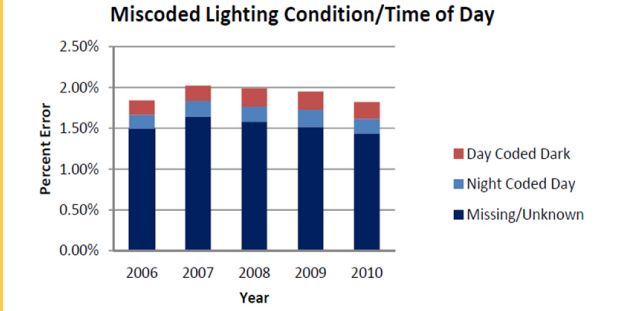
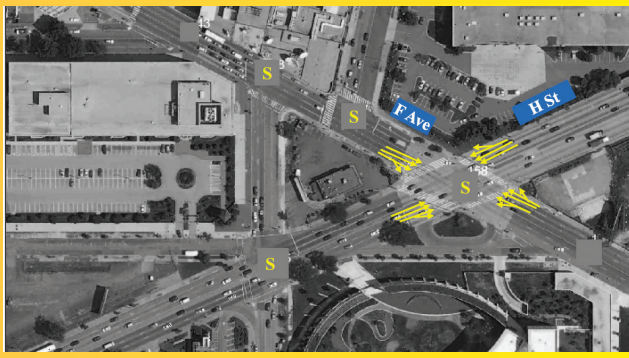
Crash Data Improvement Program (CDIP) Final Report





FHWA Safety Program

U.S. Department of Transportation
Federal Highway Administration

Safe Roads for a Safer Future
Investment in roadway safety saves lives

http://safety.fhwa.dot.gov

FOREWORD

The Safe Accountable Flexible Efficient Transportation Equity Act: a Legacy for Users (SAFETEA-LU) authorizing legislation stressed that states should develop and implement measures to assess how well their safety data systems are performing and to provide a baseline measure to assess future improvements in data quality. In response, the Federal Highway Administration (FHWA) developed the Crash Data Improvement Program (CDIP) to provide states with a means to measure the quality of the information within their crash database. It provides states with an approach that can be used to establish quantified measures of data quality for the timeliness, accuracy, completeness and consistency of their crash data. Additionally, it provides measures to assess a state's ability to integrate crash data with other safety databases and how well the state makes the crash data accessible to users.

The CDIP also familiarizes the collectors, processors, maintainers and users of crash data with the concepts of data quality and how quality data helps to improve safety decisions. Through the CDIP, a technical assistance team, comprised of crash data system experts, works with a state to conduct a data quality workshop and a technical transfer session with the administrators, managers and technicians of the state's crash database to identify issues and opportunities for the state to improve their crash data system. The CDIP technical assistance team develops a report providing the state with specific recommendations for improvement.

CDIPs have been conducted with states since 2008. During this time, 21 states and the District of Columbia have participated in the program. This report reflects the findings from the CDIP's conducted between 2010 and early portion of 2013. Some states have made notable progress in improving their crash data systems and the quality of their crash data as a result of their participation the CDIP.

After serving as a partner to FHWA in the CDIPs for a number of years, the National Highway Traffic Safety Administration (NHTSA) will take over the administration of this program beginning in Fiscal Year 2014. FHWA, recognizing the value and importance of quality crash data, will continue to support efforts to improve the quality of crash data and crash data systems through the CDIP and other programs.

Michael S. Griffith
Director, Office of Safety Technologies

Notice

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document.

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this report only because they are considered essential to the objective of the document.

Quality Assurance Statement

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

TECHNICAL DOCUMENTATION PAGE

<table>
<tr><td colspan="2">1. Report No.
FHWA-SA-14-001</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient's Catalog No.</td></tr>
<tr><td colspan="3" rowspan="2">4. Title and Subtitle
Crash Data Improvement Program (CDIP) Final Report</td><td colspan="2">5. Report Date
January 2, 2014</td></tr>
<tr><td colspan="2">6. Performing Organization Code</td></tr>
<tr><td colspan="3">7.Author(s)
Robert A. Scopatz, Jack Benac, and Nancy Lefler</td><td colspan="2">8. Performing Organization Report No.</td></tr>
<tr><td colspan="3" rowspan="2">9. Performing Organization Name and Address

Vanasse Hangen Brustlin, Inc. (VHB)
8300 Boone Blvd., Suite 700
Vienna, VA 22182-2626</td><td colspan="2">10. Work Unit No.</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
DTFH61-10-D-00022 (VHB)</td></tr>
<tr><td colspan="3" rowspan="2">12. Sponsoring Agency Name and Address
Federal Highway Administration Office of Safety
1200 New Jersey Ave., SE
Washington, DC 20590</td><td colspan="2">13. Type of Report and Period
Draft Final Report, October 2010 – October 2013</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code
FHWA</td></tr>
<tr><td colspan="5">15. Supplementary Notes
The contract manager for this project was Robert Pollock.</td></tr>
<tr><td colspan="5">16. Abstract
This report provides an overview of the Crash Data Improvement Program (CDIP) covering the period from October 2010 through May 2013 during which time 10 States participated. Common findings from the CDIPs include that States lack formal data quality management programs and do not routinely report data quality measurements in most of the “six pack” attributes of timeliness, accuracy, completeness, uniformity/consistency, integration, and accessibility. Common recommendations include developing formal data quality management practices; creating a single, centralized crash data records system; increasing the detail in descriptions of key crash data improvement programs listed in the Strategic Plan for Traffic Records Improvement, and creating routine reports of a complete set of data quality measurements. Recommendations for the future of the CDIP program are also provided, including expanding the focus of the CDIP to go beyond safety engineering concerns to address law enforcement and injury control as well.</td></tr>
<tr><td colspan="2">17. Key Words:
safety data, roadway inventory data, traffic data, data collection, crash data, data quality</td><td colspan="3">18. Distribution Statement
No restrictions.</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classif. (of this page) Unclassified</td><td>21. No. of Pages
79</td><td colspan="2">22. Price</td></tr>
</table>

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
	NOTE: volumes greater than 1000 L shall be shown in m^3			
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	$candela/m^2$	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	$candela/m^2$	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

TABLE OF CONTENTS

List of Tables

List of Figures

ACRONYMS

CDIP	Crash Data Improvement Program
CODES	Crash Outcome Data Evaluation System
DDACTS	Data Driven Approaches to Crime and Traffic Safety
DOT	Department of Transportation
E911	Enhanced 911 (emergency call-in system)
eCrash	Electronic crash reporting
EMS	Emergency Medical Services
FARS	Fatality Analysis Reporting System
FFY	Federal Fiscal Year
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
GIS	Geographic Information System
IT	Information Technology
LEA	Law Enforcement Agency
LRS	Linear Reference System
MMUCC	Model Minimum Uniform Crash Criteria
NHTSA	National Highway Traffic Safety Administration
PDO	Property-damage-only crash
SAFETEA-LU	Safe Accountable Flexible Efficient Transportation Equity Act-a Legacy for Users
SHSP	Strategic Highway Safety Plan
TAT	Technical Assistance Team
TRCC	Traffic Records Coordinating Committee
YTD	Year-to-date

EXECUTIVE SUMMARY

The Crash Data Improvement Program (CDIP) provides States with a means to measure the quality of the information within their crash database, and provides recommendations for improvement. It can provide States with measures to address the six data quality attributes of timeliness, accuracy, completeness, uniformity/consistency, integration, and accessibility. Additionally, the CDIP helps familiarize the collectors, processors, maintainers and users with the concepts of data quality and how high quality data helps to improve safety decisions.

This report covers the period of October 2010 through May 2013, during which time ten States participated in the program. None of the States had data quality performance measurements for more than one attribute—that is, States that measured timeliness, for example, were not *also* measuring accuracy, completeness, or other data quality attributes. More importantly, none of the States operate a formal, comprehensive data quality management program. Only two of the States had enough of a start on such a program that the CDIP Technical Assistance Team (TAT) could recommend additions to the existing processes. In the other eight States, the TAT recommended the State begin by developing a formal, comprehensive data quality management program.

In the CDIP reports to States, the TAT has recommended a gradual approach in establishing and formalizing a data quality management program. This is important because (a) States should establish the program that makes the most sense for their own situation, and (b) it is costly and resource-intensive to set up such a program while most of the effort will be manual in nature. The better course is to work toward the more formal data quality management program by including it in the Traffic Records Strategic Plan as a project (or series of projects) and to link those projects to specific upgrades to the crash data management system (software and analytic capabilities) over time. Automating the components of a data quality management program—especially the components related to calculating data quality performance measures—helps to lower costs and increases the likelihood that the State will be able to maintain the effort in the long term.

States are also very likely to need help with strategic planning. The CDIP includes a review of the documents the State provides, including their most recent Traffic Records Strategic Plan update and the Strategic Highway Safety Plan (SHSP). In some States, the two plans are not coordinated. The CDIP has also uncovered that some States do not include sufficient details in their plans to fully describe the process they will use to reach key goals—such as the goal to increase electronic submission of crash reports by law enforcement agencies (LEAs). If a State lacks knowledge of the technological capabilities of the LEAs, for example, it cannot effectively plan the future of electronic submissions from those agencies. Where States have detailed, up-to-date information on the capabilities of LEAs to collect and submit crash data electronically,

staff can also determine how best to meet the needs for interfaces, data submission guidelines, a certification process, and test procedures. Several States need assistance in setting up an effective process.

Perhaps the most critical need seen among the CDIP States is for a single official repository of crash data. Many States are faced with a situation in which it is easier and cheaper for key users of the crash data to simply make changes to a copy taken for their own use, largely relegating the official crash data system to the role of repository of the original source data. Unfortunately, this "cheaper, quicker" method has unrecognized costs associated with its long-term use. Most notably, States in this situation often find that the highest quality crash data are not in the official crash data repository but in some other system maintained by a key user agency. Worse, there may not be a single "best" version of the data. A single centralized system that includes both the as-submitted and corrected/enhanced crash data would also support most of the processes of a formal, comprehensive data quality management program.

States will not achieve this ideal centralized system quickly or without cost. The CDIP reports include advice to States that they should add these capabilities as they are redesigning or upgrading their crash records systems. At that time the data quality support features can be added to the system for a relatively small incremental cost. Grafting these capabilities onto an existing system, conversely, is likely to be difficult and expensive.

The CDIP is well accepted by States and has been successful in responding to States' comments and suggestions for improvement. The value of a focus on engineering uses of crash data is appreciated by the participants from the crash data management operation and law enforcement as much as it is by the engineers themselves. These are definitely positive aspects of the program and should be retained as it moves from the Federal Highway Administration (FHWA) to the National Highway Traffic Safety Administration (NHTSA).

There are opportunities to expand the CDIP to reach a broader audience and to address a wider range of issues. For example, there are key users in law enforcement whose needs mesh well with the engineering needs already being addressed because LEAs need location data. Injury Surveillance Programs' needs also mesh well with the engineering uses already focused on in the CDIP. Some of the examples in the CDIP Workshop are drawn from injury surveillance analyses; the Crash Outcome Data Evaluation System (CODES) information included in the section on Data Integration. The utility of CODES and similar programs in the engineering context stems from their ability to produce valid information on the outcomes of crash-related injuries both in terms of severity and cost—two things that engineers normally have to estimate, often using sources they know are not as accurate as the data from State healthcare records.

INTRODUCTION

The Crash Data Improvement Program (CDIP) provides States with a means to measure the quality of the information within their crash database, and provides recommendations for improvement. It can provide States with measures to address the six data quality attributes of timeliness, accuracy, completeness, uniformity/consistency, integration, and accessibility. Additionally, the CDIP helps familiarize the collectors, processors, maintainers, and users with the concepts of data quality and how high quality data helps to improve safety decisions.

This report covers the years 2010 through 2013. The CDIP began in 2008 and has been delivered in 18 States, 10 which are under the current contract between VHB, Inc. and the Federal Highway Administration (FHWA) Office of Safety. Figure 1 shows the eight States that were completed before the start of the current contract, the 10 states completed at the time this report was developed, and four more States that are scheduled for completion before the end of the contract in 2013. This will bring the total number of CDIP States to 22 by the end of 2013.

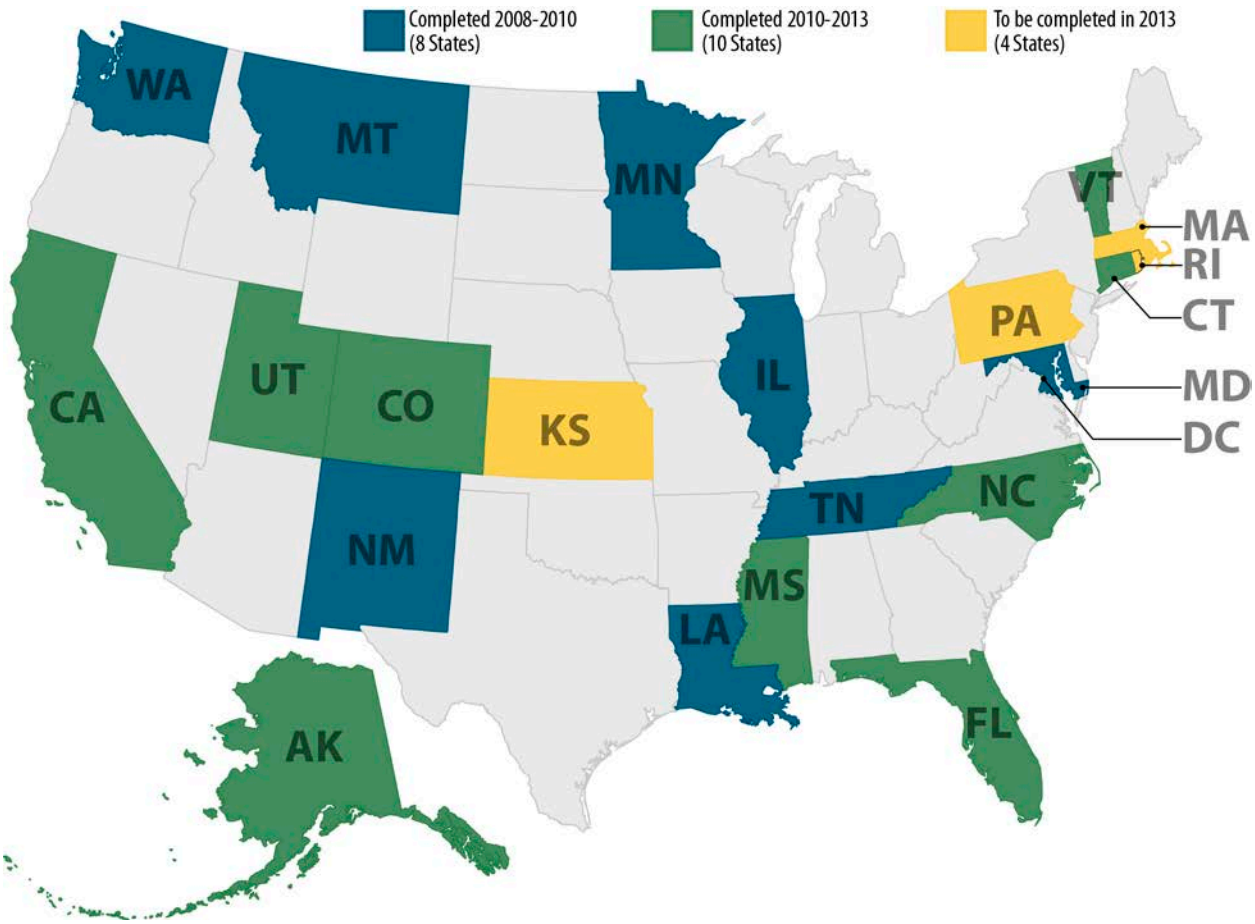


Figure 1: All CDIP States 2008-2013.

CDIP started in response to weaknesses in States' submissions to the Section 408 grant program (Traffic Records Improvement Funding) under the preceding Transportation Authorization Bill: SAFETEA-LU (Safe Accountable Flexible Efficient Transportation Equity Act—a Legacy for Users). The panel reviewing States' grant applications observed that many States had few (if any) quantitative measures of crash data. Many of the data quality measurements that States reported provided only project-specific views of data quality rather than the desired global, system-wide perspective that was called for in the legislation and the Section 408 grant application guidance. It was clear from the quality of the submissions, and from Traffic Records Assessments, that many States lacked a formal, comprehensive crash data quality management system. Moreover, it appeared that few States truly understood the methods and importance of data quality measurement. The CDIP has been intent on providing States with guidance on assessing and improving crash data quality. The delivery mechanism to States has used the expert advice of a two- or three-person Technical Assistance Team (TAT). The team's mandate is to provide States with detailed, readily-implemented recommendations that will lead directly to improved crash data quality measurement and management.

In October 2010, the new contract kicked off with a call to modify the CDIP process, update the pre-site visit materials, and revise the training content. At the request of FHWA and several States, the CDIP team has increased the specificity of the presentations and recommendations —using the State's own data and analyses by preference. Wherever possible, the CDIP recommendations take the form of a worked-out example of how the State should calculate the performance measures recommended by the TAT. If the State has provided the TAT with analytic support or a crash data file that supports calculation of the performance measures, the TAT produces the example performance measures using the State's own data. If the TAT cannot use State data, examples are drawn from other States, including those used in the CDIP presentations. All examples are based on real data.

Beginning in Federal Fiscal Year (FFY) 2013, the National Highway Traffic Safety Administration (NHTSA) joined FHWA on the CDIP management and oversight team. This is a prelude to NHTSA taking full control of the CDIP program in FFY 2014. This change is in parallel with changes NHTSA is implementing in the Traffic Records Advisory and Assessment process. One goal expressed by NHTSA is to use CDIP as one of several types of *GO Team* support supplied to States at the States' request. In preparation for this transition, the CDIP materials were revised in 2013 to bring them into agreement with the NHTSA Advisory materials so that the same terminology is used in both. In addition, a revised information request and questionnaire now identifies tie-ins to the questions States already answered in the Assessment. In the future, as more States have gone through the new Assessment process, the CDIP team will be able to pre-fill much of the CDIP questionnaire thus reducing the burden placed on States and avoiding having to ask a State for information they have already supplied to NHTSA. States will still need to review and update their answers to the questions.

CDIP PROCESS AND METHODOLOGY

The CDIP project team consists of the FHWA Office of Safety Project Manager, the contractor Project Manager, and the TAT members. The CDIP team has marketed the CDIP to States through direct contact and at conferences and meetings. States may request a CDIP through their FHWA Division Office and, more recently, through the NHTSA regional staffs or the Traffic Records Team in the National Center for Statistics and Analysis. Information describing the program is also available via the CDIP website where users may obtain a summary of the program, the CDIP brochure, and the 2010 *CDIP Program Guide (http://safety.fhwa.dot.gov/cdip/)*. During the transition to NHTSA, these materials will be updated and the website will move. States that express interest in having a CDIP are contacted by the CDIP team to begin the scheduling and information exchange process that precedes CDIP on-site work. In the months leading up to the TAT's site visit the State must schedule suitable conference rooms for three meetings on three successive days, respond to the *CDIP Information Request* and complete the *CDIP Pre-Site Visit Questionnaire* (see Appendix A). Sufficient lead time is required to enable the State to respond and for the TAT to conduct analyses and tailor the CDIP presentation in reaction to the documents and answers the State provides.

The CDIP process follows a standard sequence:

- Initial contact with State.
- Key contacts to settle logistics for the workshop (location, dates, venue).
- Preparation:
    - State completes information request and questionnaire.
    - State supplies crash data or completes requested analyses.
    - TAT reviews State-supplied materials.
    - TAT analyzes State data or reviews State's analyses.
    - TAT revises presentation materials tailoring to the State's needs.
- On-site in the State:
    - Day 1: one-day CDIP Workshop.
    - Day 2, morning: roundtable session with the crash data file manager/administrator and key staff, plus NHTSA, FHWA and Federal Motor Carrier Safety Administration (FMCSA) regional/division staff.
    - Day 2, afternoon: optional short tour of the crash data processing facility.
    - Day 2, evening: the TAT produces the Matrix Summary Report.

  - Day 3: report out to all interested parties.
- Formal CDIP Final Report due 30 days after the site visit.
- Follow-up at three, six, and 12 months after the CDIP Final Report.

Attendance sheets are collected at each of the three CDIP meetings to identify participants. Evaluations are collected at the end of the Workshop on Day 1. Feedback on Day 2 activity is gathered during the out-brief meeting on Day 3 and via the State's comments on the *Matrix Summary Report.* A *CDIP Final Report* is provided to the State in draft form and revised based on the State's comments.

PRE-SITE VISIT MATERIALS AND ACTIVITIES

The *CDIP Information Request* provided in Appendix A presents a list of documents and data that States supply including, but not limited to, their Strategic Plan for Traffic Records, their most recent Traffic Records Assessment, relevant sections of their Strategic Highway Safety Plan (SHSP), a copy of the crash report, the crash reporting instruction manual, the crash database data dictionary, crash data quality edit checks and business rules, descriptions of projects designed to improve crash data quality, and existing crash data quality reports. States are also asked to produce a series of standard data quality analyses referenced in the CDIP. As an alternative, they may provide three or more years of their most recent crash data (redacted to remove personal identifiers) for the TAT to conduct the analyses.

The *CDIP Pre-Visit Questionnaires* provided in Appendix B and C present a series of questions covering the crash data custodial responsibilities, laws and policies, data quality management program, and specific data quality measurements of timeliness, accuracy, completeness, uniformity/consistency, integration, and accessibility. Each of these data quality attributes is defined in the questionnaire and information request. Appendix B is the original questionnaire used through 2012. Appendix C is the updated questionnaire used during 2013.

The State uploads all materials to the contractor extranet site assigned for the project. This site allows two-way sharing of documents and data files between the State and CDIP team. The CDIP team reviews all the documents and the answers to the questionnaire in the weeks leading up to the CDIP site visit. If the State chooses to produce their own analyses—calculations of standard data quality measures—those will be reviewed at this time. If the CDIP team conducts the analyses there are additional contacts with the State in order to ensure the TAT analyst understands the database structure and contents and that items such as the total number of crashes are in close agreement with the numbers already reported by the State. To facilitate this process, States are asked to provide a crash data users guide along with the data extracts.

When the CDIP team's reviews and analyses are completed, they revise the presentation materials (workshop slides and talking points) to tailor the session contents to the unique features of the State's crash data management process and to include State-specific examples of data quality measurements. The final materials are provided to the State in advance of the CDIP site visit.

ON-SITE CDIP MEETINGS AND ACTIVITIES

The CDIP site visit typically sets Monday as a travel day which allows the team to arrive, unpack materials, and review the presentations. The next three days are a combination of meetings with State personnel and post-session work by the TAT. The CDIP contents are designed to appeal to an audience including data collectors, data managers, and data users. The first day is designed as a workshop training session to explain and promote data quality measurement. It typically has the largest attendance of the three day CDIP process. The State is encouraged to gather a group of 20-30 individuals from engineering, law enforcement, crash data management, the State's Traffic Records Coordinating Committee (TRCC) membership, and other stakeholders, but the decision of who to invite is left up to the State. The second day is intended to be a detailed discussion with the crash data manager and a select group of data collectors and key data users. Typically the State safety engineer and a representative of State-level law enforcement will attend along with the crash data custodian and key staff responsible for data entry, electronic data transfer, database management, information technology (IT) support, FARS, SafetyNet, and data analysis. The third day is set aside for the report out in which the TAT briefs attendees on their findings and recommendations. Typical attendees include most of the people who participated in the discussions on the second day, plus any additional invitees. Some States choose to invite senior management and agency executives or schedule a separate briefing with these upper-level decision makers.

The CDIP contents are designed to appeal to an audience including data collectors, data managers, and data users.

On Day 1, the TAT conducts the CDIP Workshop. The Workshop is divided into the following sections:

- **Section 1: Introduction.** The TAT and the State participants introduce themselves and discuss the goals for the day and for the CDIP effort overall.

- **Section 2: Opening Exercise.** Participants are divided into groups and asked to discuss their use of crash data, unmet needs, and impressions of overall crash data quality.
- **Section 3: CDIP Introduction.** The TAT presents details of the CDIP goals and process, followed by an overview of the six data quality attributes to be discussed in Sections 4A and 4B.
- **Section 4A: Quality Attributes, Part 1.** The TAT presents definitions and examples of measurements of *timeliness, accuracy, and completeness*.
- **Section 4B: Quality Attributes, Part 2.** The TAT presents definitions and examples of measurements of *uniformity, integration, and accessibility*.
- **Section 5: Group Exercise.** Participants break into groups to conduct a safety review of a complex intersection based on one of two versions of the crash data. One set of groups receives a relatively complete dataset with detailed information about each crash and motor vehicle and pedestrian traffic in the intersection. The other set of groups receives a less complete data set which lacks several key pieces of information. Both groups are asked to identify problems at the intersection and to propose solutions, as well as to discuss any problems they saw in the data. At the end of the exercise, the groups present their findings and recommendations, and the facilitators reveal that not everyone had a complete dataset. The full meeting reconvenes to discuss the value of complete and accurate data.

On Day 2, the TAT facilitates a half-day round-table discussion with the select group of attendees who are most responsible for administering/managing/maintaining the crash database. The goal of the session is to confirm and expand on conclusions the TAT has reached based on the pre-site visit questionnaire answers, their review of documents supplied by the State, and the results of any analyses conducted in support of the CDIP site visit. This meeting is also used to discuss the feasibility of any preliminary recommendations the TAT is considering in order to gauge the State's ability to implement the suggestions. In this way, the TAT is able to more fully understand the existing processes and the barriers to implementing improvements. After the meeting, the TAT works alone to produce the *Matrix Summary Report*. The report is submitted to the FHWA program manager for review and finalized that evening or early the next morning.

On Day 3, the TAT delivers the *Matrix Summary Report* to the primary State contact for the CDIP. The report's contents are reviewed and discussed in a meeting typically lasting two to three hours. This meeting is often attended by the key personnel responsible for crash data management, safety engineering, and law enforcement. Participants are encouraged to respond verbally to the findings and recommendations as presented and to provide their comments after reviewing the *Matrix Summary Report*. The TAT then presents the timeline and activities

required for completion of the CDIP, including the State providing comments on the *Matrix Summary Report* and the TAT incorporating those comments into an expanded *CDIP Final Report.* At the State's request, Day 3 may also include a separate briefing for key agency executives.

MATRIX SUMMARY REPORT

The *Matrix Summary Report* is designed as a simplified report listing the TAT's findings and associated recommendations. These are presented in a table where each row addresses a specific data quality management process or a data quality attribute measurement. The section headings are:

- **Administration (crash database):** This section identifies the crash data custodian by name, agency, and office/section.
- **Crash Data Processing:** This section covers the **crash data production process** (methods of creating the central database), the **crash location process** (methods of assigning location codes to individual crash reports), and the **data quality management process** (standard procedures the State uses to assess and improve data quality).
- **Timeliness:** This section covers measurements of overall crash data timeliness and the timeliness of component processes (submissions from law enforcement, location coding, data quality management, and other post-submission data management processes).
- **Accuracy:** This section covers measures designed to identify internal disagreement among data fields in the crash report and errors resulting in incorrect information in the crash report. Typical inclusions are measures of the accuracy of key data fields, the proportion of crashes that can be successfully located on the roadway network, and the State's tracking of errors and corrections.
- **Completeness:** This section covers measures designed to identify under-reporting and missing data. Typical inclusions are comparisons of current year-to-date (YTD) versus historic average reporting levels, the proportion of serious crashes (injury and fatal) of all crashes reported, and the percentage of reports with missing data in key fields.
- **Uniformity:** This section covers measures designed to identify inconsistencies among law enforcement agencies' (LEAs) interpretation of the requirements for reporting and the definitions of terms in the crash report. Typical inclusions are measures comparing current year reporting patterns to historical averages at the level of individual LEAs. This section also covers measures of the crash report form's level of compliance with the Model Minimum Uniform Crash Criteria (MMUCC) guideline. States were encouraged to provide whatever MMUCC compliance assessment information they had available—the TAT did not perform an independent review. The TAT reported the

current status as "measured" or "not measured" in those States that provided information.

- **Integration:** This section covers measures designed to assess the level of success the State has in merging crash data with other sources of traffic records information. Typical inclusions are measures assessing the level of integration of the crash database with roadway inventory, driver/vehicle, citation, and injury surveillance (emergency medical services (EMS), trauma registry, emergency department, hospital discharge, and vital records) data.
- **Accessibility:** This section covers measures designed to count the users and uses of crash data. Typical inclusions are the number of requests for data or analysis received and fulfilled, the uses of any web-based analytic resources, and user satisfaction survey results.

CDIP FINAL REPORT

The *CDIP Final Report* is due to the State approximately one month after the completion of the CDIP site visit. It has the same sections as the *Matrix Summary Report* and represents an expansion of that document to provide additional background information; more detailed findings, and expanded explanations of each recommendation. For example, in the *Matrix Summary Report* there might be a recommendation that the State develop a measure based on the ratio of serious (fatal and injury) crashes to total crashes reported by each LEA. That same recommendation in the *CDIP Final Report* would discuss the reasons for the recommendation, the method of calculation, and provide a table (or extract of a table) presenting sample data for the State. The *CDIP Final Report* also includes, at the State's request, the CDIP team's suggested priorities. These are based on the TAT's understanding of the critical needs in the State as well as the State's preparedness to implement each of the recommendations in the report.

The *CDIP Final Report* is delivered in draft form to the State. A final draft is produced addressing each of the State's comments on the draft version.

CDIP FOLLOW-UP

The CDIP project team follows-up with each CDIP State at prescribed intervals. Three, six, and twelve months after delivery of the final report, the State is contacted to obtain feedback and a status update on each of the recommendations. There are no penalties for failure to implement a recommendation. The follow-up phone call or email is a way for FHWA to gauge the practical utility of the CDIP report. States are asked for suggestions to improve the project and specifically which of the recommendations have proven useful, or problematic. Another form of follow-up with CDIP States has been scheduled sessions/peer exchanges at the Traffic Records Forum. FHWA has sponsored CDIP States' meetings at Traffic Records Forums and

taken the opportunity to gather feedback and suggestions for improvement. These sessions also serve as useful marketing opportunities when other interested States attend.

CDIP RESULTS

This section presents the results of the ten CDIPs conducted between December, 2010 and May, 2013. Figure 2 shows the States that have participated in a CDIP during this period.

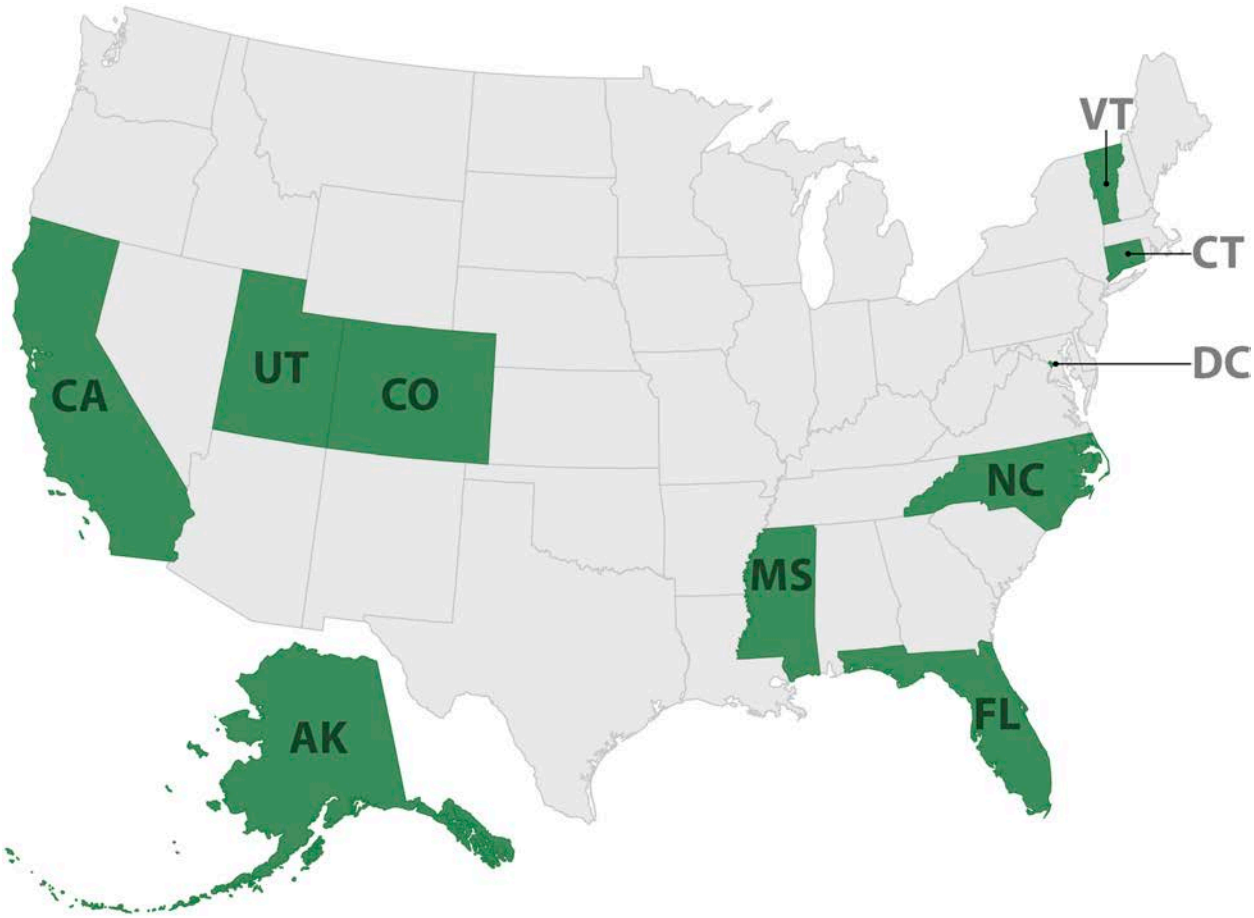


Figure 2: CDIP States 2010-2013.

Since 2010, the following ten States have participated in a CDIP:

- Alaska.
- California.
- Colorado.
- Connecticut.
- District of Columbia.
- Florida.
- Mississippi.
- North Carolina.
- Utah.
- Vermont.

CDIP WORKSHOP EVALUATIONS

At the end of Day 1, participants in four of the 10 States completed an evaluation form. The form is divided into segments to match the portions of the CDIP Workshop, as follows:

- Section 1: Introduction.
- Section 2: Opening Exercise.
- Section 3: CDIP Introduction.
- Section 4A: Quality Attributes, Part 1.
- Section 4B: Quality Attributes, Part 2.
- Section 5: Group Exercise.

Participants rated each segment using a five point scale (five being the best) on each of the following five dimensions:

- Coverage of the topic.
- Information quality.
- Illustrations.
- Visual impact.
- Applicability.

The form also included space for comments on each segment and overall for the entire Day 1 program.

Fifty-nine participants completed an evaluation. The results of the evaluations show that participants generally rated the CDIP Day 1 program highly. Figure 3 shows the overall distribution of scores for each of the five dimensions and overall for the entire program. Each set of five bars (corresponding to the ratings of five to one) sums to 100 percent.

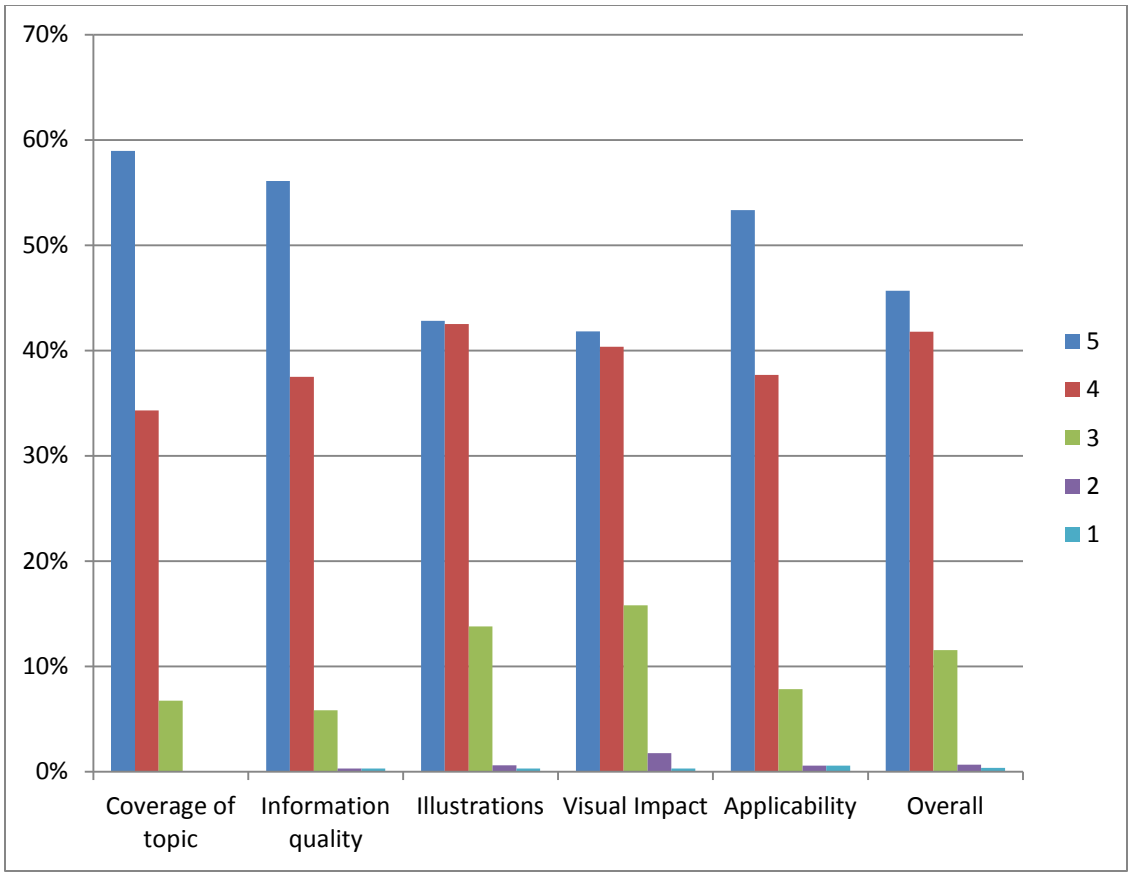


Figure 3: Overall distribution of ratings (five-point scale) for the CDIP workshop.

As seen in the figure, Workshop participant ratings of five (best) and four predominate in each dimension and for the CDIP overall. In all cases the top two responses account for between 80 to 90 percent of the ratings. There are some differences in ratings across dimensions, with participants rating "Coverage of topic" and "Information quality" most highly, followed by "Applicability". "Illustrations" and "Visual Impact", which were less highly rated than the other dimensions, still received high marks.

COMMON FINDINGS

This section presents the most common findings and recommendations included in the ten CDIP Final Reports produced from October, 2010 through May, 2013. Any finding that appeared in at least five of the reports is included in this section. The text following each list includes selected less frequent findings and recommendations that have general importance for FHWA, the States, or recommendations for the future of CDIP.

Findings

In the ten CDIP reports, an average of 24 findings were referenced in the recommendations, the list below presents the most common findings. Numbers in parentheses are the number of States (out of 10) for which the finding was noted in the CDIP Final Report.

Crash Data Processes:

- There is already widespread use of electronic crash data collection software (7).
- The State is working to increase electronic submission (7).
- Multiple copies of the crash database exist in multiple agencies (7).
- Errors are corrected centrally, but not logged (6).
- eCrash reports are printed for manual central data entry (5).

Crash Location Process:

- The roadway inventory lacks some (or all) local roadways (7).
- Crash location codes are assigned manually (6).

Data Quality Assessment:

- There is no formal data quality management program (8).
- Data users correct errors but corrections are not captured in the official file (7).
- Reports returned to the LEA for correction are not tracked (6).

Timeliness:

- There are no measures of the timeliness of submission by LEAs (9).
- There are no measures of the timeliness of component processes (9).
- There are no measures of overall timeliness (6).

Accuracy:

- There are no measures of accuracy (10).
- Edit checks have not been reviewed and appear inadequate (9).

None of the ten States had routine measures in two or more data quality attributes.

Completeness:

- There is strong evidence of under-reporting, especially of property damage only (PDO) crashes (9).
- There are no measures of completeness (8).

Uniformity/Consistency:

- There are no measures of uniformity other than MMUCC compliance (8).
- MMUCC compliance has not been measured (5).

Integration:

- Data integration exists and is working (10).
- There are no measures of data integration (9).

Accessibility

- There are no measures of data accessibility (7).

The findings, in general, indicate that States are not engaging in formal data quality management and, in particular, almost all States lack sufficient data quality performance measures for timeliness, accuracy, completeness, uniformity, integration, and accessibility. Few States log the errors that are noted in crash reports. This refers to both the data entry processes *and* errors noticed later by users. In the most typical case, when a user identifies an error in a crash record, they will fix the problem in their own copy of the data but then either (a) do not communicate the error and correction to the central crash data manager, or (b) the central crash data management process does not result in an update to the official record. Only one State kept an error log, but the contents of that log were not used as part of the data quality management process—that is, the information in the log was not used to develop training content, change the form or data collection manual, or develop new edit checks to address the most frequent errors. As implemented, the error log only applied to electronic data and did not include a way to log corrections to reports once they were accepted into the central crash database.

Roadway inventory files generally lack local roadway data. That means that when States measure the accuracy of location coding for crashes the number is, at best, incomplete. States in this situation may map 100 percent of crash locations, but they can only "land" the crashes for locations in their linear reference system (LRS). Depending on how the State counts the local roadway crashes in its indicator of success, this number could either over- or under-

estimate the accuracy of location information. If the State reports the percentage of crashes that are assigned geographic coordinates and appear in the crash layer of its geographic information system (GIS), then the result is usually at or near 100 percent. The proportion of crash records that link to a roadway inventory record is often much lower, corresponding to those roadway segments the State maintains plus the subset of local roads on which the State gathers data for Federal reporting purposes.

States clearly need guidance not just on what to measure, but how to incorporate data quality as part of the day-to-day management of the crash data resource.

In reviewing States' ability to measure the "six pack" of data quality attributes, the most common finding is that the attributes are not measured in a systematic way. Many States *can* produce measures, but the finding stands that the measures are not produced and reported *on a regular basis*. States that do measure data quality performance in one attribute typically have one measurement defined rather than a series of related measures. Such States are likely to have no measures of other data quality attributes. Thus, six States have measures of "overall timeliness" but only one State has a measure of crash data submission timeliness. That same State lacks measures of accuracy, completeness, uniformity/consistency, integration, and accessibility. *None of the ten States had routine measures in two or more data quality attributes.* This has important implications for data quality management in several ways, not the least of which is that most States are not providing the data collectors in law enforcement any quantitative feedback on the quality of the reports they are submitting.

One State had a key agency that failed to report any crashes at all; however, systematic under-reporting is a serious problem affecting every State that participated in the program. Three States have adopted a crash reporting threshold that contributes directly to the under-reporting problem either by setting the damage threshold too high or because law enforcement officers are confused as to the exact criterion to apply. Two of the ten States address under-reporting by using reports submitted by drivers involved in crashes (operator reports). While not a recommended practice (drivers provide inaccurate data on some fields of the form), operator reports can help a State at least gauge the level of under-reporting by law enforcement.

Nine of ten States have not reviewed their edit checks in several years, and none do so on a periodic basis. For these nine states, the edit checks had not been reviewed since the most recent system update and sometimes not even then. The time spans ranged from five to twenty years or more. Edit check review is a process that would benefit from information contained in an error log, if such a log existed.

Some less frequent findings are also worth noting because they have an important impact on States' ability to manage their crash data:

- As noted, five of the seven States that were described as having "widespread use of eCrash" also made at least some of those LEAs print their crash reports and send them in for manual data entry by the State. In every case, the State has a plan for increased electronic submission by law enforcement. In three of *those* States, the plan lacked details such as a timeline and interim measures of success.
- Three States are unable to identify which LEA is responsible for a specific crash report. As a result, it is impossible for the crash data managers to provide specific feedback to the agencies. This eliminates much of the value of a performance measurement system for crash data quality as the most crucial measures relate directly to data collection.
- Three States have a crash report submission/entry backlog of more than a year. That makes it nearly impossible to implement any data quality improvement project that relies on feedback to law enforcement. Timely feedback is required for activities including (a) returning crash reports to the originating officer for correction; (b) providing agency-specific tallies of most common errors; and (c) discussions of deficiencies in edit checks and supervisory reviews.

Recommendations

The ten CDIP reports contained an average of 34 recommendations. The following list shows the most frequent issues addressed in the recommendations mirror the list of most frequent findings. Note that in some cases the recommendation appears in a different section of the report; the same key issues are addressed.

Crash Data Process:

- Greater specificity is required for plans to increase electronic submission (7).

Crash Location Process:

- Automate crash location collection through a combination of technology deployed in the field (e.g., smart map) and post-processing algorithms and tracking (9).
- Expand the LRS/roadway inventory database to include all public roads (5).

Data Quality Assessment:

- Create error logs to track corrections to crash reports as requested by key users (10).

- Establish formal data quality management processes (8).
- Establish a single official resource reflecting *as submitted* and *final* data including all corrections and additions (8).
- Track the return/resubmission of corrected reports by law enforcement (6).
- Formalize the year-end close-out analysis process (5).
- Analyze and report data quality throughout the year (5).
- Transition resources away from data entry and toward data quality (5).

Timeliness:

- Measure and report the timeliness of submissions by each LEA (10).
- Establish a measure of overall timeliness (9).
- Establish a measure of timeliness for each component of the crash data management process (9).

Accuracy:

- Develop measures of accuracy (10).
- Track the percentage of crash reports "landed" in the LRS (9).
- Report accuracy for each LEA specifically and in aggregate to the TRCC (9).
- Create an automated review of the accuracy of critical data elements (8).
- Compare as-submitted to final corrected data (8).
- Review, revise, and expand the list of edit checks (8).
- Develop and/or improve communication with law enforcement (6).

Completeness:

- Develop measures of completeness (10).
- Measure the proportion of missing and "unknown" in critical data elements (9).
- Compare LEAs on the proportion of serious crashes using (fatal+injury)/total crashes as a measure of under-reporting (9).
- Report completeness for each LEA specifically and in aggregate to the TRCC (9).
- Compare YTD to prior years' reporting levels (8).

- Encourage more LEA reporting of crashes through use of the measurements (8).

Uniformity:

- Use the (fatal+injury)/total ratio as a measure of uniformity (7).
- Measure and/or increase the level of MMUCC compliance (7).

Integration:

- Develop macro-level integration measures to show which files are merged (10).
- Develop micro-level integration measures to gauge the success of merger (10).

Accessibility:

- Develop measures of accessibility based on requests received and fulfilled (10).
- Measure user satisfaction (8).
- Develop a web portal for users to access data and analytic tools (8).

The recommendations are somewhat surprising in the frequency with which States were advised to create basic, top-level measurements for five of the six data quality attributes. For at least the past five years (the most recent cycle of Traffic Records Assessments), States have been required to develop data quality performance measures in order to qualify for funding under Section 408 of SAFETEA-LU. This requirement has continued under Section 405 of the new Transportation Reauthorization (MAP-21: Moving Ahead for Progress in the 21st Century). For timeliness, accuracy, completeness, integration, and accessibility, every CDIP report included a recommendation to create basic measurements assessing data quality. This is the result of the finding that none of the ten States had a data quality management program that included systematic measurement and reporting of more than one data quality attribute—most had no such measures.

The States do not engage in formal, comprehensive data quality management processes. The recommendations in the crash data management, location, and data quality assessment processes take on greater importance as a result. States clearly need guidance not just on *what* to measure, but how to incorporate data quality as part of the day-to-day management of the crash data resource. Measurement alone would move the States forward, but measurement in the context of a formal data quality management program would make it clear that data quality improvement is part of the job of data collection, data administration, and data use. The system to do this is not in place. States need both a series of measurements *and* a process to interpret

and use those measurements within the context of a program that is designed to address the deficiencies that the numbers serve to uncover.

In every final report, the CDIP team acknowledged the costs associated with data quality management and made suggestions on the State-specific priorities and ways to fund the effort. One frequent suggestion applies to those States that still invest resources in managing a paper crash reporting process (either with staff or through a contract). In five States, the CDIP team recommended that the crash data custodian begin the process of transitioning away from a traditional crash data entry role to a crash data quality management role. In particular, the resources currently devoted to managing the paper reporting process can be reduced as electronic data *submission* increases. The States planning for increased electronic submission are advised to retain a portion of the resources rather than offer the entire data entry/paper management operation as a budget cut. The remaining resources should be allocated to support the recommended increases in data quality management activities. In some States, the same staff and job titles can be retained as the work transitions to focus on data quality management. In others, the transition will require a change in personnel and/or job titles, or a different type of vendor contract.

Some lower-frequency recommendations are important to consider for what they say about current State needs and in recommendations for the future of CDIP. These are:

- Develop/deploy a web-based crash reporting capability. This was recommended in four States. The CDIP findings indicated that many States do not have a detailed plan for how they will achieve their highest possible level of electronic submissions. One technology that could help is web-based crash reporting. Especially for small agencies that write only a few crash reports each year, web-based reporting is an economical option that can avoid the cost of in-vehicle equipment by allowing officers to complete a crash report using a desktop computer online.
- Implement the LRS in the statewide GIS. This recommendation applied in four States. Most States already do this to some extent (at least for the State-maintained roadways), but for those that have not already done so, GIS offers an elegant solution to the problem of merging crash and roadway data. In order to work with all of the State's other roadway data, it is imperative that the LRS be mapped in the GIS. The States that received this recommendation are already aware of the power of GIS and want to implement their LRS in the GIS environment. The formal recommendation helps support that effort. (The formal mapping of all public roads into a state's LRS is now a requirement from the FHWA per the August 7, 2012 Guidance Memorandum, **Geospatial Network for All Public Roads.**)
- Develop a unique identifier for each LEA in the crash database. As noted under the findings, this is a crucial recommendation for the three States to which it applies.

Without it, they will not be able to implement many of the most effective recommendations presented in their CDIP report. Crash data quality improvement efforts rely heavily on the knowledge and skills of the data collectors in law enforcement. Timely feedback serves an instructional role, helping officers and supervisors spot errors before the reports are submitted.

FOLLOW-UP CONTACT RESULTS

The FHWA CDIP program manager followed up with the six States that participated during 2010 and 2011. These States were:

- Alaska.
- Colorado.
- Connecticut.
- Florida.
- Mississippi.
- Utah.

Of the remaining four States, two (Vermont and District of Columbia) have been contacted by the contract team to obtain preliminary follow-up results, and the other two (California and North Carolina) were too recent to be able to comment on project implementations based on the CDIP recommendations. Because of the preliminary nature of the follow-up results obtainable from these four States, their results are not reflected in the following discussion.

Table 2 presents a summary of all results, showing the status of recommendations in each of the seven main sections of the CDIP reports.

Table 1: Status of recommendations in six States.

Status	Crash Production Process	Timeliness	Accuracy	Complete-ness	Uniformity	Integration	Accessibility	Total
Implemented/ Partial	23	9	4	5	6	3	4	54
Not Implemented	13	4	5	5	0	3	7	37
Plan to Implement	21	11	11	10	10	8	6	77
Total	57	24	20	20	16	14	17	168

Figure 10 shows the percentages of recommendations for each CDIP section that have been implemented, those that the State plans to implement, and those that the State has decided *not* to implement.

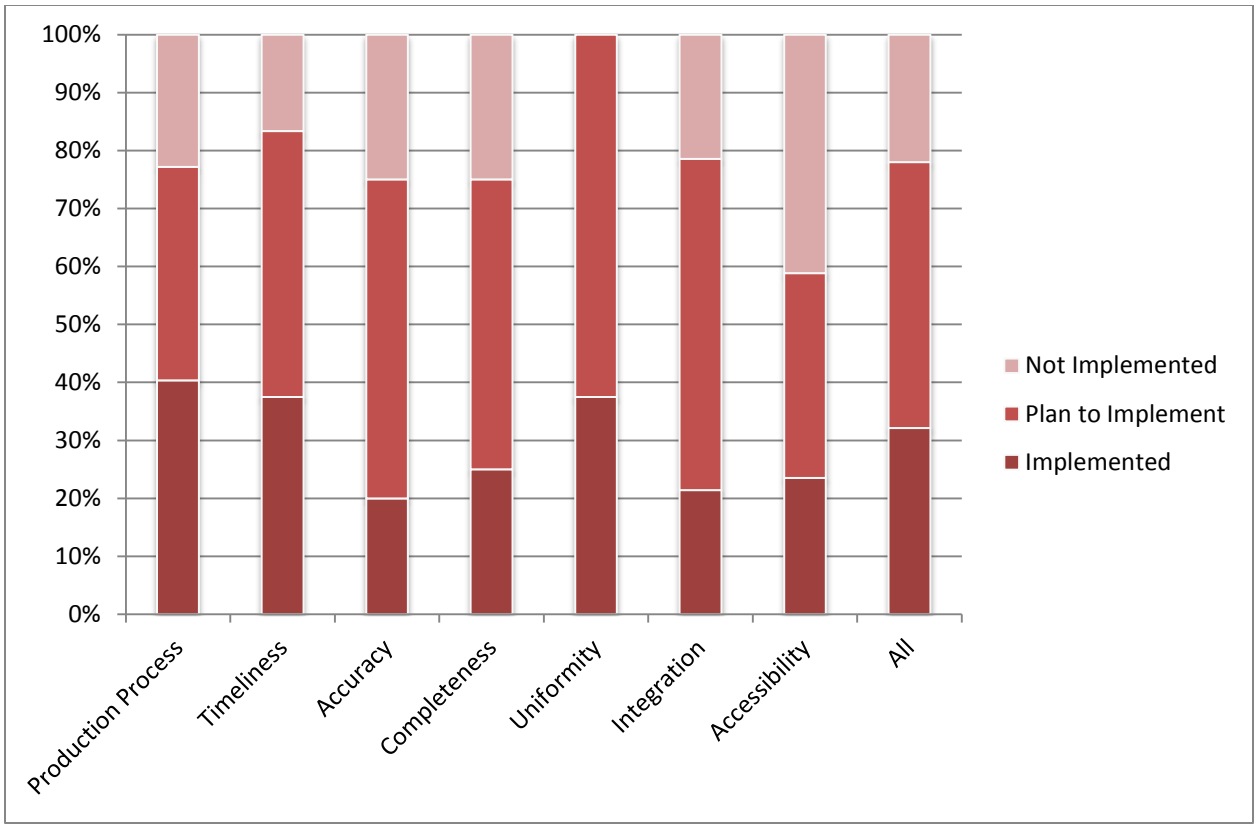


Figure 4: Status of recommendations.

The figure shows that 78 percent of recommendations are either implemented (32 percent) or the State has plans to implement (46 percent). Within the categories corresponding to the sections of the CDIP reports, the recommendations related to the crash production process were most likely (40 percent) to have been implemented by the States, while those related to accuracy (20 percent) were least likely. Including States' plans for implementation shows that the uniformity recommendations (100 percent) were the most likely to be implemented eventually, while those related to accessibility were the least likely to be implemented (41 percent were listed by States as "not implemented").

States provided comments for each recommendation explaining their decision to implement, plan to implement, or not implement. These are too numerous to present individually. A review of the State's responses leads to the following general conclusions:

- As expected, some of the already implemented recommendations are items that the TAT recommendation endorsed or reinforced existing State efforts and urge that they be completed.

- States are generally finding it easier to implement process-oriented recommendations than to implement recommendations that require new analyses or analytic resources.

- All but one of the six States planned a comprehensive process to review crash data quality measurement overall and work toward establishing a data quality management process to include new data quality measurements. The one State that had no plans expressed uncertainty over how it could implement the measurement-related recommendations in general.

- The higher implementation percentage for uniformity recommendations is, in part, due to the number of recommendations related to MMUCC compliance reviews—an easy recommendation to implement (or plan to implement) because the work is done through NHTSA rather than the State's personnel.

- Of the recommendations that States said they were *not* implementing, most were rejected due to resource constraints combined with a perception that the State would reap diminishing benefits over time. Several responses indicated that the State had decided to use its limited resources to make progress on other actions that would remove the need for the recommendation in question—by reducing the number of paper crash reports received.

- Recommendations requiring changes to centralized crash reporting systems are the hardest for States to implement, but in practically every case the States reported that they plan to implement those recommendations as time and resources permit. Some of the responses included statements indicating that the State had already sought funding to proceed with implementation.

Overall, the results of the CDIP follow up contacts provide some encouraging news—States are moving forward with existing plans that the CDIP TAT endorsed and are planning to create more formal data quality management programs, including increased data quality measurement. The barriers noted by States relate most often to a lack of resources, rather than a lack of interest or desire to improve data quality. The tendency to reject recommendations that relate to outmoded paper-management process makes sense in the context of resource constraints. It does have one unfortunate downside which the States do acknowledge. Without measurement of the current paper-based process (especially of timeliness, accuracy, and completeness), the State will not have a good baseline against which to assess progress as the percentage of reports collected and submitted electronically increases. This is unfortunate because the comparison of paper and electronic reports can help to justify further investment in electronic data collection and submission. Ultimately, as long as a State is making progress in establishing formal data quality management processes, it will be able to quantify the benefits of its data improvement programs. In the long run, it is more important that the States establish and maintain a formal data quality management process than that they establish, in every case, a baseline measurement of the older paper-based processes.

It will also be interesting to see over the course of three to five years how many States develop the more formal data quality management processes. In the longer term, as statewide crash data systems are replaced or upgraded, the States should find it easier (and less resource intensive) to measure data quality. In fact, the CDIP TAT often included in its recommendations the idea that data quality measurements should be built into the next system update in lieu of doing it now when the effort would be largely manual, and thus resource intensive.

SUMMARY, KEY FINDINGS AND FUTURE CONSIDERATIONS

The CDIP is completing its fifth year of operation; third year with the current contractor. Over that time it has been modified and updated in response to feedback from the participants. In particular, States asked for more relevant examples. The CDIP contents and process were revised to include the State's own data in the Workshop and Technical Transfer sessions, where possible. This change is significant because it also helped the TAT better understand the technical barriers facing analysts who would be charged with calculating the measures of data quality performance recommended in the CDIP reports. Only one State among the ten opted to conduct their own data analyses—in that State's CDIP, the TAT used the results tables provided by the State as content in the Workshop.

Two States have requested follow-up assistance through their FHWA Division Office and NHTSA Regional Office representatives. In both cases, the need was for help in combining multiple strategic plans into a single comprehensive document. One of the States opted to create a business plan that showed the linkages among several projects designed to improve the crash data. This helped the State identify task dependencies that might delay high-priority projects because a crucial activity in a different project was not completed in time. The other State needed help combining its Traffic Records Strategic Plan with its SHSP to produce one document that upper-level management in the affected agencies could review and endorse. This effort went beyond the CDIP in that it included all of the traffic records components; however, the single largest category of projects in both plans was that related to crash records improvement.

Other States have said that one of the reasons they scheduled the CDIP was because they wanted to integrate CDIP findings in their Strategic Plan for Traffic Records and their SHSP. The other thing that has happened in the aftermath of a CDIP is that the States see the need to integrate their plans more thoroughly—that projects they've been pursuing on a separate course may in fact interact in ways that can and should be part of the plan.

The CDIP has evolved into a planning tool for States. As such, it also has taken on an important role in supplementing the Traffic Records Assessments conducted by NHTSA. In recognition of this fact, FHWA and NHTSA have agreed to transition the CDIP from FHWA to NHTSA. Federal Fiscal Year (FFY) 2013 is a transition year in which both administrations have worked cooperatively to manage the program. In FFY 2014, NHTSA will take full control of the CDIP. In anticipation of that event, the contract team worked with NHTSA and FHWA to modify the CDIP guide, the pre-site questionnaire, the pre-site visit information request, and the contents of the CDIP Workshop presentations. Most of the changes were made to more closely align the CDIP terminology with the wording used in the NHTSA Traffic Records Advisory and Assessment. In addition, the pre-site visit information request and questionnaire

were cross-referenced to show States which items they may have already provided or answered as part of their Assessment. In the future, this will be a useful feature because it is expected that the CDIP will become part of NHTSA's follow-up with States requesting help in response to recommendations in their Traffic Records Assessment. States can request help, in the form of Go Teams, for a variety of traffic records improvement projects. The CDIP TAT will be one example of a NHTSA Go Team.

Throughout the period of 2010 through 2013, the CDIP TAT advised States to establish formal, comprehensive data quality management programs. The description of what such a program might look like has become a shared item among FHWA and NHTSA. The 2011 revision of the crash section of the NHTSA Traffic Records Advisory includes the same description of a formal, comprehensive data quality management program as is presented in the CDIP reports. That description was also used in the last two-to-three years of Traffic Records Assessments (i.e., as the 2011 revision was being developed, but before it was first used in a State Traffic Records Assessment). As a result of this coordination, there is now one well-documented list of what a State's crash data quality management program should include. The components are:

- **Automated edit checks and validation rules** that ensure entered data fall within the range of acceptable values and is logically consistent between other fields. Edit checks are applied when data are added to the record. Many systems have a two-tiered error classification system, distinguishing critical errors that must be corrected before submission and non-critical error warnings that may be overridden.
- **Limited State-level correction authority** is granted to quality control staff working with the statewide crash database to amend obvious errors and omissions without returning the report to the originating officer. Obvious errors include minor misspellings, location corrections, and directional values. Obvious omissions include missing values that can easily be obtained from the narrative or diagram. States may opt to grant no editing authority to quality control staff and still meet this description—the point is that the authority must be *limited*.
- **Processes for returning rejected crash reports** are in place to ensure the efficient transmission of rejected reports between the statewide data system and the originating officer as well as tracking the corrected report's submission.
- **Performance measures** are tailored to the needs of data managers and address the concerns of data users. Measures can be aggregated from collectors, users, and the State TRCC. The crash data should be timely, accurate, complete, uniform, integrated, and accessible. These attributes are tracked using State-established quality control measures. The State is encouraged to develop additional measures that address their specific needs.

- **Numeric goals**—or performance metrics—for each performance measure are established and regularly updated by the State in consultation with users via the TRCC.
- **Performance reporting** provides specific feedback to each LEA on the timeliness, accuracy, and completeness of their submissions to the statewide crash database relative to applicable State standards.
- **High-frequency errors** are used to generate new training content and data collection manuals, update the validation rules, and prompt form revisions.
- **Quality control reviews** comparing the narrative, diagram, and coded report contents are considered part of the statewide crash records system's data acceptance process.
- **Independent sample-based audits** are conducted periodically for crash reports and related database contents. A random sample of reports is selected for review. The resulting reviews are also used to generate new training content and data collection manuals, update the validation rules, and prompt form revisions. At a minimum, these audits occur on an annual basis.
- **Periodic comparative and trend analyses** are used to identify unexplained differences in the data across years and jurisdictions. At a minimum, these analyses occur on an annual basis.
- **Data quality feedback** from key users is regularly communicated to data collectors and data managers. This feedback will include corrections to existing records as well and comments relating to frequently occurring errors. Data managers disseminate this information to law enforcement officers as appropriate.
- **Data quality management reports** are provided to Department Administrator(s) and TRCC for regular review. The TRCC used the reports to identify problems and develop countermeasures.

Eight of the ten CDIP implementations discussed in this report included the finding that the State lacked a "formal, comprehensive data quality management program." This is not to imply that the other two States had such a program *in its entirety*. Rather, the CDIP TAT found, in those two States, that the existing quality control program was functioning sufficiently well that the recommended quality control improvements (especially in the area of measurement) could be added to the existing management program.

It is likely, based on experiences in CDIP and in the Traffic Records Assessments over the most recent five-year period, that very few States (likely less than 20 percent), have a crash data quality management program. Most will have heard that their existing data quality management program lacks specific components from the list presented above. It is also likely that a majority of States will need help to establish such a program. In the CDIP reports, the TAT

very often recommends that the State proceed gradually in establishing and formalizing a data quality management program. This is important because (a) States should establish the program that makes the most sense for their own situation, and (b) it is costly and resource-intensive to set up such a program when most of the effort will be manual in nature. The better course is to work toward the more formal data quality management program by including it in the Traffic Records Strategic Plan as a project (or series of projects) and to link those projects to specific upgrades to the crash data management system (software and analytic capabilities) over time. Automating the components of a data quality management program—especially the components related to calculating data quality performance measures—helps to lower costs and increases the likelihood that the State will be able to maintain the effort in the long term.

As noted earlier in this section, States are also very likely to need help with strategic planning. The CDIP includes a review of the documents the State provides, including their most recent Traffic Records Strategic Plan update and the SHSP. In some States, the two plans are not coordinated. The CDIP has also uncovered that some States do not include sufficient details in their plans to fully describe the process they will use to reach key goals—including the goal to increase electronic submission of crash reports by LEAs. If a State lacks knowledge of the technological capabilities of the LEAs, for example, it cannot effectively plan the future of electronic submissions from those agencies. Where States have detailed, up-to-date information on the capabilities of LEAs to collect and submit crash data electronically, staff can also determine how best to meet the needs for interfaces, data submission guidelines, a certification process, and test procedures. Several States appear to need some assistance in setting up an effective process, especially if they are trying to accommodate multiple different crash data collection or records management systems in use among LEAs.

Perhaps the most critical need seen among the CDIP States is for a single official repository of crash data. Many States are faced with a situation in which it is easier and cheaper for key users of the crash data to simply take a copy (a data extract) for their own use, largely relegating the official crash data system to the role of repository of the source data. Users in this situation will make improvements to the copy of the data that they maintain. Unfortunately, this "cheaper, quicker" method has unrecognized costs associated with its long-term use. Most notably, States in this situation often find that the highest quality crash data is not in the official crash data repository, but in some other system maintained by a key user agency. Worse, there may not be a single "best" version of the data. The corrections made in a State DOT are different from those that would be made in support of an injury surveillance program, for example. The highest quality data would be achieved only if the corrections applied by all key users were available for analysis by others. States that can transition their legacy centralized crash databases to serve multiple users with complete and corrected data will ultimately save money by reducing redundancy and sharing analytic knowledge and resources. Although this benefit has not yet been measured in any State, the reduction in redundant

databases and the improved communications implicit in a formal data quality management program should generate measurable savings and data quality improvements.

In the ideal centralized system, the crash database would have a record of the data as it was originally submitted along with a log of all changes that have been made to the data. That change log (or transaction table) can be used to roll-forward or roll-back a crash record to access any desired set of changes, from none at all to every change of a particular type to every change made for any purpose. This type of system is ideal for automating the calculation of crash data quality measurements of accuracy and completeness. Once the relevant reports are created, the crash data managers could obtain at any time, and for any time period, a report that shows (a) the most frequent errors, (b) the LEAs making the errors, and (c) the most frequent corrections to specific errors. This information not only describes the quality of the data in the crash system, it also supports several of the components of a formal, comprehensive crash data quality management program. With this information, a State could update its crash report form and data collection manual, add specific examples of likely errors to training and messages sent to law enforcement in general, and provide specific feedback to individual LEAs on their own most frequent errors. In addition, some of the errors can be eliminated through imposition of new edit checks and validation rules so that field data collection software and the data intake process for the statewide database can catch the errors before they end up in the system.

It should be clear that States may not achieve this ideal system quickly or without cost. The CDIP reports include advice to States that they should add these capabilities as they are redesigning or upgrading their crash records systems. At that time, the data quality support features can be added to the system for a relatively small incremental cost. Grafting these capabilities onto an existing system, conversely, is likely to be difficult and expensive. Follow-up with CDIP States in three to five years after they received the recommendations is likely to show substantial progress in formalizing data quality management practices, especially in those States that have upgraded their crash data management systems.

RECOMMENDATIONS

The CDIP is well accepted by States and is successful in responding to States' comments and suggestions for improvement. The value of the focus on engineering uses of crash data is appreciated by the participants from the crash data management operation and law enforcement as much as it is by the engineers themselves. These are positive aspects of the program and should be retained as it moves from FHWA to NHTSA.

There are opportunities to expand the CDIP to reach a broader audience and to address a wider range of issues. For example, there are key users in law enforcement whose needs mesh well with the engineering needs already being addressed. As more LEAs adopt the Data Driven Approaches To Crime and Traffic Safety (DDACTS), or similar analysis-based resource allocation models, they will rely more heavily on detailed location data for crashes, citations, and crime data. This fits well with the uses already covered in the CDIP and could easily be incorporated in the examples presented in the Workshop.

Injury Surveillance Programs' needs also mesh well with the engineering uses already focused on in the CDIP. Some of the examples in the Workshop are drawn from injury surveillance analyses; for example, Crash Outcome Data Evaluation System (CODES) information using hospital discharge data is included in the section on data integration. The utility of CODES and similar programs in the engineering context stems from their ability to produce valid information on the outcomes of crash-related injuries both in terms of severity and cost—two things that engineers normally have to estimate, often using sources known to be inaccurate. The appeal of the CDIP program to users from the injury surveillance program staffs is limited because most of the presentations rely on how engineers could better use the injury surveillance data, and not vice versa.

Altering the CDIP to appeal to a broader audience, including injury surveillance, court, and driver or vehicle records staffs is likely to be a challenge. The material as it is today does not touch on these users' needs. These needs are different enough from the kinds of issues covered in the CDIP's engineering-focused examples and exercises that accommodating them would require large changes in the program. This could be accomplished with a longer workshop, for example, or by cutting back on the amount of presentation material in order to allow time for additional group exercises. The new exercises could focus on the needs of one or more of these other user groups by examining uses of crash data for driver control, epidemiological research, or as a factor in adjudication. In addition, examples of data quality performance measures could be expanded to include items that specifically relate to these other, non-engineering users of the crash data. Examples of key data elements and measurements of their accuracy and timeliness, in particular, would be useful in broadening the

audience to include these other groups. As was done for the existing CDIP materials, it would be a good idea to test any new materials with a select audience of representative participants.

As the CDIP transitions from FHWA to NHTSA, there are further opportunities to reinforce the messages from both the CDIP and the NHTSA Traffic Records Advisory/ Assessment process, the CDIP could also help to reinforce the need for more detailed strategic plans. One way to do this would be to set aside time in the Day 2 Technology Transfer session to discuss the State's strategic planning efforts and how they are coordinated. The CDIP includes a review of the documents that the State provides, including the Traffic Records Strategic Plan and the SHSP. These are both also reviewed as part of the Traffic Records Assessment. States that need help coordinating these two plans (and perhaps other strategic planning documents as well), could also request additional help from NHTSA in the form of a different Go Team. However, it is a good idea to intentionally incorporate a brief discussion of strategic planning on Day 2 of the CDIP. If nothing else, this will give the TAT a good idea of what the State personnel think is part of their existing plans and how they see those plans unfolding over the next few years.

APPENDIX A: CDIP INFORMATION REQUEST

CRASH DATA IMPROVEMENT PROGRAM
PRE-SITE VISIT INFORMATION COLLECTION

PLEASE PROVIDE THE FOLLOWING RESOURCES:

- The most recent Traffic Records Assessment (can also be obtained thru NHTSA – with permission of the State).
- The most current version of their Traffic Records Strategic Plan used for Section 408 applications (can also be obtained thru NHTSA – with permission of the State).
- The CDIP questionnaires that are sent to the state for completion (attached).
- A current version of their crash production process flow chart (this is one of the items requested in the CDIP questionnaire).
- Any other materials the state feels may be relevant, e.g. research studies that highlight difficulties with using state crash data, or any previous documentation on measuring crash data quality.

CRASH DATA ANALYSIS

In addition to the above resources, the FHWA CDIP project team would like to opportunity to review the State's crash data prior to the site visit. Below is a list of crash data analyses to be conducted by the state. If the State does not have resources to conduct the analysis, please provide 3-5 years of historical crash data to the project team, and they can conduct the analyses.

Timeliness:

- A: Overall: Event date to availability on the statewide database (ready for use/analysis) - annual averages.
- B: If available, specific components of timeliness:
    - Law Enforcement submission: Time from Event to receipt of crash reports- annual averages.
    - Data Entry: Time from receipt to completion of primary data entry - annual averages.

- Location Coding: Time from completion of primary data entry to application of location codes - annual averages.
- Other: Any other processes (QC, data cleansing, financial responsibility, other?) required prior to data availability for use/analysis - annual averages.

Accuracy:

- A: Locations: Percent of reported crash locations that automatically code (without human intervention) to the appropriate state locating coding system (LRS, Rte # & milepoint, coordinate-based, whatever the appropriate system is). If applicable, the percentages of successful location coding can be provided separately for state-maintained (on-system) and local (off-system) roadways. Additionally, if the state tracks both initial automated location coding and a final percent coded successfully (with human intervention/correction), reporting both numbers is welcome.
- B: Form-level reject rate: Percent of crash reports rejected (i.e., failing to pass mandatory—fatal error—edit checks) upon initial submission. NOTE: this may not be relevant for your system or may only be relevant to those reports submitted on paper (if any).
- C: Specific Field Edits and Logic Checks:
  - Agreement between time of day and lighting condition (e.g., percent of reports where Lighting_Condition and Time_of_Day are not in agreement.
  - Agreement between number of vehicles/units & number of drivers (adjusting count for hit & run or recorded phantom vehicles) recorded
  - Agreement between persons injured/killed in the person records and total injured/killed as reported at the crash level (this may be a calculated field in which case this analysis is moot)
  - Agreement between highest severity level injury at the person level (individuals) and the crash-level injury severity (may also be a calculated field, in which case this analysis is moot)
  - Results of any other data cleansing analysis/accuracy metric that are run routinely.
  - VIN (if recorded, what percent decode to a valid vehicle year, make & model)

Completeness:

- A: Narrative/Diagram: Percent of reports received lacking either the narrative, the diagram, or both.
- B: Location: Percent of reports missing location information (e.g., GPS coordinates if required to report, or information such as on street name/identifier, nearest/intersecting street name/identifier, distance from street/landmark, or other required location information).
- C: Percent of other key fields left blank: Making allowances for hit & run or phantom vehicles (if recorded), look for unexplained blanks/missing data in select fields such as:
    - Driver name/address
    - VIN
    - Harmful Event
    - Environment, Vehicle & Driver contributing factors
    - EMS data (if recorded)
- D: Under-reporting:
    - Comparing across the most recent three complete years of data (e.g., 2007, 2008, and 2009), are there any law enforcement agencies that have seen a significant drop in crash report submissions that is (for example) double the year-to-year drop for the state as a whole?
    - Are there any medium-to-large agencies (in terms of annual submissions historically) that have turned in either zero or an uncharacteristically low number of crash reports for 2010 YTD?

Consistency:

- A: Calculate the ratios to compare across years:
    - Percent PDO crashes of total crashes
    - Percent Injury crashes of total crashes
    - Percent Fatal crashes of total crashes

The above should be calculated overall for the state and individually for each law enforcement agency—large agencies that are significantly at variance with statewide proportions may indicate a lack of uniform application of the reporting criteria.

Data Integration and Accessibility:

- If you have information on something like a CODES data linkage project (proportion of records linked and the strength of the matching achieved) for *integration*, or the percentage of linkage/matching you have when linking the crash database with roadway inventory or other traffic records databases. For accessibility, perhaps there is a record of number of requests, analyses run, web-enabled data analysis page hits, or any other indicator you might have.

APPENDIX B: ORIGINAL CDIP PRE-VISIT QUESTIONNAIRE

CRASH DATA IMPROVEMENT PROGRAM QUESTIONNAIRE FOR COMPLETION PRIOR TO STATE VISIT

These questions should be distributed to and completed by manager(s) or individuals most knowledgeable about the State's crash data and systems.

ADMINISTRATIVE

1. Which Department/Agency is the main custodian of the state's crash database?

2. Which section/office within that department has the principal responsibility for managing/maintaining the crash database?

3. Who is the individual responsible for administering/managing the crash database? What is their position/job title? Provide their contact information (mailing address, phone, fax, email)?

PART I – CRASH DATA COLLECTION AND REPORTING PROCESS OVERVIEW

1. Identify each step in the process and flow of data from the crash event to the completion and review of the crash report, through the data entry process in the State's crash data system. Where relevant, please describe the differences in processing of paper versus electronic crash reports. (If available, please provide a flow chart showing the processes.)

2. How many crash data system personnel are there in your office and what are their roles (data entry, supervision, location coding, data validation and correction, other)?

3. Are other offices are involved in the Crash Reporting process? What are their roles (for instance, locating crashes to a statewide base map based on location information from the crash form)? What type of coordination is there among the offices involved in the process?

4. Are there any changes to this process being considered?

5. Does your state have a relational database for crashes?

6. What type of data structure does your state currently have (e.g. SQL, Oracle, etc.)?

7. In your opinion, how can the crash data collection processes be improved in your State (refer to steps in the flow chart)?

PART II. CRASH DATA COLLECTION SPECIFICS

Reportable Crashes

1. What is the law or statute that requires law enforcement officers to investigate and report on fatal, injury and property damage only crashes? Does the law or statute specify a time limit within which the crash must be submitted to the state crash database? Are there any reasons that a law enforcement agency might have an incentive for not providing crash reports in a timely manner? (for example, municipality charges fees to the public for providing crash report)

2. What is the minimum reporting threshold ($, towaway, other?) for a property damage only crash?

3. How many law enforcement agencies within your State are responsible for investigating crashes and submitting reports?

4. Does the state have a standard Police Accident Report (PAR) form that is used by all agencies in the state? If multiple forms are used, how many different versions of the PAR form are being used?

5. Are any law enforcement agencies collecting crash data electronically at the crash scene? If 'yes', how many (or what percentage) agencies? Do the electronically collected data identically match the data elements and attributes on the paper PAR elements?

6. Does the state have a single electronic crash software product for use by all law enforcement? If not, is there a standard for data submissions to be attained by all electronic reporting systems? Does the state have the capability of receiving and downloading data collected on software other than the state supported 'standard' software?

7. Are any data verified electronically at the crash scene? If so, which data are verified? What technologies are used for the verification? What are the data verified against (e.g. driver license file, vehicle registration file)?
8. By what process(es) are PARs from law enforcement agencies submitted to the crash database (mail, internet, secure electronic upload, CDs)?

9. Does the state have the capability to accept electronic uploads of PARs directly from law enforcement agencies? Are there edit checks/validations run by the

crash database on the PARs submitted electronically, prior to uploading the data? Do the field data collection software products require the same set of edit checks for completion of a crash report as are required for that report's acceptance into the State's crash data system? (if not, how do the field data collection edit checks differ from the centralized system's checks?)

10. Does your state use a supplemental form to collect specific crash data information (for instance, information on commercial motor vehicles involved in crashes or on BAC test results) or is the data collected on a single form? If 'yes', what determines whether the supplement is filled out at the scene?

11. Are there any plans to modify/update the crash or supplemental forms in your State? If 'yes', when and who will be involved?

12. What feedback has your state received regarding problems with either the paper or electronic PAR from the following persons:
    a. Police Officers
    b. Crash from reviewer (officer supervisor)
    c. Data entry person
    d. Other

PART III. STATE DATA QUALITY ASSESSMENT PROCEDURES

1. Does the state currently have a process(es) for assessing the quality of the information within the crash database in terms of its (Please indicate the appropriate response):

    i. Timeliness: Yes No

    ii. Accuracy: Yes No

    iii. Completeness: Yes No

    iv. Uniformity/Consistency: Yes No

    v. Integration (of data with other databases): Yes No

    vi. Accessibility: Yes No

2. Does the state conduct separate quality assessment procedures for PARs that are submitted on paper versus those submitted electronically?

Timeliness

Overall Measures of Timeliness

1. What output measure(s) of timeliness are currently reported for your state? (e.g. average days for reports to be entered into the database, percentage of cases reported within 30 days, percentage of cases over 90 days etc.)

2. Are these output measure(s) routinely calculated? If 'yes' please describe how this calculation is derived. **If available, please provide sample reports of this output measure.**

Component Measures of Timeliness

4. Does the state assess the time it takes to complete each step in the data entry process? (Refer back to the 'Flow Chart' in Part 1, Question 1).

5. If no flow chart was provided, a listing of possible data entry component steps is provided below. Please identify which, if any, of these component steps are used by the state, does the state calculate the time each of these component steps takes to complete

    a. Time from crash occurrence until receipt at the state crash database
    b. Time for edit or validation checks on the data included in the PAR.
    c. Time for coding of locations
    d. Time for other processes required prior to data entry onto the state crash database (e.g., some states delay processing to ensure driver reports are collected or to allow time for verification of financial responsibility).

Assessing the Timeliness of Individual Agency Report Submission

8. Does the state measure the time it takes to get the PARs from each individual law enforcement agency? If' 'yes', how is this measure calculated?

9. Has the state developed a 'year-to-date' measure that calculates how many PARs were submitted for each law enforcement agency in the same time-period during the prior year?
10. If the state has developed a 'year-to-date' measure, does the state have a procedure for reporting to law enforcement agencies on the timeliness of their crash report submissions? If 'yes', what is the period or frequency with which the agencies are notified about their timeliness?

Accuracy

Overall Measures of Accuracy

1. What output measure(s) of accuracy are currently reported for your state? (e.g. less than five errors per PAR, number of PARs returned to the submitting agency for corrections, zero errors on 'critical' data elements, etc.)

2. Are these output measure(s) routinely calculated? If 'yes' please describe how this calculation is derived. **If available, please provide sample reports of this output measure.**

Component Measures of Accuracy – Content Assessment

4. Does the state have a 'data dictionary' or other documentation that identifies and/or defines all of the data elements that are reportable to the crash database and their attributes, range of allowed values, and logical consistency among various data elements?

5. Does the state run a procedure to assess the content (attributes) for each of the data elements within the crash database (e.g. a descriptive analysis reporting the frequencies or percentages of responses)?

6. Does the state compile a listing of the data elements with the error type and frequency (or percentage) of error results reported? (e.g. number of errors per data element)

Component Measures of Accuracy – Relational Assessment

7. Does the state run a procedure to assess the relational content of information? (e.g. business edits and validations to assess the logical relations between date elements)

8. Does the state compile a listing of the type and frequency of relational errors? (e.g. number of errors identified by the business edits and data element validations)

Assessing the Accuracy of Individual Agencies

9. Does the state measure the accuracy of reports from individual law enforcement agencies submitting crash reports? If 'yes', how is this measure calculated?

10. Does the state have a procedure for reporting to law enforcement agencies on the accuracy of their crash report submissions? If 'yes', what is the period or frequency with which the agencies are notified about their accuracy?

Completeness

Overall Measures of Completeness

1. What output measure(s) of completeness are currently reported for your state? (e.g. less than five total omissions per PAR, less than five omissions of critical elements per PAR, the number of PARs returned to the submitting agency for the additional information, etc.)

2. Are these output measure(s) routinely calculated? If 'yes' please describe how this calculation is derived. **If available, please provide sample reports of this output measure.**

Component Measures of Completeness – Content Assessment

3. Does the state run a procedure to assess the presence of content (attributes) for each of the data elements within the crash database?

4. Does the state compile a listing of the data elements with their error frequency (omissions) results reported? (e.g. number and/or percent of errors per data element)

Component Measures of Completeness – 'Unknowns' Assessment

5. Does the state run a procedure to assess the frequency with which 'unknown' is the reported value for each of the data elements within the crash database?

6. Does the state compile a listing of the data elements identifying the frequency with which 'unknown' is the reported response? (e.g. a descriptive analysis listing the frequency and/or percentage of 'unknowns' per data element)

Assessing the Completeness of Individual Agencies

7. Does the state measure the completeness of reports from individual law enforcement agencies submitting crash reports? If 'yes', how is this measure calculated?

8. Does the state have a procedure for reporting to law enforcement agencies on the completeness of their crash report submissions? If 'yes', what is the period or frequency with which the agencies are notified about their completeness?

Consistency

Overall Measures of Consistency

1. What output measure(s) of consistency are currently reported for your state? (e.g. all law enforcement agencies reporting crashes use a standard crash report form, all law enforcement agencies report identical data elements with identical data attributes, number of law enforcement agencies not using the standard PAR)

2. Does the State track consistency of reporting by law enforcement agencies (i.e. tracking ratios for fatal, injury, property damage only by year to date)? If 'yes', please describe how?

3. Are these output measure(s) routinely calculated? If 'yes', please describe how the calculation is derived. **If available, please provide sample reports of this output measure.**

Integration

Overall Measures of Integration

1. What output measure(s) of integration are currently reported for your state? (e.g. a count of the number of databases the crash database is linked to, for example the crash database linked to the driver licensing database and/or the vehicle registration database)

2. Are these output measure(s) routinely calculated? If 'yes', please describe how the calculation is derived. **If available, please provide sample reports of this output measure.**

Accessibility

Overall Measure of Accessibility

1. What output measure(s) of accessibility are currently reported for your state? (e.g. a count of the number of requests for crash data fulfilled by the agency/department that maintains the crash database, number of agencies that have on-line access to the crash database)

2. Are these output measure(s) routinely calculated? If 'yes', please describe how the calculation is derived. **If available, please provide sample reports of this output measure.**

PART IV – CRITICAL DATA ELEMENTS

Crash Location

1. Does the state have a process to locate each crash onto a statewide base map of all of the roads in the state? Is there one standard for location coding used on all public roads, or does the location coding for state-maintained roads differ from that used for local roads?

2. Does the crash database custodial office perform the 'crash locating' function or is this effort managed by another office within the state?

3. Has the state assessed what percentage of reportable crashes are not able to be located onto the statewide base map? Has the state assessed how long this crash locating process takes? If yes, please provide the corresponding measures of accuracy/completeness and timeliness.

Other Critical Crash Variables

1. Does the state crash database include the vehicle identification numbers of crash involved vehicles?

2. Does the state crash database include the Blood Alcohol Concentration (BAC) test results of crash involved drivers?

3. Does the state crash database include the officer reported restraint system (or helmet use for motorcycle operators and riders) for all of the occupants of crash involved vehicles?

4. Does the state crash database include the investigating officer's identification of first harmful event and/or the sequence of events involved in each crash?

5. Are a narrative and diagram required on all crash reports? Does the state track the percentage of reports received with inadequate, incomplete, or missing narratives or diagrams?

APPENDIX C: UPDATED CDIP PRE-VISIT QUESTIONNAIRE

CRASH DATA IMPROVEMENT PROGRAM QUESTIONNAIRE

These questions should be assigned to individuals most knowledgeable about the State's crash data and systems. NOTE: Where necessary, additional information detailing the question's *standard of evidence* is indicated with the (SOE:) notation. Where applicable, questions that overlap with the NHTSA Traffic Records *Assessment Questionnaire* are indicated with the (AQ-#) notation. ***If the State has gone through a recent Assessment, it may save time and effort to provide the CDIP Team with a copy of answers provided to the Assessment Questionnaire.***

The remainder of this document comprises the Crash Data Improvement Program Questionnaire. This questionnaire is intended for states to complete well in advance of the CDIP Technical Advisory Team (TAT) visit .This will help the TAT members prepare for their visit and arrive on site with a clearer understanding of the current status of the crash system. On-time completion of the questionnaire also allows for a round of follow-up questions prior to the TAT State visit. The value of this during the in-person portions of the CDIP is enormous. With this information, the TAT can tailor the presentations to address needs that the State has and avoid those issues for which the State already has a solution in place. To give just one example, if your State no longer uses paper crash report forms, the TAT can concentrate their presentation on the issues related to electronic data collection and transmission.

There are multiple sections of the questionnaire.

Part I deals with administrative details and is best answered by the crash data custodian. Part II deals with top-level processes of data collection and reporting, while Part III touches on specifics of those same processes. Part IV addresses the needs of key users of the crash data, and how those needs are met. Parts V and VI address the crash data quality management program in general and the specifics of data quality performance measurement, respectively.

In order to complete the questionnaire, the State CDIP coordinator (the person working with the CDIP team from the State's side) should identify the person or persons best able to answer each question, assign the questions appropriately, and make sure that the answers are provided in a timely manner. While it is not necessary that the State CDIP coordinator review every answer from every respondent, the TAT will be likely to ask follow-up questions about any disagreement among the various sources of information as part of their preparation for the on-site visit. The State CDIP coordinator has the option of seeking multiple answers to any of the questions and to ask additional people to respond to any follow-up questions.

PART I. ADMINISTRATIVE

The State crash system ideally contains, at a minimum, basic information about every reportable motor vehicle crash in the State. ("Reportable" is defined by the applicable State statute.) The available data should be sufficient to permit decision-makers to draw valid conclusions about the crash experience in their State. Ideally, all State crash data is consolidated into one, generally accessible, database with a clearly defined organizational custodian. The crash system provides both an official record of the crash and data for analytic purposes. The crash system documents the characteristics of a crash and provides the following details about each incident:

- Who: Information about the drivers, occupants, and non-motorists involved in a crash (e.g., license status, age, sex);
- What: Information about the type of vehicle involved in a crash (e.g., make, model, body type, vehicle registration);
- When: Information detailing the time a crash occurred (e.g., time of day, day of week);
- Where: Information about the crash location (e.g., location name, coordinates, type, attributes);
- How: Information describing the sequence of events and circumstances related to a crash— up to and including the first harmful event through the end of a crash and its consequences (e.g., damage, injury);
- Why: Information about the interaction of various systems that may have contributed to the crash occurrence (e.g., weather, light conditions, driver actions, non-motorist actions) and/or the crash severity.

Ideally, crash data reflecting all levels of severity (including fatal, injury, and property damage only) is collected and used to support safety analysis.

In most states a single agency is designated the official crash records custodian. This agency is ultimately responsible for gathering data on reportable crashes into a Statewide database to serve as a resource for all traffic safety data users. The custodian is also generally considered to have oversight of the crash data quality management processes although some of the actual data management may be distributed among collectors and key data users.

Our purpose in this section is to identify the lead agency and other key stakeholders in the crash data management processes, including management of crash data quality. We hope to understand the relationships between the crash data custodian and the other key agencies (data collectors and users included) and the capabilities of the Statewide crash system to meet the needs of all stakeholders.

4. Which Department/Agency is the custodian of the State's crash database of record? (AQ-37)

5. Which section/office within that department has the principal responsibility for managing/maintaining the crash database? (AQ-37)

6. Does State law require that the crash database of record retain the original crash record as submitted by law enforcement? (SOE: provide relevant law or policy.)

7. Who is the individual responsible for administering/managing the State's crash database of record? What is their position/job title? Provide their contact information (mailing address, phone, fax, email)? (AQ-37)

8. Does your State have a relational database for crashes?

9. What type of data structure does your State currently have (e.g. SQL, Oracle, etc.)? (AQ-36) (SOE: provide a description of Statewide database and specify how the data is consolidated)

10. If other entities maintain Statewide crash databases other than the database of record (e.g., files used for engineering, public health, driver control analysis), who maintains them and what are they used for?

PART II. CRASH DATA COLLECTION & REPORTING OVERVIEW

The State should maintain accurate and up-to-date documentation—including process flow diagrams—detailing the policies and procedures for key processes governing the collection, submission, post-processing (e.g., location coding), posting, and maintenance of crash data. This should include provisions for submitting fatal crash data to the State FARS data collection unit and commercial vehicle crash data to SafetyNet.

Process flow diagrams document key processes including interactions with other data systems. Ideally, each diagram should be annotated to show the time required to complete each critical step. The process flow diagram also includes the processes for managing errors and incomplete data (e.g., returning crash reports to the originating officer or department for correction and resubmission). The documentation accounts for both paper and electronic process flows.

An ideal crash system is linked with other traffic records systems to enhance data quality and support the crash system's critical business processes. System *interface* describes a timely, seamless relationship and a high degree of interoperability between systems.

In practice, system interface is useful when circumstances require relationships between traffic records data systems that need to be connected and accessible at all times. These interfaces occur throughout a crash record's lifecycle: data collection, submission, processing, posting, and maintenance. Ideally, such interfaces improve the efficiency and cost effectiveness of the crash system.

Ideally, the State also maintains standards for all traffic records applications and databases, and the data dictionary should include consistent definitions for all elements—particularly those common across applications and databases.

11. Identify each step in the process and flow of data from the crash event to the completion and review of the crash report, through the data entry process in the State's crash data system. Where relevant, please describe the differences in processing of paper versus electronic crash reports. If available, please provide a flow chart showing the processes. (AQ-56)

12. How many crash data system personnel—State and contract employees—are there and what are their roles (data entry, supervision, location coding, data validation and correction, other)?

13. Are other offices involved in crash report processing (for instance, locating crashes to a Statewide base map based on location information from the crash form)? If so, where are they based and what are their roles?

14. What type of coordination is there among the offices involved in crash report processing?

15. Are any changes to this process being considered? If so, what are they and why are they being considered?

16. In your opinion, how can the crash data collection processes be improved in your State (refer to steps in the flow chart requested in question 8)?

PART III. CRASH DATA COLLECTION SPECIFICS

The crash data collection system contains basic information about every reportable (as defined by State statute) motor vehicle crash on any public roadway in the State. Further, all data collection systems (electronic and paper) should be based on a uniform crash report containing standard data elements. There are several guidelines available to States wishing to build and maintain appropriate crash data systems. The Model Minimum Uniform Crash Criteria (MMUCC) provides a suggested minimum set of crash-related data elements that enable valid statistical analysis*(4).* MMUCC is, however, a minimum guideline and States are expected to adopt additional standard data elements and attributes as dictated by their specific data needs.

When creating or updating crash systems, States can also consider two applicable manuals published by the American National Standards Institute. ANSI-D16, the *Manual on Classification of Motor Vehicle Traffic Accidents*,*(5)* is a standard for statistical motor vehicle traffic accident classifications for nationwide use and provides a common language for crash data reporters, classifiers, analysts, and users.

ANSI-D-20, the *Data Element Dictionary for Traffic Records Systems*,*(6)* provides standard terminology and coding instructions designed to facilitate representations of data elements communicated between jurisdictions. (http://aamva.gorg/aamva/DocumentDisplay.aspx?id={3D25B551-1E16-4EF5-A74C-F321DE19D6E5})

Ideally, crash data should be collected electronically in the field by all jurisdictions using a uniform, efficient approach (e.g., question or scenario-based software) that is consistent with the Statewide database's validation rules. Data is subject to validation checks as it is collected in the field and upon receipt and submission to the Statewide database.

In addition, crash system documentation indicates if edits and other steps are accomplished manually or electronically. The State ideally has documented retention and archival storage policies that serve the needs of safety engineers and other users with a legitimate need for long-term access to the reports.

Uniform crash reporting or under reporting of property damage crashes can be affected by the use of short forms, operator reports and supplemental forms. When these data collection systems are used the State should institute processes to understand their impact and assure uniform crash reporting.

The CDIP TAT needs to fully understand the crash data collection processes in the State. This may in turn require understanding of the mix of paper reporting processes and use of one or more field data collection software solutions, as well as law enforcement agency records management systems and their interface(s) with the State's centralized crash records database.

17. What is the law or policy that requires law enforcement officers to investigate and report on fatal, injury, and property damage only crashes? (SOE: Provide full text of law or policy.)

18. Does the law or policy specify a time limit within which the crash must be submitted to the State crash database?

19. Are there any reasons that a law enforcement agency might have an incentive for not providing crash reports in a timely manner (for example, municipality charges fees to the public for providing crash report)?

20. What is the minimum reporting threshold (monetary, tow-away, other) for a property damage only crash?

21. How many law enforcement agencies are responsible for investigating crashes and submitting reports to the State? (AQ-52)

22. Does the State have a standard Police Accident Report (PAR) form that is used by all agencies in the State? (AQ-53) (SOE: provide a copy of the PAR)

23. If multiple versions of the PAR are used, how many different versions of the PAR form are being used? (SOE: provide a copy of all PARs currently in use; and under what circumstances/ jurisdictions they are used)

24. Is there a short form used for PDO crashes? (SOE: provide the form.)

25. Does the State require crash-involved drivers to submit an operator report? If yes, does the State include operator-submitted data to the database. (SOE: provide the form if used in the crash database.)

26. Does the State use a supplemental form to collect specific crash data information (for instance, information on commercial motor vehicles involved in crashes or on BAC test

results) or is the data collected on a single form? (SOE: provide any supplemental forms and the guidance for their use.)

27. Are there any plans to modify/update the crash or supplemental forms in the State? If 'yes', when and who will be involved?

28. Are any law enforcement agencies collecting crash data electronically at the crash scene? If 'yes', how many (or what percentage of) agencies collect crash data electronically? (AQ-53)

29. Statewide, what percentage of crash reports is collected electronically? (AQ-53)

30. Does the electronically collected data identically match the data elements and attributes on the paper PAR elements?

31. Does the State have a single electronic crash software product for use by all law enforcement?

32. If not, is there a standard for data submissions to be attained by all electronic reporting systems?

33. By what process(es) are PARs from law enforcement agencies submitted to the crash database (mail, internet, secure electronic upload, CDs)?

34. Does the State crash database accept any crash data electronically? If so, what percentage is accepted electronically?

35. Are there law enforcement agencies collecting crash data electronically that do not submit their data to the Statewide crash database electronically?

36. Is any data verified electronically at the crash scene via real-time interface with other data systems (e.g., driver information from the DMV)? If so, which data is verified? (AQ-54)

37. What technologies are used for the verification?

38. What are the data verified against (e.g. driver license file, vehicle registration file)?

39. Are there edit checks/validations run by the crash database on the PARs submitted electronically, prior to uploading the data? (AQ-54)

40. Do the field data collection software products require the same set of edit checks for completion of a crash report as are required for that report's acceptance into the State's crash data system? If not, how do the field data collection edit checks differ from the centralized system's checks? (AQ-54)

41. What feedback has your State received regarding problems with either the paper or electronic PAR from the following persons: police officers, crash form reviewing supervisory officer, data entry person, other?

PART IV. DATA FOR DECISION-MAKING

A State crash system assists the traffic safety community in implementing programs and countermeasures that reduce motor vehicle crashes, deaths, and injuries. Data-driven improvements rely on a State's crash system to identify opportunities to improve highway safety, measure progress, and systematically evaluate the effectiveness of countermeasures. It is critical that a State's crash system include or have access to the critical data to address the diverse safety problems such as:

- Engineering – accurate crash locating to support integration of crash data with roadway characteristic data.
- Enforcement - ensure driver/vehicle compliance (i.e., graduated driving licensing, alcohol, and speeding).
- Education – human behavioral issues to address seat belt usage, distracted driving, driving under the influence of drugs and alcohol, and motorcycle, bicycle, and pedestrian safety.
- Emergency Response (Injury Surveillance) – data collection and integration of emergency medical services, ambulatory care, acute care, trauma and rehabilitation facilities, and vital records data with other Traffic Records Systems.

42. Does the State have a process to locate each crash onto a base map? (AQ-60, 167, 168)

43. Is the base map inclusive of all public roads in the State? (AQ-60, 167, 168)

44. Based on location information from the crash report, can crashes be assigned a location code that matches the location coding in the base map? (SOE: Describe the process for assigning matched location codes. If the processes differ for State/Federal aid and local roads, please describe both.) (AQ-60, 167, 168)

45. How long, on average, it takes to assign a crash location to the basemap? (AQ-56) (SOE: provide the timeliness measure for the crash location coding process. If it differs between State/Federal Aid and local roads, please describe both.)

46. Does the crash database custodial office perform the 'crash locating' function or is this effort managed by another entity within the State? (AQ-56)

47. What percentage of reportable crashes is locatable on the Statewide base map? (AQ-385)

48. What percentage of reportable crashes is successfully located automatically? (AQ-385)

49. Does the State crash database include the vehicle identification numbers of crash involved vehicles? (AQ-59)

50. Does the State crash database include the Blood Alcohol Concentration (BAC) test results of crash involved drivers? (AQ-50)

51. Does the State crash database include all officer-reported restraint systems (helmet for motorcycle operators and riders) availability and usage for all of the occupants of crash involved vehicles? (AQ-50)

52. How are crash-involved pedestrians and bicyclists recorded in the Statewide crash database?

53. Does the State crash database include data elements related to distracted driving?

54. What is the injury severity scale used on the crash report?

55. How does the State define “serious injury” on the crash report?

56. Are a narrative and diagram required on all crash reports? (AQ-50)

57. Are the narrative and diagram for all crashes retained in the crash database?

58. Does the State track the percentage of reports received with inadequate, incomplete, or missing narratives or diagrams?

PART V. STATE DATA QUALITY MANAGEMENT PROGRAM

THE NHTSA 2012 UPDATE TO THE TRAFFIC RECORDS PROGRAM ASSESSMENT ADVISORY (P. 26-28) DESCRIBES A FORMAL, COMPREHENSIVE DATA QUALITY MANAGEMENT PROGRAM FOR THE CRASH SYSTEM. SUCH A PROGRAM WOULD BE DESIGNED TO COVER THE ENTIRE CRASH DATA PROCESS—THE COLLECTION, SUBMISSIONS, PROCESSING, POSTING, AND MAINTENANCE OF CRASH DATA. AN IDEAL SYSTEM WOULD INCLUDE THE ASPECTS LISTED BELOW.

Automated edit checks and validation rules that ensure entered data falls within the range of acceptable values and is logically consistent between other fields. Edit checks are applied when data is added to the record. Many systems have a two-tiered error classification system, distinguishing critical errors that must be corrected before submission and non-critical error warnings that may be overridden.

Limited State-level correction authority is granted to quality control staff working with the Statewide crash database to amend obvious errors and omissions without returning the report to the originating officer. Obvious errors include minor misspellings, location corrections, and directional values. Obvious omissions include missing values that can easily be obtained from the narrative or diagram.

Processes for returning rejected crash reports are in place to ensure the efficient transmission of rejected reports between the Statewide data system and the originating officer as well as tracking the corrected report's submission.

Performance measures are tailored to the needs of data managers and address the concerns of data users. Measures can be aggregated from collectors, users, and the State TRCC. The crash data should be timely, accurate, complete, uniform, integrated, and accessible. These attributes are tracked using State-established quality control measures.

Numeric goals —or performance metrics—for each performance measure are established and regularly updated by the State in consultation with users via the TRCC.

Performance reporting provides specific feedback to each law enforcement agency on the timeliness, accuracy, and completeness of their submissions to the Statewide crash database relative to applicable State standards.

High-frequency errors are used to generate new training content and data collection manuals, update the validation rules, and prompt form revisions.

Quality control reviews comparing the narrative, diagram, and coded report contents are considered part of the Statewide crash database's data acceptance process.

Independent sample-based audits are conducted periodically for crash reports and related database contents. A random sample of reports is selected for review. The resulting reviews are also used to generate new training content and data collection manuals, update the validation rules, and prompt form revisions. At a minimum, these audits occur on an annual basis.

Periodic comparative and trend analyses are used to identify unexplained differences in the data across years and jurisdictions. At a minimum, these analyses occur on an annual basis.

Data quality feedback from key users is regularly communicated to data collectors and data managers. This feedback will include corrections to existing records as well and comments relating to frequently occurring errors. Data managers disseminate this information to law enforcement officers as appropriate.

Data quality management reports are provided to the TRCC for regular review. The TRCC used the reports to identify problems and develop countermeasures.

59. Does the State have automated edit checks and validation rules that ensure entered data falls within the range of acceptable values and is logically consistent with other fields? If so, please describe. (SOE: or provide a listing of the edit checks and validations used by the state.)

60. Does the State grant limited correction authority to quality control staff working with the Statewide crash database in order to amend obvious errors and omissions without returning the report to the originating officer? If so, please describe.

61. Does the State have processes in place for transmitting and tracking rejected reports between the Statewide crash system and the originating officer? If so, please describe.

62. Does the State establish numeric goals—performance metrics—for each performance measure? If so, please describe.

63. Does the State provide performance reporting to law enforcement agencies regarding the timeliness, accuracy, and completeness of their submissions to the Statewide crash database? If so, please describe.

64. Are high-frequency errors tracked and used to generate new training content and training manuals, updated validation rules, and prompt form revisions? If so, please describe.

65. Does the State's data acceptance process include quality control reviews that compare the narrative, diagram, and coded report contents? If so, please describe.

66. Does the State periodically conduct independent, sample-based audits of crash reports and related database contents? If so, please describe.

67. Does the State periodically conduct comparative and trend analyses to identify unexplained differences in the data across years and jurisdictions? If so, please describe.

68. Does the State regularly collect data quality feedback from key users and share it with data collectors and managers? If so, please describe.

69. Does the State regularly provide the TRCC data quality management reports? If so, please describe.

PART VI. STATE DATA QUALITY PERFORMANCE MEASUREMENT

The CDIP concentrates explicitly on crash data quality performance measurements as a component of the formal, comprehensive data quality management program. As noted earlier, performance measures are tailored to the needs of data managers and address the concerns of data users. Measures can be aggregated from collectors, users, and the State TRCC. The crash data should be timely, accurate, complete, uniform, integrated, and accessible. These attributes are tracked using State-established quality control measures. The measures in Table 1 are examples of high-level quality management indicators. The State is encouraged to develop additional measures that address their specific needs.

Table 1: Example Quality Control Measures for Crash Data Systems

Timeliness	• **The median or mean number of days from (a) the crash date to (b) the date the crash report is entered into the database.** • **The percentage of crash reports entered into the database with XX* days after the crash. *e.g., 30, 60, or 90 days.**
Accuracy	• **The percentage of cash records with no errors in critical data elements. Example: crash severity.** • **The percentage of in-State registered vehicles on the state crash file with VIN matched to the State vehicle registration file.**
Completeness	• **The percentage of crash records with no missing critical data elements.** • **The percentage of crash records with no missing data elements.** • **The percentage of unknowns or blanks in critical data elements for which unknown is not an acceptable value.**
Uniformity	• **The number of MMUCC-compliant data elements entered into the crash database or obtained via linkage to other databases.**
Integration	• **The percentage of appropriate records in the crash database that are linked to another system or file.**
Accessibility	• **Identify the principal users of the crash database. Query the principal users to assess (a) their ability to obtain the data or other services requested and (b) their satisfaction with the timeliness of the response to their request . Document the method of data collection and the principal users' responses.**

70. Does the State currently have a process(es) for assessing the quality of the information within the crash database in terms the data quality attributes listed in Table 1? (Please indicate the appropriate response):

71. Does the State conduct separate quality assessment procedures for PARs that are submitted on paper versus those submitted electronically?

Timeliness

The CDIP Guide, using information from a NHTSA-sponsored training program titled *Traffic Records 101* defined timeliness as "when the needs of users are met." It goes on to say that crash data are considered timely when the typical time interval from crash event to use of the data via an electronic database is consistent with the time required by state-of-practice methods of data capture, reporting, processing, and editing. Timely data are considered important from a data quality management perspective because progress in many of the other data quality attributes relies on the State's ability to provide feedback to data collectors. If that feedback is delayed because the process of getting the data into the system takes too long, it becomes less useful in helping data collectors do a better job.

There are two aspects of timeliness that are generally recognized as relevant to crash data:
A) **For an individual report**, timeliness is indicated by the time between a traffic crash event and the time when the information about that event is placed into an accessible database. The crash record than can then be retrieved from the electronic database, studied as a single event, and related to other crashes and events in an analysis.

B) **For a crash dataset**, timeliness is indicated by the time between the time between the occurrence date for the most recent crash record contained in the dataset and the time when the dataset is available for analysis. This definition allows for timeliness of the dataset to be assessed throughout the year and after the end of a year when, for example, some states generate a final dataset for use in annual analyses.

The Traffic Records Assessment Advisory also recommends that states examine timeliness of the various processes from initial report writing to final availability of the record in the State's centralized crash database. There are many processes that could be reviewed including:

- Initial report submission by the officer within the law enforcement agency
- Supervisory review
- Data entry into a departmental database (if done as a precursor to State submission)
- Submission to the Statewide centralized data custodian

- Initial acceptance into a "pending" queue
- Quality review in the centralized data management process
- Pre-processing, scanning, sorting, and pre-review by other entities (e.g., DMV)
- Data entry
- Location coding
- Final acceptance

Of these, the CDIP has most prominently focused on initial submission from law enforcement agencies to the State, location coding, and final acceptance. It is, however, useful for the State to maintain an annotated flow chart showing each step in the crash reporting process and the average or expected duration of each step.

72. What output measure(s) of timeliness are currently reported for your State (e.g. average days for reports to be entered into the database, percentage of cases reported within 30 days, percentage of cases over 90 days etc.)? (AQ-66) (SOE: Provide baseline current data and trend if available)

73. Does the State assess the time it takes to complete each step in the data entry process? (Refer to the flow chart in question 8). (SOE: provide annotations on the flow chart of the time it takes for each step OR provide the timeliness measures for each major step in the process.)

74. Does the State measure the time it takes to get the PARs from each individual law enforcement agency? (SOE: provide the measurement)

75. Has the State developed a 'year-to-date' measure that calculates the timeliness of PAR submissions compared to the same time-period during the prior year for each law enforcement agency?

76. If the State has developed a 'year-to-date' measure, how does the State report to law enforcement agencies on the timeliness of their crash report submissions? (SOE: provide a copy of a recent report and specify the frequency of issuance.)

Accuracy

Accuracy is a measure of how well crash *data* agrees with the "ground truth" of what actually happened in crashes. On an individual crash report, accuracy is a measure of how precisely the information about the person(s), vehicle(s) environment and circumstances of that crash are reported. On an aggregate basis, accuracy is a measure of how well the information in the crash

database reflects correct information about the people, vehicles, locations, and circumstances involved in all the crashes reported for a particular jurisdiction or the whole State. Measuring accuracy is important because it gives users a valid way to determine if the data are suitable to support an intended analysis. Data collectors and managers can use accuracy measurements to identify problems, implement solutions (such as training or process changes), and monitor the effectiveness of those solutions.

There are two approaches to measuring accuracy, through *internal* (within crash report) edit checks and through *external* validation checks. With internal edits, data is assessed to assure that the values reported for the data elements are within a range of acceptable responses (e.g. BAC is reported as a value between 0.00 and 0.99) or one data element is cross-checked with another to see if the pair makes sense. If the crash vehicle action is *ran traffic signal*, and the crash location is reported to be on a non-intersection segment of rural highway, then at least one of the two data elements is likely incorrect. Such cross-field edit checks are often referred to as checks for logical consistency or internal consistency. External validation is any process in which data from within a crash report is checked against an external data source such as when the crash location is verified using information from the State's roadway inventory database or GIS. Likewise driver or vehicle information may be validated by checking for agreement with the State's driver licensing or vehicle registration databases. One way that systems improve accuracy is through use of external sources. Where authoritative data already exist in a different data set, the crash database can be enhanced through linkage to the appropriate existing data rather than requiring officers to collect the same data elements at the scene.

77. What output measure(s) of accuracy are currently reported (e.g., less than five errors per PAR, number of PARs returned to the submitting agency for corrections, zero errors on 'critical' data elements, etc.)? (SOE: Provide baseline current data and trend if available, and specify frequency of issuance) (AQ-67)

78. How does the State assess the content (attributes) for each of the data elements within the crash database (e.g. a descriptive analysis reporting the frequencies or percentages of responses)?

79. How does the State compile a listing of the data elements with the error type and frequency (or percentage) of error results reported? (e.g. number of errors per data element)

80. How does the State assess the relational content of information? (e.g. business edits and validations to assess the logical relations between date elements)

81. How does the State compile a listing of the type and frequency of relational errors? (e.g. number of errors identified by the business edits and data element validations)

82. How does the State measure the accuracy of reports from individual law enforcement agencies submitting crash reports? (SOE: Provide a copy of the report and its frequency of issuance.)

Completeness

Completeness is the property of having information in all required data fields for all reportable crashes. There are two main types of measures of completeness. For individual crash reports, completeness addresses whether or not the report is missing any required information. At the aggregate level, completeness addresses whether or not the Statewide crash database contains a record of every crash that should have been reported. Measuring completeness is important to users because, like measures of accuracy, measures of completeness tell them if the data are sufficient to support an intended analysis. For data collectors and managers, measures of completeness help to identify patterns of incomplete reports and under-reporting.

Measures of completeness of individual reports typically count the number of times that required data fields are left blank. Completeness measures may also be designed to assess overuse of non-specific values in the data (such as “unknown”, “not applicable”, or “other”). While these values are allowed in the data (they would pass an edit check), inappropriate use of these values can indicate a failure of the data collectors to gather all required information.

Measures of completeness of report submissions typically count the number of reports and compare the proportion of serious crashes to the overall number of reports submitted. The total number of reports submitted by individual law enforcement agencies can help managers spot those agencies that have, for whatever reason, changed their crash reporting policies without notification. The ratio of serious crashes to all reported crashes can help data managers identify those agencies that have a set a reporting policy that does not align with the State’s laws or policies defining a reporting threshold.

83. What output measure(s) of completeness are currently reported for your State (e.g., less than five total omissions per PAR, less than five omissions of critical elements per PAR, the number of PARs returned to the submitting agency for the additional information, etc.)? (SOE: Provide baseline current data and trend if available, and provide sample reports) (AQ-68)

84. How does the State assess the presence of content (attributes) for each of the data elements within the crash database?

85. How does the State compile a listing of the data elements with their error frequency (omissions) results reported? (e.g. number and/or percent of errors per data element)

86. How does the State assess the frequency with which 'unknown' is the reported value for each of the data elements within the crash database?

87. How does the State identify the frequency with which 'unknown' is the reported response? (e.g. a descriptive analysis listing the frequency and/or percentage of 'unknowns' per data element)

88. How does the State measure the completeness of reports from individual law enforcement agencies submitting crash reports? (SOE: provide the measurements)

89. How does the State report to law enforcement agencies on the completeness of their crash report submissions? (SOE: provide a copy of the report and frequency of issuance)

Uniformity

Uniformity is the property of being externally and internally consistent with established standards. External uniformity of crash data is measured against the *Model Minimum Uniform Crash Criteria* (MMUCC) guideline, the Fatality Analysis Reporting System (FARS) and Motor Carrier Management Information System (MCMIS)/SafetyNet data system reporting requirements, and the ANSI D-16.1 and D20 crash reporting standards. Internal uniformity is typically measured as the degree of consistency with which law enforcement officers and their agencies' policies and practices conform to the State's guidance on reporting crashes. This guidance, in turn, generally comes from sources such as State laws and policies, as well as the State's crash reporting data collection manual given explicit direction on which crashes to report and how to interpret each field on the official crash report form. Measurement of uniformity is important because it tells data users if comparisons across years and among jurisdictions are validly supported by the dataset. Data collectors and managers can use measures of uniformity to identify problems at an individual, agency, or aggregate level, design measures to address those problems, and evaluate the effectiveness of the solutions.

Measures of external uniformity include the percent compliance with the MMUCC guideline, as well as the SafetyNet, and FARS reporting requirements. Measures of internal uniformity include reviews of completeness and accuracy over time or comparing among jurisdictions. Thus it is important to note that the data quality attributes of accuracy, completeness, and uniformity are interrelated.

90. What output measure(s) of uniformity are currently reported (e.g., all law enforcement agencies reporting crashes use a standard crash report form, all law enforcement agencies report identical data elements with identical data attributes, number of law enforcement

agencies not using the standard PAR)? (AQ-69) (SOE: Provide baseline current data and trend if available)

91. How does the State track uniformity of reporting by law enforcement agencies (i.e. tracking ratios for fatal, injury, property damage only by year to date)?

92. How are these output measure(s) calculated? (SOE: provide sample reports and their frequency of issuance)

Integration

Integration is the degree to which the crash data can be merged with other traffic-safety related data to create a richer resource for analysis. Integration is typically measured based on the number of databases that can be merged and, more importantly, by the extent of merger of records among two or more datasets in comparison to an expected level (the best possible). Measuring data integration can tell users whether or not the crash data have been successfully combined with another data set of interest and, more importantly, it can tell them if the resulting merged dataset is sufficient to support a specific analysis. Data collectors and managers may also use measures of integration to assess the quality of the crash data in relation to other related datasets. Failures to achieve an expected level of merged data very often point to problems in data definitions, accuracy, completeness, or uniformity. As noted in the CDIP Guide, data integration:

- Provides safety managers and analysts with a greater range of data elements to analyze and examine for factors that may be affecting safety.
- Identifies redundant datasets and/or superfluous duplication.
- Verifies or validates the accuracy of the information in each of the databases.
- Indicates better management of the quality of the data in the databases that are integrated.
- Detects unintended consequences when component files of the integrated system are changed, improved, and modernized.
- Improves confidence in using integrated files for data-driven decision making.

At a macro level, data integration is measured by tallying the number of traffic safety-related databases that can be merged, with one point for each successful pairing. More informative measures of data integration assess the quality of the merged dataset, especially the comparison of the number of merged records to the expected or theoretical maximum number. For example, if crash and EMS data are merged, it is expected that every crash-involved person who was coded as having been transported by EMS to the hospital will have a corresponding patient care report in the EMS database. Likewise, every person in the patient care run report

database who is identified as having been transported by EMS because of injuries sustained in a motor vehicle crash should be also identified as a crash-involved person in the crash database. If, in the process of merging the crash and EMS data, it is found that some records in either or both datasets remain unmatched, that would serve as an indication of data integration problems. Researching those individual "orphaned" cases can be helpful in efforts to refine the data merging process, but may also have implications for data element definitions, data quality edit checks, training, and the data collection and management processes as well.

93. What output measure(s) of integration are currently reported for your State (e.g., a count of the number of databases the crash database is linked to, for example the crash database linked to the driver licensing database and/or the vehicle registration database) (AQ-70) (SOE: Provide baseline current data and trend if available)

Accessibility

Accessibility is the extent to which users and potential users of crash data can effectively complete tasks requiring data or analytic support. It is typically measured by counting the number of users who can obtain data extracts and summary reports, and by counting the total number of analyses performed on behalf of users. Measures of accessibility can encourage expanded use of the data by alerting potential users to the methods and modes of data access. Data collectors and managers can use measures of accessibility to identify the degree to which the data is serving its purpose of supporting data-driven decision-making. It is widely recognized that the more crash data is used, and the more varied the uses made of it, the more support there will be for its improvement. Thus, measuring accessibility can also serve as a surrogate measure to assess a State's commitment to improving crash data quality. If the measures are developed with an eye toward quantifying the level of usage among a variety of user types, they can also help to identify under-served uses and users.

Many states measure accessibility in multiple ways. For example, the State may measure the number of copies of an annual crash summary report that are delivered or downloaded. It may also measure the number (and variety) of analytic requests and data extract requests that it fulfills each year. As a companion to this, it may also measure how long it takes (on average) to fulfill a request for data or analytic results. States may also choose to measure the satisfaction of data users—were they able to get the data or analytic results they needed? Were there any questions they could not answer with the level of access they had to the data? In this way, states may also use accessibility measurements to identify gaps in the data that could potentially be filled either through expanded access or by changing the data collection and data management processes.

94. What output measure(s) of accessibility are currently reported (e.g., a count of the number of requests for crash data fulfilled by the agency/department that maintains the crash database, number of agencies that have on-line access to the crash database) (AQ-71) (SOE: Provide baseline current data and trend if available)

For More Information:

Visit: http://safety.fhwa.dot.gov/rsdp/

FHWA, Office of Safety

Robert Pollack
Robert.Pollack@dot.gov
(202) 366-5019

FHWA-SA-XX-XXX