Data Quality for Aggregation and Dissemination of DOT Traveler Information: An Analysis of Existing System Best Practices

by

Leann Koon
Research Associate II

And

Douglas Galarus
Program Manager / Senior Research Scientist

Western Transportation Institute
College of Engineering
Montana State University

For

COATS Phase VI
Western States Rural Transportation Consortium

June 17, 2016
(Revised and Finalized August 17, 2016)

DISCLAIMER

ACKNOWLEDGEMENTS

The research team thanks the California Department of Transportation for funding this research, with the support of the Western States Rural Transportation Consortium. The authors also thank the project steering committee, specifically Sean Campbell and Ian Turnbull of Caltrans, for their input to this work. Thank you goes to David Veneziano (Iowa State University) and Dan Richter (WTI) for their contributions to this research and related work such as the One-Stop-Shop. Thank you to student research assistant Fredric Vollmer for his work on the literature review.

Thank you to the following individuals for participating in the survey of Department of Transportation practitioners and being open and forthcoming with their responses.

- Lisa Idell-Sassi, Alaska DOT&PF
- Mike Jenkinson, California Department of Transportation (Caltrans)
- Dennis Jensen, Tony Ernest; Idaho Transportation Department (ITD)
- Brandi Hamilton, Montana Department of Transportation (MDT)
- Ismael Garza, Jim Whalen, Israel Lopez, Seth Daniels; Nevada DOT
- Galen McGill, Chris Wright; Oregon DOT
- Lisa Miller, Utah DOT
- Ron Vessey, Monica Harwood, Tony Leingang, Jay Wells, Chris Thomas, Hal Weiblen, Tom Stidham, Kerry Jorgensen, Mike Kress; Washington State DOT

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AASHTO	American Association of State Highway and Transportation Officials
API	Application Program Interface
ATIS	Advanced Traveler Information Systems
Caltrans	California Department of Transportation
CCTV	Closed Circuit Television
CFR	Code of Federal Regulations
CHP	California Highway Patrol
CMML	Canadian Meteorological Markup Language
CMS	Changeable Message Sign
CWWP2	Commercial Wholesale Web Portal, Version 2
DRISI	Division of Research Innovation, and System Information (Caltrans)
DOT	Department of Transportation
DOT&PF	Department of Transportation and Public Facilities
ESS	Environmental System Sensors
FHWA	Federal Highway Administration
FTP	File Transfer Protocol
IQR	Interquartile Range Spatial Test
IT	Information Technology
ITD	Idaho Transportation Department
ITE	Institute of Transportation Engineers
ITS	Intelligent Transportation Systems
IVR	Interactive Voice Response
KML	Keyhole Markup Language
MADIS	Meteorological Assimilation Data Ingest System
MDT	Montana Department of Transportation
NCEP	National Centers for Environmental Prediction (NWS)
NDEX	Nevada Data Exchange
NDOT	Nevada Department of Transportation
NERON	NOAA Environmental Real-Time Observation Network
NOAA	National Oceanic and Atmospheric Administration
NTCIP	National Transportation Communications for ITS Protocol
NWS	National Weather Service
ODOT	Oregon Department of Transportation
OSS	One-Stop-Shop
POR	Period of Record
QA	Quality Assurance
QC	Quality Control
QFLAG	MesoWest Quality Control Flag
RBA	Results Based Alignment
RSS	Really Simple Syndication
RTMC	Regional Transportation Management Center (Caltrans District 3)
RTSMIP	Real-Time System Management Information Program
RWIN	Road Weather Information Network (Canada)
RWIS	Road Weather Information Systems

RWMP	Road Weather Management Program (FHWA)
SAFETEA-LU	Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users
TMC	Transportation Management Center
TOC	Transportation Operations Center
UDOT	Utah Department of Transportation
USDOT	United States Department of Transportation
VDT	Vehicle Data Translator
VMS	Variable Message Sign
WSDOT	Washington State Department of Transportation
WSRTC	Western States Rural Transportation Consortium
WTI	Western Transportation Institute
XML	Extensible Markup Language

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The quality of data is a crucial consideration for the provision of meaningful traveler information. When drivers access traveler information that is up to date, correct, and accessible every time they need it, they will use it to make travel decisions which ultimately impact traffic management effectiveness. On the other hand, if for example, travelers see old or stale, incorrect information, they may be less likely to make travel decisions based on the traveler information or even access the information in the first place. This can significantly diminish the effectiveness of traffic management efforts (29).

The goal of this task was to analyze and document existing system best practices for data quality for the aggregation and dissemination of state department of transportation traveler information. To achieve this goal, the research team conducted a survey of DOT practitioners in western states, as well as a literature review on data quality within the transportation field. "Best practices" were documented, and recommendations and next steps were formulated based on applicability to Caltrans traveler information data and processes.

The research aimed to compile a collection of best practices for the aggregation and dissemination of quality traveler information. At the onset of the task, it was recognized that there may not be any so called "best practices" established for traveler information data quality. Neither the survey of DOT practitioners nor the literature review found a comprehensive, well-defined plan for unified, multi-dimensional approaches to quality assurance of traveler information. However, all of the DOT practitioners that were surveyed, as well as the literature reviewed relative to data quality in transportation, indicated in some way that quality data was important for safe, efficient operation of the transportation system, including provision of traveler information that is accurate, timely, and reliable. This observation is especially valid given the current environment that is increasingly focused on performance measurements, accountability and "smarter" operation of roadways.

Procedures and considerations relative to data quality were documented from the literature and the DOT practitioner survey. For example: automated sensor tests, maintenance and calibration, data sampling, well-defined measures and standards, operator observations, user error reporting, etc. Recommendations and next steps were made specific to Caltrans traveler information data and processes. There is an apparent need to define where the state DOT is in the traveler information environment. Furthermore, it was recommended that preliminary steps be taken to establish a data governance model that defines who owns what data, as well as uses and associated thresholds for the specified data. Relevant quality metrics and requirements should be clearly defined. Common statewide standards for data quality, performance, maintenance, and calibration should be established. It was also recommended that implementation of additional automated feeds should be investigated.

This task was sponsored by Caltrans under the auspices of the Western States Rural Transportation Consortium as a technology incubator project. Incubator projects are smaller research efforts that serve as a "proof of concept". Based on the results of an incubator project, further research on a larger scale may be pursued. This incubator project merely scratched the surface of the challenge of traveler information data quality. Additional research is needed to more thoroughly evaluate and establish best practices.

1. INTRODUCTION

The transportation industry has shifted from building more infrastructure to “smarter” operation of existing roadways in order to better manage challenges such as congestion, inclement weather, maintenance, increased traffic volumes, etc. This shift has been accompanied by a rapid advance in technologies and an increased demand for more high quality, real-time traveler information. As discussed in a review of the state of the practice for traffic data quality and quoted from the Federal Highway Administration’s (FHWA) first National Summit on Operations: “As more transportation agencies move aggressively toward system operations and performance measurement, the need for comprehensive quality data becomes imperative”(17, 30) .

Data quality for traveler information has generally been handled on an ad-hoc basis, with little or no provision for error notification other than perhaps through user-reporting of observed errors. The quality of data - for example, whether it is accurate, timely, and reliable - is a crucial consideration for the provision of traveler information. When drivers access traveler information that is up to date, correct, and accessible every time they need it, they will use it to make travel decisions which ultimately impact traffic management effectiveness (29). However, if for example, travelers access traveler information and see old camera images, obviously incorrect weather information (e.g., subzero temperatures when it is warm everywhere else in near proximity), or misspellings on sign messages that change the meaning of the message, then users are less likely to make travel decisions based on the traveler information or even access the information in the first place. This can significantly diminish the effectiveness of traffic management efforts. Even worse, if drivers use incorrect information to make travel-related decisions, more serious consequences could result.

2. BACKGROUND

Weather-related systems such as MADIS (Meteorological Assimilation Data Ingest System), MesoWest and *Clarus* have applied quality checks to weather sensor data, but these checks don't necessarily transfer to other sensor and data types. Further, these checks may not be applicable to department of transportation Road Weather Information System (RWIS) sites in the absence of data from additional sites. Some, including Caltrans District 2, have implemented measures of reliability based on network and file transfer performance. The District 2 Information Relay and the Caltrans Commercial Wholesale Web Portal (Version 2) (CWWP2) efforts have also included some checks for bad data in CCTV and other feeds. However, there does not appear to be unified, multi-dimensional approaches to data quality for aggregation and dissemination of DOT traveler information.

The process of providing accurate, timely, and reliable traveler information that effectively impacts traffic management, safety, and operations, is complex and rife with challenges. One of the core problems is determining which data quality descriptors to apply and how they should be used relative to traveler information, e.g. determining what to measure and how to measure it. In September of 2000, ITS America and the U.S. Department of Transportation established guidelines for traveler information data collection and quality. In the introduction to the report, the authors comment that the early vision for traveler information was simple – data would be collected by public agencies and disseminated to various devices and media outlets. However, "In hindsight, it is clear that the difficulty of collecting good complete and timely data, transforming data into information, then packaging, marketing and communicating that information to people and devices was underestimated" (32). In the era of big data and rapidly changing technology, this statement still holds true, possibly more so now than ever.

The quality of data is a crucial consideration in the process of aggregating and disseminating meaningful traveler information. Potential issues and problems with traveler information data and metadata include:

- Meta data – field element location, timestamps
- Old or frozen CCTV images
- Partial CCTV image, dark CCTV image
- Indication of camera being unavailable
- Duplicated cameras in different locations
- Misspellings on sign messages
- Camera settings visible to the public
- Color display (affects interpretation of road conditions)
- Incidents incorrectly located
- Chain-up control requirements
- RWIS data – e.g., surface temperature, error, missing, displaying a value when there is no sensor, depth and water level
- Incorrect data (temperature, precipitation, etc.)

To illustrate the challenge of providing quality data for traveler information, below are several examples of traveler information from the One-Stop-Shop (http://oss.weathershare.org/):

Figure 1 shows a CMS message from the Arizona DOT that includes two misspellings. While the message can still convey the necessary meaning to travelers, its effectiveness may be diminished by these misspellings. Even worse, a single dropped letter can potentially change the meaning of a message. See Figure 2.

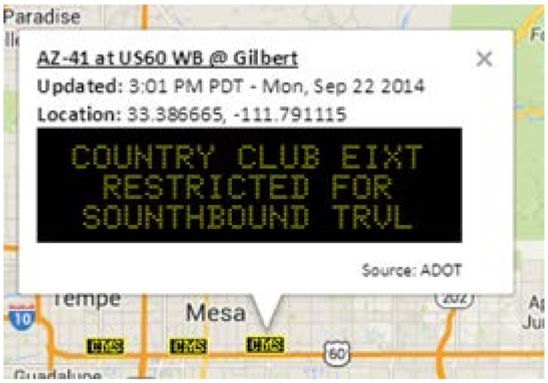

Figure 1: Arizona DOT CMS Message with Two Misspellings

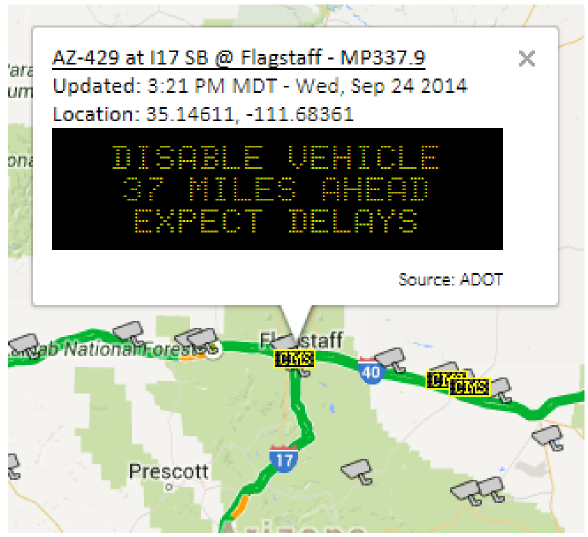

Figure 2: Arizona DOT CMS Message with a Missing Letter that Dramatically Changes the Meaning of a Message

Figure 3 shows a CMS message from Caltrans District 1 that is verified by Figure 4, which shows a camera image of the CMS showing this message. Caltrans District 1 has deployed a number of cameras in proximity to CMS to help verify that messages are correctly displayed on CMS. This example is shown not because camera verification of sign messages is a preferred approach, but to demonstrate the concern from DOT staff that messages may not be properly displayed on the physical signs.

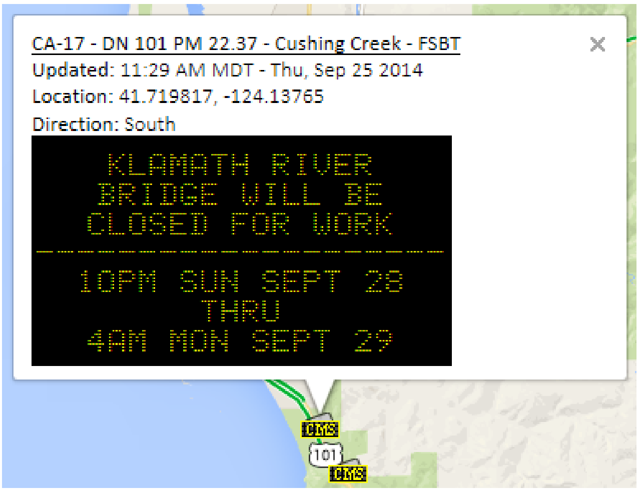

Figure 3: Caltrans District 1 CMS Message

Figure 4: Caltrans District 1 CCTV Image to Confirm CMS Message

Figure 5 and Figure 6 show cameras for which location metadata is obviously bad. In Figure 5, a WSDOT camera is incorrectly located several miles from its actual location along Interstate 90. In Figure 6, a Montana DOT camera is incorrectly located approximately 20 miles south of its actual location along Interstate 90. As a result, automated placement of these cameras on map-based traveler information systems is problematic, and detection likely requires user recognition of the problem.

Figure 5: Bad Location for WSDOT CCTV Image

Figure 6: Bad Location for Montana DOT CCTV Image

Figure 7, Figure 8, Figure 9 and Figure 10 show additional problems with CCTV images. Figure 7 shows Caltrans District 1 camera settings that should not be viewed by the public. Further, the camera image does not appear to show road conditions. At the very least, the camera is pointed away from the roadway, and there may be further color problems with the image. Figure 8 shows a WSDOT placeholder image indicating that an image from that CCTV location is not available. This image is helpful to the users of a traveler information system, to recognize that an image of the roadway is not available. Figure 9 shows a dark image from Caltrans District 4. While such an image is possible at sites with no nighttime illumination, this image is shown in the middle of the day. Figure 10 shows a partially rendered image from Caltrans District 10.

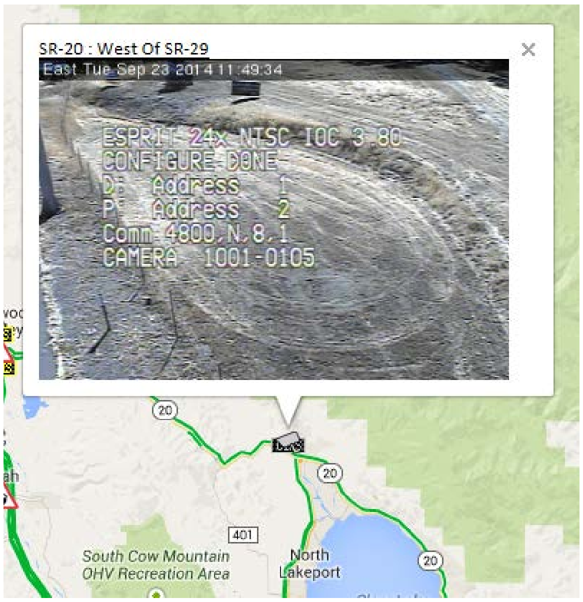

Figure 7: Caltrans District 1 CCTV Showing Camera Settings and a Strange Image

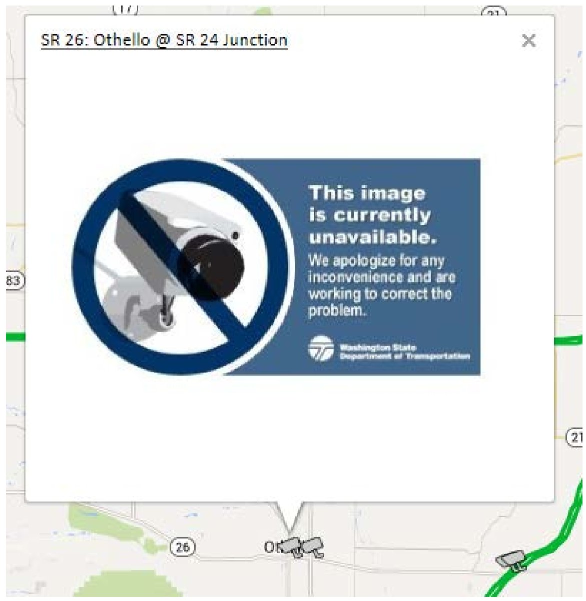

Figure 8: WSDOT Unavailable CCTV Image

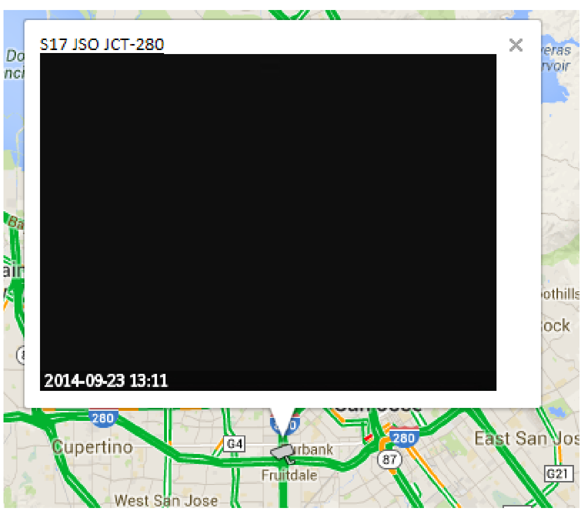

Figure 9: Caltrans D4 Dark Image

Figure 10: Caltrans D10 Partial Image

Figure 11 shows a complete image from Caltrans District 10, but there is an obvious problem with the colors in the image. The problem is severe enough that the weather condition (sunny, rain, etc.) can't be readily determined from the image. At the time this image was taken, it was raining, but one can't tell that from the image.

Figure 11: Caltrans D10 Image with Color Problem

Figure 12 shows an incorrectly located CHP incident. The coordinates are given as 0 degrees Latitude, 0 degrees Longitude, which places the incident in the Gulf of Guinea near Africa. Incidents like this one have gone largely unnoticed in the One-Stop-Shop since they result in points on the map in areas not normally viewed by typical users. Server-side checks to identify this data as erroneous have not been implemented. Regardless, there isn't a straight-forward option available to point out the problem to CHP.

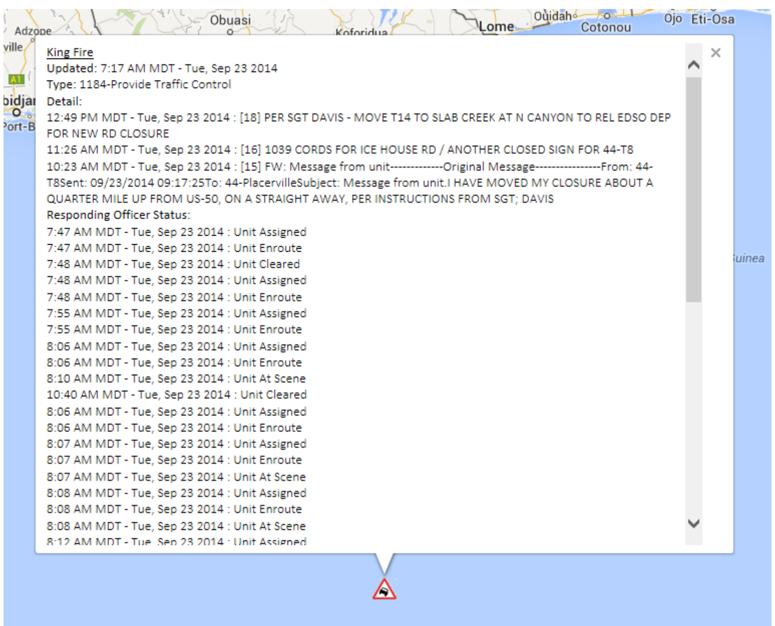

Figure 12: Incorrectly Located CHP Incident

Figure 13, Figure 14 and Figure 15 show apparent errors in RWIS data. Figure 13 shows a surface temperature of 0° F while the ambient air temperature is 63° F. Taken in the middle of the day in late September and in the absence of other nearby cold readings, this reading is almost certainly incorrect. Further, it results in the incorrect display of a "freezing" icon in the One-Stop-Shop. Figure 14 and Figure 15 show apparently bad values for "Depth" and "Water Level" at two different Caltrans sites. These values may correspond to indications of sensor error or missing values, in which case, the One-Stop-Shop should display a more user-friendly result.

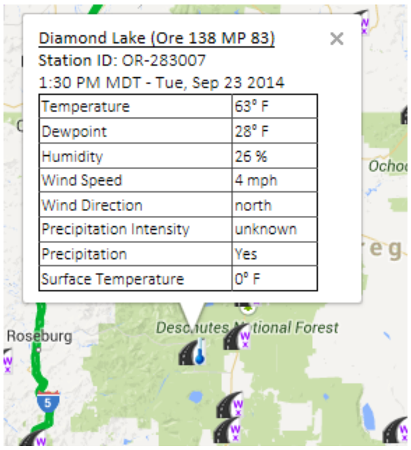

Figure 13: Incorrect ODOT RWIS Surface Temperature

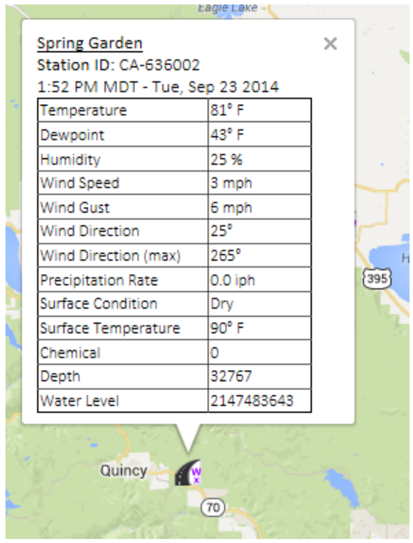

Figure 14: Caltrans D2 RWIS Depth and Water Level Values Reflect Error or Missing Values, Not Actual Sensor Readings

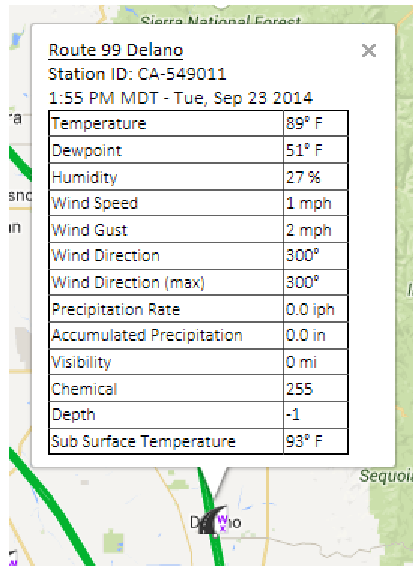

Figure 15: Caltrans D10 RWIS with Invalid Depth Value

There is an apparent need for the development of best practices for data quality for the aggregation and dissemination of state department of transportation (DOT) traveler information, as demonstrated by erroneous data in existing feeds. Taken individually, the presence of errors may be considered anecdotal. But recognizing that many errors can be present at an arbitrary time, there may be a more systemic problem in regard to data quality of DOT-provided traveler information that affects all state DOTs, with numerous potential points of failure. In light of this, the goal of this research was to analyze and document existing system best practices for data quality for the aggregation and dissemination of state department of transportation traveler information.

This task was sponsored by the California Department of Transportation (Caltrans) through the Western States Rural Transportation Consortium (WSRTC). Caltrans has been working on the challenge of traveler information data quality relative to QuickMap (http://quickmap.dot.ca.gov/) and the One-Stop-Shop (http://oss.weathershare.org/) traveler information websites, among other projects. Based on these experiences, they believe that systemic problems exist regarding data quality and feel it is a critically important issue that needs to be addressed. Other state departments of transportation anticipate benefit from this work as well.

To achieve the goal of this task, the research team conducted a survey of DOT practitioners in western states, as well as a literature review on data quality within the transportation field. "Best practices" were documented, and recommendations and next steps were formulated based on applicability to Caltrans traveler information data and processes.

3. METHODOLOGY

This research consisted of six tasks: *Project Management, Survey of DOT Practitioners, Literature Search, Document Best Practices, Recommendations and Next Steps,* and *Final Report.* This section includes a summary of the methodologies used for each task. More detailed descriptions can be found in the individual sections of the report and related documents which are referenced below.

Project management included quarterly reports to the Project Manager on both the progress of the work and the budget expended to date. The research team provided a status update to the members of the WSRTC at the annual meeting. The task was described on the WSRTC website (http://www.westernstates.org/Projects/Default.html) and quarterly updates on progress provided on the COATS project pages. A presentation was also given at the 2015 National Rural ITS Conference in Snowbird, Utah.

For Task 2 *Survey of DOT Practitioners*, the research team developed a survey instrument that could be used as either a guide for a telephone interview or to collect responses electronically. The survey questions were designed to identify the current state of the practice for quality control of traveler information data and investigate the accuracy of data and metadata. Practitioners were asked about quality control measures currently being employed by their state DOT and how data is handled throughout the process from collection to dissemination. Essentially: How is the traveler information provided; what is done to make sure the information is right; and how are errors/problems detected and fixed? The questions are outlined in Section 4 Survey of DOT Practitioners. The complete survey instrument, the introductory email correspondence, and a one page information sheet are included in the Appendix.

The research team developed a list of DOT personnel contacts for administration of the survey. Contacts were the person or people responsible for traveler information in their state. An introductory email message was sent to the contact list. Most of the contacts opted for a telephone interview and included other appropriate individuals on the call. The rest simply answered the questions in a Microsoft Word document and returned their responses electronically.

The research team conducted a literature search (Task 3) for best practices regarding quality control of traveler information data. To bound the task and adhere to the scope of work, the search was limited to information on data quality within the transportation field.

To document best practices (Task 4) for the aggregation and dissemination of quality traveler information, the research team combined the results of the Survey of DOT Practitioners (Task 2) and the Literature Search (Task 3). The best practices documented in Task 4 were analyzed for applicability specific to Caltrans.

Finally, this report fulfills Task 6 *Final Report*. This document summarizes the work completed for this task and includes data collection methodology, survey instruments, a summary of the information collected, and the presentation of related analysis, recommendations and next steps.

4. SURVEY OF DOT PRACTITIONERS

The Survey of DOT Practitioners was designed to investigate the accuracy of data and metadata, what quality control measures are used, and how the data are handled throughout the process from collection to dissemination. An introductory email message was sent to identified contacts or a general (traveler) information contact for state DOTs in the Western States Region. Those contacts were identified as being the individual responsible for traveler information in their state. Telephone interviews were conducted and/or written responses to the survey questions were collected. Representatives from the following state DOTs participated in the survey: Alaska, California, Idaho, Montana, Nevada, Oregon, Utah, and Washington. Other contacts were made with individuals in Arizona, Colorado, New Mexico, and Wyoming. It should be noted that there are varying levels and complexity of ITS deployments and associated traveler information around the region and within individual states. This relates to the volume of data in question, the number of field elements deployed, and the number and type of traveler information resources made available to the public. The extent of traveler information and ITS implementation is heavily impacted by funding, personnel, and need.

The following sections are a summary and analysis of the information collected through the survey. Detailed responses are included in Appendix E: Detailed Survey Responses.

4.1 General Responses

All of those surveyed were forthcoming in their responses and expressed significant interest in this task. With that said, there were as many different answers to the survey questions as those interviewed. Each state DOT is organized differently and manages and uses their various ITS assets differently; often there are significant differences even between districts or regions within a state. However, there was some agreement that the problem of data quality may be endemic. Taken a step further, California/Caltrans specifically commented that they have no statewide defined procedures for data quality control and issue resolution. Though not stated directly, this is likely true for other states as well. While Transportation Management Centers (TMC) or 511 traveler information websites for example, may implement some form of data quality control procedures to varying degrees, there did not appear to be a focused effort for data quality from the state DOT. That is not to say that quality control is not being done at the state level, but rather in many cases it is being done ad-hoc based on the application or implementation. While the point could be made that this is not necessarily bad or ineffective, it does present different challenges for statewide or regional traveler information.

This begs two more related questions or issues that were mentioned or alluded to. One is the question of data ownership. In one state, an office other than the traveler information office owns the data. In some cases, urban areas or regions have their own systems so the state DOT doesn't monitor that data. When issues started to surface in this instance (e.g., CMS signs located in China), the traveler information office worked with each of the data owners to develop internal data quality control processes. The question of data ownership brings to light other challenges as well. The use of the data determines what thresholds are set and different users may have different thresholds. Or this issue becomes as one survey respondent said, "what one would call good data, another would call bad data."

Secondly is a question of where the DOT should be in the traveler information environment. As one state responded, the DOT has limited flexibility or ability to move quickly. This makes it challenging to quickly design and publish an app for example, because it would already be obsolete by the time it was out for public use. In this case, the survey respondent indicated that the DOT's "niche" may be to provide quality data to others and allow the traveler information environment to sustain it from that point.

With a few exceptions, there is a general lack of automation for quality checking data that is used in traveler information systems. More automated processes could help with the obvious problems such as stale CCTV images or invalid sensor readings. Automated tools could also more proactively help identify a piece of equipment that is failing or reaching the end of its lifecycle by providing a data trend over time.

4.2 Traveler Information Data from Collection to Dissemination

As noted earlier, there is variety among and even within states as to the type of traveler information data that is collected and made available to the public. This is due to a number of different factors, including need, funding, and staffing. In one state for example, survey respondents said that "three regions are heavily invested, three regions do the best they can." As an example, in those regions with fewer ITS resources, if a sensor is not reporting, it is just turned off. In this state, if a region wants to improve something, they must do so with their own funds (vs. state funds.) Some of the variety can surely be attributed to rural versus urban road networks, although that does not necessarily track directly to need. Montana for example is mostly rural with a few small-sized cities. It has no transportation management center. In contrast, California has several major metropolitan areas in addition to rural areas, and 12 TMCs or operations centers.

The responses to how traveler information data gets from the field to the public varied accordingly. In general terms, for several states, data is collected at a district or local level and then input to a central system which in turn publishes it for public consumption. One state feeds all of its data to a third party vendor which in turn makes it available to the public. A few states skip the local collection step and instead data go directly to a central database system which publishes it for traveler information.

Traveler information outlets include but aren't limited to:

- 511 phone system,
- 511 website,
- Statewide traveler information websites,
- Road condition maps,
- Social media (Facebook, Twitter),
- Mobile websites,
- Media outlets such as local TV stations,
- Link to transit operators,
- Smart phone apps,
- Changeable Message Signs, and
- Highway Advisory Radios.

Survey Question: Generally and briefly, how is DOT data used for traveler information handled from collection to dissemination? How does data get from the field to the public view (online, 511, TV/radio, social media)?

- **Alaska:** "All of the information on 511 is manually input except for the RWIS data and camera images, NWS [National Weather Service] weather watches and Nixle.com feeds from local law enforcement. Those are all automatic. We do offer XML feeds of the 511 driving conditions, planned events and alerts for anyone interested in sharing the information with their users or travelers." "The system is maintained, operated and hosted offsite."
- **California:** Data from field elements is aggregated in the district TMC (could be multiple systems) by respective ITS unit. The information is shared by publishing to the Commercial Wholesale Web Portal 2 (CWWP2). DOT traveler information office pulls the data from the portal and repackages it for the web page and the Interactive Voice Response (IVR) phone system. The data is the same for Caltrans and the public.
- **Idaho:** Multiple sources of information, both auto feeds and manual entry, are all fed to a third party vendor which creates events. Outputs include 511 phone system, websites (mobile version, customizable high bandwidth version, low bandwidth version for commercial vehicle operators), link to transit operators, social media (Twitter), and a smart phone app. The data can be pushed to the public or the public can passively receive it.
- **Montana:** Drivers radio road conditions to division offices. Division/district staff enter the information into a database which disseminates it to various outlets (511, web, mobile, social media).
- **Nevada:** Central System Software collects information from ITS devices and populates the 511 system as well as public device interfaces. Data is published to an FTP site. With the newly created Nevada Data Exchange, devices are polled by the transportation management system, sent to the NDEX, and then distributed everywhere else.
- **Oregon:** The system is highly automated and integrated, very little is done manually. Once the information is in the system, it goes everywhere (TripCheck, 511, TV). Those entries that are done manually are reviewed by a TOC operator before it goes out. When the TOC operator enters something, the TOC software builds the message for them and it can be edited or left as is.
- **Utah:** "UDOT's traveler information outlets include Twitter, Facebook, 511 phone line, CMS, a website, a mobile website and a smartphone app. We also share camera feeds with local media. UDOT has a fiber optic network of over 1800 miles which helps get communication/messaging to rural devices."
- **Washington:** In the Northwest Region, data is collected from the field and sent to the central server with central software which has multiple pulling and polling functions. The data is repackaged and published to a web accessible database that can be queried. The data is then pushed out to the various outlets. In the South Central region which is more rural, data from the mountain passes is manually collected and keyed in by operations personnel. It is then pushed to the public (i.e., 511 system) via the same database. RWIS/weather data are not on social media. The Twitter feed is automated for the Olympic and Southwest Regions.

4.3 Issues with Traveler Information Data Quality

As expected, DOTs have experienced a variety of issues related to traveler information data quality as listed below.

Survey Question: What issues have you had, if any, with data quality and traveler information? (Describe.)

- CMS feeds (e.g., misspellings or typos, wrong abbreviations, multicultural challenges, sign limitations)
- Equipment failure or equipment end of life
- Cameras: offline, communications, power issues, potential old or partial picture, pixels out.
- Field element location
- Timeliness (e.g., old or stale information, conditions being reported are inaccurate, miss a winter road report, wrong start date on construction project)
- Operator input error
- Sensors, sensor calibration
- Connectivity and communications
- Missing information regarding an incident, construction, lane closures, etc.

4.4 Traveler Information Data Quality Process

The survey sought to outline the process used by state DOTs in the western region to identify and resolve problems with traveler information data or metadata. It addressed who handles quality control issues, whether the process was automatic or manual, and a brief description of the process. Survey respondents were asked how problems are resolved and how long it generally takes to reach a resolution once a problem is identified.

In many cases, a problem is identified by the traveling public who call, email, or use social media to report it. *Depending on the problem*, it is directed to the appropriate office or individual. With more fully automated systems, software generates an auto alert for a potential problem which is directed to traveler information, TMC, and/or system personnel. In some cases, a third party vendor is checking for sensor values that exceed pre-set thresholds. In other instances, a problem is caught by operations or maintenance personnel, or others such as state patrol troopers. In one state, RWIS technicians monitor their own equipment. One survey respondent noted that elements and data feeds are monitored more closely after system changes are implemented. Another state added that each of the TOCs have their own quality assurance processes and each month a sample of incidents is reviewed for accuracy and procedure.

Quality control processes for traveler information data range from completely manual, to combination automatic and manual, to mostly automated. The level of automation is dependent on resources (i.e., staff, funding, etc.) as well as type and complexity of systems. Overall, manual quality control processes were more common, but survey respondents expressed or implied a desire for more automated quality control and/or indicated the state DOT was moving in that direction.

How often manual quality control processes are conducted and by whom again varies depending on resources as well as system type and complexity. In some cases, data is checked a couple times per day every day. In other cases, it is done once per day or one staff member is assigned to monitor the system weekly. In California, some manual quality checking is done in "responsive mode," for example, manual checks are done when a series of problems occur over a short period of time. Several states emphasized the importance of training for consistency and accuracy, particularly related to CMS messages. Some respondents indicated that manual data entry processes are simplified as much as possible, noting that even those manual entries have canned information that can be used. Again, particularly related to CMS messages, as a matter of standard practice TMC/TOC operators review manual entries – "Read twice, post once." Those entering data manually are "self-policing and embarrassed if they make a mistake." A couple states commented on spell checker functionality within the system and one state mentioned that their construction/road maintenance information system includes a method that provides a "public preview" option that shows how the public will see the information they've entered.

Once a problem is identified and defined, the type of problem generally determines how it gets resolved and how long it takes to do so. In some cases, operators fix a problem directly (i.e., sign message misspelling) or take the element out of service while the problem is corrected. *Depending on the problem,* a work order may be issued or the problem handed over to a subject matter expert and/or a technician. Problems could be directed to maintenance contractors, regional ITS personnel, IT maintenance department, third party vendors/contractors, or other appropriate personnel. One survey respondent said, "One user can start/stop everything; everyone knows what to do and can fix it themselves." Another commented that they have no maintenance personnel dedicated to their RWIS field elements and that in rural areas, secondary road elements often have a lower priority. One state did mention that their traveler information systems operate redundantly and that systems are not completely taken down while a problem is resolved. Again, *depending on the problem,* resolution could be done "instantly" or within minutes while more complicated issues could take up to a week or longer. In some cases, the public may not even know there was a problem.

Traveler information data control is a concern to state DOTs, with one state commenting that they "make sure that stale information is not presented to the public." Certainly, some proactive quality control procedures are utilized, particularly related to a handful of automated feeds. However, the traveler information element, system type and complexity, determine the level and form of quality control processes - ad hoc application of quality control. There does not appear to be unified, multi-dimensional approaches to data quality for aggregation and dissemination of DOT traveler information.

4.5 Quality Dimensions

Each state DOT was asked to describe any quality dimensions that are used to assess system and site performance. General dimensions include accuracy, timeliness and reliability. For example:

- Accuracy – is the data correct and how do you know it is correct?
- Timeliness – is the data made available in a timely fashion and is the data current?
- Reliability – percent up-time (the percentage of time the data meeting the accuracy and timeliness criteria is available for use).

In response to this question, most states mentioned the Real-Time System Management Information Program (RTSMIP) (Section 1201 SAFETEA-LU, 23 CFR 511) (8, 9) and the challenges associated with meeting its reporting requirements. The federal rule includes minimum requirements for traffic and travel conditions, but as noted by those interviewed, there are no specific metrics or definitions for data quality for RWIS or CMS for example. As one state commented relative to quality dimensions and meeting the RTSMIP requirements, "This has been problematic - how to measure accuracy of … e.g., incident data … we're reporting what we know but how do we know if that's all." In other words, the requirements can be met, but if the quality of the information can't be quantified or defined, the requirements can be met with garbage data.

In general, no formal processes are in place to characterize system/site performance. "We are very short on metrics for traveler information performance measures," stated one survey respondent. One commented that it was an "operations thing" and others noted a common theme that performance measures are applied at varying levels depending on district/region versus statewide. Still another said "If it isn't measured, then it isn't fixed." California called for a data governance model and a common statewide standard, and expressed a need to automate and define relevant metrics. Another state mentioned that public feedback is collected and a public survey is conducted every two years. On the other hand, one state did respond that the DOT was working on performance measures across the organization and that data quality was an important component of that program. "… is moving toward a Results Based Alignment for service delivery framework from which we will evaluate (measure) the contribution (efficiency and effectiveness) of the services we deliver in support of our mission. Data quality is an essential ingredient of the RBA. We look at six elements for data quality (three are noted above): Accuracy, Accessibility, Completeness, Coverage, Timeliness, and Validity/Reliability."

Specifically in regard to accuracy, timeliness, and reliability, survey participants responded with the following comments.

Survey Question: What quality dimensions are used to characterize the traveler information system/site performance? Describe. E.g., accuracy, timeliness, reliability.

<u>Accuracy</u>

- Visual confirmation.
- Check portable license plate readers/travel times one to three times per year.
- No way of tracking through 511.
- District centric.

<u>Timeliness</u>

- Update times based on federal requirements.
- Many data feeds updated every 15 minutes.
- "Very good."
- Valid image every 20 minutes.
- CMS update every 5 minutes or are blank depending on time of day. Do not keep track as there are too many elements to monitor. Do check situationally.
- "We are very short on metrics for traveler information performance measures. We do not have any metrics on the time it takes to disseminate information. This would be a great improvement in our department."
- "Supposed to be 85% up."

Reliability

- "Recently, traveler information system is "exceedingly reliable" and near 100%. Data entry is 98% reliable. Downtime is usually for scheduled maintenance, server or system upgrades."
- "We have some metrics for this through an asset management program. The reliability metrics are mostly hardware related and not message related."
- One region (in one state) indicated upper 90's for reliability.
- One state compiles a database of outages with the number of devices and uses it to extrapolate percent up and down time.
- CFR minimums and general guidelines. This metric is required for federal funding requests.
- A need to automate and define metrics.
- 75% statewide considered "good", but terrible to traveler information staff.
- "Want to do timestamps."
- Copper theft an issue.
- Funding an issue.

4.6 Other Quality Control Processes

A logical next question asked state DOTs about other quality control processes they are employing to ensure that field elements are operating properly and disseminating associated data as intended. Verifying CMS messages with a CCTV camera, ground-truthing and sensor calibration, and user reporting of errors, are some possible examples. Flagging questionable or bad weather sensor readings or posting a status message, such as "Image unavailable" or "Camera down for maintenance", are other examples.

As listed below, the state DOTs surveyed for this research have implemented a number of other processes to corroborate proper field element operation and data distribution:

- Operator observations and human interactions
- Review TMC logs
- Go through an extensive checklist at TMC shift change which includes a review of what is up and what is down.
- Check CMS messages:
    - CCTV camera
    - Public facing traffic website
    - Communications channel within the sign software
- One state indicated that they had "resisted" the concept of verifying CMS messages with a CCTV camera because the "software was better than that." Instead, it is done remotely and logs checked for compliance.

- For another state, the "… software for displaying an error message for a broken camera does not trigger a work order or any other type of notification to the developers. We are working to fix this oversight. Also, no notification comes in when servers need to be reset to continue disseminating traveler information, which leaves finding errors to chance. This is an area that needs improvement."
- "User reporting is ad-hoc."
- Post status messages (i.e., "Camera unavailable" or "Camera down for maintenance") or remove an icon from display.
- "Limited RWIS data quality control program that prevents observations with NTCIP 1204 missing data codes or cases where the observations are out of sensor range."
- Two people in each TMC always looking at all of the ITS field elements.
- A contractor is tracking how often field elements are not available.
- Multiple images available at each RWIS site.
- In one state, four times per month all data entry is tracked. A sample is taken and errors noted. Error frequency is about 2% - 98% or better accuracy. Similarly, each TOC in another state has their own quality assurance processes which includes a monthly sample of incidents with a review of response relative to accuracy and procedure.
- Contractor does remote calibration when needed.
- Rigorous testing of field elements as they are brought online before going live.
- Preventive maintenance on all RWIS systems.
- Work with maintenance to look at something specific while in the field (i.e., frozen camera, bad bulbs, etc.). Very manual process.
- Specific to RWIS data, one state responded: "… we are procuring consultant services right now to develop a QC/QA system in place that will notify us of RWIS sensors or cameras not performing or reporting incorrect data. The QC processes will be based on the Clarus conversion to MADIS (Meteorological Assimilation Data Ingest System) for Road Weather Information System data. We really have no automated system in place to be notified for 511 or RWIS. Putting the one into place for RWIS is a start."
- One state said they had no ITS solutions for processes such as CMS message verification, although section staff might verify a sign message.

4.7 Additional Relevant Information

Finally, many states mentioned future additions and improvements to their traveler information data and systems and/or indicated they were "trying to improve." Others shared relevant information based on their experiences with traveler information data systems. For interest and the sake of documentation, these comments are listed below if they have not been included elsewhere in this document.

Survey Question: Any other relevant information you would like to share?

- "We are in the process of applying for research funding to test a road condition reporting app developed [by another state DOT]. We have been looking for a more efficient process. Reporting to 511 from the trucks would improve our reporting times and the information that goes, as well as make it easier on our operators."
- Currently "… deploying non-intrusive pavement sensors to provide pavement temperature and pavement state. Many other RWIS states are doing the same to avoid the deterioration of in-pavement sensors. The pavement temperature measurements provided by these sensors are reasonable when compared with the traditional in-pavement sensors. The pavement condition (contaminant depth, friction, and pavement state) needs to be further evaluated."
- Currently "… involved in a vigorous seasonal weight restriction program. Vertical thermistor strings are deployed from the pavement surface down through six feet. These thermistor strings are polled daily and the vertical temperature profiles presented on [the road weather information system]. Users can look at the current or past monthly, weekly, or daily profiles. We have recently added to the weight restriction program by adding seasonal freeze/thaw profiles for each thermistor site. While not tied directly to the traveler information program, these temperature data probes can be used by our pavement engineers to institute/remove weight restrictions and by the commercial trucking enterprise to view the likelihood of weight restrictions."
- In addition to the state DOT, a number of different public entities within the state (e.g., metropolitan planning organizations) provide traveler information. Thus, when considering unified statewide data quality processes, it isn't enough to just look at the state DOT.
- Anticipate having VMS on the traveler information website in the next year.
- Data flows are property of the state DOT. The real time feeds do not keep meaningful records for historical purposes.
- One contractor is responsible for the data flow from the RWIS systems as well as their maintenance. Another contractor handles the 511 website, 511 phone system, data entry system, modular traveler information system, and the ingestion of auto feeds into the system.
- Currently conducting a project for traveler information program assessment.
- Focusing on training at the section level.
- "Communications staff is thin."
- New data exchange system includes a visualization module that can graphically represent the status and performance of all devices and their associated data, as well as visually portray any of the current and historical data in the system.
- More VMS and travel time information being integrated into the statewide traveler information system website.

- The "… Traffic Operations Center is a statewide traffic management center. My position of Traveler Information Manager is unique among DOT's. I am a part public information officer, part traveler information quality control in house at the TOC. I find this very valuable and helpful in keeping data consistent, timely and complete."
- Avoid storing / archiving data and recording video because the public disclosure requests and retention policies were "distracting" (e.g., one region was receiving two to five legal video requests per week and that was becoming overwhelming). Traffic studies are acceptable. Traffic consultants and universities do use images.

4.8 Data Repositories

Several states have some sort of publically accessible data repository, e.g., California's Commercial Wholesale Web Portal 2 (CWWP2). The table below lists the link and any corresponding descriptive information as reported by the survey respondents. For reference and the purpose of documentation, a list of traveler information websites for the states interviewed is included in Appendix F: Traveler Information Websites.

Survey Question: Does the DOT have a data repository that allows public and/or third-party access to the raw data? If so, what is the site? How is the data accessible and how is it documented?

Table 1: List of DOT Data Repository Sites

State	Data Repository Site(s)
Alaska	For 511 (XML feeds): Road Conditions: https://511ride.alaska.gov/rss2?type=roadcondition Alerts: https://511ride.alaska.gov/rss2?type=incident Planned Events (e.g. road work or planned closures): https://511ride.alaska.gov/rss2?type=plannedevent RWIS web application: http://roadweather.alaska.gov
	Information on how to access the 511 XML feeds is under "Website FAQs" on www.511.alaska.gov. Access to raw RWIS data requires a formal request. The RWIS web application allows for queries (plots and data exports) by the day or week. Users can also request a period of record (POR) data export.
California	Commercial Wholesale Web Portal 2 (CWWP2): http://www.dot.ca.gov/cwwp2/
	Provides individual district ITS data in a uniform manner. The datasets are in four different documented formats. The listed link includes a description, data format and file layout, field element file locations, miscellaneous file locations (i.e., route charts, District map, etc.), and tools for CCTV and CMS.
Montana	RSS Feeds and Google Earth: www.mdt.mt.gov/travinfo/rss.shtml
	Includes XML, KML, and RSS feeds in several formats for road conditions by segment, road closures, and incident updates. An FTP site is available for particular projects. A data base output file is produced every 10 minutes.
Nevada	NVRoads Data Exchange Site: http://wiki.511nv.com/index.php/DE:DEHome
	API that provides access to raw data for selected elements used by the 511 website and phone system. Requirements matrix developed using ITE and AASHTO guidelines; detailed XML schema.
Oregon	TripCheck Traveler Information Portal: http://www.tripcheck.com/TTIPv2/ISP/login.aspx
	Provides access to a variety of transportation data used in the TripCheck system. XML data file with self-describing schema documents.
Utah	UDOT Traffic, "Center to Center" data feed http://udottraffic.utah.gov/
	Center to Center data feed provides access to crash data, weather data, lane closures, and other information.
Washington	Traveler Information API: http://webpub3qa.wsdot.wa.gov/traffic/api/
	Provides a single gateway to all of WSDOT's traveler information data. Includes documentation, RSS feeds, KML files, etc.

5. LITERATURE SEARCH

A review of the literature was conducted to identify what processes are employed regarding quality control of traveler information data. To bound the task and adhere to the scope of work, the search was limited to information on data quality within the transportation field. The Federal Real-time System Management Information Program is discussed along with some early work on ITS traffic data which defined "traffic data quality" and provided the impetus for a set of tools for assessing data quality. Several weather data systems have implemented quality assurance procedures which are applicable to RWIS and other traveler information such as advisories and alerts. Also included are a few studies that relate to a specific type of ITS field element (e.g., CCTV) and a brief overview of the envisioned data translator for the Connected Vehicle of the future. Specific to California, quality control processes implemented through the WeatherShare system are discussed.

5.1 Federal Real-Time System Management Information Program, 23 CFR 511, Section 1201

In November 2010, a final rule was published establishing the Real-Time System Management Information Program (23 CFR 511) in accordance with Section 1201 of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU). "The Real-Time System Management Information Program is to provide the capability to monitor in real-time the traffic and travel conditions of the major highways across the U.S. and provide a means of sharing these data with state and local governments and with the traveling public" (8). It provides a foundation for basic traveler information and data exchange formats. The program established minimum requirements for real-time traffic and road condition information for construction activities, road or lane blocking incidents, road weather observations, travel times, information accuracy, and information availability. Of particular relevance, a real-time information program is to be 85 percent accurate at a minimum, or have a maximum error rate of 15 percent (9).

Methods for measuring accuracy or other quality metrics are not included. (E.g., how do you know the requirement is being met with "good" data versus "garbage" data?) The program also does not define metrics for specific elements such as RWIS or CMS. Instead, individual states have the "flexibility to use methods appropriate to systems and processes used to acquire information and data" (33). The program is to be implemented for all interstates and metropolitan "Routes of Significance" (identified by states and local agencies) (36).

5.2 ITS Data, Data Quality

In 2000, ITS America's Advanced Traveler Information Systems (ATIS) Committee published guidelines intended to aid public agencies and private firms in producing and utilizing data for traveler information. The guidelines addressed issues of content including type, coverage, and quality of the data, as well as issues associated with access to the data for development purposes (32).

At the end of 2002, three white papers were published by the Federal Highway Administration's Office of Policy in support of a series of workshops on data quality to be presented the following year. Margiotta (17) characterized the state of the practice for traffic data quality from operations (urban TMC activities) and planning perspectives. While not fully dedicated to traveler information data, discussing ITS traffic data as a whole is certainly relevant. After identifying different types of traffic data and potential applications for it, he broadly discussed possible sources of data quality issues including equipment type, environmental conditions, installation methods, calibration, inadequate maintenance, communications, and equipment failure. He also noted that planning personnel were likely to perform more in-depth data checks such as range checks, equipment checks, or cross-checks. At the time the report was published, operations personnel generally did not need highly accurate data for their applications. However, Margiotta pointed out that transportation agencies were moving towards system operations and performance measurement which made comprehensive quality data vital to both planning and operations activities. "Perhaps the best way to influence the quality of ITS-generated traffic data is to foster the development of more sophisticated operational response strategies that require more accurate and timely data" (17). Margiotta suggested data sampling, sharing maintenance resources, private sector data collection, and standards for maintenance, calibration, and performance, as potential solutions to address the issues of traffic data quality.

In *Defining and Measuring Traffic Data Quality* (34), Turner recommended that traffic data quality be defined as "the fitness of data for all purposes that require it. Measuring data quality requires an understanding of all intended purposes for that data." Six quality measures were listed as essential to measuring data quality in traffic data applications: accuracy, completeness, validity, timeliness, coverage, and accessibility. Turner further recommended that goals or target values for these measures be determined at the jurisdictional/program level due to the significant variation in need and resources. The paper identified three traffic data consumer groups and summarized the data quality measures suggested or used in each of these groups. Considering discussion relevant to traveler information, one of the groups identified included those performing real-time traffic monitoring and control (e.g., TMCs). Some of these agencies/groups used simple data quality checks from field data collection hardware and software, tracked failed field equipment to measure coverage or completeness, evaluated accuracy of new equipment types before widespread implementation, and provided indication of data timeliness on traveler information websites (i.e., timestamps). A second group consisted of operations and users of ITS data archives including traveler information systems. Turner noted a general reluctance to use archived data due to data quality issues (real or perceived), adding that the field was in its infancy. However, some data quality processes/measures included data completeness (e.g., number of data samples in a summary statistic), data validity flags, data quality mapping utility, data validation checks (i.e., "business rules", software-based), and data integrity checks on data files and station integrity. One agency paid a data collection contractor based on quality measures of data validity and data completeness. As a final recommendation, if data quality is measured, Turner suggested a data quality report be included in metadata associated with the raw dataset.

The third white paper in this group presented techniques for enhancing data quality with quality control procedures (20). Of interest for traveler information applications, the authors discussed paying contractors based on data quality, standards development (i.e., for traffic monitoring devices), and detailed training for equipment and procedures. The benefits and value of sharing data between agencies and states were emphasized.

A significant outcome from the FHWA traffic data quality workshops was a set of methods and tools for assessing traffic data quality. These tools included a sequential framework and guidelines for addressing "… technical issues related to the data quality standards, data sharing, estimates of the level of effort required for measuring and reporting data quality, and specifies procedures for using metadata" (3).

5.3 Weather

The effects of weather on the transportation system are well-documented in the literature. The Federal Highway Administration has dedicated significant resources to weather responsive traffic management in order to improve mobility, safety, and reliability of the transportation system during inclement weather. Traveler information is a key component in this initiative. Similarly, weather information is a crucial aspect of traveler information systems.

In 2004, the FHWA Road Weather Management Program completed a preliminary gap analysis for weather information services and products for surface transportation users, operators, and maintainers (28). The authors found that weather data quality was not uniform – sensor siting criteria, test procedures, processing software, and multiple display formats were all contributing factors. Among other needs, "Rigorous standards for sensor siting and calibration, data validation, as well as data processing and presentation are needed to address deficiencies and inconsistencies in the quality of weather data used by transportation managers" (28). Pisano and Goodwin also identified research needs for weather-responsive traffic management that included how road weather information is delivered to different types of travelers (24). Baseline characterizations of road weather information were later established by Hart et al (12, 13, 2). The anticipated outcomes from ongoing monitoring related to traveler information included improved traveler advisory messages, improved responsiveness for implementing weather related traffic controls, and incentives for equipment and processing improvements for low quality resources (1).

The FHWA Road Weather Management Program launched the *Clarus* Initiative in 2004 with the objective of mitigating the effects of adverse weather (e.g., fatalities, injuries, delay) by providing nationwide, standardized weather data. The intent was "… to develop and demonstrate an integrated surface transportation weather observing, forecasting and data management system, and to establish a partnership to create a Nationwide Surface Transportation Weather Observing and Forecasting System" (26). The *Clarus* system itself was "a network for assimilating and exchanging *quality-controlled* environmental data related to surface transportation" (25). Data quality was an essential factor in the development and implementation of the *Clarus* System. One key component was that data flagged as anomalous through the quality control processing was transferred back to the source agency so that it could be further investigated and any issues could be resolved (25). Limber, Drobot, and Fowler described the quality checking tests designed and implemented into the *Clarus* System (15). These procedures included:

- Sensor Range Test – "detects sensor readings that fall outside the range of sensor hardware specifications or theoretical limits (i.e., a maximum and minimum value)."
- Climate Range Test – "detects sensor readings that fall outside predetermined climate range values."
- Time Step Test – "detects sensor readings whose values change by more than a predefined variable-specific or station-specific rate over a thirty minute (past) and five minute (future) configurable period."
- Like Instrument Test – "detects sensor readings whose values differ from the average of all sensor values obtained from the same station with the same weather parameter type by more than a predefined variable-specific threshold."
- Persistence Test – "detects sensor reading whose values remain constant for a predefined variable-specific period of time."

- Interquartile Range (IQR) Spatial Test – "a method for checking whether a sensor reading is consistent with its neighboring sensor readings. It detects sensor readings that differ by more than a predefined threshold from an expected value within a neighborhood of the target sensor reading."
- Barnes Spatial Test – compares sensor readings with those from neighboring stations. "Like the IRQ test, the neighboring stations used for spatial comparison are determined by a formula based on configurable tolerance bounds. Unlike the IRQ test, neighboring sensor readings are weighted according to their distance from the original sensor, with the weight decreasing exponentially with the distance from the station."
- Dew-point Temperature Test – "detects air temperature and relative humidity sensor readings whose corresponding derived dewpoint temperatures do not pass a Barnes spatial analysis."
- Sea Level Pressure Test – "a method for checking whether an atmospheric pressure measurement is consistent with its neighboring sensor readings, when both the target pressure sensor reading and its neighbors have been reduced to sea level pressure. It detects reduced pressure sensor readings that differ by more than a predefined threshold from an expected value within a neighborhood of the target sensor reading."
- Precipitation Estimation Test – "algorithm compares a target ESS [Environmental Sensor Station] precipitation accumulation report to nearby [National Centers for Environmental Prediction] Stage II or IV grid values."

Fowler et al (11) further described and presented test results on *Clarus* data for spatial quality control tests. The researchers developed an automated process for making spatial comparisons and tested the algorithm on RWIS data from the *Clarus* network. Though limited to stations with a minimum of five neighbors, the proposed robust algorithm performed as well as traditional algorithms and better in cases where neighboring measurements were erroneous.

Around the same time the *Clarus* Initiative was kicking off, Canada also prototyped the Road Weather Information Network (RWIN) to enhance the country's highway system with improved maintenance operations, more accurate traveler information, and enhanced monitoring of conditions, among other goals. The RWIN was a centralized repository of data from provincial environmental system sensors (ESS) and instrumented vehicles. It performed "…Quality Control (QC) services, sending real-time alerts when deterioration or outage of roadside components [was] detected, and deliver[ed] QC'ed data in real-time to the Provinces, their agents, and their designated recipients in CMML [Canadian Meteorological Markup Language] and other formats" (14). Data quality checks were automatically performed based on predetermined thresholds for ESS and pavement data. All data were collected in a central repository which maintained ESS and instrument status reports and generated alerts. Data field statistics, including counts, frequencies, missing and extreme values, were available through a data auditing/profiling feature (14).

Operated at the NWS National Centers for Environmental Prediction (NCEP), the Meteorological Assimilation Data Ingest System, or MADIS, is a database of meteorological observations and a data delivery system. It collects weather data from numerous sources, decodes, and reformats it to a common format including uniform observational units and time stamps (23). MADIS observations are subjected to both static and dynamic quality control checks. Static checks include validity, internal consistency, and vertical consistency. To address the challenge of "statistically reasonable but invalid data", dynamic checks use other available information to check position consistency, temporal consistency, and spatial consistency (22). Results are summarized for each station after all the checks are run.

MesoWest is another database of weather observations from government agencies, private firms, and educational institutions. Quality control processes include "range checks" which flag values that fall outside maximum or minimum thresholds. Observation time is also checked (e.g., an observation would be flagged with "Suspect Time" if a time reported is well ahead of the current time). Regression statistical checks are conducted on available temperature, dew point, and pressure data. Statistical checks are also conducted on available wind data that check for stagnant wind observations. MesoWest data include a QFLAG variable with each observation that defines data quality from a particular station (not a specific sensor reading) (19).

The Oklahoma Mesonet is a network of environmental monitoring stations. The Oklahoma Climatological Survey at University of Oklahoma receives the data, verifies data quality, and then provides the data to Mesonet customers. The Oklahoma DOT provides links to the Mesonet relative to traveler information. Quality Assurance consists of (18, 31):

- Laboratory calibration – ensures that all sensors deployed are calibrated to the same high standards;
- On-site intercomparison – annually, sensor accuracy is verified with a portable system of calibrated reference sensors;
- Automated Quality Assurance – software algorithms (e.g., Barnes Spatial Test);
- Manual Quality Assurance – meteorologists use manual techniques such as analyzing monthly statistics to detect sensor drift or bias and marking the actual start time for problems. Meteorologists are also responsible for coordinating issue resolution with field technicians.

While not directly related to traveler information, the National Oceanic and Atmospheric Administration's (NOAA) Environmental Real-Time Observation Network (NERON) also included a quality assurance process in order to provide high quality weather information to users. The NERON system consisted of approximately 100 automated weather sites in several northeastern states. Automated quality assurance processes were based on the Oklahoma Mesonet and included filters (e.g., failed range tests) and a number of algorithms that performed spatial tests, step tests, persistence tests, and step-to-normal tests. Quality Assurance meteorologists reviewed the automated quality assurance results daily, investigated erroneous values, and issued trouble tickets to resolve problems (10).

The Caltrans District 3 Regional Transportation Management Center (RTMC) implemented a weather alert notification system to provide timely traveler and road weather information particularly related to fog, wind, and frost conditions. Some findings and lessons learned relevant to traveler information data quality are listed below (6):

- In regard to alert timeliness, alerts should be issued when conditions exceed the pre-determined threshold, or at the start of the event.
- The alert system helped operators post messages more appropriately with regard to event duration. The addition of an "all clear" signal would be useful.

- Operator training and clear, consistent procedures are essential.
- Operators "uniformly said it was critical to confirm either data from sensors with human observations or readings from other sensors, or human observations with sensor readings. This follows management guidance that all weather data be verified with one or more additional sources before making a decision to post a travel advisory message."
- Quality checking procedures such as those implemented in the FHWA *Clarus* system could be used to identify those sensors that were performing poorly.
- The FHWA ESS Siting Guidelines (16) could be used to determine preferred locations for collecting quality data.
- In regard to quality and timeliness of weather related decision making in the RTMC, one operator commented, "We are as good as the information we receive" (6).

5.4 WeatherShare

The Phase 2 WeatherShare System implemented three levels of quality control. Level 1 consisted of range checks, defined for eleven data elements. Level 2 consisted of temporal checks for either no change or for large changes in data within a 24-hour time period, and was defined for two data elements. The Level 3 quality control check was considered experimental and was only applied to air temperature values. Observations were compared to predictions from a global, multivariate linear regression model and observations that showed a large deviation from that predicted by the model were further compared to a local model. If an observation showed a large deviation compared to the prediction from the local model, then it was flagged as failing the Level 3 check.

Perhaps the most useful quality control mechanism implemented in the WeatherShare project was user error reporting. Users were provided with a link to invoke a form that included a drop-box of typical problem types and an open text field that allowed for users to provide a description of the problem. Common problems reported included incorrect station location as well as observations that didn't match the known condition. In most cases, the only option available to the project team for mitigation was disabling the display of the problem point since the data came from third-party providers. Proactive monitoring of error reports and, where appropriate, responding to individuals who make error reports with resolution were considered necessary components of the user error reporting functionality (27).

5.5 Other

The literature revealed a few studies and publications related to data quality for particular traveler information elements such as CMS, CCTV, and travel times. For example, in Minnesota DOT's 2012 CMS Manual of Practice, messages were to be verified through CCTV cameras or field staff, and operators were not to rely solely on electronic verification from the software/computer system. The CMS control software documented timestamps, message placement and termination (21). Chavan et al reported that overlaying travel time messages on speed plots provided a useful means of comparing sign messages with the actual travel conditions (5). Turner et al described procedural guidelines for evaluating the accuracy of traveler information including average route travel time (35). Others have presented techniques and methods for detecting anomalies or corrupted images from traffic cameras (37, 4).

A number of initiatives and research projects related to Connected Vehicle technologies have been conducted or are currently underway. One such project was the Vehicle Data Translator (VDT) software system. As envisioned by the overall Connected Vehicle initiative, the system was designed to incorporate "…vehicle-based measurements of the road and surrounding atmosphere with other weather data sources" (7). Numerous quality control routines were necessarily a crucial part of the VDT, many based on the quality routines in the *Clarus* system. *Clarus* based tests included Anticipated Range Test, Persistence Test, Step Test, Spatial Tests (replaces Neighboring Station Test), and Climate Range Test. Non-*Clarus* based tests included Data Filtering Test, Model Analysis Test (Enhanced), Neighboring Vehicle Test, and Combined Algorithm Test (7).

5.6 Summary

In general, concern for data quality is expressed throughout the literature in regards to traveler information as well as operations and maintenance. The Federal Real-Time System Management Program established minimum requirements for data quality, but left it up to the states and local jurisdictions to determine how to meet those requirements. A shift in the transportation industry to improved management of existing resources and infrastructure as compared to building more roadways has focused agency resources on performance based metrics, which necessarily includes some treatment of data quality. This is particularly evidenced in the literature relative to road weather management.

Robust and continuous data quality procedures were fundamental to the operation of weather information systems such as *Clarus*. Other weather data systems also employ various quality control measures. Some documentation and research exists that addresses specific traveler information elements such CMS messages, CCTV, and travel times. However, while the literature documents a need, there does not appear to be any form of comprehensive plan for uniform, multi-dimensional traveler information data quality.

Other domains may have work relative to data quality. For instance, quality control processes used for security and surveillance may be applicable to CCTV or other traveler information elements. Documenting such work was outside the scope of this research task however.

6. BEST PRACTICES

Task 4 aimed to compile a collection of best practices for the aggregation and dissemination of quality traveler information. At the onset of the task, it was recognized that there may not be any so called "best practices" established for traveler information data quality. Neither the survey of DOT practitioners nor the literature review found a comprehensive, well-defined plan for unified, multi-dimensional approaches to quality assurance of traveler information.

However, all of the DOT practitioners that were surveyed, as well as the literature reviewed relative to data quality in transportation, indicated in some way that quality data was important for safe, efficient operation of the transportation system, including provision of traveler information that is accurate, timely, and reliable. This observation is especially valid given the current environment that is increasingly focused on performance measurements, accountability and "smarter" operation of roadways.

A number of procedures and/or considerations relative to data quality were repeated in the literature or the DOT practitioner survey and are important to mention here. Listed below are some of the techniques, tests, or considerations employed to address traveler information data quality.

- Automation - Automated procedures could more quickly identify obvious problems such as stale CCTV images or invalid sensor readings. The more quickly a problem is detected, the more likely it can be resolved before it impacts traveler information. Automated tools could also more proactively help identify a piece of equipment that is failing or reaching the end of its lifecycle by providing a data trend over time. Automated processing can also mitigate challenges associated with staffing resources (i.e., overloaded/understaffed).
- Regular and adequate preventive maintenance, including calibration, for all field elements is essential. Maintenance resources and responsibilities could be shared.
- Rigorously test field elements before they are brought online. Laboratory calibration can help ensure that all sensors deployed are calibrated to the same high standards.
- Test field elements for accuracy before widespread deployment.
- Specific quality control processes, tests, information, etc.
- Timestamps
- Data validity flags
- Automated sensor tests
    - Sensor range tests
    - Climate range tests
    - Time step test
    - Like instrument test
    - Persistence test
    - Interquartile range spatial test
    - Barnes spatial test

  - Dew-point temperature test
  - Sea level pressure test
  - Precipitation estimation test

- Data sampling – Review a sample of data for accuracy, timeliness, procedure, etc. Regularly analyze statistics to detect sensor drift or bias.
- Computing reliability – compile a database of outages with the number of devices and extrapolate percent up and down time.
- Post status messages (e.g., “Camera unavailable.”). Remove icons or prevent display when data is flagged as erroneous.
- The ability to remotely calibrate sensors and restart systems as needed are both useful tools.
- Operate traveler information systems redundantly. This could mean that while one issue is being resolved other parts of the system are less likely to be affected. This could also mean providing multiple sources of information on the same incident so that if one source is bad, other sources can still provide information about the situation.
- TMC operations – checklists used during shift changes include a review of which elements and/or systems are up or down. Consider staff responsibilities (i.e., public information officer also tasked with traveler information quality control). Develop clear, consistent procedures for handling traveler information.
- Simplify manual data entry processes as much as possible. Spell checker functionality and a “public preview” option can be useful tools.
- Operator observations and human interactions are still important, even with automated systems.
- Training – For consistency, accuracy, policy, procedure, etc.
- Define performance measures and standards for data quality, data validation, data processing and presentation, performance, maintenance, sensor calibration, and sensor siting.
- Centralization – Having some sort of central data aggregator/processing system that can consistently format (e.g., common format, uniform observational units, timestamps, etc.) the data for multiple uses and users seems beneficial.
- User error-reporting.

7. RECOMMENDATIONS AND NEXT STEPS

The recommendations and next steps outlined below are made in light of Caltrans traveler information data and processes. Some of the practices identified in the previous section may be preferable to others, and some may only be applicable in certain situations.

- There is a need to define where the state DOT is in the traveler information environment. As a survey respondent commented, the DOT has limited flexibility or ability to move quickly. This makes it challenging to quickly design and publish an app for example, because it would already be obsolete by the time it was out for public use. In this case, the survey respondent indicated that the DOT's "niche" may be to provide quality data to others and allow the traveler information environment to sustain it from that point.
- It is recommended that preliminary steps be taken to establish a data governance model. This model should clearly define who owns what data, as well as uses and associated thresholds for the specified data.
- Relevant quality metrics and requirements should be clearly defined. This includes how to determine that requirements are being met with quality data.
- Common statewide standards for data quality, performance, maintenance, and calibration should be defined and established using an engineering approach. These standards should be tied to all specific uses of the data.
- It is recommended that implementation of additional automated feeds should be investigated.
- This research was sponsored by Caltrans under the auspices of the Western States Rural Transportation Consortium as a technology incubator project. Incubator projects are smaller research efforts that serve as a "proof of concept". Based on the results of an incubator project, further research on a larger scale may be pursued. This incubator project merely scratched the surface of the challenge of traveler information data quality. Additional research is needed to more thoroughly evaluate and establish best practices.

8. APPENDIX A: RESEARCH TASK INFORMATION SHEET

Data Quality for Aggregation and Dissemination of DOT Traveler Information: An Analysis of Existing System Best Practices

Western Transportation Institute, Montana State University
California Oregon Advanced Transportation Systems (COATS), Phase 6:
www.westernstates.org/Projects/COATS/Default.html

Data Quality Incubator Project: www.westernstates.org/Projects/Default.html
Western States Rural Transportation Consortium (WSRTC): www.westernstates.org

Project Description and Background

Data quality for traveler information has generally been handled on an ad-hoc basis, with little or no provision for error notification other than perhaps through user-reporting of observed errors. The quality of data, specifically whether it is accurate, timely, and reliable, is a crucial consideration for the provision of traveler information. When drivers access traveler information that is up to date, correct, and accessible every time they need it, they will use it to make travel decisions which ultimately impact traffic management effectiveness. However, if for example, travelers access traveler information and see old camera images, obviously incorrect weather information (e.g., freezing temperatures when it is warm everywhere else in near proximity), or misspellings on sign messages that change the meaning of the message, then users are less likely to make travel decisions based on the traveler information or even access the information in the first place. This can significantly diminish the effectiveness of traffic management efforts. Even worse, if drivers use incorrect information to make travel-related decisions, more serious consequences could result.

There is an apparent need for the development of best practices for data quality for the aggregation and dissemination of state department of transportation traveler information, as demonstrated by erroneous data in existing feeds. Taken individually, the presence of errors may be considered anecdotal. But recognizing that many errors can be present at an arbitrary time, we believe that there is a more systemic problem in regard to data quality of DOT-provided traveler information that affects all state DOTs, with numerous potential points of failure. In light of this, the goal of this project is to analyze and document existing system best practices for data quality for the aggregation and dissemination of state department of transportation traveler information. The types of traveler information to be investigated include the following:

- CCTV images
- RWIS data
- CMS messages
- Chain Requirements
- Incidents
- Construction and/or Lane Closures
- Optional
    - Commercial Vehicle Information
    - General Road Information including Road Closures

- Road Conditions (Human-reported, point and segment-based)
- HAR Messages
- Rest Areas, including status and amenities
- Truck Scales / Weigh Stations

This project is sponsored by the California Department of Transportation (Caltrans) because they think problems exist and feel the work could be beneficial to other state DOTs.

9. APPENDIX B: INTRODUCTORY EMAIL FOR SURVEY OF DOT PRACTITIONERS

Subject: Traveler information data quality project
Hello ______,

I am contacting you as part of an effort to analyze and document best practices for traveler information data quality. The Western Transportation Institute is conducting the project for the California Department of Transportation (Caltrans) through the Western States Rural Transportation Consortium (WSRTC).

We're investigating the accuracy of data and metadata, what quality control measures are used, and how the data is handled throughout the process from collection to dissemination. Essentially, how is the traveler information provided, what is done to make sure it is accurate, and how errors/problems are detected and fixed. Some of the types of traveler information to be investigated include CCTV images, RWIS data, CMS messages, chain requirements, incidents, and construction or lane closures.

Some examples of potential data quality issues include:

- Meta data – incorrect location of field elements, inaccurate timestamps
- Old or frozen CCTV images
- Misspellings on sign messages
- Camera settings visible to the public
- Incidents incorrectly located
- RWIS data – e.g., surface temperature, error, missing, displaying a value when there is no sensor, depth and water level
- Incorrect data (temperature, precipitation, etc.)

Caltrans has been working on this issue in relation to QuickMap and the One-Stop Shop traveler information websites, among other projects. They believe that systemic problems exist regarding data quality and think this work could be beneficial to other state DOTs.

Are you the best person to talk with regarding ________ DOT's approach to data quality for traveler information? If so, would you be available in the next several weeks for a brief telephone call? If you are not the appropriate person to talk with, would you point me toward a contact within the agency that can answer some brief questions about traveler information data?

For more information about this project and the Western States Rural Transportation Consortium, please visit www.westernstates.org and www.westernstates.org/Projects/Default.html.

Thank you in advance for your assistance with this project. I look forward to talking with you.

Sincerely,

Leann Koon
Montana State University
Western Transportation Institute

10. APPENDIX C: SCRIPT FOR SURVEY OF DOT PRACTITIONERS

Hello. This is Leann Koon from the Western Transportation Institute at Montana State University. We're working on a project for the California Department of Transportation (Caltrans) involving data quality for DOT traveler information. Are you the best person to talk to regarding this?

Looking for the person responsible for traveler information. Others might do the technical stuff. Data quality and provision of traveler information, how is data quality dealt with.

If no: Will you point me toward a contact within the agency that can answer some brief questions about traveler information data?

If yes:

Do you have a few minutes to answer some brief questions regarding data quality and traveler information?

For some background on the project, we're analyzing and documenting existing system best practices for data quality of traveler information. We're investigating the accuracy of data and metadata, what quality control measures are used, and how the data is handled throughout the process from collection to dissemination. (Basically, how is the traveler information provided, what is done to make sure it is right, and how are errors/problems detected and fixed?) The project is sponsored by Caltrans through the Western States Rural Transportation Consortium. Caltrans has been working on this issue in relation to QuickMap and the One-Stop Shop traveler information websites, among other projects. They believe that systemic problems exist regarding data quality and think this work could be beneficial to other state DOTs.

11. APPENDIX D: SURVEY OF DOT PRACTITIONERS

Contact Information	
Name:	
Title:	
State:	
Agency:	
Phone number:	
Email address:	
Traveler Information Website(s):	
Data Feed Website(s), FTP Server, API, or similar:	

Questions

1. How do you find out about problems with your traveler information data or metadata? What do you do (as the individual responsible for traveler information) in regard to these things?
    a. Who finds out?
    b. Is it an automatic or manual process?
    c. Describe the process.
    d. If it is done manually, who does it and how often?
    e. How are problems resolved?
    f. How long does it take to resolve a problem?
    g. What issues have you had, if any, with data quality and traveler information? (Describe.)

Some examples:

- Meta data – field element location, timestamps
- Old or frozen CCTV images
- Partial CCTV image, dark CCTV image
- Indication of camera being unavailable
- Duplicated cameras in different locations
- Misspellings on sign messages
- Camera settings visible to the public
- Color display (affects interpretation of road conditions)
- Incidents incorrectly located
- Chain-up control requirements
- RWIS data – e.g., surface temperature, error, missing, displaying a value when there is no sensor, depth and water level
- Incorrect data (temperature, precipitation, etc.)

2. What quality dimensions are used to characterize the traveler information system/site performance? Describe.
    a. E.g., accuracy, timeliness, reliability.
        i. Accuracy – is the data correct and how do you know it is correct?
        ii. Timeliness – is the data made available in a timely fashion and is the data current?
        iii. Reliability – percent up-time (the percentage of time the data meeting criteria i and ii is available for use).

3. What other quality control processes are in place to ensure proper operation of field elements AND provision/dissemination of their associated data?
    a. E.g., verification of CMS messages with a CCTV camera; ground truthing and sensor calibration; user reporting of errors; etc.
    b. Status messages (i.e., camera down for maintenance, flag weather sensor readings to indicate if they are questionable or bad.)

4. Generally and briefly, how is DOT data used for traveler information handled from collection to dissemination? How does data get from the field to the public view (online, 511, TV/radio, social media)?

5. Does the DOT have a data repository that allows public and/or third-party access to the raw data? If so, what is the site? How is the data accessible and how is it documented?

6. Any other relevant information you would like to share?

7. Is there anyone else that I should talk with, either within (state) or in other states?

8. Would you like a copy of the project’s findings?

12. APPENDIX E: DETAILED SURVEY RESPONSES

1. How do you find out about problems with your traveler information data or metadata? What do you do (as the individual responsible for traveler information) in regard to these things?

   - Public call in, email, social media
   - Auto alerts, software
   - Third party vendor
   - Operations personnel
   - 511 phone line comment module
   - Data owners

   a. Who finds out?
      - Depends on the problem. (RWIS, communications, software, etc.)
      - Traveler information person, state communications office
      - TMC, TOC
      - Operations personnel, dispatch, district office
      - Third party vendor
      - ITS Maintenance, Regional ITS personnel
      - IT Maintenance
      - In one state, RWIS techs monitor their own equipment.

   b. Is it an automatic or manual process?
      - Level of automation ranges from completely manual, to combination automatic and manual, to mostly automated.
      - Depends on resources (staff, funding) as well as type and complexity of systems.
      - Desire for more automated quality control.

   c. Describe the process.
      - Depends on the problem, and how/by who the problem is discovered.
        - Public calls (email, Facebook, Twitter, etc.) in, problem directed to the appropriate person (i.e., traveler information office, ITS personnel, IT department, third party vendor/contractor, local maintenance office, technicians, etc.).
        - If camera down or image is stale, auto alert sent to phone.
        - Software indicates that data files are not up to date, email, text, phone call to TMC.
        - Third party vendor checking for readings outside thresholds, automatically generates service report. Notifications sent to the DOT contact.
      - Elements and data feeds are checked more closely after system changes.

- When a data station or chain of data is lost, then it is filtered and flagged as "bad."
- The state patrol and maintenance are relied upon for reporting issues with VMS.
- Service reports from third party vendors and reports from regional offices can also be a part of the process.

d. If it is done manually, who does it and how often?
- A) Human observations 7 days/week, a couple times per day. B) One time per day, one or two people look at each CCTV image. C) One staff following each week.
- Traveler information coordinator checks when a series of problems occur over a short period of time. CA: responsive mode.
- "It's a group effort – the 511 Working Group is made up of state DOT and State Trooper Dispatch staff."
- Read twice, post once. Those entering data manually are "self-policing and embarrassed if they make a mistake."
- Several states emphasized the importance of training for consistency and accuracy.
- Manual data entry processes simplified as much as possible.
- Even manual entries have canned information that can be used.
- Spell checker functionality.
- "Public preview" option that shows how the public will see the information entered.

e. How are problems resolved?
- Depends on the problem.
- Operator interaction – fix or take out of service.
- Define the problem and then hand off to subject matter expert and/or a technician.
- Issue work order.
- Notify third party vendor.
- Maintenance contractors, regional ITS personnel, IT maintenance, technicians, etc.
- Close communications between maintenance and law enforcement during incidents.
- In California, urban areas/regions (MPOs) have their own internal processes so the DOT is not directly monitoring the data feeds.
- System redundancy.
- "One user can start/stop everything; everyone knows what to do and can fix it themselves."

f. How long does it take to resolve a problem?
   - Depends on the problem.
   - "Most within two hours. Some that are more complicated take up to a week."
   - "…, once detected, less than one minute (fast, seconds). 24-48 hours."
   - "If it is just a reboot then it is fast. If a camera is broken, then it takes longer. Pretty quick. Takes priority."
   - "Communications issues – relatively short time, less than one hour. If device failure, days to a week."
   - "One minute to days or weeks."
   - "Ten minutes to one week."
   - "Data feed, fixed within 10 minutes. Quite often the public might not even know there was a problem. One to three days for other issues."
   - "Depends on the severity of the issue. Could be instantly fixed, or a few hours."
   - "For RWIS, depends on the issue; could be a week or more. No dedicated RWIS maintenance people. In rural areas, secondary road elements have a lower priority."

g. What issues have you had, if any, with data quality and traveler information? (Describe.)
   - CMS feeds (e.g., misspellings or typos, wrong abbreviations, multicultural challenges, sign limitations)
   - Equipment failure or equipment end of life
   - Cameras: offline, communications, power issues, potential old or partial picture, pixels out.
   - Field element location
   - Timeliness (e.g., old or stale information, conditions being reported are inaccurate, miss a winter road report, wrong start date on construction project)
   - Operator input error
   - Sensors, sensor calibration
   - Connectivity and communications
   - Missing information regarding an incident, construction, lane closures, etc.

2. What quality dimensions are used to characterize the traveler information system/site performance? Describe.

   a. E.g., accuracy, timeliness, reliability.

      i. Accuracy – is the data correct and how do you know it is correct?

      ii. Timeliness – is the data made available in a timely fashion and is the data current?

      iii. Reliability – percent up-time (the percentage of time the data meeting criteria i and ii is available for use).

- Real-Time System Management Information Program (Section 1201 SAFETEA-LU, 23 CFR 511)
    - Challenges associated with meeting its reporting requirements.
  - Includes minimum requirements, but there are no specific metrics or definitions for data quality for RWIS or CMS.
  - "This has been problematic - how to measure accuracy of … incident data … we're reporting what we know but how do we know if that's all."
  -
- "…moving toward a Results Based Alignment for service delivery framework from which we will evaluate (measure) the contribution (efficiency and effectiveness) of the services we deliver in support of our mission. Data quality is an essential ingredient of the RBA. We look at six elements for data quality (three are noted above):
  - Accuracy
  - Accessibility
  - Completeness
  - Coverage
  - Timeliness
  - Validity/Reliability"
- Characterizing system/site performance is an "operations thing" and there is no formal process.
- Collect public feedback and do a public survey every two years.
- "If it isn't measured, then it isn't fixed."
- California indicated a need for a data governance model and a common statewide standard.

<u>Accuracy</u>

- Visual confirmation.
- Check portable license plate readers/travel times one to three times per year.
- No way of tracking through 511.
- District centric.

<u>Timeliness</u>

- Update times based on federal requirements.
- Many data feeds updated every 15 minutes.
- "Very good."
- Valid image every 20 minutes.
- CMS update every 5 minutes or are blank depending on time of day. Do not keep track as there are too many elements to monitor. Do check situationally.
- "We are very short on metrics for traveler information performance measures. We do not have any metrics on the time it takes to disseminate information. This would be a great improvement in our department."

- “Supposed to be 85% up.”

Reliability

- “Recently, traveler information system is “exceedingly reliable” and near 100%. Data entry is 98% reliable. Downtime is usually for scheduled maintenance, server or system upgrades.”
- “We have some metrics for this through an asset management program. The reliability metrics are mostly hardware related and not message related.”
- Upper 90’s.
- Compile a database of outages with the number of devices and use it to extrapolate percent up and down time.
- CFR minimums and general guidelines. This metric is required for federal funding requests.
- A need to automate and define metrics.
- “75% statewide considered “good”, but terrible to traveler information staff.”
- “Want to do timestamps.”
- Copper theft an issue.
- Funding an issue.

3. What other quality control processes are in place to ensure proper operation of field elements AND provision/dissemination of their associated data?
    a. E.g., verification of CMS messages with a CCTV camera; ground truthing and sensor calibration; user reporting of errors; etc.
    b. Status messages (i.e., camera down for maintenance, flag weather sensor readings to indicate if they are questionable or bad.)

- Operator observations and human interactions
- Review TMC logs
- Go through an extensive checklist at TMC shift change which includes a review of what is up and what is down.
- Check CMS messages:
    - CCTV camera
    - Public facing traffic website
    - Communications channel within the sign software
- Have “resisted” the concept of verifying CMS messages with a CCTV camera because the “software was better than that.” Instead, it is done remotely and logs checked for compliance.

- The "… software for displaying an error message for a broken camera does not trigger a work order or any other type of notification to the developers. We are working to fix this oversight. Also, no notification comes in when servers need to be reset to continue disseminating traveler information, which leaves finding errors to chance. This is an area that needs improvement."
- "User reporting is ad-hoc."
- Post status messages (i.e., "Camera unavailable" or "Camera down for maintenance") or remove an icon from display.
- "Limited RWIS data quality control program that prevents observations with NTCIP 1204 missing data codes or cases where the observations are out of sensor range."
- Two people in each TMC always looking at all of the ITS field elements.
- A contractor is tracking how often field elements are not available.
- Multiple images available at each RWIS site.
- Four times per month all data entry is tracked. A sample is taken and errors noted. Error frequency is about 2% - 98% or better accuracy.
- Each TOC has their own quality assurance processes which includes a monthly sample of incidents with a review of response relative to accuracy and procedure.
- Contractor does remote calibration when needed.
- Rigorous testing of field elements as they are brought online before going live.
- Preventive maintenance on all RWIS systems.
- Work with maintenance to look at something specific while in the field (i.e., frozen camera, bad bulbs, etc.). Very manual process.
- Specific to RWIS data, "… we are procuring consultant services right now to develop a QC/QA system in place that will notify us of RWIS sensors or cameras not performing or reporting incorrect data. The QC processes will be based on the Clarus conversion to MADIS (Meteorological Assimilation Data Ingest System) for Road Weather Information System data. We really have no automated system in place to be notified for 511 or RWIS. Putting the one into place for RWIS is a start."
- Have no ITS solutions for processes such as CMS message verification, although section staff might verify a sign message.

4. Generally and briefly, how is DOT data used for traveler information handled from collection to dissemination? How does data get from the field to the public view (online, 511, TV/radio, social media)?

- **Alaska:** "All of the information on 511 is manually input except for the RWIS data and camera images, NWS weather watches and Nixle.com feeds from local law enforcement. Those are all automatic. We do offer XML feeds of the 511 driving conditions, planned events and alerts for anyone interested in sharing the information with their users or travelers." "The system is maintained, operated and hosted offsite."

- **California:** Data from field elements is aggregated in the district TMC (could be multiple systems) by respective ITS unit. The information is shared by publishing to the Commercial Wholesale Web Portal 2 (CWWP2). DOT traveler information office pulls the data from the portal and repackages it for the web page and the IVR phone system. The data is the same for Caltrans and the public.
- **Idaho:** Multiple sources of information, both auto feeds and manual entry, are all fed to a third party vendor which creates events. Outputs include 511 phone system, websites (mobile version, customizable high bandwidth version, low bandwidth version for commercial vehicle operators), link to transit operators, social media (Twitter), and a smart phone app. The data can be pushed to the public or the public can passively receive it.
- **Montana:** Drivers radio road conditions to division offices. Division/district staff enter the information into a database which disseminates it to various outlets (511, web, mobile, social media).
- **Nevada:** Central System Software collects information from ITS devices and populates the 511 system as well as public device interfaces. Data is published to an FTP site. With the newly created Nevada Data Exchange, devices are polled by the transportation management system, sent to the NDEX, and then distributed everywhere else.
- **Oregon:** The system is highly automated and integrated, very little is done manually. Once the information is in the system, it goes everywhere (TripCheck, 511, TV). Those entries that are done manually are reviewed by a TOC operator before it goes out. When the TOC operator enters something, the TOC software builds the message for them and it can be edited or left as is.
- **Utah:** "UDOT's traveler information outlets include Twitter, Facebook, 511 phone line, CMS, a website, a mobile website and a smartphone app. We also share camera feeds with local media. UDOT has a fiber optic network of over 1800 miles which helps get communication/messaging to rural devices."
- **Washington:** In the Northwest Region, data is collected from the field and sent to the central server with central software which has multiple pulling and polling functions. The data is repackaged and published to a web accessible database that can be queried. The data is then pushed out to the various outlets. In the South Central region which is more rural, data from the mountain passes is manually collected and keyed in by operations personnel. It is then pushed to the public (i.e., 511 system) via the same database. RWIS/weather data are not on social media. The Twitter feed is automated for the Olympic and Southwest Regions.

5. Does the DOT have a data repository that allows public and/or third-party access to the raw data? How is the data accessible and how is it documented?

   See Table 1: List of DOT Data Repository Sites.

6. Any other relevant information you would like to share?

- "We are in the process of applying for research funding to test a road condition reporting app developed [by another state DOT]. We have been looking for a more efficient process. Reporting to 511 from the trucks would improve our reporting times and the information that goes, as well as make it easier on our operators."
- Currently "… deploying non-intrusive pavement sensors to provide pavement temperature and pavement state. Many other RWIS states are doing the same to avoid the deterioration of in-pavement sensors. The pavement temperature measurements provided by these sensors are reasonable when compared with the traditional in-pavement sensors. The pavement condition (contaminant depth, friction, and pavement state) needs to be further evaluated."
- Currently "… involved in a vigorous seasonal weight restriction program. Vertical thermistor strings are deployed from the pavement surface down through six feet. These thermistor strings are polled daily and the vertical temperature profiles presented on [the road weather information system]. Users can look at the current or past monthly, weekly, or daily profiles. We have recently added to the weight restriction program by adding seasonal freeze/thaw profiles for each thermistor site. While not tied directly to the traveler information program, these temperature data probes can be used by our pavement engineers to institute/remove weight restrictions and by the commercial trucking enterprise to view the likelihood of weight restrictions."
- In addition to the state DOT, a number of different public entities within the state (e.g., metropolitan planning organizations) provide traveler information. Thus, when considering unified statewide data quality processes, it isn't enough to just look at the state DOT.
- Anticipate having VMS on the traveler information website in the next year.
- Data flows are property of the state DOT. The real time feeds do not keep meaningful records for historical purposes.
- One contractor is responsible for the data flow from the RWIS systems as well as their maintenance. Another contractor handles the 511 website, 511 phone system, data entry system, modular traveler information system, and the ingestion of auto feeds into the system.
- Currently conducting a project for traveler information program assessment.
- Focusing on training at the section level.
- "Communications staff is thin."
- New data exchange system includes a visualization module that can graphically represent the status and performance of all devices and their associated data, as well as visually portray any of the current and historical data in the system.
- More VMS and travel time information being integrated into the statewide traveler information system website.

- The "… Traffic Operations Center is a statewide traffic management center. My position of Traveler Information Manager is unique among DOT's. I am a part public information officer, part traveler information quality control in house at the TOC. I find this very valuable and helpful in keeping data consistent, timely and complete."
- Avoid storing / archiving data and recording video because the public disclosure requests and retention policies were "distracting" (e.g., one region was receiving two to five legal video requests per week and that was becoming overwhelming). Traffic studies are acceptable. Traffic consultants and universities do use images.

7. Is there anyone else that I should talk with, either within (state) or in other states?
    - Jack Stickel, Geospatial Engineering Services Manager, Alaska ISSD (Information Systems & Services Division)
    - For perspective from the user side on what the DOT is doing, ask Metro 511 providers. (California)
    - Phil Braun – VMS (Idaho)
    - Theresa Bousliman – RWIS (Montana)
    - Maricopa County DOT – Nevada Data Exchange based on their system.
    - Washington State DOT (mobile presence and multi-modal build out), Amtrak Cascade
    - Missouri DOT – social media outreach program
    - Minnesota DOT – web and 511 systems
    - Tina Greenfield, Iowa DOT – Aurora project, robust existing RWIS network.
    - Dennis Jensen, Idaho – winter maintenance, algorithms.

8. Would you like a copy of the project's findings?
    - Yes.

13. APPENDIX F: TRAVELER INFORMATION WEBSITES

Table 2: State DOT Traveler Information Websites

State	Website	Description
AK	Alaska 511 511.alaska.gov/	Alerts, Driving Conditions, Weather Watches, Weather Cameras, Planned Events, Future Planned Events, Message Boards, Travel Not Advised- No Winter Maintenance.
CA	Caltrans QuickMap quickmap.dot.ca.gov/	Traffic Speeds, Lane Closures, CHP/CHIN Incidents, Message Signs, Cameras, Chain Controls, Full Road Closures, High Speed Rail Full Closures.
ID	Idaho Transportation Department hb.511.idaho.gov/	Road Reports, Winter Driving, Traffic Speeds, Cameras, Weather Stations, Electronic Signs, Mountain Passes, Rest Areas, Social Media.
MT	Montana RoadReport roadreport.mdt.mt.gov/travinfomobile/	Road Incidents, Construction, RWIS, RWIS with Camera, Other Camera, Road Conditions, Radar, Rest Areas.
NV	511 Nevada NVRoads nvroads.com/	Incidents, Construction, Weather, Special Events, Cameras, Freeway Message Signs, Weather Sensors, Rest Areas, Welcome Centers, Truck Parking, Road Conditions, Live Traffic Speeds.
OR	ODOT TripCheck tripcheck.com/	Road Cameras, Road Closures, Transit Delays, General Information, Weather Hazards, Weather Warnings, Snow Zones, Weather Stations, Construction, Truck Restrictions, Local Road Conditions, Special Areas, Parking Lots, Crashes, Traffic.
UT	UDOT Traffic udottraffic.utah.gov/	Traffic, Cameras, Signs, Incidents, Construction, Lane Closures, Weather.
WA	WSDOT wsdot.com/traffic/	Traffic, Travel Alerts, Ferries, Mountain Passes, Weather, Cameras.

14. REFERENCES

1. Alfelor, Roemer M. "Baselining Current Road Weather Information: Characterizing Sources and Evaluating Attributes." Washington, D.C.: Road Weather Management, U.S. Department of Transportation, 17 September 2010.
2. *Baselining Current Road Weather Information: Results of the 2010 Quality and Importance Survey Final Report*. Washington, D.C.: Research and Innovative Technology Administration, U.S. Department of Transportation, 2011.
3. Battelle, Cambridge Systematics Inc., Texas Transportation Institute. *Traffic Data Quality Measurement Final Report*. Washington, D.C.: Office of Highway Policy Information, Federal Highway Administration, 2004.
4. Charoensripongsa, Thiraphat, Wasan Pattara-atikom and Sukree Sinthupinyo. "A Histogram-Based Detection of Corrupted Images from Traffic Surveillance Cameras." *2011 8th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. Khon Kaen, Thailand: Institute of Electrical and Electronics Engineers, 2011. 480-483.
5. Chavan, Priya, et al. "Extending the Use of Archived ITS Data As a Potential Management Tool to Evaluate Traveler Information on Dynamic Message Signs." *Transportation Research Board 87th Annual Meeting Compendieum of Papers*. Washington, D.C.: Transportation Research Board, 2008.
6. Cluett, Chris and Fred Kitchener. *Implementation and Evaluation of the Sacramento Regional Transportation Management Center Weather Alert Notification System*. Washington, D.C.: Office of Operations, Federal Highway Administration, 2010.
7. Drobot, Sheldon, et al. *The Vehicle Data Translator V3.0 System Description*. Washington, D.C.: FHWA Office of Operations, Research and Innovative Technology Administration, U.S. Department of Transportation, 2011.
8. Federal Highway Administration. *Real-Time System Management Information Program*. 20 October 2015. Web. February 2015. <http://www.ops.fhwa.dot.gov/1201/>.
9. —. "Real-Time System Management Information Program [1201] Fact Sheet." 8 November 2010. *Real-Time System Management Information Program.* February 2015.
10. Fiebrich, Christopher A., et al. "An End-To-End Quality Assurance System for the Modernized Coop Network." *15th Conference on Applied Climatology, 13th Symposium on Meteorological Observations and Instrumentation*. Savannah, Georgia, 2005. J3.3.
11. Fowler, Tressa, et al. "Robust Spatial Quality Control of Road Weather Sensor Measurements." *15th Symposium on Meteorological Observation and Instrumentation*. Atlanta, GA: The 90th American Meteorological Society Annual Meeting, 2010. 3.2.
12. Hart, Robert, et al. *Baselining Current Road Weather Information: Summary Report*. Washington, D.C.: Research and Innovative Technology Administration, U.S. Department of Transportation, 2009.

13. Hart, Robert, Leon Osborne and Steve Conger. *Baselining Current Road Weather Information: Final Report*. Washington, D.C.: Research and Innovative Technology Administration, U.S. Department of Transportation, 2009.

14. Koonar, Awtar, Paul Delannoy and Dave Denault. "Building a Road Weather Information Network for Integrating Data from Heterogeneous Sources." *Proceedings of the Third International Conference on Information Technology and Applications*. Sydney, Australia: Institute of Electrical and Electronics Engineers, 2005. 501-507.

15. Limber, Martha, Sheldon Drobot and Tressa Fowler. *Clarus Quality Checking Algorithm Documentation Report*. Washington, D.C.: Research and Innovative Technology Administration, U.S. Department of Transportation, 2010.

16. Manfredi, John, et al. *Road Weather Information System Environmental Sensor Station Siting Guidelines*. Washington, D.C.: Office of Transportation Operations, Road Weather Management Program, Federal Highway Administration, 2005.

17. Margiotta, Rich. *State of the Practice for Traffic Data Quality: White Paper*. Washington, DC: Office of Policy, Federal Highway Administration, 2002.

18. Mesonet. *Quality Assurance*. n.d. April 2016. <http://www.mesonet.org/index.php/quality_assurance/quality_assurance>.

19. MesoWest, University of Utah. *MesoWest Quality Control Information Help Page*. n.d. April 2016. <http://mesowest.utah.edu/html/help/qc.html>.

20. Middleton, Dan, Deepak Gopalakrishna and Mala Raman. *Advances in Traffic Data Collection and Management, White Paper*. Washington, D.C.: Office of Policy, Federal Highway Administration, 2002.

21. Minnesota Department of Transportation. *2012 CMS Manual of Practice*. Minnesota Department of Transportation, 2012.

22. National Oceanic and Atmospheric Administration. *MADIS Quality Control*. 12 November 2014. April 2016. <http://madis.noaa.gov/madis_qc.shtml>.

23. —. *MADIS: Meteorological Assimilation Data Ingest System, Home Page*. 28 January 2015. April 2016. <http://madis.noaa.gov/index.shtml#>.

24. Pisano, Paul A. and Lynette C. Goodwin. "Research Needs for Weather-Responsive Traffic Management." *Transportation Research Record: Journal of the Transportation Research Board* 1867 (2004): 127-131.

25. Pisano, Paul A., et al. "Concept of Operations for Clarus – The Nationwide Surface Transportation Weather Observing and Forecasting System." *TRB 85th Annual Meeting Compendium of Papers*. Washington, D.C.: Transportation Research Board, 2006.

26. Pisano, Paul and James Pol. "Clarus - America's 21st Century Surface Transportation Weather Observing and Forecasting System." Washington, D.C.: Road Weather Management, Federal Highway Administration, 2004.

27. Richter, Daniell, Shaowei Wang and Douglas Galarus. *WeatherShare Phase 2 Final Report*. Sacramento, California: Division of Research and Innovation, California Department of Transportation, 2010.

28. Road Weather Management Program, Office of Transportation Operations, Federal Highway Administration. *Weather Information for Roadway Transportation Preliminary Data Gap Analysis*. Washington, D.C.: Office of the Federal Coordinator for Meteorology, 2004.

29. Robinson, Emanuel, et al. *Deployment, Use, and Effect of Real-Time Traveler Information Systems*. Washington, D.C.: National Cooperative Highway Research Program, Transportation Research Board, November 2012. June 2016. <http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_w192.pdf>.

30. Schuman, R. "Summary of Transportation Operations Data Issues." *National Summit on Transportation Operations*. Ed. A O'Brien. Washington, D.C.: Federal Highway Administration, 2001.

31. Shafer, Mark A., et al. "Quality Assurance Procedures in the Oklahoma Mesonetwork." *Journal of Atmospheric and Oceanic Technology* 14 June 1999: 474-494.

32. Steering Committee, ATIS-related data collection and data quality, ITS America ATIS Committee. *Closing the Data Gap: Guidelines for Quality Advanced Traveler Information System (ATIS) Data*. Washington, D.C.: U.S. Department of Transportation, ITS America Advanced Traveler Information Systems Committee, 2000.

33. Trachy, Larry, Ricky Via and Scott Cowherd. "SAFETEA-LU Section 1201 Real-Time System Management Information." *Introductory MPO Collaborative Meetings*. Operations Division, Virginia Department of Transportation, 2013. Presentation. 2016. <http://www.mwcog.org/uploads/committee-documents/a11aXldc20130807135940.pdf>.

34. Turner, Shawn. *Defining and Measuring Traffic Data Quality, White Paper*. Washington, D.C.: Office of Policy, Federal Highway Administration, 2002.

35. Turner, Shawn, et al. "Guidelines for Evaluating the Accuracy of Travel Time and Speed Data." 2011.

36. United States Department of Transportation. "Real-Time System Management Information Program, 23 CFR Part 511." *Federal Register, Rules and Regulations* 8 November 2010.

37. Wang, Yuan-Kai, Ching-Tang Fan and Jian-Fu Chen. "Traffic Camera Anomaly Detection." *22nd International Conference on Pattern Recognition*. Ed. IEEE Computer Society. Stockholm, Sweden: Institute of Electrical and Electronics Engineers, 2014. 4642-4647.