definitions (including data collection instrumentation and methodology as well as recommended procedures for calculating traffic statistics). In its present form, the standard includes conventionally gathered traffic data, data collected directly from ITS systems, and travel time data obtained from probe vehicles.

The development of any safety application, and in particular those analyzed in this report, requires basically four types of data: traffic, crash, weather, and roadway information. The ADUS traffic monitoring standard focuses on the first source of data. It can therefore have potentially large impacts on the development of safety applications if the archived information does not meet certain requirements in terms of granularity, as well as spatial and temporal distributions. Those requirements, derived from the studies reviewed in this report, provide the basis for the following recommendations regarding the archiving of traffic monitoring data as a source of information necessary for the development of safety applications.

1.-Traffic monitoring data needs to be archived at the most detailed granularity provided by the sensors.

This prevision, suggested in the National ITS Architecture and the Standard *Guide for Archiving and Retrieving Intelligent Transportation System-Generated Data*, is critical for the development of safety applications since all of them rely on some measure of the traffic conditions at the time and location of the crash. Aggregated measures such as AADT and hourly traffic volumes had been used in many safety applications such as the identification of countermeasures for hotspots, before-and-after studies, and others, simply because there were no other more disaggregated statistics. Other safety applications such as those reviewed in this report (e.g., automated collision notification, real-time hazardous condition identification) could only be developed and studied because of the availability of fine resolution traffic data.

The latter type of safety applications need to take into consideration not only the average values of traffic parameters (crash precursors) but also their variability within small intervals of time before and after the crash. The variability of these crash precursors (i.e., speed, density, and volume) is usually assessed using the standard deviation of the traffic parameters under consideration within 5- to 15-minute intervals. Typically, data is archived at this level of aggregation (see Appendix A), but unless the time of crash occurs exactly at the beginning/end of one of these archived time intervals, the information would not be adequate for the development of those safety applications. They need averages and variability measures within time intervals that are centered on the time of the crash, and this can only be obtained if the traffic data is archived at its highest possible resolution.

There is another important reason to archive the data at a high resolution, especially for the development highway safety applications. The time of crash, a critical piece of information in any safety analysis, has to be very accurate in many cases. One way of checking the actual time of crash against the information posted on police reports is by means of traffic flow theory (i.e., identifying shockwaves and other measures occurring as a consequence of a crash), which can only be achieved if there is enough resolution in the traffic monitoring data.

2.-Raw sensor data should be archived for a minimal period of time that allows the collection of statistically significant safety data information.

The National ITS Architecture (Information Management User Services) and the ASTM E2259-03a standard (sections 10.15.1 and 10.16.5.2) suggest that raw data be archived for a minimum amount of time, leaving the decision of determining that minimum period to the discretion of the archive manager.

Crashes are low probability events, and therefore the expected time to observe one at a give location is very long. This alone should not determine the length of time for which the raw traffic monitoring data should be archived. Indeed, there are two other considerations that have to be taken into account when deriving that minimum time. The first deals with the development of safety applications and associated requirement of knowledge of traffic conditions at the location of a crash not only around the time of its occurrence, but also on the same day of the week and at the same time of the day when there were no accidents. For example, if a crash occurred at a given location on a Tuesday at 3:24 PM, traffic conditions (averages and variability factors) need to be known at that place not only around that time and date, but also on other “Tuesdays” and around the same time. Considering this as a “rule of thumb” (and under certain assumptions about the data) it is necessary to have approximately 30 observations to achieve statistically significant results. It would therefore be necessary to have 30 “Tuesdays” worth of archived data available for the example above. This imposes a minimum archiving time of 30 weeks, assuming that there are no seasonal variations in traffic conditions.

The second factor deals with the availability of the crash information. The estimated minimum archiving time derived above assumes that the occurrence of a crash is known instantaneously. In cases where there are incident detection systems deployed (usually, urban freeways), this is a reasonable assumption if the crash-occurrence information is also stored in the archive. However, there are other cases (e.g., crashes at surface street intersections) where awareness of an incident occurs when the police accident report becomes available; which in many states could be several months. For such cases, the minimum time for the archiving of the raw data should take into account this reporting delay as well.

3.-The spatial distribution of traffic information should be maintained in the archive.

As in the case of temporal distribution of the traffic information, its spatial distribution also plays a key role in the development of safety applications, especially those geared towards real-time hazardous condition identification. Spatial distribution of the traffic information along the direction of travel allows the determination of the space-related variability of traffic parameters (e.g., speed differentials between consecutive detector stations) which some safety applications use as crash predictors. Spatial distribution is also important in the determination of the time of crash when it is derived through methodologies that are based on traffic flow theory (e.g., shockwave analysis). Variability across lanes is also important and is used in many safety applications. Therefore, the aggregation of data in either direction (i.e., longitudinally along the direction of travel or transversal across lanes) for archiving purposes would result in the loss of valuable information required for the development of safety applications.

The archiving of disaggregated traffic information is also important for the development of new safety technologies that provide collision countermeasure systems (CCS). Such technologies may rely on detectors located at different points upstream of the surface street intersections. Aggregating the information provided by these spatially distributed detectors would hamper the development of these applications.

4.-The standards should be able to archive new traffic parameters, such as travel time generated from pseudo probe-vehicles.

The latest development of real-time safety applications was only possible due to the recent availability of high resolution (temporally and spatially) traffic parameters such as speed, volume, and density. At the present time, this "traditional" traffic information is the only type of data that the ITS infrastructure collects. However, new technologies are being deployed that provide other types of traffic information (e.g., travel time), which in the past was either not available or not very accurate. This type of information, once it becomes readily available, would play a very important role in the development of new safety applications.

The *Standard Specifications for Archiving ITS-Generated Traffic Monitoring Data*, in its present version, covers the archiving of traditional traffic monitoring information (i.e., obtained from roadside detectors); as well as traffic parameters derived from probe-vehicles furnished with GPS devices. There are other technologies not covered in the Standard that allow the computation of travel times by "creating" probe-vehicles out of regular cars. One of these technologies is license-plate readers that, by matching the license plates of vehicles at different locations of the roadway, allow the computation of travel time on those road segments. There is also a new generation of inductive detectors that can potentially provide very detailed vehicle signatures, allowing the identification of specific vehicles at downstream locations, thus creating de-facto probe-vehicles. Also, the US DOT Vehicle-Infrastructure Integration (VII) initiative focuses on vehicles communicating with other vehicles as well as with the infrastructure through dedicated communication channels. Under this initiative, vehicles would serve as data collectors and anonymously transmit traffic and road condition information from every major road within the transportation network.

It appears that the Standard, as it is currently specified, could allow archiving the information generated by these new technologies in a similar manner as with conventional probe-vehicles (i.e., by defining links through the position of consecutive sensors –license-plate readers, inductive sensors, VII beacons– instead of using landmarks such as intersections or ramp gores). However, in its present version the Standard does not contain previsions to specify how the travel-time information was generated (i.e., probe-vehicle, license plate reader, inductive sensors, etc.). More importantly, it is not possible to link the travel-time information gathered by these pseudo probe-vehicles to the detectors that contributed to these measurements. As opposed to probe-vehicles furnished with GPS information gathered from satellites, these new technologies and systems rely on measurements obtained from sensors deployed along the roadways which may or may not work properly. The Standard does not connect these two types of entities (probe-vehicles and sensors) which would be necessary to determine the quality of the measurements obtained.

Upcoming Standards. The Subcommittee is considering the development of two new data archiving standards one for transit and the other one for incident management. The latter, which has critical implications for the development of safety applications, may be based on the NHTSA Model Minimum Uniform Crash Criteria (MMUCC) Guideline (2nd Edition, 2003). These guidelines encourage states to adopt standardized data elements for crash reporting, as collected by police officers investigating crashes.

From the view point of the development of safety applications, the exact location and time of the crash are perhaps the most important pieces of information (together with the type of crash, and the number and type of vehicles involved). The other key requirement is the ability of linking the data to other sources of information such as traffic monitoring data, roadway geometry, and weather information.

References

Abdel-Aty, M. and R. Pemmanaboina, 2005, "Advances in Transportation Studies: Assessing Crash Occurrence on Urban Freeways Using Static and Dynamic Factors," *Journal of Advances in Transportation Studies* (forthcoming).

Abdel-Aty, M., N. Uddin, and A. Pande, 2005, "Split Models for Predicting Multi-Vehicle Crashes during High- Speed and Low-Speed Operating Conditions on Freeways," *Transportation Research Record CD*, TRB Paper No. 05-0026.

Abdel-Aty M. and M. F. Abdalla, 2004, "Linking Roadway Geometrics and Real-Time Traffic Characteristics to Model Daytime Freeway Crashes Using Generalized Estimating Equations for Correlated Data," *Transportation Research Record CD*, TRB Paper No. 04-3174.

Abdel-Aty M., N. Uddin, M. F. Abdalla, A. Pande, and L. Hisa, 2004, "Predicting Freeway Crashes Based on Loop Detector Data Using Matched Case-Control Logistic Regression," *Transportation Research Record CD*, TRB Paper No. 04-2829.

Advances in Traffic Data Collection and Management, White Paper prepared by Battelle, Texas Transportation Institute, and Cambridge Systematics, Inc. for FHWA, USDOT; December 2002a.

AMPO Survey: Use of Archived Data from Intelligent Transportation Systems. Association of Metropolitan Planning Organizations (AMPO), available at the AMPO web site: http://www.ampo.org/survey/ArchivedDatafromITS-Results.pdf, Accessed July 2005.

CCMPO AADT Traffic Database. Chittenden County (VT) Metropolitan Planning Organization. Available at: http://www.ccmpo.org/data/ccmpo_aadt_metadata.html Accessed July 2005.

Content Standards for Digital Geospatial Metadata. FGDC-STD-001-1998, Federal Geographic Data Committee; June 1998.

Data Element Definitions for Transportation Sensor Systems (NTCIP 1209). National Transportation Communications for ITS Protocol published by the American Association of State Highway and Transportation Officials (AASHTO), in conjunction with the Institute of Transportation Engineers (ITE) and the National Electrical Manufacturers Association (NEMA); v01.18d, April 2004.

Defining and Measuring Traffic Data Quality, White Paper prepared by Battelle, Texas Transportation Institute, and Cambridge Systematics, Inc. for FHWA, USDOT; December 2002b.

Gardner, D. and R. Bostrom, "Ohio and Kentucky Approach to Data Archiving in Cincinnati," presented at the TRB 82nd Annual Meeting, Archived Data User Service (ADUS): System Designs and Implementation Examples Workshop, January 12, 2003, Washington DC.

IEEE Standard for Data Dictionaries for Intelligent Transportation Systems (P1489). IEEE-SA Standards Board; June, 1999.

iFlorida Work Plan - Surface Transportation Security and Reliability Information System Model Deployment. FL Department of Transportation for FHWA, USDOT; August 2003. http://www.itsdocs.fhwa.dot.gov//JPODOCS/REPTS_TE//13838.html Accessed July 2005.

Improving Crash Data for Safer Roads: MMUCC Minimum Uniform Crash Criteria Guideline. National Center for Statistics and Analysis (NHTSA/USDOT), Second Edition (2003).

ITS Data Archiving Five-Year Program Description. FHWA, 2000.

ITS Deployment Tracking Survey. ITS Joint Program Office, Federal Highway Administration FHWA, August 2003.

ITS Enhanced Data Collection for Safety Evaluation: Work Plan. SAIC for FHWA, USDOT; April 2003 (a).

ITS Enhanced Data Collection for Safety Evaluation: Stage 1 Report (Draft). SAIC for FHWA, USDOT; August 2003 (b).

ITS Standards Advisory web page. US Department of Transportation. http://www.its-standards.net/Documents/ADUS_advisory.htm.

ITS User Services Document. Architecture Development Team for FHWA, USDOT; January 2005.

Lee, C., F. Saccomanno, and B. Hellinga, 2002, "Analysis of Crash Precursors on Instrumented Freeways," *Transportation Research Record*, 1784, pp. 1-8.

Lee, C., B. Hellinga, and F. Saccomanno, 2003a, "Proactive Freeway Crash Prevention Using Real-Time Traffic Control," *Canadian Journal of Civil Engineering*, 30, pp. 1034–41.

Lee, C., B. Hellinga, and F. Saccomanno, 2003b, "Real-Time Crash Prediction Model for Application to Crash Prevention in Freeway Traffic," *Transportation Research Record CD,* TRB Paper No. 03-2749.

Lee, C., B. Hellinga, and F. Saccomanno, 2004, "Assessing Safety Benefits of Variable Speed Limits," *Transportation Research Record CD,* TRB Paper No. 04-4835.

Lord D., A. Manar, and A. Vizioli, 2004, "Modeling Crash-Flow-Density and Crash-Flow-V/C Ratio Relationships for Rural and Urban Freeway Segments," *Transportation Research Record CD,* TRB Paper No. 04-3365.

Ma J., K. M. Kockelman, and C. B. Luce, 2004, "Anticipating Injury and Death: Controlling for New Variables on Southern California Highways," *Transportation Research Record CD,* TRB Paper No. 04-4379.

National Transportation Communications for ITS Protocol, www.ntcip.org. Accessed July, 2005.

Pande, A., M. Abdel-Aty, and L. Hsia, 2005, "Spatio-Temporal Variation of Risk Preceding Crash Occurrence on Freeways," *Transportation Research Record CD,* TRB Paper No. 05-0010.

Report of the Verification, Validation & Certification (VV&C). USAF, Tiger Team; June 1998.

Skabardonis, A., "PeMS: Moving to Statewide Deployment," presented at the 11th Annual PATH Research Meeting, October 25, 2002, RFS UC Berkeley, CA.

Smith, B. L., W. T. Scherer, and J. H. Conklin, 2003, "Exploring Imputation Techniques for Missing Data in Transportation Management Systems", *Transportation Research Record CD*, 2003.

Smith B. L. and S. Babiceanu, 2004, "An Investigation of Extraction Transformation and Loading (ETL) Techniques for Traffic Data Warehouses," *Transportation Research Record CD*, TRB Paper No. 04-3719.

Standard Guide for Archiving and Retrieving Intelligent Transportation Systems-Generated Data (ASTM E2259-0). American Society for Testing and Materials (ASTM), West Conshohocken, PA, United States; 2003.

Standardized Procedures for Personal Travel Surveys (NCHRP Project 8-37). Louisiana Transportation Research Center, Baton Rouge, LA. (Scheduled completion date: January 31, 2005.)

State of the Practice for Traffic Data Quality, White Paper prepared by Battelle, Texas Transportation Institute, and Cambridge Systematics, Inc. for FHWA, USDOT; December 2002c.

Strategic Plan for the Development of ADUS Standards, Prepared by Cambridge Systematics, Inc. for FHWA, USDOT; May 2000.

Surface Transportation Security and Reliability: Information System Model Deployment (Draft). Florida Department of Transportation for FHWA, USDOT; January 2004.

The Case for Standardizing Household Travel Surveys, NCHRP Research Results Digest 261. Transportation Research Board, National Research Council; Washington, DC.; January 2002.

Traffic Data Quality Measurement Final Report, ITS America, http://www.itsa.org/itsnews.nsf/key/CA0A?OpenDocument. Accessed July, 2005.

Transport information and control systems — Requirements for an ITS/TICS central Data Registry and ITS/TICS Data Dictionaries (ISO 14817). International Organization for Standardization, Geneva, Switzerland, 2002.

Transportation Safety Information Management System, http://tsims.aashtoware.org. Accessed July, 2005.

Turochy, R.E. and B.L. Smith, 2000, "A New Procedure For Detector Data Screening in Traffic Management Systems," *Transportation Research Record*, 1727, pp. 127-31.

University of Kentucky, Kentucky Transportation Center, Intelligent Transportation Systems web page, http://www.ktc.uky.edu/ITS/Research.htm. Accessed July, 2005.

Varaiya, P., "PeMS Overview," TRB 82nd Annual Meeting, Archived Data User Service (ADUS): System Designs and Implementation Examples Workshop, January 12, 2003, Washington DC.

XML Schemas for Exchange of Transportation Data (NCHRP Project 20-64). Transportation Research Board, National Research Council; Washington, DC; in progress.

ITS User Services Document. Architecture Development Team for FHWA, USDOT; January 2005.

Appendix A: Review of Data Archiving Deployments

This section presents an update on several on-going data archiving projects, some of which are sponsored by FHWA. While ADUS standards are being developed, many traffic operation centers and other organizations that have a real and practical need to archive data are developing their own systems. There is no doubt that some of these on-going data archiving projects will play a significant role in the development of the ADUS standards, and therefore a review of them can shed some light on the form that those still-to-be-developed standards may take.

Virginia ADMS

Smith and Babiceanu (2004) provide a detailed description of all the procedures and steps in archiving traffic data in the state of Virginia. Based on the premise that archived data management systems cannot be created by simply exporting data from an operations system, the authors describe the methodology used in developing the Virginia ADMS. This methodology calls for three steps which must be carefully performed: the extraction, transformation, and loading of information from an operational data store servers to the data warehouse (i.e., the ADMS).

The paper defines extraction as the process of selecting and obtaining data from the operational data store servers (e.g., transportation management systems servers) to use within the data warehouse. This operation can take place in real-time (i.e., data is extracted from the servers to populate the information warehouse on a nearly continuous process), or in a delayed manner (e.g., once a day all of the data accumulated at the operational server is uploaded to the data warehouse). The main objective of the transformation step is to prepare the data extracted from the operational data stores so it can support analysis. In the final step (loading) the manipulated data is added to the data warehouse. The ETL (extraction, transformation, and loading) process can also be viewed as composition of four steps: data aggregation, data quality assessment, data imputation, and data characterization.

In the case of Virginia, the operational data stores collect field traffic data (volume, occupancy, speed) every 20 seconds from 471 freeway detector locations. For temporal aggregation, the Virginia ADMS uses a one minute interval (i.e., three loop-detector readings), which makes it easier to further aggregate at multiples of one minute (e.g., 1, 5, and 15 minutes, 1 hour, 1 day, 1 week, 1 month). Spatially, the Virginia ADMS data warehouse stores data at two aggregation levels, detector and station, in two different tables. For any given roadway and direction, one table stores data for all lanes separately, while the other one does it for a set of detectors on lanes across a transversal section of the roadway. The advantage of using two tables is increased flexibility.

The next critical step in the process is the determination of quality of the data. Data quality can be differentiated between quality assurance, which attempts to fix or identify defective data instead of ensuring the reliability and accuracy of the equipment, and quality control which emphasizes good data by ensuring selection of the most accurate detectors. Issues related to data quality have been extensively researched in a trilogy of papers sponsored by FWHA (Battelle, Texas Transportation Institute, and Cambridge Systematics, 2002a, 2002b, 2002a). In the case of Virginia, the data quality assessment procedure implemented in the ADMS is based on the Turochy and Smith approach (see Turochy and Smith, 2000), which allows to screen defective data.

Virginia has implemented a two-tiered approach to data imputation, the next step in the ETL process. A table that keeps historical averages is used to provide averages as imputed values in real-time. After that every night the Expectation Maximization technique (see Smith et al., 2003) is run in a batch mode to refine the imputations. The objective of the final step of the ETL process (i.e., data characterization) is to calculate and store a measure of the level of normality (i.e., the difference between the values of the current record from what would be expected at that particular date, time, and location) of each data record. Virginia uses the multivariate statistical quality control methodology to develop a single measure across a continuous range with respect to the difference from normal (i.e. mean conditions) for each record.

iFlorida

Another project that will certainly play an important role in the development of ADUS standards is the FHWA sponsored Surface Transportation Security and Reliability Information System Model Deployment. In March of 2003, the Florida Department of Transportation (FDOT) was selected to participate in this pioneering model deployment project –called iFlorida– which has as main goal *"to demonstrate the wide variety of operational functions that are enabled or enhanced by a surface transportation security and reliability information system"*.

Two objectives of this project, related to the ADUS standards development, are to expand and integrate existing data collection and monitoring systems, and to collect and share data. To accomplish these objectives, the University of Central Florida (UCF) has received over $1 million in funding to design and implement a Central Florida Data Warehouse (CFDW) for the storage and study of traffic information and the analysis of the impacts that the access to real-time traffic data has on drivers' behavior (FDOT, 2004). The UCF has been archiving loop detector-based speed data along I-4 since 1993, and is currently working on the expansion of that database —which collects, aggregates, processes, disseminates, and archives traffic data— to incorporate data from other sources (e.g., weather information).

This project was aiming at having a web site —http://www.iflorida.org— operational by end of 2003; however, as of today (July 2005), that web site has not been implemented. At this time, and to the best of ORNL's knowledge, no direct literature exists describing the implementation of this database and its data archiving capabilities and features. Nevertheless, some inferences can be made about the way the traffic data is currently archived by means of the papers by Abdel-Aty et al. and Abdel-Aty & Abdalla discussed in the previous section, which use the information stored in this database to develop their methodology. In fact, during the U.S. DOT sponsored ITS Data Quality Workshop held in Houston, TX, in February 2004, it was announced that the iFlorida model deployment would perform a variable speed limit trial on I-4 in Orlando during the following year. For this trial, the iFlorida project would integrate many types of data (such as for example weather forecasts and probe-vehicle information) into an expanded data warehouse, and would also put into practice a new *condition system* application. This new application, that would drive many of the new traveler information services, would recommend messages to be displayed on variable message signs as well as variable speed signs used for the VLS trial. Regarding the weather information, the CFDW will archive all the raw data collected as well as the processed data (i.e., 15 minutes to 48 hours time-sliced forecasts for each road segment covered).

ARTIMIS

Another project, ARTIMIS (Advanced Regional Traffic Interactive Management and Information System), provides information about traffic conditions, construction, and transit within the Cincinnati- Northern Kentucky Regional Area. Traffic information, covering 88 miles of highways, is collected through an array of different sources such as sensors (1,100 detectors, including inductive loops, video imaging and radar detectors; and 80 CCTV cameras), probe vehicles, a dedicated airplane, freeway service patrols, police, fire departments, and construction personnel.

The information collected by ARTIMIS is archived at different temporal and spatial resolutions. The "Segment Archiver" and the "Ramp Archiver" store speed, volume, and lane occupancy at 15 minute increments for all the system freeway segments and ramps, respectively, while the "Segment/Ramp snapshot file" does it at 30 second increments. The Segment/Ramp archivers run on a per corridor basis (57 corridors), collecting and storing data internally for every 15-minute interval, and creating a data file on a daily basis. Information is also archived in FHWA TMG format, including vehicle volume and classification at 60-minute increments, vehicle volume at five-minute increments, vehicle length in 15 bins with five-foot increments from 0-15ft to >70ft, and vehicle speed in 15 bins with five-mph increments from 0-20 mph to >85 mph. These files are stored on a server running the RedHat Linux system and can be accessed through an ftp (file transfer protocol) site.

The Kentucky Transportation Center is currently developing a prototype ADMS that will receive, process, and store the ARTIMIS and TRIMARC[3] data. The ADMS will provide an interface so that users can access the data and perform queries online, and will have a GIS engine to add spatial analysis capabilities. Data quality algorithms will be also incorporated into the ADMS to assure data quality control.

PeMS

Another important project is the freeway Performance Measurement System (PeMS) which is being conducted by the Department of Electrical Engineering and Computer Sciences at the University of California, at Berkeley, with the cooperation of the California Department of Transportation. The aim of this project is to collect historical and real-time freeway data from California freeways to compute freeway performance measures.

PeMS started in 1998 as a research project, and has between two to three terabytes of data available online with two gigabytes added everyday (Varaiya, 2003). The system collects traffic data from all districts with traffic management centers (23,138 loops), the California Highway Patrol CAD in real-time, and also includes TASAS (Caltrans Traffic Accident Surveillance and Analysis System) accident data for all districts.

As opposed to other systems, PeMS provides both quality assurance and quality control of the information stored in its ADMS. For example, PeMS produces daily diagnosis of every loop detector in the system and generates a report indicating which loops are not working and why (e.g., communications problems, detector station

[3] TRIMARC is a cooperative effort between the Kentucky Transportation Cabinet, the Indiana Department of Transportation and FHWA to improve the performance of the existing freeway system in the Metropolitan Louisville and Southern Indiana area.

failure, or loop malfunction). The report, plus the ability of accessing the data online, allows the loop maintenance personnel to determine almost immediately if a malfunctioning detector has been fixed. The data is also tested for completeness and accuracy. As each sample arrives at the database, PeMS determines in real time if a data element is missing or if its value is suspect. Although each missing or incorrect sample is replaced with a new value imputed from adjacent samples, the raw data is still retained.

The PeMS website, which has an average of 4,000 hits per day, provides a series of built-in analytical capabilities to support a variety of uses at different levels of data access. The system is composed of a bottom layer, which has as its main function the administration of the database. The middle layer —which works in real time— aggregates 30-second flow and occupancy values into five-minute intervals for each lane; computes the speed for each lane; aggregates lane-by-lane flow, occupancy, and speed across all lanes at each detector station; calculates performance measures such as delay, vehicle-miles-traveled, vehicle-hours-traveled, and travel times. Finally, the top layer comprises many built-in applications, including graphs and contour plots.

Other Implementations of ITS Data Archiving

An informal survey conducted by the ASTM ADUS Committee revealed several ITS data archiving implementations that are currently in operation across our nation. These include the following:

State of California, City of San Jose

System: TRADAS

Data: loop detector data is gathered from their TransCore 2000 Urban Traffic Management System, and then grouped as "pseudo-ATRs" (Automatic Traffic Recorders). The system performs quality control, summarization, factor calculations, etc., and puts the results in an Oracle database. The statistics from the ITS system are used to calculate adjustment factors to correct short-term counts, which are also placed in the database. In summary, the system is a fusion of an ITS ADMS and a traditional traffic counting program.

State of Washington through the University of Washington

Currently Deployed: freeway ADMS, in active use since 1999.

Research Projects:

- *CVISN Data Archive:* tag data to be used to monitor long distance travel times, and for monitoring urban freeway travel times in the Vancouver, WA metro area.
- *Arterial Signal Operations Data Archiving:* to capture arterial data to manage the performance of the arterial.
- *City of Lynnwood (north of Seattle):* similar system as the previous one but with different types of detection and with a Transit Signal Priority (TSP) project evaluation effort.
- *Probe-vehicle Database:* volunteer trucks carry GPS devices; the data collected is used to monitor roadway performance in the Puget Sound region (i.e., an ITS Archive, but it doesn't fit the definition of an ADMS).

Future Deployments: The Eastern Region office of WSDOT is planning to implement a version of PeMS to store and analyze some arterial and freeway data from Spokane, WA. WSDOT headquarters group is studying using PeMS as a central repository for statewide traffic data for the statewide emergency response center.

State of Washington (2)

City of Portland (through the Portland State University): Tri-Met, transit data.
City of Seattle (through the University of Washington): King County Metro, transit data.

State of Washington (3)

King County Metro
Data: archives APC (Automatic Passenger Counter) data since 15 years ago; and AVL (Automatic Vehicle Location) data since 4 years ago.
System:

- *APC Data:* Oracle database.
- *AVL Data:* Informix.
- *Data Mining:* the data can be, and is, mined for many purposes using various tools, including SQL, Web Focus, and Access, as well as GIS tools that allow accessing the data spatially.

State of Alaska

Data: Alaska is archiving WIM (Weigh-In-Motion) data.
System: Oracle database with a web interface (allows for data extracts of vehicle classification and volume data for HPMS and other federal reports).
Standards: used ASTM standard for this development (started small with 2 WIM sites and will addition other WIM sites data and CVISN data in the future).

State of Arizona through the University of Arizona

City of Tucson: archiving of ITS data in the form of transit operating data (vehicle locations, passenger counters, etc.), for the Sun Tran system.

State of Missouri (Missouri DOT)

Status: currently designing a statewide ADMS (just held the stakeholder meetings in accordance with the ASTM standard).

State of Wisconsin

Milwaukee Traffic Operations Center (MTOC): similar to San Jose, MTOC makes pseudo ATR files from the raw data downloads of the traffic controllers.
System: the data is processed through TRADAS and generates standard traffic statistics which are then stored in an Oracle database.
Other Data Archiving Activities: MTOC also was archiving (but non-permanently) the summarized counts from their sampling of the installed sensors. A comparison of this data to the data later downloaded from the controllers indicated a high level of error in the sampling database. At the present, MTOC is assessing the feasibility of adding a quality control system to their normal program.

Appendix B: ITS Deployment Tracking Survey

This section includes a summary and analysis of the data archiving practices of freeway and arterial management centers across the 78 major metropolitan areas of the nation. The archiving information presented below was obtained from the FHWA-sponsored 2002 ITS Deployment Tracking database.

Of the 151 freeway agencies that participated in the 2002 survey, 111 responded to the question of whether or not they had in place an ADMS system as described in the National ITS Architecture. Forty-six agencies (41.4%) responded affirmatively. The same question was asked to 526 arterial management agencies. Out of the 239 that answered this question, 30 (12.6%) said yes and 209 (87.4%) no.

As discussed before, one of the critical components of any data archiving system is its information quality control subsystem. When asked if an assessment of the data quality was performed, 28.9% of the freeway agencies which had an ADMS in place said "yes", while only 7.3% of the agencies with no ADMS in place responded affirmatively. These percentages were 8% and 4.1% for arterial agencies with and without an ADMS. Figure B1 shows these results graphically.

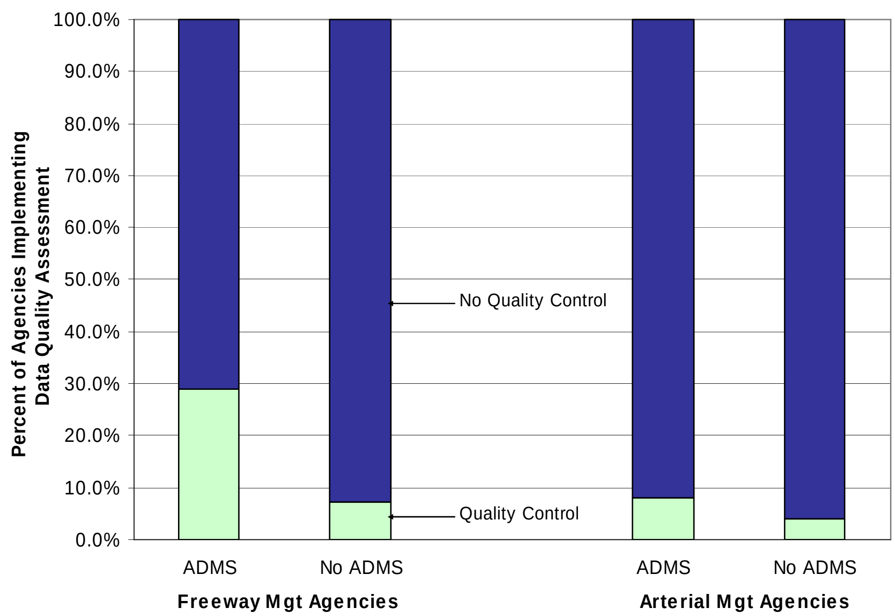

Figure B1. Data Quality Assessment among Freeway and Arterial Management Agencies with and without an ADMS Deployed

The agencies surveyed also provided information regarding whether or not they processed the data before archiving it. Out of the 46 freeway agencies with an ADMS deployed, 76.1% archived raw data only or both raw and processed data, while this percentage was only 30.8% for agencies without an ADMS in place. For arterial management agencies, the figures were 50.0% and 12.9%, respectively. Table B1 below presents this information in more detail, while Figure B2 shows the same results when the quality control dimension was added. In that figure it is possible to see that several agencies indicated that they save processed data only and do so without quality control, which is not an advisable practice to follow.

Table B1. Percent of Total Freeway and Arterial Management Agencies (with and without ADMS) by Type of Data Archiving Manipulation.

Type of Data Manipulation	Freeway Mgt Agencies		Arterial Mgt Agencies	
	with ADMS	w/o ADMS	with ADMS	w/o ADMS
Archives Raw and Processed Data	47.8%	15.4%	36.7%	7.2%
Archives Processed Data Only	19.6%	20.0%	36.7%	13.4%
Archives Raw Data Only	28.3%	15.4%	13.3%	5.7%
Neither	4.3%[a]	49.2%	13.3%[a]	73.7%

[a] These agencies (2 freeway and 4 arterial management agencies) are either not archiving the data yet — although they acknowledge they have an ADMS in place—, or did not understand the question correctly.

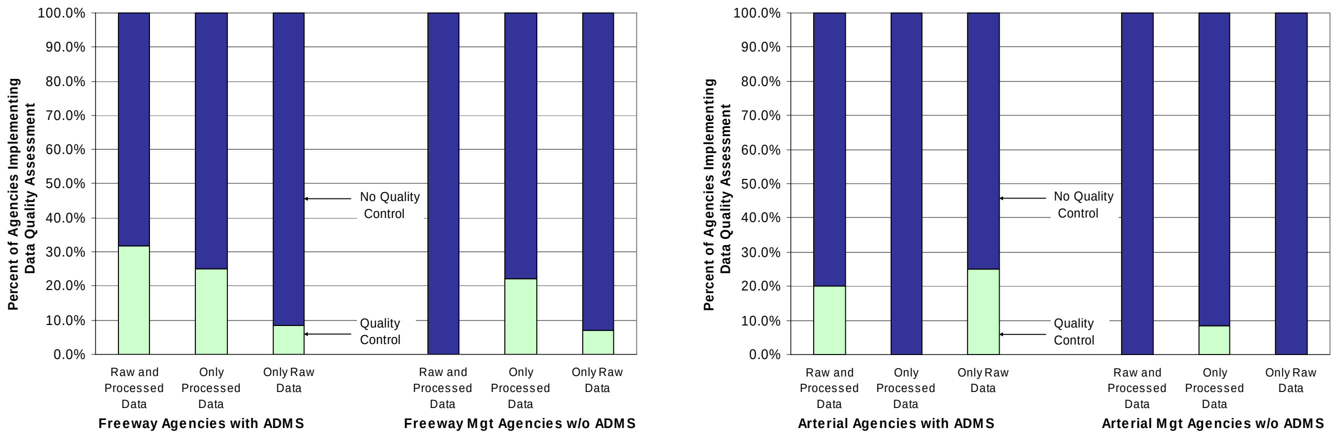

Figure B2. Data Processing before Archiving and Data Quality Assessment among Freeway and Arterial Management Agencies with and without an ADMS Deployed

The ITS Deployment Tracking survey also contained questions regarding the type of media used in archiving the data. Of the 46 freeway agencies with an ADMS in place, 71.7% archived data in digital form, while the percentage was only 30.8% for agencies without an ADMS deployed. For arterial management agencies, the figures were 63.3% and 10.5%, respectively. Table B2 below presents this information in more detail, while Figure B3 shows the same results when the quality control dimension was added. That figure illustrates the fact that those agencies that save data in several types of media are also more likely to perform quality control assessments.

Table B2. Percent of Total Freeway and Arterial Management Agencies (with and without ADMS) by Type of Data Archiving Media.

Type of Data Archiving Media	Freeway Mgt Agencies		Arterial Mgt Agencies	
	with ADMS	w/o ADMS	with ADMS	w/o ADMS
Archives Data Digitally Only	39.1%	6.2%	33.3%	3.3%
Archives Data on Paper Only	17.4%[a]	16.9%	23.3%[a]	21.1%
Archives Data Digitally and on Paper	32.6%	24.6%	30.0%	7.2%
Archives Data on Other Media	10.9%	52.3%	13.4%	68.4%

[a] These agencies (8 freeway and 7 arterial management agencies) are either not archiving the data yet — although they acknowledge they have an ADMS in place—, or did not understand the question correctly.

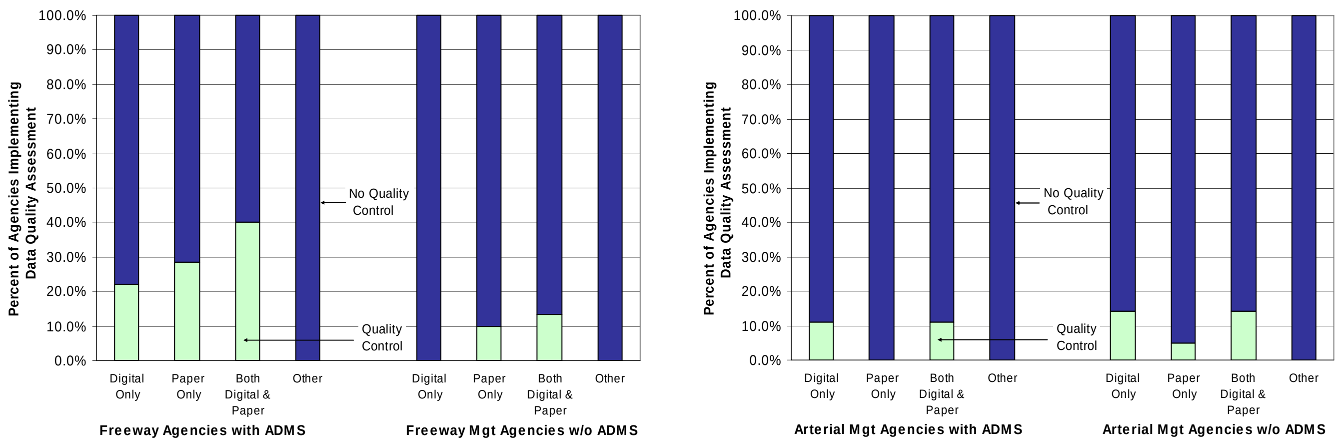

Figure B3. Data Archiving Media and Data Quality Assessment among Freeway and Arterial Management Agencies with and without an ADMS Deployed

As discussed in the main body of this report, in the development of real-time traffic safety applications for freeways researchers used mostly information about traffic volumes, traffic speed, and lane occupancy. This is explained, to a high degree, by the fact that this is the only type of traffic information available (SAIC is currently concentrating in determining what other traffic parameters and measures of effectiveness may be of relevance in the development of real-time safety applications). The ITS Deployment Tracking survey asked the participating agencies what type of data they collected and archived. Of the 111 freeway management agencies that answered the survey, 64.9% collected and archived traffic volumes, 54.1% archived traffic speeds, 40.5% lane occupancy, and 42.3% collected and archived vehicle classification information. For the 239 arterial management agencies surveyed, the four types of data that were archived by most agencies were traffic volumes (31.0% of the agencies), traffic speeds (20.9%), traffic signal phases (23.0%), and turning movements (16.7%). Figures B4 and B5 show this information in more detail for agencies with and without an ADMS in place and with and without quality control assessments.

Other type of information was also collected by the freeway and arterial management agencies as shown in Table B3 below. Some of the information collected is relevant to the real-time traffic safety applications that were discussed in the main body of this report (e.g., weather information).

Figure B4. Type of Data Most Frequently Archived and Data Quality Assessment among Freeway Management Agencies with and without an ADMS Deployed

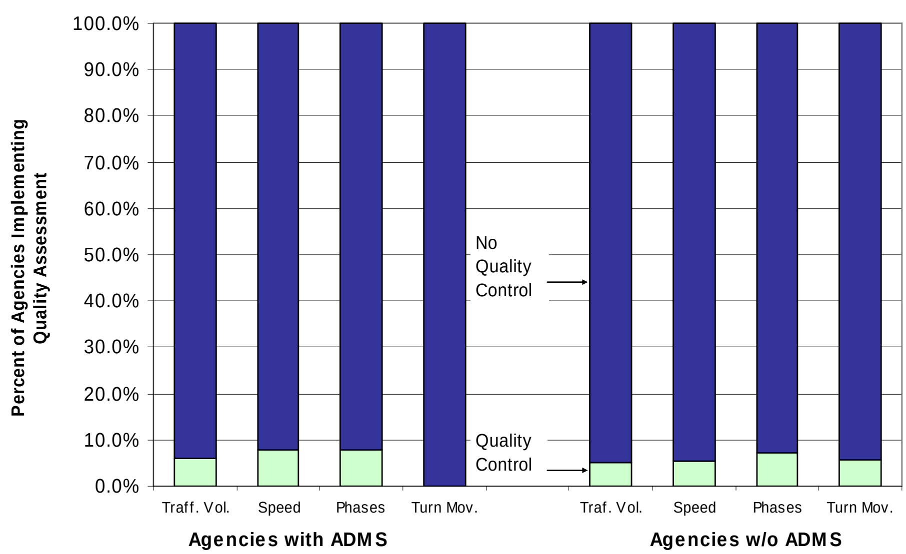

Figure B5. Type of Data Most Frequently Archived and Data Quality Assessment among Arterial Management Agencies with and without an ADMS Deployed

Table B3. Percent of Total Freeway and Arterial Management Agencies (with and without ADMS) by Type of Data Archived.

Type of Data	Freeway Mgt Agencies		Arterial Mgt Agencies	
	with ADMS	w/o ADMS	with ADMS	w/o ADMS
Traffic Volumes	93.5%	44.6%	56.7%	27.3%
Traffic Speeds	84.8%	32.3%	43.3%	17.7%
Lane Occupancy	73.9%	16.9%	26.7%	10.0%
Vehicle Classification	54.3%	33.8%	10.0%	8.1%
Travel Time Information	30.4%	1.5%	10.0%	1.9%
Road Conditions	30.4%	15.4%	10.0%	1.9%
Weather Conditions	26.1%	13.8%	3.3%	2.9%
Video Surveillance	8.7%	1.5%	---	---
Turning Movements	---	---	16.7%	16.7%
Queues	---	---	6.7%	1.9%
Phasing/Cycle Lengths	---	---	43.3%	20.1%
Emergency Veh. Signal Preemption	---	---	10.0%	11.5%
Transit Veh. Signal Preemption	---	---	6.7%	1.0%
Incidents	---	---	16.7%	9.1%

About a year after the 2002 ITS Deployment Tracking survey was completed, the Association of Metropolitan Planning Organizations (AMPO) conducted a survey among its members related to ITS data archiving (AMPO, 2003). Out of the 60 MPOs that responded to the survey, 60% answered that ITS data archiving was a practice in their organizations. Forty percent of the respondents indicated that they had an ADMS in place that was consistent with the National ITS Architecture (46% said yes to this question in the 2002 ITS Deployment Tracking survey). Regarding the type of data archived, 35.0% saved traffic volumes, 26.7% archived traffic speeds, 10% lane occupancy, and 16.7% vehicle classification. These figures are somehow smaller than those reported in the ITS Deployment Tracking survey.

Appendix C: ADUS Standards Committee Activities (Jan 2004-Jul 2005)

One of the main tasks of this project consisted in following the activities of the ASTM ADUS committee (specifically, the ASTM Task Group E17.54.02.1) to identify areas in the data archiving standards being developed that may have an impact on future developments of highway safety applications. ORNL closely tracked the discussions of the ASTM Committee E17.54 during most of FY04 and FY05. The following subsections present detailed descriptions of these activities/discussions, which were included in three partial reports submitted to FHWA in Jan '04, Jul '04, and Jan '05. To summarize, as of July 2005 the ADUS Committee has published its first standard, the *ASTM E2259-03a*, *Standard Guide for Archiving and Retrieving Intelligent Transportation System-Generated Data* (ASTM, 2003), has released the metadata standard specification for archiving ITS-generated data, which is in its final reviewing stages, and it is currently developing the standard specification for archiving ITS-related traffic monitoring data (this document has just recently being released internally by the committee).

FY04 Activities

The schedule of activities proposed in the 2000 ADUS strategic plan included, among other tasks, the development of metadata standards. During 2004, the ASTM ADUS committee was expected to work on the metadata standard specification for archiving ITS-generated data, which was the companion of the ASTM E2259-03a guideline. The metadata standard would provide specifications of the description of the required ADUS data (in addition to those attributes required by other ITS data dictionaries). Parallel to that effort, the ASTM Task Group E17.54.02.2 was to develop a standard specification for archiving ITS-related traffic monitoring data, with a first draft scheduled to be released at end of 2004.

Committee Discussions Directly Related to the Metadata Standard

During the first half of 2004, the ASTM ADUS Committee chair along with some members distributed several existing documents on standards and metadata information as a starting point for the committee discussions. These discussions focused on the development of the metadata standard specification for archiving ITS generated data, as well as the standard specifications for archiving ITS-related traffic monitoring data.

The documents submitted to support these activities included the *NCHRP Project 8-37, "Standardized Procedures for Personal Travel Surveys"*, which has as the main objective the development of standardized procedures for improving the conduction, evaluation, and reliability of personal travel surveys. The committee was particularly interested in the way the geocoding of the information is addressed in this study. The NCHRP Project 8-37 is scheduled to be completed at the end of January, 2005; as way of preliminary project findings, TRB published the *NCHRP Circular 261*, *The Case for Standardizing Household Travel Surveys*, which includes some information about geocoding, although it seems to be at a level that is too coarse for ADUS purposes.

Another document submitted to the committee was the 1998 USAF white paper entitled *Report of the Verification, Validation & Certification (VV&C).* Specifically, the committee members were asked to concentrate on the section dealing with the data quality metadata template. Among other considerations, this section suggests that quality metadata should be included with the data in a way that makes capturing and subsequent access to the information most efficient, and also that the metadata associated with data system quality should be textual in format and its design should be sufficient to describe data inputs, internal data processes, transformations, and outputs to the user.

Combining metadata and geospatial information, a document describing the Chittenden County, VT, MPO Metadata was submitted to the committee member list. The document outlined data documentation for CCMPO's distributed AADT traffic database using the FGDC *Content Standard for Digital Geospatial Metadata* (FGDC-STD-001-1998).

A fourth document submitted as reference to the committee members was the *NTCIP 1209*, a data dictionary standard which defines data elements used for controlling and monitoring transportation sensor system (TSS) devices (i.e., those sensors that can detect and communicate certain traffic parameters, obtained not only from detection of automobiles, but also light-rail vehicles, pedestrians, and other modes of travel).

Related to the exchange of information among transportation applications, a reference to the NCHRP Project 20-64, *XML Schemas for Exchange of Transportation Data*, was submitted to the members of the committee. The main objective of this NCHRP project was to develop public domain XML schemas for exchange of transportation data, while providing a framework for the development, validation, dissemination, and extension of current and future schemas. Four main areas would constitute the initial focus of this project: a) Survey/Roadway Design, b) Transportation Construction/Materials, c) Highway Bridge Structures, and d) Transportation Safety**.** Afterward, the project would be expanded to encompass a broader set of schemas, including ITS.

As an example of the type of philosophy that the committee should adopt in developing the ADUS Standards, the NHTSA *Model Minimum Uniform Crash Criteria (MMUCC) Guideline* was submitted to the members. These guidelines are an attempt to encourage states to adopt standardized data elements for crash reporting, as collected by police officers investigating crashes. Since every state is different, and they have largely resisted adopting a national standard, the MMUCC's provides guidance on how to revise the crash data collection when that need arises for the states. In this way, a form of national standardization can be achieved over a period of time. Because the MMUCC is a guideline, it only provides very basic information (e.g., data elements, definitions, and valid values). The thinking of the ASTM ADUS Committee regarding the development of the standards at hand was on something slightly more detailed, but not exceeding a level of complexity that could detract users from adopting it.

Two options were proposed regarding the documentation of the data dictionary: the new ISO 14817: *Transport information and control systems — Requirements for an ITS/TICS central Data Registry and ITS/TICS Data Dictionaries* and the *IEEE Standard for Data Dictionaries for Intelligent Transportation Systems* (P1489). Both documents specify the framework, formats, and procedures used to define information exchanges within the ITS and Transportation Information and Control Systems (TICS) sector, as well as defining the content of the ITS/TICS central Data

Registry and Data Dictionaries, and the registration process to enter data concepts into the Data Registry.

An example of the P1489 format dealing with the probe-vehicle portion of the TMDD was circulated among the committee members for discussions at the scheduled Kansas City committee meeting in June 2004[4]. However, due to a very low rate of committed attendants, this meeting was postponed to coincide with the NATMEC conference held in San Diego at the end of June. For the California meeting, an outline of the standard (temporarily named *Standard Practice for Archiving ITS-generated Traffic Data*) was also provided to the committee members. The outline consisted of the following sections: Introduction, 1. Scope, 2. Referenced Documents, 3. Terminology, 4. Summary of Practice, 5. Significance and Use, 6. Archived Traffic Data Dictionary (ATDD), 7. Basic Data Structure, 8. Metadata Specifications, 9. Data Processing Procedures, and Annexes.

Other Committee Discussions

Other discussions carried out during the first six months of 2004 included topics such as how to handle missing data and whether or not new analysis tools could lead to the widespread usage of archived data. Regarding missing data, three possible courses of actions were proposed: a) omit those records; b) include those records with the proper timestamp but code a value for missing in the data fields; or c) include some sort of dummy record for each missing record. The consensus (three respondents out of four) was that option b) was the best approach.

In relationship to the whether or not the availability of new data manipulation tools promotes the usage of archived data, the opinions were divided. Some members concluded that the lack of a tool was "very much a hindrance to archive (data) use"; others indicated that the obstacle was the lack of data rather than the lack of a tool; while for some members the real impediment was to convince their district offices that archived data has real value.

FY05 Activities

The main activity of the ASTM E17.54 Subcommittee in the second semester of 2004 was a meeting held on December 6th and 7th, 2004 in Washington, D.C. (Note: the main conclusions of this meeting are presented later in this section). During the first part of the semester, the subcommittee was not funded and only minor activities were carried out. In September 2004, ASTM was asked by the Joint Program Office at FHWA to submit a project plan for the subcommittee to complete, in coordination with ITE and AASHTO, the Metadata and Traffic Monitoring Standards. On September 29th, the plan was approved and funds in the amount of $134,298 were sent to ASTM for the ASTM E17.54 ADUS Subcommittee to develop both standards.

The Work Plan document provides some detail regarding these two standards. The purpose of the *Standard Specification for Metadata to Support Archived Data*

[4] The sample files also included information on detectors (e.g., variables describing road, detector station, lane, and detector specific information) and traffic measurement (e.g., variables describing the detector from which the measurements were taken, date, time, type of measurement, etc).

Management Systems is to develop a common set of definitions and terminology for documenting the quality, lineage, content, and other characteristics of archived data generated by ITS sources. It is intended to be used by all levels of government and the private sector to document archived traffic monitoring data.

The Standard will establish the names and definitions of single and compound data elements to be used for these purposes, as well as the values that are to be provided for the data elements. Three basic types of metadata will be considered in the Standard. Those include archive structure metadata (i.e., descriptive information about the structure of the data archive and the archived data); processing documentation metadata (i.e., information describing the processes applied to the data from its source to its storage); and data collection system metadata (i.e., information regarding the equipment that provides the data as well as procedures and conditions under which the original source data were observed, surveyed, measured, gathered, or collected).

The other standard included in the Work Plan, the *Standard Specifications for Archiving ITS-Generated Traffic Monitoring Data*, will cover two forms of ITS travel monitoring data. The first one will be related to traffic information (including volume, speed, occupancy, headway, density, and vehicle classification, among other traffic characteristics) gathered by roadway equipment, while the second one will deal with travel time data measured by probe-vehicles and personal devices (such as cell phones and other wireless devices, satellite tracking, etc.).

The *Data Dictionary for Archiving ITS-Generated Traffic Monitoring Data Standard* will cover both metadata and supporting data to adequately describe the collection, processing, and storage of primary data elements (i.e., those data elements gathered by the ITS sources supporting the archive.). The metadata will include the attributes of data elements specified in IEEE P1489, as well as additional metadata. The elements in the data dictionary will cover aspects such as data measurement status (i.e., how directly the data element relates to measured conditions), initial collection source (including make and model of the equipment that collected the data), data quality control procedures used to process the data, data quality control results (e.g., error flags), data imputation to fill in missing data or to replace erroneous data, aggregated/summarized data statistics, and transformed data statistics. The data dictionary may also include recommended data records (logical combinations of data elements) as an aid in constructing databases for the ADMS.

Washington, D.C. Meeting (December ’04)

In early December 2004, the ASTM E17.54 Subcommittee met in Washington, D.C. to discuss courses of action to be taken in developing both the Metadata and the Traffic Monitoring Standards described above. Twelve members attended the meeting, with one person from FHWA, four state DOT representatives, two persons from local agencies, three from private companies, one from an academic institution, and one from ASTM. The subcommittee spent the first day discussing the Metadata Standard, while the following morning of was devoted to the Traffic Monitoring Standard.

Regarding the Metadata Standard, the subcommittee goal was to have the standard balloted by April 15th, 2005 at the subcommittee (17.54) and committee (E17) levels, which requires the next draft of the metadata standard to be finalized by late January to early February. The discussions focused on potential copyright issues associated with using the FGDC metadata standard as the basis to develop the ADUS Metadata Standard (which will be a copyrighted ASTM standard). On the technical side, the subcommittee discussed issues related to the documentation of linear referencing systems in the proposed standard, changing “data element” to “metadata element” throughout the document, and developing metadata for some selected datasets to test the draft standard (four members volunteered to develop FGDC metadata for Transit Data, Incident Data, Intersection/Network Data, and Traffic Data).

The Traffic Monitoring Standard was briefly discussed at this meeting. Some major points included: 1) naming this a “standard practice” (although the data dictionary portion would be a “standard specification”); 2) allowing for the archiving of vehicle-level records for detector data as well as probe data; 3) allowing for the archiving of turning movements at signalized intersections; 4) considering that currently ITS field equipment aggregates data at 20- or 30-second intervals to reduce the communication burden, but this may not be the case in the future; and 5) researching data processing (e.g., quality control) before they can be included in the standard. Some materials related to this Standard were distributed to the members at the meeting for further discussion. Those included data dictionaries for traffic location and traffic measurements and traffic data structure diagrams.

While preparing for their meeting in December ’04, the ASTM E17.54 Subcommittee looked at the NTCIP 1206 and NTCIP 1209 standards and other type of information that could be relevant to this project. The NTCIP 1206 standard –*Object Definitions for Data Collection and Monitoring (DCM) Devices*– defines the data elements used for the configuration control and status monitoring of transportation data collection devices. The DCM equipment processes sensor signals to gather information regarding the traffic passing over an array of sensors which is stored within the equipment as data files for future retrieval. The DCM equipment may be portable (e.g. to be deployed at a site for a data collection period as short as one day), or permanently installed for continuous monitoring.

The NTCIP 1209 Standard –*Data Element Definitions for Transportation Sensor Systems*– defines the data elements used for controlling and monitoring transportation sensor system (TSS) devices (i.e., sensors capable of detecting and communicating certain traffic parameters). A TSS could use a single loop detector, could be a video image processing system, or could use other sensing technology for detecting automobiles, light-rail vehicles, pedestrians, and other modes of travel. Included in this standard are Data Dictionaries that define data elements for communication between system nodes (e.g., central management application and field controllers).

Both NTCIP 1206 and 1209 standards have been reviewed and have been approved by the NTCIP Joint Committee. Their impact in data archiving is expected to be significant, since they will induce the creation of data collection devices that have an open architecture for communication and data retrieval. With the data being available from the collection device in an open and known format, users and archivists should be able to obtain and utilize traffic data without the need for a multitude of proprietary data readers and converters. This will play a very important role for safety related applications, particularly at the intersection level.

The ADUS Subcommittee also discussed considering the AASHTO's TSIMS (Transportation Safety Information Management System) as another source of information for their activities. TSIMS is an Enterprise Safety Data Warehouse aimed at enhancing and extending the safety analysis capabilities of current state crash records information systems. This would be achieved by integrating crash records with other safety information maintained by each state. The project has also produced a draft Model Highway Data Dictionary that could become a de-facto standard for State DOTs.

Santa Fe, NM Meeting (March ’05)

The following subcommittee meeting was held on March 23rd and 24th, 2005, in Santa Fe, NM. The focus of this meeting was primarily to discuss the development of the metadata standard, which was becoming ready to be sent out for balloting on April 15th to the subcommittee members who are also ASTM members (non-ASTM subcommittee members cannot vote although they can provide comments). Discussions centered on the data quality section of the Standard, the requirement-level clause (i.e., mandatory, mandatory if applicable, and optional) associated with different elements in the Standard, the need to cross-reference to the Metadata Standard and other ADUS standards, and the inclusion of a Reader’s Guide at the front of the document referencing examples illustrating the application of the Standard (two examples, one dealing with traffic and the other with transit data, can be found in the appendix of that document; other examples will be included in a supplemental document). Emphasis was also given to the need of adding language to the Standard that would promote more complete metadata documentation for systems that are supposed to be interoperable.

The second part of this meeting was devoted to the Traffic Monitoring Standard. It was decided to include in this Standard information gathered from roadway-based detectors as well as probe vehicle information. Specific discussions focused on 1) the need to specify mandatory location referencing for detector locations and links, 2) how to specify link geometry (a basic requirement for archiving of data collected through probe-vehicles), 3) the inclusion in the standard of traffic monitoring data for signalized arterials, 4) the need to reference the NMEA (National Marine Electronics Association) standard for GPS devices, 5) the usefulness of imputations to fill in missing data, 6) the implementation of aggregation levels for traffic data (lowest level, as received from the source; either 5- or 15-minute intervals, and hourly), and 7) the type of statistics and performance measures (AADT, PHF, and travel time/delay performance measures) that should be included the standard.

At this meeting the subcommittee also discussed incident management and transit information (both of them important to the ADUS community) as the next standards to be considered.

Reno, NV Meeting (June ’05)

On June 13th, 2005, the ASTM E17.54, Subcommittee reconvened in Reno, NV, for a two-day meeting. The Metadata Standard ballot was examined (there were 10 affirmatives and one abstention with comment), and it was decided to send the draft standard to a few more knowledgeable individuals outside of ASTM for reviewing. Some small changes were suggested to be incorporated in the standard, including a better definition of “entity” and “attribute”, and the use of the FGDC valid values for the update frequency (in case the update frequency of an archive is different from those, it can always be specified as text within the FGDC rules).

Most of the meeting focused on the Traffic Monitoring Standard, specifically, the data dictionary. The discussions concentrated on the standard adopting the locally-defined location referencing system for the archive, the nomenclature for what constitutes the aggregation of traffic measurements in a direction (the consensus was to use the term "Station"), the proposed structure for the data dictionary, the incorporation of data related to traffic measurements at traffic signals and freeway ramps, how to handle reversible lanes since they may cause problems for detector station definitions, the development of Concept of Operations (ConOps) which specifies how the standard will be used, and the inclusion of specifications for traffic flow-based performance measures within the Standard.

Other Information

During the period covered in this report, two FHWA announcements were made that could have an impact on the findings and recommendations that the ASTM E17.54 Subcommittee will formulate. In September '04, FHWA announced that USDOT funding support for the Data Registry effort was discontinued. The decision was made based on issues such as problems with the harmonization effort, intellectual property rights, standard-development organizations' lack of interest in using the completed data registry or in continuing with the present data registry effort, and low benefit/cost ratios (i.e., significant financial resources requirements with modest utility achieved over the years).

The other announcement was the release of the FHWA final report on Traffic Data Quality Measurements on January 31st, 2005. This report provides a framework for deciding on the data quality measures for different applications; as well as guidelines and standards for defining and measuring traffic data quality and determining acceptable levels of quality. Data quality is an important issue for the ADUS Subcommittee and this report will certainly be included in their discussions of the Traffic Monitoring Standard.

Scheduled Activities

For the rest of the 2005 calendar year, the subcommittee has scheduled the following major activities:

- August: closure of the Traffic Monitoring Subcommittee ballot; working group meeting to discuss changes from the June meeting to be incorporated into the Metadata Standard and to review the results from the Traffic Monitoring Subcommittee ballot.
- September – December: redrafting the traffic-monitoring standard based on feedback from the August meeting.
- October: issuing of the revised metadata draft for E17 Main Committee ballot.
- November: the Metadata Standard Main Committee ballot closes.
- December: review of the comments from Main Committee ballot by the subcommittee and issuing final recommendations of clarifications and editorial changes before publication of the Metadata Standard.

TRB Subcommittee on ITS Traffic Data Archiving Meeting

The Subcommittee on ITS Traffic Data Archiving (part of the TRB Committee on Highway Traffic Monitoring AFD30/A2B08) met at the 84th TRB Annual Meeting on January 10th, 2005. Twelve persons attended the meeting, including one person from FHWA, one state DOT representative, one person from a local traffic agency, five from academic institutions, three persons from the private industry, and one from a national laboratory (ORNL).

The subcommittee discussed several issues. Included among them was the deletion from the subcommittee scope of the sentence *"the use of traffic data for real-time purposes is only incidentally covered"* since it was agreed that ITS data archiving should support any purpose, including real-time purposes (such as real-time safety applications). The subcommittee also agreed that a short Research Circular would be the best way to frame the issues related to ADUS research as well as building support for such efforts (two of the members agreed to develop a draft of that circular). The Chair of the ASTM E17.54 ADUS Subcommittee, Rich Margiotta, provided an update on the subcommittee activities (see above).

Appendix D: Standard Guide for Archiving and Retrieving Intelligent Transportation Systems-Generated Data

Note: The hard-copy version of this document contains the *Standard Guide for Archiving and Retrieving Intelligent Transportation Systems-Generated Data* published by the American Society of Testing and Materials (ASTM) International here.