Crash Data Improvement Program Guide

U.S. Department of Transportation
National Highway Traffic Safety Administration

NHTSA

DISCLAIMER

Suggested APA Format Citation:

Scopatz, R., Brown, R., Zhou, Y., Benac, J., Peach, K., Bryson, M., & Lefler, N. (2017, December). *Crash data improvement program guide* (Report No. DOT HS 812 419). Washington, DC: National Highway Traffic Safety Administration.

Technical Report Documentation Page

1. Report No. DOT HS 812 419	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Crash Data Improvement Program Guide		**5. Report Date** December 2017
		6. Performing Organization Code
7.Authors Bob Scopatz, Richard Brown, Yuying Zhou, Jack Benac, Kara Peach, Meg Bryson, and Nancy Lefler		**8. Performing Organization Report No.**
9. Performing Organization Name and Address Vanasse Hangen Brustlin, Inc. 8300 Boone Blvd., Suite 700 Vienna, VA 22182-2626		**10. Work Unit No.**
		11. Contract or Grant No. DTNH22 14 D 00342L/0001
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration 1200 New Jersey Avenue SE. Washington, DC 20590		**13. Type of Report and Period** Final Report, March 2015 – October 2016
		14. Sponsoring Agency Code NHTSA
15. Supplementary Notes The contract manager for this report was Sarah Weissman Pascual.		
16. Abstract The Crash Data Improvement Program (CDIP) was developed to provide States with a means to assess and measure the data quality of their crash databases. As part of the overall CDIP program, NHTSA developed a revised CDIP Guide. The CDIP Guide provides an overview of crash data quality and guidance on how to measure and improve quality in terms of the six data quality characteristics - timeliness, accuracy, completeness, uniformity, integration, and accessibility. It also provides guidance for States on how to map their State's police crash reports to the Model Minimum Uniform Crash Criteria. The CDIP guide is intended for State crash database administrators and managers, State Traffic Records Coordinating Committee \ members, State highway safety offices and State DOT safety office personnel, local traffic safety personnel (e.g., law enforcement, traffic engineers, city and county planners), and other Federal, State, and local traffic safety professionals.		
17. Key Words: safety data, roadway inventory data, traffic data, data collection, crash data, data quality performance measures, traffic records	**18. Distribution Statement** Document is available to the public from the National Technical Information Service www.ntis.gov.	

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	164	

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

Table of Contents

Chapter I: Crash Data Improvement Program 1
1.1 Introduction 1
1.2 Improving Safety Decisions by Improving Data Quality 5
1.3 The Traffic Records System and Its Component Databases 7
1.4 The Crash Data Component 8
1.5 State Crash Data Collection, Management, and Use 10
1.6 CDIP Process 14
1.7 Measuring and Improving Crash Data Quality 16
1.8 Relationship of Crash Data to Other Traffic Records System Components 19
1.9 Summary and Conclusions 20
Chapter 2: Timeliness 22
2.1 Purpose of Measuring Timeliness 22
2.2 Defining Timeliness 22
2.3 Assessing Timeliness 24
2.4 Developing Timeliness Performance Measurements 25
2.5 Examples of Timeliness Performance Measurements 27
2.6 Final Considerations 37
2.7 Questionnaire 37
Chapter 3: Accuracy 39
3.1 The Purpose of Measuring Accuracy 39
3.2 Defining Accuracy 39
3.3 Assessing Accuracy 42
3.4 Developing Accuracy Performance Measurements 45
3.5 Examples of Accuracy Performance Measurements 48
3.6 Final Considerations 56
3.7 Questionnaire 56
Chapter 4: Completeness 57
4.1 The Purpose of Measuring Completeness 57
4.2 Defining Completeness 57
4.3 Assessing Completeness 59
4.4 Developing Completeness Performance Measurements 61
4.5 Examples of Completeness Performance Measurements 62
4.6 Final Considerations 67
4.7 Questionnaire 67

Chapter 5: Uniformity ... 68
5.1 Purpose of Measuring Uniformity ... 68
5.2 Defining Uniformity ... 68
5.3 Assessing Uniformity ... 69
5.4 Developing Uniformity Performance Measurements ... 70
5.5 Examples of Uniformity Performance Measurements ... 70
5.6 Final Considerations ... 73
5.7 Questionnaire ... 73
Chapter 6: Integration ... 74
6.1 Purpose of Measuring Integration ... 74
6.2 Defining Integration ... 74
6.3 Assessing Integration ... 78
6.4 Developing Integration Performance Measurements ... 79
6.5 Examples of Integration Performance Measurements ... 80
6.6 Final Considerations ... 81
6.7 Questionnaire ... 84
Chapter 7: Accessibility ... 85
7.1 The Purpose of Measuring Accessibility ... 85
7.2 Defining Accessibility ... 85
7.3 Assessing Accessibility ... 86
7.4 Developing Accessibility Performance Measures ... 86
7.5 Examples of Accessibility Performance Measures ... 87
7.6 Final Considerations ... 91
7.7 Questionnaire ... 92
Chapter 8: MMUCC Mapping Program Guide ... 93
8.1 Introduction ... 93
8.2 Process for Mapping a State PCR to MMUCC ... 94
8.3 Process for Mapping State Crash Database to MMUCC ... 117
8.4 Summary ... 126
References ... 127
Appendix A: Pre-Site Visit Information Collection ... 129
Appendix B: Assessment Questionnaire ... 133

List of Tables

Table 1.1. Components of a Traffic Records System 7

Table 1.2. Example Crash System Data Quality Performance Measures 16

Table 2.1. Example Data for Timeliness.... 27

Table 2.2. Example Data for Timeliness Calculation, Sorted by Number of Days 28

Table 2.3. Example Data Showing Timeliness of Paper PCR Submissions by LEAs 30

Table 2.4. Example Data Showing Timeliness of Electronically Submitted PCR Data by LEAs 31

Table 2.5. Example Data Showing Timeliness of Manual Location Coding for Reports Submitted by LEAs 34

Table 2.6. Example Data Showing Law Enforcement Timeliness in Correcting and Resubmitting Crash Reports 35

Table 3.1. MMUCC Attributes for the MMUCC Light Condition Data Element 41

Table 3.2. Example Data for Accuracy Calculation C-A-1 49

Table 3.3. Example Data for Calculating Percentage of Matching VINs 50

Table 3.4. Number of Errors per Paper Form Report in 2014 52

Table 3.5. Number of Errors per Electronically Submitted Report in 2014 53

Table 3.6. Percent of Decodable VINs in Crash Reports – 2014.... 54

Table 3.7. Percentage of Crash Locations Successfully Mapped 55

Table 4.1. Percent of Complete Crash Reports Received in 2009-2013.... 62

Table 4.2. Overall Completeness Percentage 2009-2013 63

Table 4.3. Percent of Unknown or Blank in Critical Data Fields for Which "Unknown" Is Not an Acceptable Value 2009-2013 64

Table 4.4. 2013 Ratio Measure of External Completeness.... 65

Table 4.5. Annual Proportion of Serious Crashes 2009-2013 66

Table 5.1. 2013 Ratio Measure of Uniformity.... 72

Table 6.1. Example Linkage Fields for Probabilistic Crash-Hospital Linkage 77

Table 6.2. Linkage Variables by Datasets Linked and Type of Linkage 78

Table 6.3. Example Data for Integration Calculation C-I-1.... 81

Table 6.4. Example of Tracking Data Sharing Agreements 81

Table 7.1. Agency Response Survey Table From Principal Users Satisfaction Survey 90

Table 7.2. Example Website Usage Statistics for Use in Accessibility Measurement 91

Table 8.1. MMUCC Element Capability Rating Scale.... 123

Table 8.2. PCR to MMUCC Mapping Compatibility Rating Scale.... 124

Table 8.3. Crash Database to MMUCC Mapping Compatibility Rating Scale 125

List of Figures

Figure 1.1. Fatalities and Fatality Rate 1964-2013 3

Figure 1.2. Economic and Societal Cost of Crashes in Comparison to Roadway Infrastructure Spending, 2010 (Revised) 4

Figure 1.3. Traffic Records System 7

Figure 1.4. CDIP Process Flow 15

Figure 2.1. Example Crash Process Flow Diagram 23

Figure 2.2. Example Showing Improvement in Statewide Average Timeliness for Crash Reports 28

Figure 2.3. Example Showing Timeliness of the Release of an Official Annual Crash Database 36

Figure 3.1 GIS Display of Crash and Roadway Information 40

Figure 3.2. Example Change Log for a Single Crash Record 44

Figure 3.3. Critical Data Elements from NHTSA 47

Figure 3.4. Average Number of Errors per Crash Report 2004-2014 51

Figure 5.1. MMUCC Mapping Results Example from NHTSA's Spreadsheet 71

Figure 6.1. Example of Deterministic Linkage Between Crash and Roadway Data in a State 75

Figure 6.2. Crash Data Linkage Opportunities with Other Traffic Records Data Systems 78

Figure 7.1. Example User Satisfaction Survey 89

Figure 8.1. State PCR Attribute "Other" Can Be Mapped to the MMUCC 97

Figure 8.2. State PCR Attribute "Other" Cannot Be Mapped to the MMUCC 97

Figure 8.3. Data Target (Left) and Data Source (Right) 115

Figure 8.4. Example of a Mapping Table 116

Acronyms

AASHTO	American Association of State Highway and Transportation Officials
CDC	Centers for Disease Control and Prevention
CDIP	Crash Data Improvement Program
DOT	Department of Transportation
ED	emergency department
EMS	Emergency Medical Services
FARS	Fatality Analysis Reporting System
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
DUI	driving under the influence
GHSA	Governors Highway Safety Association
GIS	Geographic Information System
GCWR	gross combined weight rating
GVWR	gross vehicle weight rating
HSP	Highway Safety Plan
LEA	law enforcement agency
LRS	linear referencing system
MAP-21	Moving Ahead for Progress in the 21st Century
MMUCC	Model Minimum Uniform Crash Criteria
NHTSA	National Highway Traffic Safety Administration
PCR	police crash report
PDO	property-damage-only
PII	personally identifiable information
QC	quality control
RMS	Records Management Systems
SHSP	Strategic Highway Safety Plan
TAT	Technical Assistance Team
TRCC	Traffic Records Coordinating Committee
U.S. DOT	United States Department of Transportation
VIN	Vehicle Identification Number

Chapter 1: Crash Data Improvement Program

1.1 Introduction

The Crash Data Improvement Program was developed to provide States with a means to assess and measure the data quality of their crash databases. That program provides States with performance measures that can be used to establish the timeliness, accuracy, completeness, uniformity, integration, and accessibility of their crash data. State crash database administrators, managers, and technicians can use the resulting information about their systems to make improvements.

Purpose of the Crash Data Improvement Program Guide

This CDIP guide will help States establish baseline performance measures that reflect the current status of crash system quality characteristics and conduct periodic updates to assess progress in improving crash data quality. The CDIP guide uses examples of good practices to help illustrate the use of quality performance measures.

The CDIP guide is intended to address the following issues relating to the crash database:

- What are the data quality characteristics?
- Why is each quality characteristic important?
- What is the definition of each data quality characteristic?
- What performance measures can be used to measure the quality characteristics?
- How are the performance measures calculated or derived?
- What metrics are used as numeric goals for data quality performance?
- How closely do the State police crash report data elements map to the Model Minimum Uniform Crash Criteria data elements?
- How closely do the State centralized crash database data elements map to the MMUCC data elements?

In addition, the CDIP guide provides sample management reports that present quality performance measures for various agencies or the State as a whole. While the CDIP guide's conceptual principles will be applicable to other traffic safety databases, the specific information presented in this CDIP guide is intended to be directly applicable only to a State's crash database.

Intended Audience

The CDIP guide is intended for State crash database administrators and managers, State Traffic Records Coordinating Committee members, State highway safety office and State DOT safety office personnel, local traffic safety personnel (e.g., law enforcement, traffic engineers, city and county planners), and other Federal, State, and local traffic safety professionals. In addition, the information presented here should benefit anyone who uses State crash data.

How to Use the Guide

This CDIP guide contains a top-level description of the CDIP process (Chapter 1) as well as a section for each of the six data quality performance measures defined by the National Highway Traffic Safety

Administration: Timeliness, Accuracy, Completeness, Uniformity, Integration, and Accessibility, and a final chapter on mapping State crash data elements to the MMUCC data elements. The introductory material lists the steps and products of the CDIP and explains the relationship between crash data and the other core components of the traffic records system. The following chapters are specific to one of the six data quality attributes. The final chapter describes the process for mapping to MMUCC. States may choose to use this guide to assess the data quality management practices for their own statewide crash records system. However, they are strongly encouraged to use the services of the NHTSA-provided technical assistance team in order to receive independent expert advice and an independent MMUCC mapping.

This guide also includes the CDIP Questionnaire. States complete the questionnaire in advance of the CDIP Technical Advisory Team visit. The questions provide the CDIP TAT the ability to create the onsite presentation for the State and address the items that have no solution in place. For example, if a State no longer uses paper crash report forms, the CDIP TAT can concentrate their presentation on the issues related to electronic data collection and transmission.

These questions should be assigned to people most knowledgeable about the State's crash data and systems. 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 Program Assessment Advisory* (NHTSA, 2012) are indicated with the (AQ-#) notation. If the State has recently completed an Assessment, it may save time and effort to provide the CDIP TAT with a copy of answers provided to the Assessment Questionnaire.

The Assessment Questionnaire is divided into the following six sections.

- Part I: Administrative details; best answered by the crash data custodian.
- Part II: Top-level processes of data collection and reporting.
- Part III: Specifics of the processes from Part II.
- Part IV: 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.

Parts 1 through 6 are included throughout Chapter 1. The remaining performance measure-related questions are integrated at the end of their corresponding chapters. Both the Pre-Site Visit Information Collection and Assessment Questionnaire are provided in Appendices A and B.

Resources Used in the CDIP Guide

The CDIP guide uses three key source documents as references for definitions of terms, practices, and data quality performance measures.

- *MMUCC Guideline: Model Minimum Uniform Crash Criteria, 4th Edition* (see Reference section, FHWA, FMCSA & NHTSA, 2012);
- NHTSA's report, *Traffic Records Program Assessment Advisory* (NHTSA, 2012); and
- NHTSA's *Model Performance Measures for State Traffic Records Systems* (NHTSA, 2011).

This CDIP guide also presents real-world examples of actual State practices aimed at measuring and improving data quality. The CDIP guide presents these examples as methodologies that States may wish to adapt to their own purposes and situations. Where appropriate, the examples are supplemented by step-by-step instructions on how to perform the required calculations, as well as a sample output report showing the information expected from the analysis.

Background

NHTSA's Fatality Analysis Reporting System data shows that the United States has reached historic lows for the number of fatal crashes and the number of people killed in motor vehicle crashes. Figure 1.1 shows that in 2013, 32,719 people died in crashes on the Nation's roadways, well below the 45,645 who died in crashes 50 years ago, and below the historic high of 54,589 fatalities in 1972. Over the same period, the fatality rate (deaths per 100 million vehicle miles traveled [VMT]) dropped from 5.39 in 1964 to 1.10 in 2013. Fatalities tell only part of the safety story, however. NHTSA estimates that in 2013 there were 5.69 million crashes, 28 percent of which involved non-fatal injuries. The Centers for Disease Control and Prevention estimates that for every person killed in a motor vehicle crash, 8 people are hospitalized and 100 are treated and released from hospital emergency departments (National Center for Injury Prevention and Control, 2014).

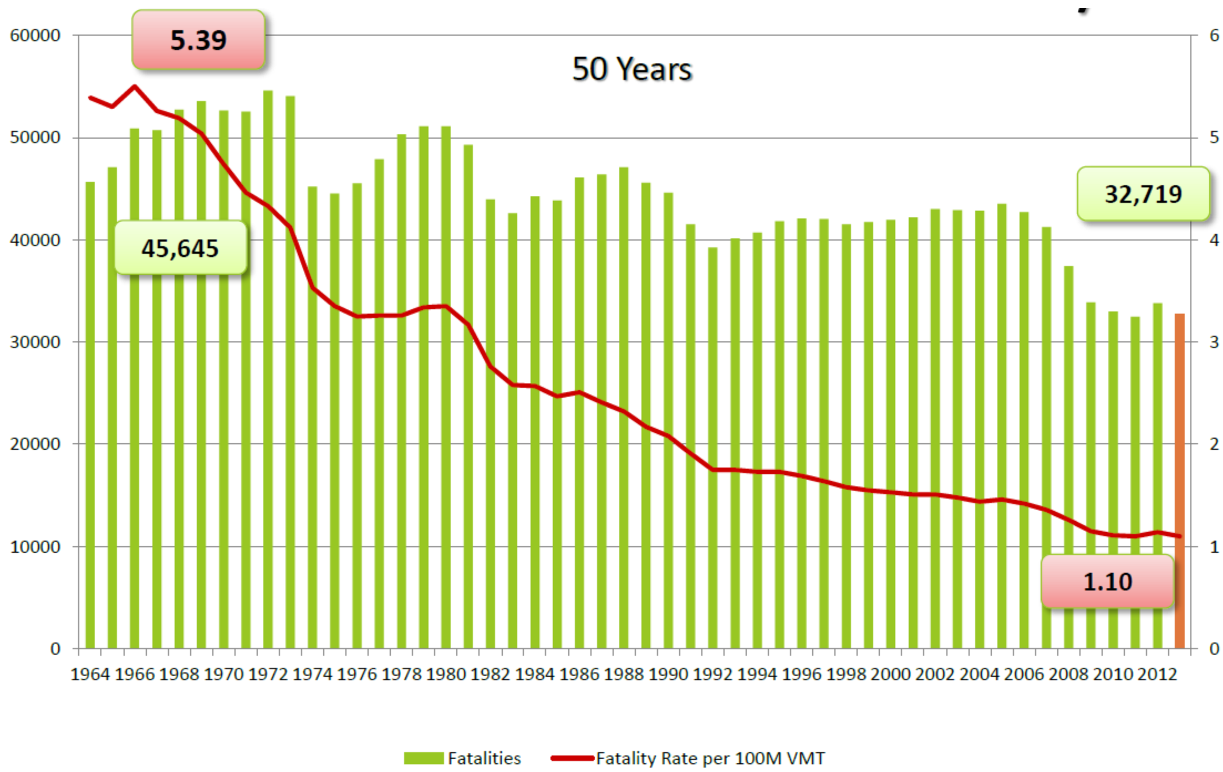

Figure 1.1. Fatalities and Fatality Rate 1964-2013

Crashes take an economic toll in terms of property damage, congestion and travel delays, lost productivity and other workplace losses, medical costs, legal and court costs, insurance administration, and emergency medical services costs. Crashes also have general societal costs expressed as the lost quality of life in years. In 2015 NHTSA published *The Economic and Societal Impact of Motor Vehicle Crashes, 2010 (Revised)* (Blincoe, Miller, Zaloshnja & Lawrence, 2015), estimating that the total economic cost of the crashes in 2010 was $242 billon; including societal costs, the total was $836 billion. This represents 1.6 percent of the U.S. GDP for that year. In contrast, all government spending (Federal, State, and local) on roadway infrastructure totaled $100.2 billion in 2010 (FHWA & Federal Transit Administration [FTA], n.d.b.). The contrast is shown in Figure 1.2.

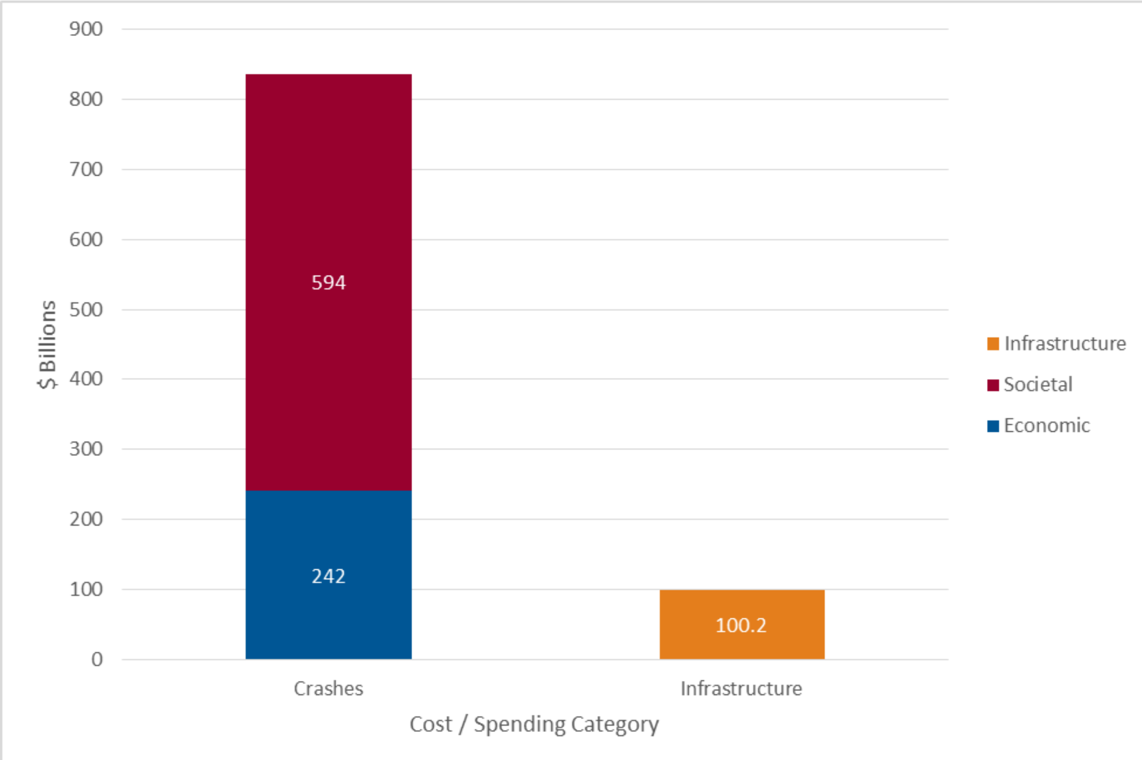

Figure 1.2. Economic and Societal Cost of Crashes in Comparison to Roadway Infrastructure Spending, 2010 (Revised)

CDIP Questionnaire Part I: Administrative

The following questions from the CDIP Questionnaire are used to identify the lead agency and other key stakeholders in the crash data management processes, including management of crash data quality. The objective is to better 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.

1. Which department/agency is the custodian of the State's crash database of record? (AQ-37)
2. Which section/office within that department has the principal responsibility for managing/maintaining the crash database? (AQ-37)
3. 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, web page links are acceptable.)
4. Who is the person responsible for administering or managing the State's crash database of record? What is their position or job title? Provide their contact information (mailing address, phone, fax, email)? (AQ-37)
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.)? (AQ-36) (SOE: provide a description of statewide database and specify how the data is consolidated into a statewide system)

7. 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?

1.2 Improving Safety Decisions by Improving Data Quality

The NHTSA *Traffic Records Program Assessment Advisory* (NHTSA, 2012) makes the case for quality data as a necessity for decision-making:

> High-quality State traffic records data is critical to effective safety programing, operational management, and strategic planning. Every State—in cooperation with its local, regional, and Federal partners—should maintain a traffic records system that supports the data-driven, science-based decision-making necessary to identify problems; develop, deploy, and evaluate countermeasures; and efficiently allocate resources. Functionally, a traffic records system includes the collection, management, and analysis of traffic safety data.

This statement fits well with the requirements set by Congress in the Moving Ahead for Progress in the 21st Century (MAP-21) legislation (MAP-21, 2012). In Uniform Procedures for State Highway Grant Programs Final Rule, (2013), NHTSA described the implementation of revised Section 402 and 405 Highway Safety Grant Programs to meet the requirements of MAP-21 including the direction that each State will submit an annual Highway Safety Plan that:

- Describes the State's highway safety program and the activities the State will undertake;
- Includes performance measures and targets as a condition for approval of the State highway safety program;
- Demonstrates continual, measurable progress in traffic safety;
- Harmonizes with other relevant safety plans, especially with respect to performance measurement; and
- Supports the U.S. DOT's efforts to analyze the linkage between safety investments and safety outcomes.

The FHWA provides States with guidance for implementing the MAP-21 requirements for data-driven decision-making in highway safety (FHWA, 2012), and for Strategic Highway Safety Plan development (FHWA, 2013). And most recently, on March 11, 2014, FHWA published a safety-related Notice of Proposed Rulemaking (FHWA, n.d.b). Included is guidance for States to:

- Use data in problem identification, countermeasure selection, and countermeasure effectiveness evaluation;
- Link multiple traffic records data sources;
- Adopt a performance-based approach (i.e., one that relies on performance measurement);
- Collect and report data on four mandatory safety performance measures for all public roadways;
- Use the 4Es (engineering, education, enforcement, and emergency services) to address safety problems;
- Identify, select, and evaluate effective countermeasures; and
- Expand their focus to *all public roads*.

The U.S. DOT administrations working most directly in highway traffic safety (NHTSA, FHWA, and FMCSA) have implemented major programs to help States adopt the data-driven approach to safety decision-making. Each agency provides States funding for data improvement and access to subject matter experts capable of advising the States on data and analysis improvements to support decision-making. In addition to the *Traffic Records Program Assessment Advisory* (NHTSA, 2012), NHTSA publishes periodic updates to the *Countermeasures That Work* report (Goodwin et al., 2013), which provides a list of behavioral program-specific countermeasures, each with indicators of effectiveness, cost, proportion of States using the countermeasure, and the time required to implement. FHWA has promoted use of the *Highway Safety Manual* (AASHTO, 2010) and related analytic tools. FHWA has also established an online Crash Modification Factor Clearinghouse (n.d.a), which is the engineering analog to NHTSA's listing of countermeasures that work.

The U.S. DOT guidance and resources related to traffic safety information and analysis are focused on data-driven decision-making. The programs and priorities are consistent in helping States identify ways to improve their data and the analyses performed to support safety decision-making. The emphasis on crash data quality in the CDIP is designed to support States' efforts to meet the U.S. DOT guidance and use the traffic safety-related resources to their best ability.

1.3 The Traffic Records System and Its Component Databases

The Traffic Records System is made up of six core component systems. These are described in Table 1.1.

Table 1.1. Components of a Traffic Records System

System	Description
Crash	The crash database includes all law enforcement-reported crashes meeting the State's reporting threshold. Sections of the crash report capture data elements related to the environment, people, and vehicles involved in crashes.
Vehicle	Vehicle databases include information on the numbers and types of vehicles registered in a State and about the owners of those vehicles.
Driver	Driver databases tell us about the license status, past convictions, and demographic attributes of drivers in the State.
Roadway	Roadway data has two key component databases: roadway attributes (inventory data) and traffic volume data.
Citation and Adjudication	Citation and Adjudication databases including information about traffic violation charges, convictions, and associated court actions.
Injury Surveillance	Injury Surveillance data include multiple databases describing emergency medical services (EMS, hospital and emergency department treatments, trauma records, and vital records (death certificate) data.

Of these, the crash database is considered central to virtually all traffic safety analyses including both behavioral and engineering- related problem identification, countermeasure selection, and effectiveness evaluation. Crash data provides the "who, what, where, when, how, and why" of safety analysis. The other databases provide key information that can refine an analysis by allowing users to create crash rates based on exposure to crash risk; identify segments of the population that are over-represented in crashes; identify roadway features associated with increased crash risk; and/or measure the outcomes of crashes in human terms (the nature and severity of injuries).

Figure 1.3. Traffic Records System

The CDIP focuses on crash data and its quality. However, it is important to recognize the value of integrating crash data with the other core traffic records system components to support decision-making. As depicted in Figure 1.3, crash data may be integrated with the other individual core traffic records system components as well as with all the components together. Crash data integrated with vehicle information can identify vehicle types (e.g., large trucks, motorcycles) that are over-represented in severe (fatal and serious injury) crashes. Crash data integrated with driver data can help to identify at-risk groups of drivers (e.g., novice teen drivers) and the relationships between driver behaviors (e.g., driving under the influence) and crash risk. Crash data integrated with roadway data is used to identify high-crash locations and to model the crash risk associated with specific roadway attributes and traffic volume levels. Crash data integrated with citation and adjudication data can quantify the crash risk associated with specific violations and identify the increased risks associated with repeat offenders. And finally, crash data integrated with injury surveillance data can help to quantify the dollar- and human-costs for injuries sustained in motor vehicle crashes.

The benefits of safety analysis using integrated data from multiple traffic records sources are proven when decision-makers set policy and design programs that are targeted to specific people, places, or event types. Safety program spending is more efficient and the States' ability to quantify the effects of selected countermeasures improves.

This CDIP guide provides examples of crash analyses integrated with other traffic records data. The focus of the CDIP is data quality measurement and improvement, but it is also important to recognize the value of the data for analysis and decision-making. This CDIP guide aims to illustrate the uses of crash data as well as how it can best be managed.

1.4 The Crash Data Component

Crash data is defined as the information gathered by law enforcement (or other sources as identified in State statute) describing the locations, circumstances, persons, and vehicles involved in motor vehicle crashes on public roadways. States may opt to go beyond this definition to include crashes on private property, crashes without motorized vehicles present (e.g., bicycle, pedestrian), crash reports submitted by the involved parties (rather than law enforcement), and crashes that fail to meet the statutory minimum reporting threshold. The crash data system description in the *Traffic Records Program Assessment Advisory* (NHTSA, 2012) was adopted for the purposes of this CDIP guide:

> The crash system not only holds the basic data critical to developing and deploying effective traffic safety countermeasures, it frequently also serves as the hub through which other systems are connected. The benefits and overall utility derived from the other traffic records systems are significantly enhanced by reliable, valid statewide crash data. Linking other systems' data with crash data enables invaluable opportunities for analysis. The resulting information drives State highway safety and injury prevention programs and has widespread applicability for all levels of government, industry, research groups, lawmakers, healthcare providers, and the public. The State crash system ideally contains—at a minimum—basic information about every reportable motor vehicle crash in the State. (Reportability 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); and
- **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.

Through linkages to other traffic records systems components, the crash data system identifies the roadways, vehicles, and people (e.g., drivers, occupants, non-motorists) involved in a crash. Data and analytic tools are broadly available so safety stakeholders can identify locations, roadway features, behaviors, driver characteristics, and vehicle characteristics that relate to crash risk. Crash data is also used to guide engineering and construction projects, prioritize law enforcement activity, and select and evaluate safety countermeasure programs. Crash data is used in analysis related to emergency response and how to maximize the level of care and the survivability associated with injuries sustained in a crash.

CDIP Questionnaire Part II: Crash Data Collection and Reporting

The following questions identify the documentation that detail the policies and procedures for the collections, submission, processing, posting, and maintenance of crash data. This documentation includes—but is not limited to—the crash data process flow diagrams, data standards, and data dictionary.

8. 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)
9. 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)? If there are people who are assigned less than full time on tasks related to the crash data system, please estimate staffing based on full-time-equivalents.
10. 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?
11. What type of coordination is there among the offices involved in crash report processing? For example, crash involved driver information may be needed for ascertaining financial responsibility, including assessment of any damages to public property—in such cases, multiple offices or agencies may be involved and information must flow between them. Location coding, in some States, is another example of inter-office or inter-agency coordination as the crash data system record is enhanced with location-specific information provided by another office. FARS and SafetyNet are also potential examples of coordinated processes.

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

13. 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)?

1.5 State Crash Data Collection, Management, and Use

Crash data collection, management, and use are activities of law enforcement, the data custodian, and analysts or decision-makers, respectively. People in these roles take on specific responsibilities related to data quality. As stated in the *Traffic Records Program Assessment Advisory* (NHTSA, 2012):

> 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 MMUCC guidelines and the statewide database's validation rules. Data is subject to validation checks at the point it is added to the record.
>
> The State maintains accurate and up-to-date documentation—including process flow diagrams—that details the policies and procedures for key processes governing the collection, submission, 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 anticipated 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.
>
> 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.
>
> 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.

The CDIP incorporates and adds detail to the guidance provided in the assessment advisory. In particular, the following components can be incorporated into a CDIP review:

- Review State's use of operator (driver) reports of crashes, especially when these are incorporated into the statewide crash database and, perhaps serve as a substitute for crashes that are not reported by law enforcement;
- Detailed examination of a State's crash data validation procedures and the content of validation rules. Specifically, CDIP includes analyses aimed at discovering the type and frequency of errors that could potentially be addressed by new validation rules;
- Examine State's process for identifying the need for new validation rules;
- Review the use of crash data in analyses to support traffic safety decision-making. This can include examination of behavior program management and engineering safety decision-making processes;

- Perform analyses using a State's crash data in creating data quality performance measures;
- Describe the Data Governance processes applied to the crash system. In particular, the data governance discussions would include involvement of the IT support staff along with the crash data managers;
- Analyze methods and effectiveness of outreach to State and local law enforcement agencies (the data collectors) with information on agency-specific data quality performance measures and the need for improvement;
- Compare data quality for paper versus electronic crash reports, and electronic versus paper submission of crash report forms/data; and
- Map the State's PCR and crash database to MMUCC data elements using a standardized methodology.

States are encouraged to discuss any special issues with their assigned CDIP contractor team leader during the planning phase of the project. This will help the team plan the CDIP to address specific concerns the State may have regarding collection, management, and use of the data to improve safety.

CDIP Questionnaire Part III: Crash Data Collection Specifics

The following questions are intended to identify how information on every reportable motor vehicle crash is collected. This section will review the data collection policies and procedures, software solutions for data collection, and information on the agencies collecting the information.

14. 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, web links are acceptable – this may have been provided in answer to question 3 earlier, but please point to the appropriate section of the law)
15. Does the law or policy specify a time limit within which the crash must be submitted to the State crash database?
16. Are there any reasons that a law enforcement agency might have an incentive for not providing crash reports in a timely manner (for example, if a municipality charges fees to the public for providing a crash report copy they may hold onto it for a time before making it available through another source such as the statewide crash reporting system)?
17. What is the minimum reporting threshold (monetary, tow-away, other) for a property damage only crash?
18. How many law enforcement agencies are responsible for investigating crashes and submitting reports to the State? (AQ-52)
19. Does the State have a standard police crash report form that is used by all agencies in the State? (AQ-53) (SOE: provide a copy of the PCR)
20. If multiple (out-of-date or specific to a municipality) versions of the PCR are used, how many different versions of the PCR form are being used? (SOE: provide a copy of all PCRs currently in use; and under what circumstances/ jurisdictions they are used)
21. Is there a short form used for PDO crashes? (SOE: provide the form.)
22. Does the State require crash-involved drivers to submit an operator report? If yes, does the State include operator-submitted data in the statewide crash database (i.e., blended with data from crash

reports supplied by law enforcement officers)? (SOE: provide the operator form if the data are used in the statewide crash database.)

23. 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.)

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

25. 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) Statewide, what percentage of crash reports is collected electronically? (AQ-53) Does the electronically collected data identically match the data elements and attributes on the paper PCR elements?

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

27. If not, is there a standard for data collection systems to be attained by all electronic reporting systems? Does the State approve vendors and/or law enforcement agencies to validate their electronic data collection software? If yes, please describe the approval process.

28. By what processes are PCRs from law enforcement agencies submitted to the crash database (e.g., mail, internet, secure electronic upload, CDs)?

29. Does the State crash database accept any crash data electronically? If so, what percentage is accepted electronically? Is there a standard for electronic data submissions? Does the State approve vendors and/or law enforcement agencies to validate their electronic data submissions? If yes, please describe the approval process.

30. Are there law enforcement agencies collecting crash data electronically that do not submit their data to the statewide crash database electronically (i.e., they collect the data electronically using software, but print the report and submit it on paper or send a static image file)?

31. 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)

32. What technologies are used for the verification?

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

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

35. 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)

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

CDIP Questionnaire Part IV: Data for Decision-Making

The following questions are intended to identify how the State uses this data in implementing programs and countermeasures that reduce motor vehicle crashes, deaths, and injuries. It is critical that a State's crash system include or have access to the key 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.

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

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

39. 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)

40. 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.)

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

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

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

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

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

46. 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)

47. How are crash-involved pedestrians and bicyclists recorded in the statewide crash database? Are they recorded as "involved units" (i.e., at the vehicle-driver level) or as "persons" (i.e., at the person level), or in some other manner?

48. Does the State crash database include data elements related to distracted driving? Please provide the data element definitions and attribute list.

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

50. How does the State define "serious injury" on the crash report?

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

52. Are the narrative and diagram for all crashes retained in the crash database? If yes, are these searchable or otherwise available for use in case selection, data aggregation, or analysis?

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

1.6 CDIP Process

When a State requests a CDIP through their NHTSA Regional Office, a process is initiated leading from initial scheduling to delivery of the technical assistance team's final report. To help States decide to participate in a CDIP, NHTSA has prepared this CDIP guide along with a CDIP pre-site-visit questionnaire and information request. States are encouraged to review these materials in advance of requesting a CDIP and to request help or further clarification as needed. The sequence of events is shown in Figure 1.4.

States are encouraged to plan well in advance. It can take time to complete the pre-site-visit questionnaire and to provide the requested information and data files. The TAT requires sufficient lead time to analyze data and prepare the site-visit slides and exercises. Once the site visit takes place, the final CDIP report can be provided to the State within a few weeks of the contractor receiving the State's comments on the draft conclusions and recommendations. States are encouraged to perform a thorough review of the draft report as this is the best time to correct any errors in the facts as presented by the TAT, and especially any conclusions that may affect the TAT's recommendations. While the TAT will take State comments into consideration where questions remain, State comments will not change the final report conclusions and recommendations. The CDIP is intended to serve as an external, objective review of a State's crash data system by outside subject matter experts.

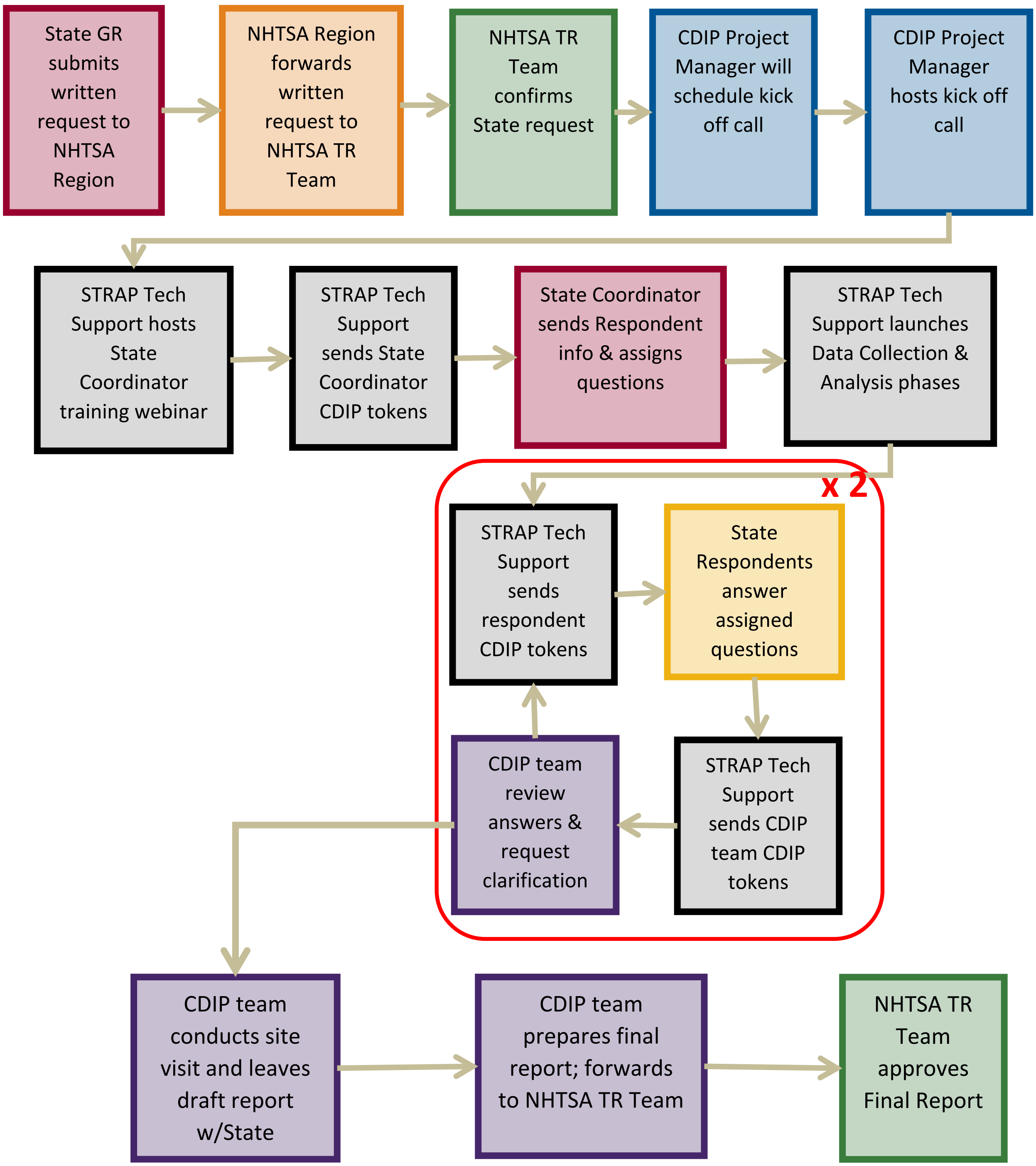

Figure 1.4. CDIP Process Flow

1.7 Measuring and Improving Crash Data Quality

The Model Performance Measures for State Traffic Records Systems (NHTSA, 2011) report defines a six-by-six matrix of six core systems and six data quality attributes (timeliness, accuracy, completeness, uniformity, integration, and accessibility. The Traffic Records Program Assessment Advisory (NHTSA, 2012) includes 10 example data quality performance measures in the assessment of the crash system, as shown in Table 1.2.

Table 1.2. Example Crash System Data Quality Performance Measures

Crash Data System Data Quality Performance Measures	
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 within XX days after the crash (e.g., 30, 60, or 90 days).
Accuracy	• The percentage of crash records with no errors in critical data elements (e.g., 95% of reports with no critical data error). • The percentage of in-State registered vehicles on the State crash file with Vehicle Identification Numbers matched to the State vehicle registration file (e.g., 95% of in-State VINs match).
Completeness	• A decrease in the percentage or number of missing data elements from the crash database for a given time period (e.g., 1 year). • The percentage of crash records with no missing data elements (e.g., 90% of reports with no missing data elements). • The percentage of unknowns or blanks in critical data elements for which unknown is not an acceptable value (e.g., 99% of reports with no unknowns or blanks in critical data elements).
Uniformity	• An increase in the number of crash data elements collected as part of a State's PCR that are mappable to MMUCC (e.g., 95% of data elements and attributes map successfully to MMUCC 4th Edition).
Integration	• The percentage of appropriate records in the crash database that are linked to another system or file. Examples: crash with in-State driver linked to driver file, crash with EMS response linked to EMS file (e.g., 85% of crash-involved persons coded as "transported" matched to EMS run report data).
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 (e.g., 90% of respondents rate their satisfaction as "satisfied" or "extremely satisfied").

The *Traffic Records Program Assessment Advisory* (NHTSA, 2012) includes these model performance measures as part of a formal, comprehensive crash data quality management program. The description of a formal, comprehensive data quality management program is derived from experiences in dozens of States throughout the decades of NHTSA traffic records assessments and the experience of several subject matter experts working with the NHTSA Traffic Records team during the 2012 update to the assessment advisory. Program components include:

- **Automated edit checks and validation rules** ensure that 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. States are advised to track changes to the data so that any alteration is tagged with a date, the change that was implemented, and the name of the person who made the change.
- **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. Performance 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 performance measures. The performance measures in:

- Table 1.2 are examples of high-level quality management indicators. The State is encouraged to develop additional performance 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 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 uses the reports to identify problems and develop countermeasures.

In coordination with the State, the CDIP team can place additional focus on one or more of the components of the data quality management program to help a State improve its practices and find ways to incorporate those components that may be missing now. One way that the CDIP can build on the Traffic Records Assessment conclusions is through the examination of the value of electronic crash data collection and electronic sharing of crash report data between LEAs and the State. The move to electronic data collection can improve accuracy, completeness, and uniformity through the application of validation rules as the officer is completing the crash report. Electronic submission can also improve timeliness. The resulting savings in data entry and error correction can often be used by the State crash data managers to implement or expand components of the formal, comprehensive data quality management program. Examples of States addressing data quality problems through the transition from paper to electronic crash reporting can be found in FHWA's *Crash Data Improvement Program (CDIP) Final Report* (Scopatz, Benac, & Lefler, 2010).

CDIP Questionnaire Part V: State Data Quality Management Program

The following questions identify if the State has a comprehensive crash data quality management program. The NHTSA 2012 update to the Traffic Records Program Assessment Advisory (p. 26-28) describes a formal, comprehensive 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.

54. 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? (AQ-63) If so, please describe. (SOE: or provide a listing of the edit checks and validations used by the state.)

55. 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? (AQ-64) If so, please describe.

56. Does the State have processes in place for transmitting and tracking rejected reports between the statewide crash system and the originating officer? (AQ-65) If so, please describe.

57. Does the State establish numeric goals—performance metrics—for each performance measure? (AQ-72) If so, please describe.

58. Does the State provide performance reporting to law enforcement agencies regarding the timeliness, accuracy, and completeness of their submissions to the statewide crash database? (AQ-73) If so, please describe.

59. Are high-frequency errors tracked and used to generate new training content and training manuals, updated validation rules, and prompt form revisions? (AQ-74) If so, please describe.

60. Does the State's data acceptance process include quality control reviews that compare the narrative, diagram, and coded report contents? (AQ-75) If so, please describe.

61. Does the State periodically conduct independent, sample-based audits of crash reports and related database contents? (AQ-76) If so, please describe.

62. Does the State periodically conduct comparative and trend analyses to identify unexplained differences in the data across years and jurisdictions? (AQ-77) If so, please describe.

63. Does the State regularly collect data quality feedback from key users and share it with data collectors and managers? (AQ-78) If so, please describe.

64. Does the State regularly provide the TRCC data quality management reports? (AQ-79) If so, please describe.

CDIP Questionnaire: 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 1.2 are examples of high-level quality management indicators. The State is encouraged to develop additional measures that address their specific needs.

Does the State currently have a processes for assessing the quality of the information within the crash database in terms the data quality attributes listed in:

65. Table 1.2. For each of the six data quality attributes, please indicate if the State has the listed data quality performance measures and list any other relevant data quality performance measures that are calculated on a statewide basis.

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

1.8 Relationship of Crash Data to Other Traffic Records

Crash data, used in conjunction with other traffic records data sources, supports decision-making by allowing analysts to produce detailed examinations of the behaviors and location attributes that play a role in either increasing or decreasing the risk of crashes, injuries, and fatalities. The *Traffic Records Program Assessment Advisory* (NHTSA, 2012) addresses these issues in the section dealing with interfaces and data integration between crash data and the other traffic records components:

- The 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. By contrast, system *integration* refers to the discrete linking of databases for analytic purposes. Data integration is addressed in Section 4 of the assessment advisory;

- 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;

- The State's crash data ideally exists in one consolidated, generally accessible database. If data is first aggregated in separate law enforcement databases or records management systems, upload to the statewide database is electronic and automatic. The statewide crash database is also capable of supplying data to law enforcement agencies' RMS; and

- Routine protocols for uploading data to FARS and SafetyNet are created to assure congruence with the State's crash data and to generate management and analysis efficiencies. Examples of useful interfaces between the crash data system and other traffic records system components are outlined in the following section.

To accomplish data interfacing or integration, all of the source databases must contain at least one common data element (a linking variable). The linkage may be accomplished using *deterministic* or *probabilistic* methods. Deterministic linkage is when the linking variables in the data sources identify unique records, people, vehicles, locations, among others (e.g., case number, driver license number, vehicle identification number, GPS coordinates). Probabilistic linkage methods are used when there is uncertainty over the exact match between records—usually when personal identifying information is redacted or key information is missing from one or more of the databases. For example, hospital discharge and EMS data may be linked with crash data based on date, time, and location of incident, as well as the age and sex of the injured party. Both deterministic and probabilistic linkage methods are valid and produce useful data.

The line between "interfacing" and "integration" may be blurred in actual practice, especially when a State has strong data governance practices applied to a majority of the traffic records system core systems. That is because one goal of effective data governance is to collect any particular data element once and share it wherever necessary—whether that is an instance of interfacing between two databases or integrating them on a record-by-record basis to support a specific analysis. Chapter 4 of *NCHRP 666: Target-Setting Methods and Data Management to Support Performance-Based Resource Allocation by Transportation Agencies* (TRB, 2010) provides transportation-data specific definitions and examples of data governance practices. The goals of data governance are to improve efficiency (less cost for the same or better quality information), improve quality (using the most timely, accurate, and complete source for each piece of information), and thus improve safety by providing decision-makers with the information they need (data *and* analytic results) focused on the issues that need to be addressed.

The CDIP can help a State identify opportunities for data interfaces and integration as well as define meaningful ways to measure the success of those efforts. In conjunction with the State, the CDIP team can focus on particular safety decisions that are of concern and provide detailed examples of how effective interfaces and data integration can support those decision-making efforts.

1.9 Summary and Conclusions

The primary focus of the CDIP is crash data. It incorporates consideration of the entire traffic records system by examining the overall management of traffic records, and by assessing the processes and measurement of data interactions between the crash data and all other components of the traffic records system (driver, vehicle, roadway, citation/adjudication, and injury surveillance).

The CDIP is aimed at improving a State's data quality management program, with its most direct impact on data quality performance measurement. The CDIP can help a State define metrics (the goals against which performance is to be measured) and the specific data quality measurements themselves. The CDIP is not limited to the set of 10 data quality performance measurements used as examples in the *Traffic Records Program Assessment Advisory* (NHTSA, 2012); however, those are acknowledged as the recommended starting point for the measurement component of a formal, comprehensive data quality management program. Following that formal program description, the CDIP also helps States identify the need for additional validation rules in collecting and managing crash data, and the data governance practices related to providing feedback on needed improvements by the data collectors. Implementing a formal, comprehensive data quality management program can be costly; therefore, the CDIP can help States identify opportunities to fund the needed quality management steps. The CDIP is also aimed at improving crash data uniformity among all States. This is supported through the MMUCC mapping portion of the CDIP and is provided to States as a way to verify the match between a State's PCR, its crash database contents, and the MMUCC Guidelines.

The overall goal of the CDIP and of Traffic Records Improvement is to provide better information with which to make safety decisions. Thus, the CDIP also examines the processes that a State uses to improve data analysis and use. All of these efforts are supported by U.S. DOT, NHTSA, FHWA, and FMCSA in their implementations of MAP-21 requirements. Under those requirements, States are pursuing data that applies to *all public roads* and are adopting a standard set of performance measures. In traffic records management, States are also formalizing their TRCCs' roles in data quality management. The *Traffic Records Program Assessment Advisory* (NHTSA, 2012) is a useful tool for States' efforts as it supports a detailed examination of how well a State manages each of the core systems, as well as the overall data improvement process. For crash data, the CDIP goes a step further by providing specific recommendations on practices, data quality performance metrics and measurements, and data management processes. This updated NHTSA CDIP guide is intended to explain how crash data management supports the safety decision-making process, and thus improves safety. State traffic records managers, crash data managers, collectors, and users are all among the intended audience for this Guide and for CDIP participation.

Chapter 2: Timeliness

2.1 Purpose of Measuring Timeliness

Crash data is timely when it meets users' needs. Users rely on timely crash data to make decisions affecting the locations, vehicles, and people involved in crashes. When data is out of date, decision-makers run the risk of addressing problems that no longer exist or are no longer a priority. Moreover, the value of feedback to data collectors is diminished. Without timely feedback, law enforcement cannot be expected to address problems on individual crash reports; nor can they be effective in addressing data quality problems. For data managers, timely data supports all other data quality improvement efforts.

The *Model Performance Measures for State Traffic Records Systems* report (NHTSA, 2011) and the *Traffic Records Program Assessment Advisory* (NHTSA, 2012) make a distinction between data quality metrics and data quality performance measures. Metrics are goals set by the State for each traffic records system and data quality attribute. Performance measures are the quantitative tools used to gauge the performance of a system in each data quality attribute. States are advised to develop data quality metrics and performance measures and, where possible, build them into the crash data management software as automated reporting functions. Comparing performance measures to timeliness metrics helps States manage crash data timeliness. By measuring timeliness, a State can:

- Detect interruptions of data flow from law enforcement agencies;
- Provide specific performance feedback to law enforcement agencies;
- Assess the impact of efforts designed to improve timeliness;
- Provide users with indicators of the relevance of the data for analysis;
- Plan release of official, year-end datasets; and
- Establish achievable metrics and identify needed improvements to achieve those goals.

2.2 Defining Timeliness

The *Model Performance Measures for State Traffic Records Systems* report (NHTSA, 2011) defines timeliness as "…the span of time between the occurrence of an event and entry of information into the appropriate database." In practice, this definition expands to capture all important events in crash data management. Figure 2.1 shows an example crash process flow diagram from the initial crash event through to storage in final databases. Each step is numbered in the sequence of data management processes. Timeliness may be defined and measured for each step. States may wish to track the timeliness of individual steps as well as the timeliness of the overall process from start to finish.

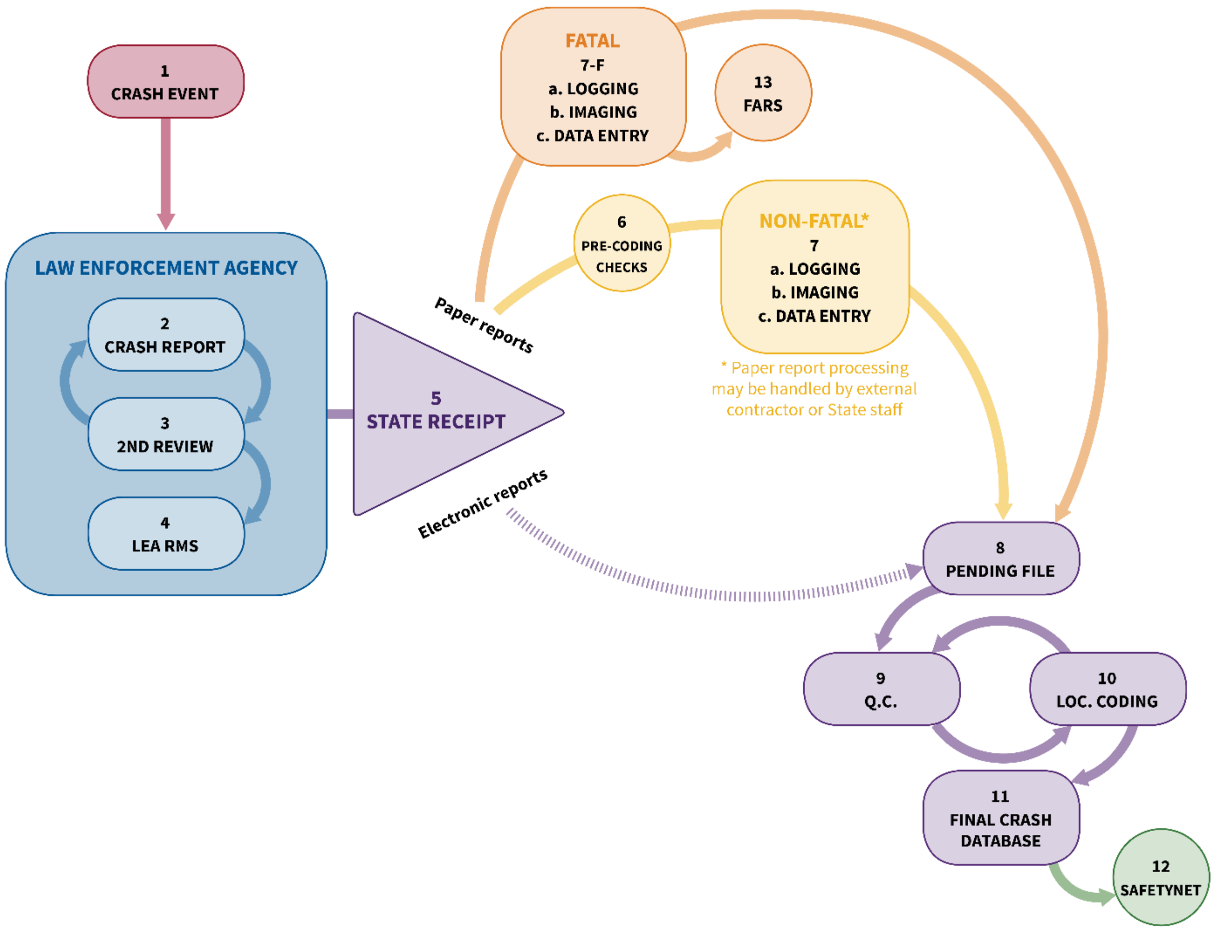

Figure 2.1. Example Crash Process Flow Diagram

The steps in the crash data management sequence are:

1. **Crash Event:** The initiating crash event that meets or exceeds the State's reporting threshold.
2. **Crash Report:** This is the crash report form or electronic record completed by the officer.
3. **Secondary Review:** This is the review by a second party or officer's supervisor to ensure the report meets the LEA's standards for submittal. Steps 2 and 3 may repeat multiple times as the secondary reviewer rejects the report and the officer corrects it.
4. **LEA RMS:** This is the entry of the crash report into the LEA's internal databases. Agencies with an RMS may have the ability to send crash reports electronically to the State.
5. **State Receipt:** This is the arrival of the crash report (or electronic crash report data) at the State office tasked with crash data management. Fatal crashes are sent to the FARS analyst for special handling.
6. **Pre-Coding Checks:** This is the State process for identifying incomplete reports or errors that would require correction before data entry. Not shown: Reports may be returned to the law enforcement agency for correction at this stage.

7. **Data Entry:** This is a multi-step process that may include logging, imaging, and manual data entry. Step 7 is shown with a box around it including all sub-steps (7a, 7b, and 7c) because, in some States, the reports are sent out to a contractor who is responsible for all of the processes but is not required to report timeliness for each sub-step separately. For reports of fatal crashes, a parallel set of processes may occur. These are labeled with the letter F (for "fatal").
8. **Pending File:** This is typically a State "holding" or development database where non-final data resides pending completion of quality control checks.
9. **QC:** This is the process where crash report data is reviewed by the State for compliance with established quality standards for accuracy, completeness, and uniformity. This may be a combination of automated checks and manual reviews. Reports that fail the QC checks may be returned to law enforcement at this stage.
10. **Location Coding:** This is the process where crash report location codes (identifying segments and intersections by number in the State's linear referencing system) are added to the crash records.
11. **Final Crash Database:** This is the process where crash report data is added to the statewide crash database and is then considered ready for access by users.
12. **SafetyNet:** This is the process where reports of crashes involving commercial motor vehicles are uploaded for validation and entry into the FMCSA's SafetyNet system.
13. **FARS:** This is the process where reports of fatal crashes are entered into FARS. Not shown: In some States, the FARS entry process is separate from entry of the same crash information into the statewide crash database. Also, for electronically submitted fatal crashes, the FARS analyst obtains the record from the crash database.

Each of these steps may be measured for timeliness, and timeliness may be defined separately for each event as well as for the process overall. In order to improve overall timeliness, States may need to focus on reducing the duration of individual steps in the process. Examples of timeliness performance measures for individual steps in the crash data management process are presented later in this chapter.

2.3 Assessing Timeliness

Timeliness performance measures are essentially the length of time between a start and an end point. Common performance measures can include: (1) the time between crash date and data submission; (2) the time between data submission and resubmission (following errors); or, (3) average time required for each step in the process of data submission and verification.

An ideal data management system should incorporate automatic timeliness performance measurements, both for the individual steps shown in Figure 2.1 and for the overall process. Ultimately, States should decide the time intervals (monthly, quarterly, etc.), levels (police department, county, etc.), and types (manual versus electronic, crash severity, etc.) of timeliness performance measures that are most important for their needs. States that do not currently have an *ideal* system in place should manually calculate the most useful timeliness reports at regular intervals.

Whether a State's timeliness reports are automatically or manually generated, they should be tailored to the needs of decision-makers. A State's TRCC can assist data managers in identifying the various data users and their needs. In establishing timeliness reports, States should recognize that the data collectors and data managers are also decision-makers and have specific needs for timeliness information.

Data Submission Type and Related Timeliness Performance Measures

The type of data submission will greatly impact the timeliness of a State's crash database. Therefore, States should measure any process that can cause delay in posting a crash record to the statewide database. Any number of performance measures that apply to a State's particular situation should be used. For instance, States that use a mix of manual and electronic data submission might benefit from developing separate performance measures, setting metrics for each, and comparing and contrasting when planning for future data improvement grant programs.

The following are examples of timeliness performance measures that can be used to determine performance for a mix of manual and electronic data submission:

- **Manual data submission by law enforcement.** States that accept manual data submission (i.e., paper crash reports mailed to an intake center or personnel) should ensure that they incorporate a report received date/time stamp. This data point can be used in a number of performance measures from the timeliness of an individual police department completing crash reports to the timeliness of the intake process (receipt of crash report to entry into a database). This can provide a basis of comparison to help quantify the advantages—both timeliness and financial—of any investments in electronic data submission.
- **Direct data entry by law enforcement officers.** States that are fully electronic and provide a user access portal for report completion and submission often have close to zero time lapsing from report submission to availability within the database. In those situations, it is still advisable to measure the time from the crash event to the completion and submission of the form.

  Systems that also incorporate supervisory review can add an additional timeliness performance measure—the time between crash occurrence and submission, with progress points at crash report completion, and supervisory approval.
- **Electronic crash reporting and submission into a pending database.** States that incorporate a large number of edit checks and data completion tasks (converting address or link/node to latitude/longitude) may use a temporary storage space or pending database for initial electronic submissions, as shown in Step 8 of Figure 2.1. The data stored here is not accessible as part of the statewide crash database. Checks can occur upon arrival, nightly, weekly, or other regular intervals. Errors discovered during edit checks can sometimes be corrected by in-house staff, or the report may be sent back to the reporting agency for review and correction.

2.4 Developing Timeliness Performance Measurements

The goal of timeliness performance measurement is to provide useful information for data collectors, managers, and users. In developing a new timeliness performance measurement, system managers should consider how the information can be used to improve timeliness. Monitoring timeliness of records in the system and providing specific feedback to LEAs on their performance are two separate purposes and should be met with two different performance measurements.

The *Traffic Records Program Assessment Advisory* (NHTSA, 2012) describes a formal, comprehensive data quality management process. This process should be reflected in the crash data process flow diagram along with other steps that may enrich the crash data, for example, adding location codes. Some of these activities (such as error checking and location coding) affect the timing of when a crash record becomes available in the statewide crash database. A State may successfully automate some of these processes so that they (generally) do not delay the appearance of a crash record in the database. It is still important to measure the processes, especially when manual intervention is required (i.e., when the automated process fails). Doing so will help to provide a comparison between the timeliness of records that are easily processed with the State's automated procedures versus those that require special handling.

For monitoring timeliness of records in the database, the date of the crash (Step 1 in Figure 2.1) should be compared to the date of entry into the final crash database (Step 11). For providing feedback to LEAs, timeliness of their submissions should be isolated from all other processes, so the performance measurement should compare the date of crash to the date of submission by law enforcement (the end of Step 4).

The *Model Performance Measures for State Traffic Records Systems* report (NHTSA, 2011) defines two performance measures of crash data timeliness. Both performance measures are based on a comparison of the date of the crash to the date that the crash record is available in the database. As stated in that report:

> Timeliness always reflects the span of time between the occurrence of some event and the entry of information from the event into the appropriate database. For the crash database, the events of interest are crashes. States must measure the time between the occurrence of a crash and the entry of the report into the crash database. The model performance measures offer two approaches to measuring the timeliness of a crash database:
>
> **C-T-1: 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 median value is the point at which 50 percent of the crash reports were entered into the database within a period defined by the State. Alternatively, the arithmetic mean could be calculated for this measure.
>
> **C-T-2: The *percentage* of crash reports entered into the database within XX days after the crash.** The XX usually reflects a target or goal set by the State for entry of reports into the database. The higher percentage of reports entered within XX days, the timelier the database. Many States set the XX for crash data entry at 30, 60, or 90 days but any other target or goal is equally acceptable.

The NHTSA *Model Performance Measures for State Traffic Records Systems* (2011) report provides the minimum performance measures for timeliness; yet more performance measures may be needed to help States meet their data quality goals. The C-T-1 performance measurement calculates the difference between Steps 1 and 11 in Figure 2.1. And, C-T-2 performance measure also uses the difference between Steps 1 and 11 in Figure 2.1, but compares each report's timeliness to a State-defined metric setting the target number of days. C-T-1 and C-T-2 are the performance measures that all States should be able to calculate and share. Other performance measures of timeliness are likely to be specific to a State's particular situation: the issues it must track and correct; the components of overall timeliness that identify bottlenecks; and, the feedback required by LEAs in order to improve their own processes.

It is also important to understand the difference between *median* and *mean* when deciding which to use in the two recommended performance measures. Calculating the mean is usually much easier than calculating the median from within most crash records systems. Another advantage of the mean is that it is easily combined across sub-groups (e.g., paper and electronic crash report timeliness) to arrive at a combined average.

The median may give a more accurate picture of expected timeliness for crash records than the mean because the median is not sensitive to unusually high or low values. The mean changes depending whether the oldest crash is 100 or 1,000 days old. The median stays the same regardless of the timeliness of the oldest or youngest crash in the database.

2.5 Examples of Timeliness Performance Measurements

The two crash data timeliness performance measurements defined in the *Model Performance Measures for State Traffic Records Systems* report (NHTSA, 2011) are calculated in the following section.

Measure 1: (C-T-1) Median or Mean Timeliness

A. **Calculate timeliness for each crash report record.** For each crash event, record the difference, in days, between the crash date and date of entry on the statewide database. The crash date is the actual date on which the event took place. The date of entry into the statewide crash database is the date when the record is first created in that database—this need not necessarily correspond to the date when the record is made available to users for analysis.

B. **Calculate the median (or mean) number of days for a specified reporting period (monthly, quarterly, annually, etc.).** For all crash records with a crash date within the specified period, calculate the mean or median number of days.

Example data for these calculations is shown in Table 2.1.

Table 2.1. Example Data for Timeliness Calculations

Record	Crash Date	Entry Date	Timeliness (Days)
1	4/1/2015	4/5/2015	4
2	4/2/2015	4/5/2015	3
3	4/3/2015	4/3/2015	0
4	4/3/2015	4/15/2015	12
5	4/8/2015	4/15/2015	7
6	4/9/2015	4/12/2015	3
7	4/10/2015	4/18/2015	8
8	4/11/2015	4/22/2015	11
9	4/19/2015	4/24/2015	5
10	4/25/2015	5/7/2015	12
Median	**6 days**		
Mean	**6.5 days**		

In this example, the median can be found by listing all of the timeliness values in order from smallest to largest and picking the value that is in the middle of the range. Since there are 10 records, this is a value half-way between the fifth largest and sixth largest timeliness numbers.

The sorted timeliness numbers are shown in Table 2.2.

Table 2.2. Example Data for Timeliness Calculation, Sorted by Number of Days

Rank	Timeliness (Days)
1	0
2	3
3	3
4	4
5	5
6	7
7	8
8	11
9	12
10	12

The fifth largest value is five. The sixth largest value is seven. The value halfway between these numbers is six. **The median timeliness for this reporting period is 6 days.** The mean is calculated by taking the sum of all the timeliness values (in this case, the total is 65) and dividing by the total number of crash records in the reporting period (in this case, 10 records). **The mean timeliness for this reporting period is 6.5 days.**

Figure 2.2 shows an example of statewide average timeliness from the date of the crashes to the date the data is available in the central crash records database. Charts like this help to demonstrate that the State's efforts aimed at improving timeliness have been effective. The baseline year (2003) shows overall timeliness of 103 days. The most recent year (2014) shows timeliness of 14 days. This represents an improvement of more than 86 percent.

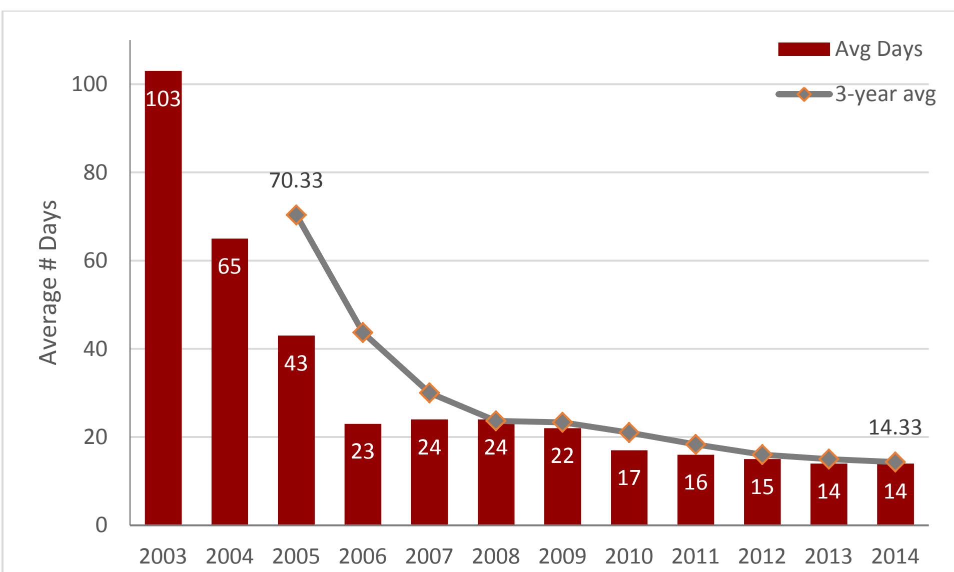

Figure 2.2. Example Showing Improvement in Statewide Average Timeliness for Crash Reports

Measure 2: (C-T-2) Percentage of Records Within Timeliness Criterion

A. **Calculate timeliness for each crash report record.** For each crash event, record the difference, in days, between the crash date (Step 1 in Figure 2.1) and the date of entry on the statewide database (Step 11 in Figure 2.1).

B. **Tally the number of crash reports entered within the criterion number of days.** For this example, the criterion has been set to 10 days. Using the same set of 10 example records, there were 7 records entered within 10 days of the crash event, and 3 records that took more than 10 days.

   The percentage of records entered into the statewide database within 10 days is 7/10 x 100 percent; or, 70 percent.

For the remaining timeliness performance measures presented in this chapter, the examples will focus on the average number of days as discussed under Measure 1. If the State prefers, they may use medians, or report percentages of reports meeting a specific timeliness metric. The procedure for calculating the average will not be repeated for each of the remaining performance measures. Instead, the remaining examples will focus on how timeliness performance measurements can be useful in providing feedback to data collectors, in identifying bottlenecks in crash data processes such as location coding, and in providing users with critical information about the status of the crash system overall.

Measure 3: Timeliness of Crash Report Submissions by LEAs

This performance measure is like Measure 1 except that the end date is not the date on which the crash report is entered into the statewide database, but rather, the date on which the centralized crash data processing office receives the crash report from the submitting LEA. The date of receipt by the State is captured separately for paper reports (Step 5 in Figure 2.1) and those received electronically (Step 8 in Figure 2.1). If a State has both paper and electronic crash report submissions, it is useful to calculate timeliness performance measures separately for both types of report submission, as well as a combined performance measure for the entire database. Note that this timeliness performance measure is independent of the date on which the crash records are entered into the statewide database. This performance measure isolates the LEAs' portion of the crash reporting process and is useful in providing those agencies (the data collectors) with feedback on how well they are meeting timeliness metrics specific to their report submissions.

Table 2.3 shows an example of paper PCR submission timeliness defined as the number of days from the crash event to the date of receipt by the central crash processing center. For example purposes, not all of the State's LEAs are shown in the table. The full report would show every LEA in the State submitting paper reports. The table shows the agency identification, reporting source (only those with "P=paper" were selected for this report), the total number of crashes submitted in 2014, and the average number of reporting days. The reporting days performance measure is the difference between the crash date (Step 1 in Figure 2.1) and the date of receipt of the paper reports by the State (Step 5 in Figure 2.1).

Table 2.3. Example Data Showing Timeliness of Paper PCR Submissions by LEAs

Agency	Reporting Source (P)	Total Crashes	Average Reporting Days
A	P	888	16.91
B	P	129	16.38
C	P	47	11.33
D	P	4	197.59
E	P	74	15.45
F	P	114	20.66
G	P	719	22.98
H	P	38	106.52
I	P	2	15.54
…			
R	P	31	17.22
S	P	9	47.15
T	P	305	40.78
U	P	561	29.28
V	P	1	29.38
W	P	564	20.88
X	P	4	4.25
Y	P	96	53.12
Z	P	8	26.68
Statewide Reported Crashes on paper PCRs and Average Reporting Days		**14,659**	**25.71**

Table 2.4 shows similar data for electronic report submissions by LEAs using electronic crash data collection software and electronic submission of crash report data. This table shows the agency, vendor, total number of crashes reported in 2014, and the average reporting days. As with Table 2.4, the reporting days performance measure is the difference between the crash date (Step 1 in Figure 2.1) and the date of receipt (entry of electronic data into the pending file in Step 8 of Figure 2.1).

Table 2.4. Example Data Showing Timeliness of Electronically Submitted PCR Data by LEAs

Agency	Reporting Vendor	Total Crashes	Average Reporting Days
ZA	A	1,498	25.69
ZB	A	496	8.25
ZC	A	209	16.44
ZD	C	479	22.57
ZE	C	3,489	21.26
ZF	C	265	25.05
ZG	C	580	19.59
ZH	C	385	19.99
ZI	C	1,502	22.68
…			
ZR	C	1,344	22.88
ZS	C	1,899	22.16
ZT	V	42	18.39
ZU	V	664	7.46
ZV	V	247	10.19
ZW	V	1,154	4.05
ZX	V	957	3.46
ZY	V	26	7.76
ZZ	V	270	63.31
Statewide Electronically Reported Crashes and Average Reporting Days		**291,815**	**12.63**

In comparing data in Table 2.3 and Table 2.4, several facts become apparent:

- Electronic submissions (on average less than 13 days) are timelier than paper submissions (on average greater than 25 days).
- In 2014 some 306,474 crash reports were submitted 14,659 + 291, 815 = 306,474.
- The State has achieved greater than 95 percent electronic submission of crash reports:

$$(291{,}815/306{,}474) * 100\% = 95.2\%$$

- Overall timeliness of all crash report submissions (paper and electronic) by law enforcement is 13.26 days as shown:

$$\frac{[(14{,}659 * 25.71)+(291{,}815*12.63)]}{306{,}474} = 13.26\ days$$

Note that the same result could be obtained by creating one master table combining Table 2.3 and Table 2.4, then calculating the average number of days for all submissions.

This type of report is useful for providing feedback to all LEAs. Some agencies included in Table 2.3 (paper submissions) take more than 1 month (and some more than 4 months) to deliver their crash reports to the State. In Table 2.4, there are electronic submitters whose average number of reporting days is higher than the average for paper report submissions. These agencies could be contacted to determine why their reporting lags behind the average and to encourage them to adopt different practices, including electronic report submission and more frequent submissions. Strategically, States may wish to prioritize the agencies with the slowest timeliness and the largest count of crash reports per year, regardless of whether they are using electronic or paper report submissions.

It is tempting to encourage all of the paper report submitters to shift toward electronic reporting and submission because that also saves time at later steps (data entry) in the crash report data management process. However, not all States will achieve 100 percent electronic reporting. A State may decide that it is more cost effective to allow a small percentage of reports to be completed manually because the challenges to automating some of the agencies are too great. Before investing heavily in automating processes for the few remaining paper report users, the State would be well served by an analysis of the project's cost-effectiveness. The State might benefit more by helping the slower agencies among the electronic submitters to improve their processes. If a State has information on the timeliness of all the steps in its crash reporting process, it can make informed decisions as to how best to use its resources to improve timeliness.

Measure 4: Timeliness of Crash Location Coding by the State DOT

Instead of measuring the data collection process in isolation (like Measure 3), Measure 4 measures the time it takes for personnel and automation to complete the location coding process (Step 10 in Figure 2.1), isolated from all other crash report data management steps. States often have two paths to completion of location coding: an automated path using software to assign location codes and a manual path requiring human intervention to arrive at a valid location code assignment. To determine when to invest in improved automated methods of location coding, a State should compare how much time the automated process takes to how much time the manual process takes. The State might also want to know what percentages of records are processed through automated means versus the percentage processed manually. These performance measures are related to companion performance measures of location coding accuracy (see Chapter 3).

States would do well to consider both timeliness and accuracy together in a coordinated manner so that both are improved.

The crash location timeliness performance measure compares the following:

- The date on which the crash record is received into the location coding process (the start of Step 10 in Figure 2.1); and
- The date on which the location coding process is completed (the end of Step 10).

By tracking whether or not a crash report is subjected to manual crash location coding, the State can easily calculate two mean timeliness scores, and an overall mean timeliness for location coding. For this example, a data table was first produced calculating the number of days required for automated location coding. The mean for 2014 crashes was .26 days (the equivalent of 6 hours and 14 minutes), a figure that is based on successful automated coding of 232,803 crash reports. In addition, there were 56,116 reports that required manual intervention in order to assign a valid location code. The manual process lasted, on average, 9.39 days. The average time required for location coding processing in 2014 was 2.03 days per crash report as shown:

$$\frac{[(232{,}803) * .26) + (56{,}116 * 9.39)]}{288{,}919} = 2.03\ days$$

The State can use these aggregate timeliness numbers to demonstrate the impact of improvements to location data collected on crash reports. If the data supplied by law enforcement are sufficient to allow the automated location coding process to succeed, then more crashes will be coded within 6 hours. In comparison, manual location coding (when required) added an average of more than 9 days of delay. Manual processing raised the average delay for location coding from about 6 hours to more than 2 days, even though it applied to only one out of every 5 crash reports.

The State may also wish to examine the reasons for delays in the manual and automated location coding processes. Table 2.5 shows the timeliness of manual location coding for each LEA. In this example, the State listed the average location coding timeliness for each LEA's crash reports in 2014. As expected, there is a great deal of variability among LEAs in the time it takes to manually code locations. This is due to differences in location data quality (completeness and accuracy). The impact of these other quality attributes shows up readily in the performance measure of timeliness for each agency.

The State can use this information to identify agencies whose crash reports are the most difficult to manually process, meaning that they frequently are missing information or contain errors, which the staff must then correct in order to assign a valid location code.

Table 2.5. Example Data Showing Timeliness of Manual Location Coding for Reports Submitted by LEAs

Agency	Total Crashes	Averaging Locating Days
AA	50	2.8
AB	18	6.46
AC	1	1.23
AD	175	2.74
AE	280	2.57
AF	23	4.76
AG	5	3.84
AH	2	12.23
AI	100	17.58
…		
AR	210	15.17
AS	165	11.75
AT	112	13.58
AU	8	14.16
AV	20	11.5
AW	1	0.24
AX	14	18.52
AY	2	5.71
AZ	3	1.24
Statewide Crashes Requiring Location Manual Intervention	**56,116**	**9.39**

Measure 5: Timeliness of Error Correction by LEAs

This performance measure is also similar to Measures 3 and 4 in calculating the mean time between the start and end dates of a process purely under the control of the LEAs. In this performance measurement, the start date is the date on which a crash report is rejected and returned to the LEA for correction. The start date is the date the report is rejected as part of the QC process and returned to the LEA (Step 9 in

Figure 2.1). The end date is the date on which the *corrected* report is received by the centralized crash data management office (Step 5 or 8 in Figure 2.1 for paper or electronic report corrections). As with Measure 3, it is useful to calculate the overall average as well as the average number of days for each LEA. This provides another piece of useful feedback to the data collectors.

Table 2.6. Example Data Showing Law Enforcement Timeliness in Correcting and Resubmitting Crash Reports

Agency #	Total Rejected Crashes	Average Return Days
A-1	15	26.2
B-2	9	18.8
C-3	40	12.2
D-4	104	26.2
E-5	12	42.3
F-6	5	24.2
G-7	3	8.6
H-8	6	12.2
…		
I-9	26	14.9
J-10	6	12.7
K-11	35	9.8
L-12	8	14.8
M-13	5	16.1
N-14	52	18.3
Statewide Rejected Crash Report Timeliness	**527**	**20.73**

Table 2.6 shows the average number of days from report rejection to report re-submission by LEAs. In this example, there were 527 reports rejected through QC processes by the crash data processing center. LEAs took on average 20.73 days to correct and resubmit those reports. The table shows that some LEAs had a much larger number of rejected reports than others—one agency accounted for almost one-fifth of the rejected reports. Thus, this report can also be used by the State to identify agencies that have data accuracy or completeness problems as well as to show the impact of erroneous and incomplete reports on overall timeliness. For these 527 reports, an extra 21 days were required before the reports were available for analysis.

Measure 6: Timeliness of an Annual Official Crash Database Release for Analytic Users

Some States create an annual master file (also known as a "close out file") which serves as the official source of crash data for the year. If crash reports come in later than the close out file's creation, the State may keep them in the database for location-based analyses, but not include them in the official tallies in annual reports. In all States—even those that do *not* create an annual close out file—users typically must wait a period of time before they can be sure that the data for the past calendar year are *reasonably* complete and thus reliable for use in analyses to support decision-making. The minimum number of days from the end of a calendar year to the date on which the crash data are complete and ready for use is set by the number of days allowed for post-crash fatalities to count in the FARS (i.e., 30 days post-crash). In practice, the minimum is much higher—often several months—because LEAs may require additional time to complete investigations. And, the State may need to complete a final set of data quality management procedures before the crash database is ready for use in annual reporting and data analyses. From the users' perspective, a useful timeliness performance measure is the number of days after the end of the calendar year that the database is ready for use. This is calculated as the difference between the end date which is the release date of the file to users and the start date of December 31 of the calendar year included in the database. It is usually reported in months, rather than days. For example, a crash database that is released on May 31, 2015, covering the 2014 calendar year is said to be 5 months old. A State may wish to track this performance measurement annually over several years as a way to demonstrate the overall improvement in timeliness in a way that matters most to users—how soon can they access the data and make use of it?

In Figure 2.3 each bar represents the statewide average number of reporting days from the end of the indicated calendar year (on December 31) to the date of the release (to users) of the annual close-out official crash database. The line shows a 3-year rolling average of the release dates indicating a clear downward trend and thus a timeliness improvement.

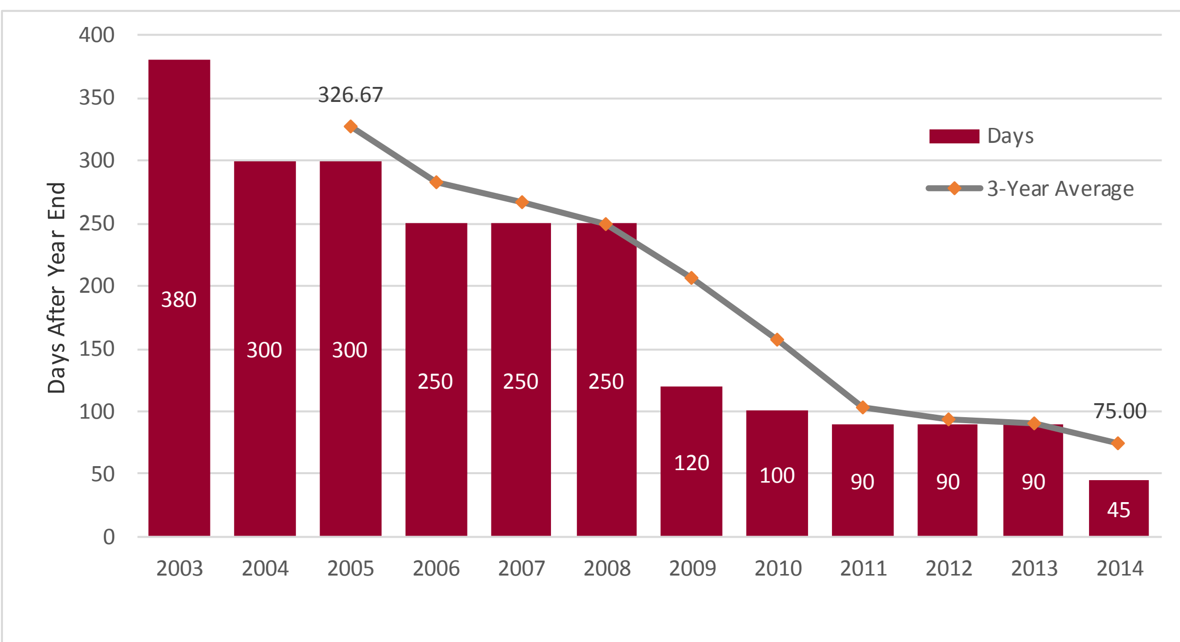

Figure 2.3. Example Showing Timeliness of the Release of an Official Annual Crash Database

States may wish to make interim releases of the data available to users throughout the year. In this case, the State can track the timeliness of those interim releases in the same way—by calculating the difference between the release date and the date of the newest records in the release. For example, if a State released an interim data file covering through March 31, 2015, on May 27, 2015, the overall timeliness is 57 days, or just under 2 months.

2.6 Final Considerations

There are many other possible timeliness performance measures. As noted earlier in this chapter, a State may wish to create timeliness performance measures for each step in the crash data management process shown in Figure 2.1 as a way to identify and remedy bottlenecks. A State may also wish to create multiple views of the same information, as is recommended in Measures 1 and 2—the same individual crash report timeliness is reported as the median, the mean, *and* as a percentage meeting the established timeliness metric.

These timeliness performance measurement calculations can be automated—as recommended in the *Traffic Records Program Assessment Advisory* (NHTSA, 2012) so that the cost of reporting timeliness remains low. Automated tools also help by providing flexible reporting across time and jurisdictions. For example, the crash database managers may want to receive daily reports of timeliness so they can tell immediately if there is a problem they need to investigate. Data collectors may want a report that presents just their individual agency's performance or presents a comparison to other agencies of a similar size or in nearby jurisdictions. Reducing the staff time required to generate a variety of reports can help the State manage crash data timeliness. A good opportunity to implement new, automated timeliness reports is when the State is updating its crash reporting system. Data quality reporting functions can be added to the system design from the start rather than attempting to add that functionality to an existing system.

NHTSA also encourages State TRCCs to periodically review crash data timeliness performance measures and metrics. The TRCC can serve in an advisory capacity as well as help prioritize funding for projects aimed at improving crash data timeliness. States transitioning from paper to electronic submission of crash reports, for example, may find it useful to present the TRCC with data on how the two processes (manual and electronic) differ in terms of crash data timeliness and other data quality performance measurements.

NHTSA also encourages States to incorporate performance measurement into a formal, comprehensive data quality management program, as described in the *Traffic Records Program Assessment Advisory* (NHTSA, 2012).

2.7 Questionnaire

The following questions are intended to identify how the State measures the timeliness of the crash data. This section will evaluate the PCR, how the State enters data into the crash data system, and how the State assesses the measurement.

67. What output measures of timeliness are currently reported for your State? (AQ-66) (SOE: Provide baseline current data and trend if available)

68. 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.)

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

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

71. 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.)

72. Does the State produce an official annual file for use in analyses and data extract releases? If so, does it track the timeliness of the availability of that file in a performance measure? (SOE: provide the overall data file release timeliness performance measure, preferably with multiple years to show the trend).

Chapter 3: Accuracy

3.1 The Purpose of Measuring Accuracy

Garbage in, garbage out. Capturing accurate crash data is critical for decision-makers who rely on the data to assess safety and make policy and programmatic decisions. This includes being able to properly describe contributing circumstances and locations, identify appropriate treatments, and evaluate the effects of the safety treatments previously implemented. Accuracy in crash data systems means that the information recorded about the crash reflects the "ground truth" of what happened, when, where, how, and to whom. When aggregated in a crash data system, the data provide an accurate summary of the crash experience for a range of subjects: a State, a political or geographic jurisdiction, specific portions of a population, times of year, and more. Accurate crash data also support decision-makers to conduct comparisons and confidently allocate resources to the target people, places, and circumstances that contribute to the State's crash experience.

Measuring crash accuracy allows data managers to:

- Identify accuracy issues among data collectors (LEAs) and provide them with meaningful feedback;
- Provide specific examples for crash reporting training, manuals, forms, and software;
- Highlight key crash data elements requiring improvement;
- Identify specific data validation checks that can be implemented to detect the most common errors before they impact system level data quality; and
- Communicate information to decision-makers about the validity of the crash data.

3.2 Defining Accuracy

According to the NHTSA (2011) *Model Performance Measures for State Traffic Records Systems:*

> Accuracy reflects the number of errors in information in the records entered into a database. *Error* means the recorded value for some data element of interest is incorrect. Error does not mean the information is missing from the record.

In a crash data system, data managers can define and validate accuracy in two primary ways: externally and internally. These are described in the following sections.

External Validation

External validation is when a data element's attributes recorded on a crash report or in a crash system database are compared to a reliable outside source result in either an "accurate" or "inaccurate" judgement. The following examples help to describe external validation for crash reporting:

- **Location Validation:** When assigning a crash to a specific location, the question is whether or not that recorded location is the true location of the crash. For example, if the officer recorded that a crash happened on Main Street, 500 feet north of First Street, the information is accurate only if the crash actually happened at that location. To verify the location, most States attempt to "locate" crashes onto the roadway network using an LRS—a sequence of defined locations along a roadway. In an ideal system, a statewide geographic information system (GIS) maps crashes using dots to indicate crash locations on the roadway network (Figure 3.1).

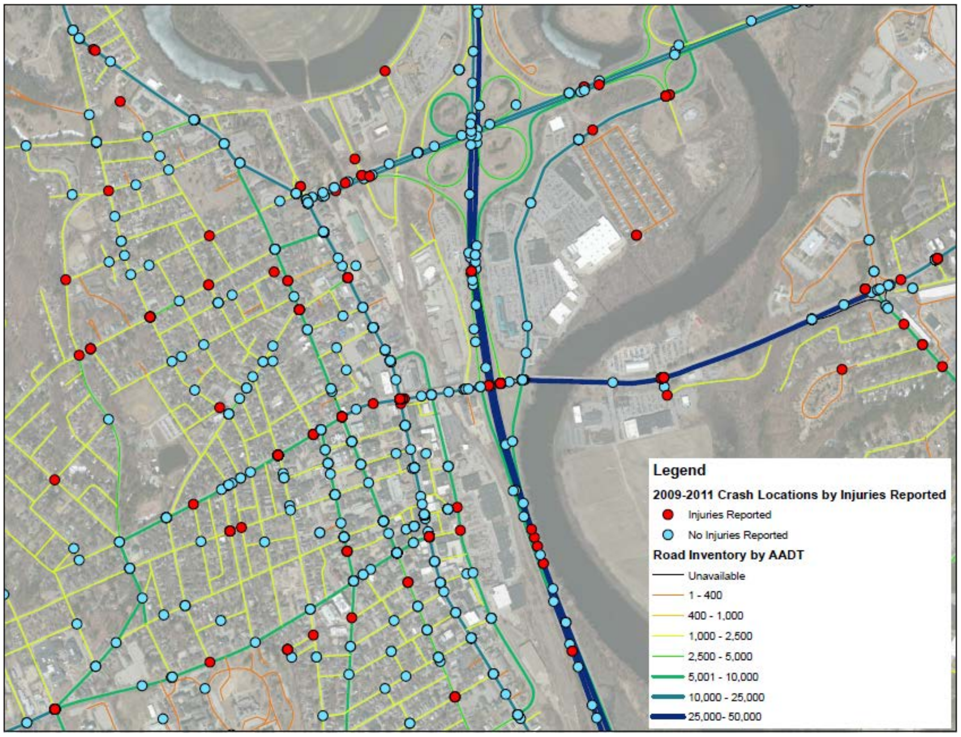

Figure 3.1. GIS Display of Crash and Roadway Information

In the example of a crash north of First Street, additional information may be presented in the officer's narrative description and diagram. For example, the narrative and diagram may show the crash at the entrance to a shopping center. With detailed records describing the roadway (or by going to the site itself), analysts may check to see if the entrance indicated is actually 500 feet from the intersection or some other distance away. Accurately placing crashes on the roadway is crucial for State and local engineers who must analyze locations and determine what treatments are warranted, if any.

- **EMS/Injury Surveillance Validation:** State crash reporting requires law enforcement officers to make judgements on the severity of injuries to any of the people involved. For persons transported by EMS and those treated in a hospital, more reliable injury data are available from the EMS, emergency department, trauma registry, and hospital in-patient databases. States may link the crash report's person records to these external sources of injury data in order to measure the accuracy of police officers' judgements. Typically, this type of linkage is done annually and may lag a year or more behind the availability of the crash data, so the performance measure may be less timely than other external validation measures.
- **Citation/Adjudication Validation:** State crash reports may contain information on citations issued to crash participants. This information is sometimes used in analyses to determine which participants were at least partially responsible for the crash. In the event of driving under the influence or other serious charges, the officer is recoding information about the suspicion of alcohol or drug involvement that may turn out to be incorrect once test results are obtained. Citation tracking, DUI tracking, and court records systems can serve as a source of external validation for violations and notations of alcohol or drug involvement. The citation tracking system should include initial and final charges, as well as the ultimate disposition of the cases related to the crash. A DUI tracking system would include test results and other information. The

court records system should include the charges and final dispositions as well. Analysts can use these external sources to verify the accuracy of the officers' at-the-scene judgments, which can, in turn, inform decision-makers about the suitability of those data elements on the crash report for use in analyses of traffic violations associated with crashes.

- **VIN Validation:** A State's crash report contains a data element recording the make of the involved vehicles. The crash report also includes space to record the license plate number and VIN. The VIN is a number assigned by the manufacturer and identifies several attributes of the vehicle, including the make, model, year, trim level, and a unique sequence number for that individual vehicle. An *external* validation using VIN and license plate information can check the crash data's accuracy against the State's vehicle registration data. States may also use external VIN decoding tools, such as NHTSA's vPIC (www.vpic.nhtsa.dot.gov/) to check the validity of the VIN recorded on a crash report, and to obtain year, make, and model information.

Internal Validation

Internal validation is when the values of data elements in the crash report are compared against a list of allowed values and checks for logical agreement among two or more data elements are conducted. The following examples help to describe internal validation for crash reporting.

- **Range Validation:** Several data elements in a typical crash report have an expected range of values. For some, the range of values is defined by the attributes from which the officer selects when filling out a field. For example, in a State's crash report form, the data element for "Light Condition" can take the eight values shown in Table 3.1 as derived from the MMUCC 4th Edition. Any data other than those eight allowed values would be an error.

Table 3.1. MMUCC Attributes for the MMUCC Light Condition Data Element

Assigned Code	Attribute Label
01	Daylight
02	Dawn
03	Dusk
04	Dark-lighted
05	Dark-not lighted
06	Dark-unknown lighting
07	Other
08	Unknown

- **Logical Agreement Validation:** Crash report data elements are often logically related so that the value of one data element limits the allowable valid responses on a second data element. There are a large number of these relationships built into every State crash report form and process. Examples include logical agreement between the time of day and lighting conditions; number of crash-involved vehicles and manner of collision data elements; and driver actions and manner of collision data elements.

3.3 Assessing Accuracy

General Guidance on Accuracy Measures and Reporting

NHTSA strongly encourages States to develop data quality performance measures. Whenever possible, these should be built into the crash data management software as automated reporting functions. States can automate the validation process in several ways. The simplest method uses validation rules built into data collection software, data transfer protocols, and data intake processes. Such automated validation rules are often called *edit checks* because they are used as part of the process of checking and editing the data prior to acceptance. If the errors are caught and corrected by law enforcement during data collection, the State typically will not have a record of the initial error. Only the corrected data is received by the central crash data management system. States should, however, keep track of all errors that are discovered in crash reports *after* they are submitted by law enforcement. The following are examples of potential edit checks.

- **Edit check for missing or invalid codes**: Display an error when Light Condition is missing or invalid (see Table 3.).

    a. If value of LIGHT_COND = display "Light Condition missing."

    b. If value of LIGHT_COND < 1 or > 8 display "Light Condition invalid."

  These two statements check for missing (empty) light condition data fields and any values outside the allowed numeric codes of 1 through 8.

- **Edit check for logical agreement**: Display an error when Crash Time and Light Condition do not logically agree.

    a. If LIGHT_COND = 1, 2 and CRASH_TIME > 22:00 or < 04:00 display "Time of Day and Lighting Condition do not match."

    b. If LIGHT_COND = 3, 4, 5, 6 and CRASH_TIME > 10:00 or < 16:00 display "Time of Day and Lighting Condition do not match."

These two statements check for agreement between lighting condition and time of day. In the first statement, the time is checked to see if crash occurred later than 10 p.m. or earlier than 4 a.m. while the lighting condition is coded as either dawn or daylight (see Table 3.1). In the second statement, the time is checked to see if the crash occurred later than 10 a.m. or earlier than 4 p.m. while one of the lighting condition codes for "dark" is selected (see Table 3.1).

A State's crash data management software should have the capability to put each crash report record through a comprehensive set of validation edit checks, such as the range and logical agreement checks shown in the examples. Chapter 8 of this document shows edit check examples for the MMUCC data elements. Crash systems should include dozens or even hundreds of such edit checks. The State should run these checks as part of the data quality control process even if those same checks are run by law enforcement prior to submission. This process should be as automated as possible, and should result in a report that shows, for each crash report, the errors that it contains. According to the *Traffic Records Program Assessment Advisory* (NHTSA, 2012), States have the option of correcting the errors centrally, rejecting the reports so that the LEA makes the corrections, or leaving the errors uncorrected in limited circumstances. States typically define two levels of errors:

- **Critical errors** are those that must be corrected before the crash report is accepted into the system. These typically result in report rejection either back to the submitting LEA or to a queue to be corrected by the State data management staff.

- **Non-critical errors** are those where the coded value or blank may be allowed under specific circumstances. In these situations, the State may program the edit check to issue a warning rather than reject the report.

Regardless of the process, the assessment advisory recommends that the errors should be tracked and the information used to:

- Notify LEAs of errors in their reports;
- Develop content describing common and serious errors in crash reporting training and instruction manuals; and
- Assess the need for additional field data collection edit checks so that the errors are trapped prior to law enforcement's crash report submission.

Accuracy performance measure reports should be available to a variety of stakeholders. The crash system data managers should have frequent updates on the accuracy of data as submitted. Data collectors in LEAs should receive periodic feedback listing the error types, the data elements involved, and the frequency of each error. Data users and the TRCC may need only annual information on accuracy. This would allow them to determine if the crash data are reliable for use in decision-making, to compare performance to established accuracy metrics, and to determine if projects aimed at improving accuracy are achieving their goals.

Guidance on Situations Requiring Detailed Measurement and Reporting

This chapter presents several crash data accuracy performance measures. States are advised to use as many of these measures as applicable. States are also advised to select the most important data elements (key fields) when reporting accuracy measures so that the focus of oversight and improvement efforts remains on those critical data elements most needed by decision-makers. Field data collection systems—especially if the system includes a set of validation rules used as edit checks prior to report submission—should positively impact accuracy. The following situations demonstrate when specific measures may apply.

- **Paper reporting by law enforcement.** For States that still accept hard copy paper reports, automated validation does not happen until the data entry process or until the centralized crash records system accepts the data. It is expected that paper-based reports contain more errors than those completed with electronic field data collection software. This is because the data collection software products typically include pick lists and edit checks that constrain the possible entries into the crash report. Comparing the errors seen in paper reports to those in electronic reports can help to demonstrate the accuracy improvements possible through use of field data collection software.
- **All law enforcement officers enter data directly into the State crash database.** This is most often applicable in situations where the State has established an online, web-based crash reporting tool used by all law enforcement officers. In this situation, the online system can include validation rules and reject any report that does not pass all of the *critical* edit checks. The software may also issue warnings for non-critical errors (violation of edit checks in non-critical data fields). In such online systems, the State should track initial errors even though they are corrected by the law enforcement officer as part of the final submission. The State can use that information to improve training, instruction manuals, and online help.
- **Ongoing accuracy validation.** It is not uncommon for users to discover errors in crash reports many months or years after the data was accepted into the statewide crash database. A user may discover that a crash from 3 years ago was located incorrectly and the location code must be

corrected. A different user may discover that several large trucks were really smaller utility vehicles miscoded on the forms. Error correction can happen at any time. In States that create a final official database for each year of crash data, 'after the close-out' corrections may not affect the official annual file. However, the State may wish to retain the original report as well as any further revisions, including those that come in after any annual file close-out. The CDIP presentation makes reference to a "change log" built into the crash data management software. Figure 3.2 shows an example of a change log for a single record in a State's traffic crash reporting system.

TRAFFIC CRASH DATA AUDIT REPORT

Crash Source = I

Ext Crash ID = 0114621

ORI = 1972500

Crash ID: 9060636

Section	**Field Name**	**Old Value**	**New Value**	**Changed**	**Changed By**
Involved Party	PRTY_KEY_7231671_PRTY_ADDR_ZIP	48820	488209140	10/14/2014	System
Involved Party	PRTY_KEY_7231672_PRTY_ADDR_ZIP	48879	488798001	10/14/2014	System
Location	CITY_TWSP_CD	60	3	10/15/2014	User
Location	INTR_STRET_NAME	BUS27	WHITMORE	10/15/2014	User
Location	INTR_TYPE	HWY	ST	10/15/2014	User
Location	LOC_ERR_CD	4	0	10/15/2014	User
Location	LOC_FLAG_CD	2	1	10/15/2014	User
Location	ORIG_DRTN_CD	SE	E	10/15/2014	User
Location	ORIG_DTNC	10	9	10/15/2014	User
Location	PRMY_PFX	W	E	10/15/2014	User

**Modifications to data may be the result of data validation errors or data correction provided by users or system batch processes.

Records 1 to 10 of 10

Figure 3.2. Example Change Log for a Single Crash Record

This example shows the changes to crash report number 9060636. It identifies the LEA, each section and data element, the old value, and the new value after the change. The date and originator of the change are also identified.

The change log is used to collect information on initial (erroneous) values in a data field, the corrected value, the date the correction was made, and the source of the correction. Storing this information is inexpensive and provides the State the ability to return to the earlier versions of a report if necessary. Some States, for example, *must* retain an official record of the report as submitted by law enforcement—this is often true if the report may be admitted in court and the officer is called upon to swear to its contents.

If the original report contained errors, however, the State should also keep a corrected version that is valid and reliable for data analysis. An annual closeout file would reflect all corrections up to the date that the file is considered closed. A change log would continue to collect error corrections so long as the crash is available for analysis. It is a good practice for the State to keep an error log that allows data managers to reproduce the report in any form—from original to most recent corrections and any status in between. Error logs also allow users to select which types of corrections they wish to ignore. There are times, for example, when different users disagree about the appropriate correction to a problem identified in a crash report. Capturing the corrections in an error log allows users to specify which corrections apply to the dataset they wish to analyze.

- **External validation through linkage.** External validation is more difficult to automate than internal validation because it requires checking values of fields on a crash report against data contained in a different database. States can link crash records with driver licensing, vehicle registration, citation, EMS, hospital, and roadway databases. When these links are real time, it is possible to use the linked data to validate the crash report information. For example, if the officer records the name, date-of-birth, address, and driver's license of a person in a crash report, linkage with the driver license database can flag any items that are not in agreement between the two sources.

  Linkage can be helpful in supplying corrections to both the crash data and the official source (e.g., license, registration). With automated field data collection systems, it is possible to automatically complete some of the fields in the crash report through linkage. However, it is still necessary for the officer to check that the automatically supplied information is correct. Similarly, it is possible to improve accuracy of crash location information through linkage to the statewide GIS using a map-based interface. The officer could then click on the map to indicate where the crash occurred and the correct location code and relevant location descriptive information will be automatically added to the crash report.

  Even when the linkage is *not* accomplished in real time, it still has value as a periodic validation step (e.g., monthly, quarterly, annually). Ideally, with periodic validation, States can check any variance between the crash report information and official source databases. When differences are detected, the State must determine whether or not to change the data in the crash records system and may alert data managers of the source files in case the source file needs updating.

3.4 Developing Accuracy Performance Measurements

The goal of accuracy measurement is to provide useful information for data collectors, managers, and users. In developing accuracy performance measures, system managers should prioritize what is most important to State stakeholders and to themselves, as managers of the system. While the goal is to identify and correct errors in the crash report, selecting the most critical data elements when assessing overall accuracy will reduce the amount of data to process for each report and increase the utility of the accuracy report.

The *Model Performance Measures for State Traffic Records Systems* (NHTSA, 2011) report defines two measures of crash data accuracy. The first relies on a pre-defined set of critical data elements and tests accuracy using internal validation. The second relies on external validation using the vehicle registration database. As stated in that report:

> **C-A-1: The percentage of crash records with no errors in *critical* data elements**. (p. 11) The State selects one or more crash data elements it considers critical and assesses the accuracy of that

element or elements in all of the crash records entered into the database within a period defined by the State.

The NHTSA report (2011) includes a list of suggested critical data elements. These are shown in Figure 3.3. The State should determine which data elements are critical by consulting with data collectors, managers, and users.

C-A-2: The *percentage* of in-State registered vehicles on the State crash file with VIN matched to the State registration file. (p. 11)

In practice, States need many more measures of accuracy than the minimum suggested in the NHTSA report (2011). C-A-1 and C-A-2 are the measures that every State should be able to calculate and share. Other measures of accuracy should be developed based on a State's needs and particular situation. If both paper and electronic reports are collected, the State may wish to track accuracy separately for the two source types. Additionally, data managers may wish to track *all* errors (not just those in critical data fields) and provide specific feedback to each LEA. Location coding failures may generate several related accuracy measures, which are important to those who manage the process and provide training and feedback. Data managers should consult with data collectors and users to determine which accuracy measures are most meaningful and relevant. Ideally, the crash data management system will calculate each measure automatically. Where automation is not possible, the State should plan to consider adding this capability in the next crash data system update.

<table>
<tr><th colspan="2">Critical Crash Data Elements Used By Many States</th></tr>
<tr><td>Environment
Record #
Location (on/at/distance from, lat/long, location code
Date, time (can calculate day of week from this, too)
Environment contributing factors (up to 3)
Location description (roadway type, location type, roadway-contributing factors—up to 3)
Crash type, severity, # involved units
Harmful events (first harmful, most harmful)</td><td>Person
Crash record #, vehicle/unit #, person #
Person type (driver, occupant, non-occupant)
Demographics (age, sex, other)
Seating position
Protective device type (occupant protection, helmet, etc.)
Protective device use
Airbag (presence, deployment: front, side, both, none)
Injury severity (if this can be sourced through EMS/trauma/hospital records
Transported to __________
Transported by __________
EMS Personal Casualty Report #</td></tr>
<tr><td>Vehicle/Unit
Crash record #, vehicle/unit #
Vin decoded sub-file of values for make, model, year, other decode values
Sequence of events (multiple codes)
Harmful events (first and most harmful for each vehicle)
SafetyNet variables for reportable vehicles/crashes, (carrier name/ID, additional vehicle codes, tow away due to damage)
Vehicle contributing factors</td><td>Administrative Tracking Variables
Agency record number (if different than crash report number)
Report completion date
Report submission date
Report accepted date
Report returned to agency for edits
Report returned corrected date
Initial errors (count by level of severity: # critical errors; # non-critical errors)
Final quality rating</td></tr>
<tr><td>Driver/Pedestrian/Pedalcyclist
Crash record #, vehicle/unit #, person #
Personal identifiers (name, driver license #, address, other)
Person type (driver, pedestrian, pedalcyclist)
License (type, endorsements & restrictions, compliance with endorsements/restrictions)
Driver maneuvers
Driver contributing factors (condition, distraction, etc.)</td><td></td></tr>
</table>

Figure 3.3. Critical Data Elements From NHTSA

3.5 Examples of Accuracy Performance Measurements

The two crash data accuracy measures defined in the *Model Performance Measures for State Traffic Records Systems* report (NHTSA, 2011) are calculated as follows:

Measure 1 (C-A-1): Percentage With No Errors

A. **Define a list of *critical* data elements.** The State should decide on a list of data elements that are most important to collectors, managers, and users. The State should set a high accuracy goal for those data elements.

B. **Define what constitutes an error for each *critical* data element.** Ideally, the critical data elements will be subjected to automated validation rules. Most likely these will be internal validations of values in a single field. However, the State may also wish to include errors that are conditional based on the values in two or more data elements simultaneously, such as when checking agreement between time of day and lighting condition. If *external* validation is included as part of this set of critical data elements, there is a possibility that the measure may be delayed in order to wait for up-to-date information in the external source file.

C. **Identify *all* errors in *any* of the critical data elements.** Each critical data element should be checked for errors in every crash report. To create the appropriate tally for measure C-A-1, flag error-containing reports once no matter how many errors the report contains in critical data elements. For example, a report with no errors is flagged with a 0; reports containing any number of errors are flagged with a 1.

D. **Tally total number of reports and total number of reports containing an error.** Do this for each time period of interest (e.g., month, quarter, year).

E. **Calculate the percentage of error-free reports.** Divide the number of *error-free* reports by total reports. Then, multiply by 100 to get the percentage of reports containing *no error* in critical fields.

Example data for this calculation is shown in Table 3.2.

Table 3.2. Example Data for Accuracy Calculation C-A-1

Report #	Reports With ≥ 1 Error	Reports With ZERO Errors
1	0	1
2	0	1
3	0	1
4	1	0
5	0	1
6	1	0
7	0	1
8	1	0
9	0	1
10	1	0
TOTALS 10 reports	**4**	**6**

Table 3.3 shows a total of 10 crashes, 4 of which contained *at least* 1 error in a critical field and 6 with no errors. The value of C-A-1 is 6/10 * 100% = **60 percent of reports contained no errors in a critical data element.**

Measure 2 (C-A-2): VIN

A. **Identify all VINs for vehicles registered in-State.** Because there can be multiple vehicles recorded in each crash, make sure to collect *all* of the VINs. If a VIN data element is left blank, that should count as an error unless the vehicle was marked as driver fleeing the scene (hit-and-run) or there are other valid reasons for the data to be missing.

B. **Check VINs against the State motor vehicle registration file.** If the VIN from the crash report matches the VIN recorded in the registration file, flag the record as correct (flag = 1); otherwise, flag the record as an error (flag = 0).

C. **Calculate the percentage match.** Divide the total number of matching VINs by the total number of in-State VINs. Example data for this calculation is shown in Table 3.3.

Table 3.3. Example Data for Calculating Percentage of Matching VINs

Report #	Vehicle # in Crash	No Match = 0 Match = 1
1	1	1
1	2	1
2	1	1
3	1	0
3	2	1
4	1	0
4	2	1
4	3	0
5	1	1
5	2	0
TOTALS **5 reports**	10 vehicles	6 matches

Table 3.3 shows that there were records for 10 in-State registered vehicles in 5 crash reports. Of the 10 vehicles, 6 had a VIN that matched the in-State vehicle registration database record.

The percentage of VINs matching the motor vehicle registration data is: 6/10 * 100 percent = **60 percent matching.**

The remainder of this section presents examples of crash data accuracy performance measures drawn from States' experiences. They are recommended for consideration whenever a State needs detailed measures of crash data accuracy.

Measure 3: Average Number of Errors per Report

The State sample shown in Figure 3.4 compares the annual average number of errors per crash report from 2004 to 2014. The figure shows that over the 11 years, this State achieved a steady reduction in errors, dropping from an average of 2.6 errors per report to 0.1 errors per report (or the equivalent of 1 error in every 10 reports).

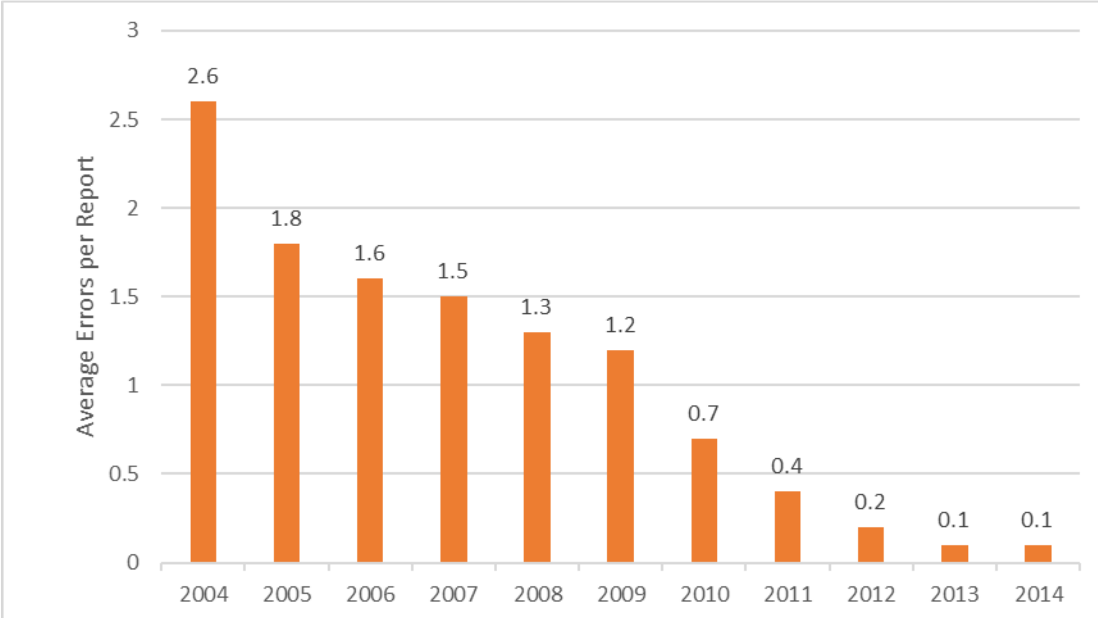

Figure 3.4: Average Number of Errors per Crash Report 2004 – 2014

During that time, the State transitioned from paper to electronic crash reporting, which likely contributed to their decline in errors. This stark contrast can be seen by comparing Table 3.4 and Table 3.5, which show the 2014 report errors for paper versus electronic reports. Table 3.6 shows errors per paper report submitted by each LEA. Paper reports had on average 1.06 errors per report in 2014. There were a total of 14,659 paper reports submitted that year.

Table 3.4. Number of Errors per Paper Form Report in 2014

Agency	Total Crashes	Data Errors	Average Errors Per Crash
A	888	525	0.59
B	129	369	2.86
C	47	89	1.89
D	4	6	1.50
E	74	198	2.68
F	114	189	1.66
G	719	531	0.74
H	38	33	0.87
I	2	0	0.00
…			
R	126	193	1.53
S	72	159	2.21
T	305	751	2.46
U	561	540	0.96
V	1	5	5.00
W	564	579	1.03
X	4	4	1.00
Y	96	130	1.35
Z	8	20	2.50
Statewide Totals	**14,659**	**15,583**	**1.06**

Table 3.4 shows the same information for electronic crash reports submitted in 2014. These are reports collected using field data collection software and submitted to the statewide crash reporting system electronically. There were 291,815 electronic reports submitted in 2014, with an average of .01 errors per report. This is because electronic reports are subjected to validation at the LEA level, before submission, using edit checks built into the field data collection software.

Table 3.5. Number of Errors per Electronically Submitted Report in 2014

Agency	Total Crashes	Data Errors	Average Errors Per Crash
A	498	0	0.00
B	82	0	0.00
C	53	0	0.00
D	1,247	14	0.01
E	926	10	0.01
F	1	0	0.00
G	73	0	0.00
H	549	0	0.00
…			
U	764	11	0.01
V	1,232	6	0.00
W	270	1	0.00
X	71	0	0.00
Y	7	0	0.00
Z	6	0	0.00
Statewide Totals	**291,815**	**3,678**	**0.01**

The overall accuracy improvement seen in Figure 3.4 is due to the growing use of electronic data collection and improvements in the validation rules standards applied at the State and local levels.

Measure 4: Percentage of VINs Decoded

This accuracy measure is related to Measure 2 (C-A-2 from NHTSA, 2011), which addresses the ability to *match* between the VIN recorded in the crash report and VINs contained within a State's vehicle registration system. Accuracy is calculated as the number of VINs successfully decoded to produce a valid year, make, and model. This requires the use of VIN decoding software and an automated validation check that compares all decoded VIN values against the vehicle attributes reported in the crash reports.

Table 3.6 shows a State's VIN decoding success rates for each LEA in 2014. The performance measure is based on all vehicles recorded in crash reports (total vehicles) as the denominator and successful decodes as the numerator * 100 percent.

Table 3.6. Percent of Decodable VINs in Crash Reports – 2014

Agency	Total Crashes	Total Vehicles	Successfully Decoded	Percent Successful
A	498	523	518	99.04%
B	82	67	66	98.51%
C	53	79	79	100.00%
D	1,247	1,542	1,510	97.92%
E	888	2,225	2,124	95.46%
F	129	220	210	95.45%
G	47	65	50	76.92%
…				
V	4	5	4	80.00%
W	96	123	113	91.87%
X	7	10	9	90.00%
Y	6	8	8	100.00%
Z	8	8	7	87.50%
Statewide Totals	**306,435**	**486,914**	**464,359**	**95.37%**

Measure 5: Location Coding Accuracy

When analysts map the locations of crashes, they typically use a combination of automated and manual processes to translate the location descriptions from the crash report into a defined location in the State's LRS. There are multiple ways for these processes to fail, some of which indicate errors in the location information provided by law enforcement. Table 3.7 shows the outcomes of location coding efforts for an example State. The data are provided for individual LEAs so that the crash data managers can provide feedback to the data collectors.

Table 3.7. Percent of Crash Locations Successfully Mapped

Agency	Total Crashes	Automated Mapped	Manually Mapped	Total Mapped	Percent Successful
A	498	200	270	470	94.38%
B	82	80	2	82	100.00%
C	53	53	0	53	100.00%
D	1,247	1,018	201	1,219	95.75%
E	888	153	235	388	43.69%
F	129	0	120	120	93.02%
G	47	14	33	47	100.00%
…					
V	4	3	1	4	100.00%
W	96	90	5	95	98.96%
X	7	5	2	7	100.00%
Y	6	2	3	5	83.33%
Z	8	2	5	7	87.50%
State Totals	**306,435**	**197,337**	**84,661**	**281,948**	**92.01%**

The State can use information like that reported in Table 3.7 to identify LEAs that may be struggling to provide complete and accurate location information. In this example, Agency E's crashes map successfully only 43.69 percent of the time. This is well below the statewide average of 92 percent. Agency F has no reports that are mapped using the automated processes. This means that although the agency provides accurate location information (mapping succeeded over 93 percent of the time), the more costly manual processes are required for every crash. Agencies Y and Z are also below the statewide average; however, with so few crash reports completed in a year, the State may decide that there is little to gain overall if their accuracy improves. There are several examples of agencies that the State could review to identify best practices. Agencies B, C, and W all have very high mapping percentages *and* almost all of their crash location mapping is accomplished through the automated processes. This saves the State time and money, and demonstrates that a high level of accuracy is achievable.

3.6 Final Considerations

There are many other possible accuracy measures. As shown in the example measures, a State may wish to measure accuracy separately for paper and electronic submissions. This provides a convenient way to demonstrate the accuracy improvements available due to validations built into the field data collection software, and overall accuracy improvements as the percentage of reports submitted electronically increases over time.

As noted, NHTSA recommends that States develop a list of critical data elements in the crash record. That list is used when measuring accuracy based on the percentage of reports with no errors in a critical data element (Measure C-A-1 from NHTSA, 2011). This is a useful concept for other accuracy measures that a State might wish to calculate. For example, the average number of errors per report could focus on just the critical data elements (rather than *all* possible errors as was done in the example in this chapter). A list of the most common errors from among those critical data elements could also be useful in defining new validation rules, training, and revised content for the crash data collection manual.

The CDIP process includes recommendations aimed at providing specific feedback to the law enforcement data collectors. Periodic (monthly, quarterly, annual) reports that demonstrate LEAs' performance on data quality measures—like percentage of VINs that decode or average number of errors per report—can help those agencies focus on necessary improvements in their data management procedures. Reporting at the LEA level can also help the State determine which LEAs are having difficulties reaching statewide accuracy targets and then focus on those agencies for increased investment or assistance.

NHTSA also encourages States to incorporate performance measure into a formal, comprehensive data quality management program, as described in the *Traffic Records Program Assessment Advisory* (2012).

3.7 Questionnaire

The following questions are intended to identify how the State measures the accuracy of the crash data. This section will evaluate how the State reports accuracy and the efforts to validate the data.

73. What output measures of accuracy are currently reported? (SOE: Provide baseline current data and trend if available, and specify frequency of issuance) (AQ-67)

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

75. 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)

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

77. 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)

78. 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.)

Chapter 4: Completeness

4.1 The Purpose of Measuring Completeness

Complete crash data is critical for decision-makers in many ways. Data gaps degrade our understanding of the places, people, vehicles, and contributing factors involved in crashes. Although a complete crash report or database has data in all fields, a more practical use for States is to understand their system's completeness. States can calculate their crash data system's completeness and work to improve those numbers. However, the relative completeness goal of a State system is one that is sufficient to support State needs through analysis, decision-making, and data integration.

Crash completeness measurement serves to:

- Identify system-wide and agency level gaps for critical data elements;
- Provide specific feedback to identified agencies;
- Identify opportunities to improve crash report training, manuals, forms and software;
- Identify and implement data validation checks for commonly omitted data to improve statewide system level data quality;
- Support efforts enhancing users' perception of data reliability for decision-making;
- Understand the limits on data integration efforts that result from missing data; and
- Prioritize the critical data elements on the crash report to improve completeness.

4.2 Defining Completeness

According to the NHTSA *Model Performance Measures for State Traffic Records Systems* (2011) report*:*

> Completeness reflects both the number of records that are missing from the database (e.g., events of interest that occurred but were not entered into the database) and the number of missing (blank) data elements in the records that are in a database. Completeness has internal and external aspects. In the crash database, external crash completeness reflects the number or percentage of crashes for which crash reports are entered into the database. It is impossible, however, to establish precisely external crash completeness as the number of unreported crashes cannot be determined. Internal completeness can be determined since it reflects the amount of specified information captured in each individual crash record.

In a crash data system, data managers can define and validate completeness in two primary ways: externally and internally. These are described in the following sections.

Internal Completeness

Internal crash completeness focuses on the data values present in individual crash reports. Internal completeness is measured based on the percentage of reports where data elements are left blank or coded *inappropriately* with a value of "missing," "unknown," or as "other" (without explanation) when the actual value should have been obtainable. Note that there is some obvious overlap between completeness and accuracy measurements as values of "unknown" or "other" (without required explanations) may be viewed as *both* inaccurate *and* incomplete.

With fully electronic reporting, States may wish to track completeness for each data element. Automating the data validations is simple and doing so can provide immediate feedback while an officer is completing a crash report. For completeness measurement purposes, States are encouraged to calculate measurements for a select list of high-priority data elements rather than examining every element in the crash database. Completeness performance measures should be calculated for individual LEAs and for the State to provide specific feedback for data collectors and reliability information for decision-makers.

Internal completeness is measured by comparing the number of times a data element is completed to the number of times it *should have been* completed, given the circumstances of the crashes. For example, every crash should have a complete and valid location. In fact, most of the crash data elements at the event level (time of day, day of week, location, weather, etc.) should be completed for every crash report.

Because some crashes involve hit-and-run drivers and "phantom" vehicles (i.e., vehicles reported as being involved in the crash but were no longer present when the officer arrived), completeness of vehicle and driver information must be handled differently. If a crash report includes a hit and run vehicle, data elements such as make, model, or plate number are expected to be blank. Similarly, the same hit and run vehicle would lack driver information (e.g., name, license number, gender, age, injury status). The report would also lack any information about the vehicle's other occupants (if any). None of these blanks should count against the measure of completeness.

External Completeness

External crash completeness is defined as the crash system having exactly as many crash records as there are reportable crashes. It is assessed based on an understanding of LEA practices. Most LEAs that respond to crashes prioritize fatal and injury crashes because of the serious nature of the crashes and their greater impact on safety and traffic flow. When resources are scarce, agencies are likely to continue to investigate fatal and injury crashes but use alternative methods to address less severe crashes. External completeness is defined in relative terms since property-damage-only (PDO) crashes, injury crashes, and fatal crashes tend to occur in proportion—there are more PDO crashes than injury crashes, and more injury crashes than fatal crashes. As the proportion of serious injury and fatal crashes increases, the data manager should check for under-reporting of PDO crashes.

Even though—as stated in *Model Performance Measures for State Traffic Records Systems* (NHTSA, 2011)—it is virtually impossible to precisely measure external completeness, it is possible to identify likely under-reporting at the individual LEA level. The measurements presented in this chapter rely on historical averages, changes in statewide reporting levels, and the proportion of serious crashes. When an LEA is under-reporting, its total number of reports submitted in a year will drop compared to historical averages. The reduction will be proportionally larger than the crash reduction experienced statewide. Additionally, the ratio of serious crashes to total crashes will be higher than previous years and higher than the same ratio calculated statewide or for similar LEAs. While none of these measurements gives a precise value for under-reporting, they do allow analysts to estimate the number of missing crash reports. The information can be shared with the LEAs as a way to encourage more complete reporting.

Measuring external completeness is especially important because it is the means by which systematic under-reporting is discovered. Comparisons to prior years' reporting levels will help to identify sudden drops such as those due to a change in an LEA's priorities. If an LEA has under-reported for a period of years, then comparison of its proportion of serious crashes to the statewide proportion will provide evidence of under-reporting minor injury and PDO crashes.

Random Versus Systematic Incompleteness

Random incompleteness refers to missing crash reports with no obvious pattern of where and when those crashes took place, or the vehicles and people involved. One example of random incompleteness occurs when crash events that do not meet the State's minimum reportable requirements are not entered into the database. The majority of States set a minimum threshold for reporting. This is typically a minimum dollar amount of damage, or a minimum threshold requiring that at least one vehicle was towed due to disabling damage. Any crash below the threshold is not to be included in the State's crash database. This is usually *not* counted as under-reporting or incompleteness because the decision was made to have a minimum cut off. Additionally, the below-threshold crashes are randomly distributed around the State, throughout the year, and are not expected to affect any group of vehicles or people more than any others. So, even though some crashes are missing, the State can still analyze its overall crash experience and reach valid conclusions. Random incompleteness is not an insurmountable problem for decision-makers because they know that they can use existing crash data to make valid decisions about where to concentrate resources in order to have the greatest possible impact on safety, even though *not every* crash is entered into the database.

In contrast, *systematic* incompleteness is a serious threat to valid decision-making. As previously noted, some agencies do not report all crashes that meet the State's threshold. If those agencies are under-reporting PDO crashes, this amounts to the agencies setting their own reporting threshold; one that is higher than the official reporting threshold set by the State. (See Chapter 5 on Uniformity for a further discussion.) When LEAs under-report crashes, the result is a specific, systematic gap in the crash data. This has a serious effect on the validity of any analyses using the data, including statewide analyses and any comparisons among geographic jurisdictions. As a consequence of these gaps, opportunities to improve safety in jurisdictions and population groups within the State are missed.

To demonstrate the impact of poor data completeness, consider the impact of a large urban police department choosing to report *only* serious injury and fatal crashes. For example, a large LEA reported a total of 100 crashes in a year. Dispatch records showed 100 calls *per day* for crash-related responses. While not all of those crashes would have required a crash report, it is clear that the department reported a fraction of the crashes that should have been reported. When LEAs decide to systematically under-report crashes, the total number of reportable crashes is artificially low for that area, county, and State. Because the data gap arose from systematic under-reporting, analysts and decision-makers cannot ignore the impact on validity. The counts are incorrect, which would lead to incorrect comparisons using the counts. The severity of crashes in the under-reporting areas is also artificially high because the few crashes they do report are those involving fatalities and serious injuries. This makes it difficult to draw valid conclusions about crashes in general because the local and statewide numbers are incorrect. When it comes time to allocate scarce safety resources, the under-reported crashes may result in fewer resources allocated for the affected area of the State.

4.3 Assessing Completeness

General Guidance on Completeness Performance Measures and Reporting

NHTSA strongly encourages States to develop data quality performance measures. For sustainability, these should be built into the crash data management software as automated reporting functions whenever possible. States can automate performance measures of internal completeness using validation rules like those discussed in Chapter 3 (Accuracy). Edit checks can identify missing data (blanks) or inappropriate use of codes for unknown values in individual crash reports. Any time an edit check identifies incomplete data, the system can produce an error report identifying the data field, crash report number, and responsible LEA. The automated system can provide summary data on how many times each data element is left blank (or is coded inappropriately as unknown). Reporting that information to LEAs can help them to improve officers' performance and result in more complete data. Using "Date of Birth" as an example, the following are potential edit checks that could be used to assess internal completeness (Please note: similar checks can be done with other data elements as well):

1. Edit check to display an error when driver date of birth is missing:
   If PERSON = "Driver," and HIT_&_RUN = "No," AGE = "", and DATE_OF_BIRTH = ""
   display "Driver Date of Birth is missing,"

   a. This statement checks for blanks in the age and date of birth data elements for drivers. If the vehicle was not coded as hit and run, the driver information should be available. Note that some States would program the edit check to issue a warning rather than reject a report with this error. This is to allow for the possibility that the officer could not interview the driver or access the driver's license at the scene.

2. Edit check to display an error when "unknown" code is used inappropriately:
   If PERSON = "Driver," and HIT_&_RUN = "No," AGE = "" and DATE_OF_BIRTH =
   "Unknown" display "Driver Date of Birth should not be UNKNOWN."

   a. This statement checks for inappropriate use of the "unknown" code in the date of birth data element for drivers. If the vehicle was not coded as hit and run, the driver information should be available. As with the first example, the error could be treated as a warning since it is possible that the driver is unavailable to be interviewed by police and no identification is provided.

To measure internal completeness, a State's crash data management software should have the capability to conduct a comprehensive set of completeness edit checks on each crash report record. The automated edit checks can verify completeness of crash report fields; although, only some of those edit checks would result in an error flagged for correction.

To measure external completeness, a State's crash data management software must be able to aggregate data for a specified reporting period. For example, a State may wish to calculate an estimated level of under-reporting on a quarterly basis to share with LEAs. These performance measures are automated in the crash records data management system as an analytic report, generated upon request. Examples later in this chapter show how this might work.

Guidelines on Data Entry Pathways into the Statewide Database

All reports should be evaluated for completeness, regardless of the contributing organization. This includes an evaluation for internal and external completeness. The evaluation methods may differ depending on the how the crash reports are entered into the statewide database.

The following situations serve as examples of the entry points into the State crash system and the ways that completeness performance measures apply.

1. **Paper reporting by law enforcement.** For States that are still accepting paper reports, personnel should verify that required, "key" data fields are complete. Ideally, this check would take place at the LEA prior to submission and then again during intake at the State. This supports early detection and rejection of reports so that the law enforcement officer can supply the missing information as soon as possible. Creating two quality control check points will increase the crash data system's completeness percentage for incoming reports. Additionally, the State should record LEA ID in order to report average levels of completeness for each agency.

2. **Law enforcement officers enter data directly into the State crash database.** This applies to States with an established online reporting system used by law enforcement officers. When a law enforcement officer completes the submission of the report, data validation can be accomplished at the point of entry—the online crash reporting software. When a required field is missing data, the report submission is rejected or a warning message is issued. Note that it is common practice

in such systems to hold submitted reports in a pending file until all State-specified critical edit checks are passed (i.e., after quality control steps are completed).

3. **Electronic crash reporting with electronic submissions from LEAs.** In many States, LEAs submit crash reports via electronic submissions involving data file transfers. A typical process flow is shown in Chapter 2 (Timeliness). Upon receipt, electronic data are placed in a pending file and later accepted into the crash records database after undergoing a series of QC checks. The validation edit checks are designed to identify errors (see Chapter 3) as well as incomplete data. When incomplete data is identified, the system should record the crash report number, missing data element, and LEA ID. Depending on the validation rules and QC procedures, the State may then reject the report and request that the agency submit a correction. Or, it may flag the report for attention by the State crash data management staff. The State can track the number of times each LEA submits a report with missing data in required fields. Field data collection software that includes identical edit checks as the statewide system will minimize such errors.

4.4 Developing Completeness Performance Measures

The goal of completeness measurement is to provide useful information for data collectors, managers, and users for improving the overall quality of the data. In developing completeness performance measures, system managers should consider what is most important or critical to the various stakeholders and for their own management needs. Selecting the State-specified critical data elements when assessing overall completeness will reduce the number of edit checks to run for each report and allow the data collectors to focus on fixing the most important data quality problems.

The *Model Performance Measures for State Traffic Records Systems* report (NHTSA, 2011) defines three performance measures of crash data completeness. Two of these advise the State to define a set of critical data elements. As stated in the report, these measures are:

> **C-C-1: The percentage of crash records with no missing critical data elements.** The State selects one or more crash data elements it considers critical and assesses internal completeness by dividing the number of records not missing a critical element by the total number of records entered into the database within a period defined by the State. (p. 14)
>
> **C-C-2: The percentage of crash records with no missing data elements.** The State can assess overall completeness by dividing the number of records missing no elements by the total number of records entered into the database within a period defined by the State. (p. 14)
>
> **C-C-3: The percentage of unknowns or blanks in critical data elements for which unknown is not an acceptable value.** This measure should be used when States wish to track improvements on specific critical data values and reduce the occurrence of illegitimate null values. (p. 14)

The NHTSA (2011) report includes a list of suggested critical data elements. The State should consult with data collectors, managers, and users to determine the most critical data elements. The TRCC is an ideal group to help develop the list.

In practice, States should go beyond the minimum completeness performance measures suggested in *Model Performance Measures for State Traffic Records Systems* (NHTSA, 2011). All States should be able to calculate and share C-C-1, C-C-2, and C-C-3. Other performance measures of completeness should be developed based on a State's needs. If both paper and electronic reports are collected, the State can track completeness separately for the two source types. In addition, the State may wish to create performance measures of external completeness in order to estimate the amount of under-reporting (as discussed in the following section).

4.5 Examples of Completeness Performance Measures

Measure 1 (C-C-1): Percentage of Reports with No Critical Elements Missing

This measure focuses on a set of data elements that the State has decided are critical to traffic safety stakeholders. Chapter 3 (Accuracy) includes a suggested list of critical data elements. For reporting the data quality measure, an annual example is used here. Within each year, all records that are missing critical elements are tallied as are the total number of records in the database. The total, minus the number with a blank in at least one critical data element, gives the number of complete records (those that are *not* missing data in critical fields). To create the percentage, the number of complete records is divided by the total number of crashes for the year. Table 4.1 provides example data for the percent of crash records not missing any critical data, summarized annually for five years.

Table 4.1. Percent of Complete Crash Reports Received in 2009-2013

Calendar Year	Records Missing at Least 1 Critical Data Element	Records Not Missing Any Critical Elements	Total Records	Percent Not Missing Any Critical Data Elements
2009	33,250	61,750	95,000	65%
2010	25,500	76,500	102,000	75%
2011	17,000	83,000	100,000	83%
2012	14,700	83,300	98,000	85%
2013	10,100	90,900	101,000	90%

In this example, the sample data shows the number of crash records missing at least one critical data element, records not missing any elements, total records, and the percent not missing any critical data elements. In 2009, there were a total of 95,000 records entered into the crash database. Of those 95,000, there were 33,250 records that were missing at least 1 data element. To obtain the percentage complete, calculate the proportion of records not missing any critical elements divided by the total records. In 2009, the proportion is 61,750 / 95,000 = 0.65 or 65 percent. If the metric for this performance measure is 90 percent, the State had room to improve from the 2006 starting value and had reached that metric by 2013. Note again that this measure covers only data elements that are deemed critical in the State's data practices.

Measure 2 (C-C-2): Percentage of Reports with No Elements Missing

This measure is very similar to C-C-1 in the way it is calculated. The difference is the data used. In contrast to just critical elements, C-C-2 covers all required/expected data elements. In order to calculate this measure, the agency will tally the number of records within the database that are missing values in any field that is expected to be completed. The number of complete records is the total number of crash records for the year minus those with missing data elements. Then the agency will divide that number by the total number of records in the database to arrive at the annual percentage. Table 4.2 illustrates the overall completeness performance measure calculation for five years.

Table 4.2. Overall Completeness Percentage 2009-2013

Calendar Year	Records Missing Data Elements	Records Not Missing Data Elements	Total Records	Percent Complete
2009	47,000	48,000	95,000	51%
2010	38,450	63,550	102,000	62%
2011	34,240	65,760	100,000	66%
2012	26,500	71,500	98,000	73%
2013	22,970	78,030	101,000	77%

In this example, the sample data shows the number of records missing at least one data element, the number of records not missing any data elements, the total number of records, and the proportion of complete records for each year from 2009 to 2013. In 2009 there were a total of 95,000 records entered into the crash database. Of those 95,000, there were 47,000 incomplete records and 48,000 complete. To obtain the percentage of complete records, calculate the proportion complete as follows: 48,000 / 95,000 = 0.51 or 51 percent. If the metric for this performance measure is 75 percent, the State was behind in 2009. By 2013, the State had achieved 77-percent complete, thus exceeding the metric.

Measure 3 (C-C-3): Percentage of Unknowns or Blanks in Critical Data Elements for Which "Unknown" is Not an Acceptable Value

This measure reports the percentage of crash reports where various non-specific data values were used incorrectly. A State can measure the inappropriate use of non-specific attributes such as "unknown" or "not available" and create a measure of completeness that specifically addresses misuse of unknown or not available as attributes. To accomplish this, the State would identify all critical data fields (as in C-C-1) and then identify the incorrect use of non-specific attributes when a specific value should have been available and reported. Table 4.3 shows an example calculation of the percentage of unknown and blank records where "unknown" is not an acceptable value.

Table 4.3. Percent of Unknown or Blank in Critical Data Fields for Which "Unknown" is Not an Acceptable Value 2009-2013

Calendar Year	Records With Invalid Use of Unknown	Total Records	Percent of Records with Invalid Unknowns
2009	1,500	95,000	1.58%
2010	1,800	102,000	1.76%
2011	1,700	100,000	1.70%
2012	2,100	98,000	2.14%
2013	1,900	101,000	1.88%

In this example, the sample data shows the number of records with invalid use of "unknown" data attributes in a critical data element, the total number of records and the percentage of invalid records for each year from 2009 to 2013. In 2009 there were a total of 95,000 records entered into the crash database. Of those, 1,500 records had at least one inappropriate use of the unknown value. To obtain the percentage of invalid records, calculate the proportion invalid as follows: 1,500 / 95,000 = 0.0158 or 1.58 percent. If the metric for this performance measure is below 2 percent, the State was over the acceptable target in 2012 but met the metric in all other years.

Measure 4: Ratio Measure of Under-Reporting

In contrast to the preceding performance measures which are used to assess internal completeness, the following is a measure of external completeness. It is based on the ratio of serious crashes to total crashes reported. This measure can be used to identify LEAs under-reporting PDO crashes based on a comparison to statewide averages and other LEAs. The logic behind this measure is that serious crashes (fatal plus injury crashes) are the least likely to be under-reported because LEAs are most likely to be called to the scene and therefore, more likely to complete a report. PDO crashes, in contrast, are more likely to be under-reported because law enforcement may not be called to the scene, or they may choose not to complete a crash report (even the State's reporting threshold requires the report). Note that the ratio measure can also be viewed as a measure of uniformity (see Chapter 5) because it can indicate when agencies are interpreting the reporting threshold in ways that are not consistent with the established statewide definition.

To calculate this measure, a State would tally the fatal, injury, and total crashes reported by each LEA and overall State. The combined fatal and injury crash totals equates to the total number of serious crashes. The completeness measure is the total number of serious crashes divided by the total number of crashes, expressed as a proportion. Table 4.4 demonstrates this measure.

Table 4.4. 2013 Ratio Measure of External Completeness

Agency	Injury Crashes	Fatal Crashes	Fatal + Injury Crashes	Total Crashes	Ratio of Fatal + Injury/Total
A	1,520	380	1,900	10,000	0.19
B	320	80	400	4,000	0.10
C	576	144	720	2,000	0.36
D	141	35	176	800	0.22
E	1,800	450	2,250	9,000	0.25
…					
W	1,560	390	1,950	3,000	0.65
X	32	8	40	400	0.10
Y	256	65	320	2,000	0.16
Z	36	9	45	900	0.05
State Total	**40,000**	**6,000**	**46,000**	**200,000**	**.23**

The injury and fatal crashes are summed for each agency and divided by the total number of reported crashes in the year. This is the ratio of serious crashes. In this example the statewide total for 2013 shows the ratio of 0.23. For measuring *external completeness*, the agencies with a ratio well above the statewide average should be reviewed. Agency W, for example, has a ratio of .65 indicating that 65 percent of its reported crashes involved an injury or fatality. It is likely that this agency is under-reporting non-injury (PDO) crashes. The importance of values below the State average will be discussed in Chapter 5 (Uniformity).

This measure can also be used to examine statewide trends and to identify LEAs that deviate significantly from the statewide average over a longer time period. As seen in Table 4.5, the statewide average is reasonably stable at around 0.23. This indicates that on average 23 percent of reportable crashes would involve injury or death.

Table 4.5. Annual Proportion of Serious Crashes 2009-2013

Agency	5-year Average Ratio	2009 Ratio Fatal + Injury/Total	2010 Ratio Fatal + Injury/Total	2011 Ratio Fatal + Injury/Total	2012 Ratio Fatal + Injury/Total	2013 Ratio Fatal + Injury/Total
A	**0.20**	0.19	0.20	0.23	0.20	0.19
B	**0.11**	0.10	0.11	0.13	0.11	0.10
C	**0.37**	0.25	0.23	0.23	0.65	0.36
D	**0.24**	0.23	0.26	0.22	0.25	0.22
E	**0.24**	0.20	0.24	0.26	0.27	0.25
…						
W	**0.63**	0.63	0.67	0.62	0.64	0.65
X	**0.09**	0.10	0.07	0.10	0.09	0.10
Y	**0.16**	0.15	0.16	0.17	0.15	0.16
Z	**0.23**	0.36	0.34	0.23	0.15	0.05
State Average	**.23**	**.22**	**.24**	**.23**	**.24**	**.23**

Looking at agency trends across years can help to identify when changes in practices occurred. For example, Agency C was reasonably close to the statewide average through 2011. The ratio increased dramatically in 2012 and then dropped in 2013, though still somewhat above the statewide average. This indicates the possibility that the agency adopted a policy of not reporting PDO crashes (in 2012) and then reversed that policy in 2013, perhaps because the State noticed the under-reporting. Agency W, in contrast, has consistently shown a high ratio of serious crashes. Additionally, Agency Z has consistently lowered the ratio over the five year period. As noted earlier, this may indicate that they are submitting non-reportable crashes that have then been erroneously entered into the statewide system. Based on the data, it is likely that the agency changed its practices beginning in 2012.

The State can use the agency-specific data to identify those which *might* have deliberately reduced their reporting of PDO crashes and encourage them to adhere to the statewide reporting threshold. This data should prompt a conversation rather than an accusation, however, because there may be legitimate reasons for the high ratios including chance fluctuations, increasing variability, and contractual agreements to assist other jurisdictions. For example, the State LEA may investigate all serious crashes for a number of smaller jurisdictions so their ratio *should* be somewhat higher. The LEAs with strikingly high or low ratios should be invited to explain why their data differs from the statewide average.

4.6 Final Considerations

There are many possible performance measures for completeness. In order to measure internal completeness, States should identify the most critical data elements and focus on those elements in the edit checks and performance measures. The ratio measure of external completeness can identify under-reporting of PDO crashes by an LEA.

These performance measures should be incorporated into feedback systems for LEAs and officers. On a periodic basis, the State should provide the information to data collectors and users to promote discussion of ways to improve completeness. NHTSA encourages using these performance measures and integrating them into a formal data quality management program, as discussed in the *Traffic Records Program Assessment Advisory* (NHTSA, 2012).

4.7 Questionnaire

The following questions are intended to identify how the State measures the completeness of the crash data. This section will evaluate how the state assesses the measurement and the efforts to increase completeness.

79. What output measures of completeness are currently reported for your State? (SOE: Provide baseline current data and trend if available, and provide sample reports) (AQ-68)

80. How does the State assess missing data for each of the data elements within the crash database?

81. 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)

82. How does the State assess the frequency with which ‘unknown’ is the reported value for each of the data elements within the crash database? (e.g., a descriptive analysis listing the frequency and/or percentage of ‘unknowns’ per data element)

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

84. How does the State report to law enforcement agencies on the completeness of their crash report submissions in terms of numbers of reports submitted? (SOE: provide a copy of the report and frequency of issuance)

Chapter 5: Uniformity

5.1 The Purpose of Measuring Uniformity

Uniformity reflects the degree to which crash reporting follows State and Federal guidelines. Uniformity applies to data elements, data element attributes, expected contents of the crash database, data collection methods, reporting criteria, data management processes, and analyses. Uniform crash data is critical to meaningful analysis and enables accurate, impactful decision-making across all jurisdictions. Lacking uniform crash data, States may not be able to develop appropriate countermeasures or properly prioritize safety dollars. Adopting recommended National data guidelines can enable agencies to significantly increase the uniformity of their crash data. Increased uniformity also provides a means for Federal agencies to conduct National data analysis, identify trending issues, and develop support for systematic safety programs.

The primary guideline for crash data collection is the MMUCC, a voluntary guide describing the minimum data elements that should be collected for each crash record. The fourth edition of the MMUCC contains 110 data elements. States are encouraged to capture the minimum plus any other State-specific elements. MMUCC has been developed by a large group of subject matter experts who have shaped and refined the model minimum list of elements and attributes. States that choose to align their crash reports with MMUCC have the added benefit of aligning more closely with other Federal datasets like FARS and SafetyNet.

Uniformity performance measures are critical for data managers to understand how closely a State's dataset aligns with State and National guidelines. Crash uniformity performance measurement serves to:

- Identify individual crash reports that do not meet State standards.
- Identify data collectors and LEAs in need of specific feedback.
- Quantify how closely the State's crash data matches National standards and guidelines.
- Recognize opportunities to improve crash reporting training, manuals, forms, and software.
- Describe how reliable the data is for decision-making.

5.2 Defining Uniformity

Uniformity in a crash data system means that the information is recorded on a single statewide crash form, using shared definitions of all elements and attributes, and entered into the database using a standardized data model. According to the NHTSA *Model Performance Measures for State Traffic Records Systems* (2011), uniformity reflects the consistency among the files or records in a database. It should be measured against an independent standard, preferably a National standard. In a crash data system, data managers can define and validate uniformity in two primary ways: externally and internally. These are described in the following sections.

External Uniformity

External uniformity is defined as adherence to a National standard or guideline. These may be written into Federal law (e.g., the Transportation Reauthorization legislation), or part of USDOT agencies' policy for grant eligibility or reporting requirements for National data systems. External uniformity is typically something that a State will examine when it updates the PCR, field data collection software, or the centralized crash database. A State will compare its current and desired data elements and their definitions to National standards such as ANSI D-16.1, FMCSA's SafetyNet data requirements, FARS, and the

MMUCC Guideline to see if adopting the definitions from those sources would result in improved detail and more utility for decision makers. States can measure external uniformity by mapping the State's data elements in the PCR and crash database to the National standards and guidelines. Because the MMUCC Guideline references data definitions from the National standards, States are encouraged to map their crash data to MMUCC as way to achieve a close match to all of the other sources.

Internal Uniformity

Internal uniformity means that the data adheres to standards set by the State. States set standards for the PCR, the data elements collected at the scene of a crash, and the threshold for a crash report to be submitted to the State. Standards are enforced by requiring all agencies to use the same PCR; implementing the same data definitions in training, officers' manuals, and field data collection software; and through edit checks and data validation analyses.

States that identify uniformity problems should: provide feedback to law enforcement, improve law enforcement training and officer instruction manuals, and improve the crash report form or crash collection software.

5.3 Assessing Uniformity

General Guidance on Uniformity Performance Measures and Reporting

NHTSA strongly encourages States to develop data quality performance measures. Where possible, States should build them into their crash data management software as automated reporting functions. For measuring external uniformity, such automated performance measurement is not feasible—the comparison of a State's PCR to MMUCC is a manual effort and requires detailed understanding of data element definitions. States can measure uniformity in an automated manner using performance measurements like those suggested later in this chapter. Because internal uniformity problems all relate to data collection problems, reporting that information to LEAs will help improve officer performance and result in more uniform data among all officers and LEAs. The following examples demonstrate how to assess internal uniformity:

- **Annual uniformity reports to LEAs.** As part of quality control checks, a State may develop performance measures of the changes in crash reporting patterns over time such as the total number of reports each agency submits, and the frequency with which specific values of critical data elements are used. Performance measures of uniformity will identify any large, unexpected changes from historical values and any agencies with data that varies from statewide averages by an unexpectedly large amount. Ideally, the uniformity performance measures may help the State identify if there is also an accuracy or completeness problem, or if some other problem is to blame. Using a uniformity report, the State provides agencies feedback on what is not being recorded properly on the crash report. Training can then be provided to those agencies on the standard practice of filling out a crash report.
- **Report of the ratio measure of (fatal + injury crashes)/total crashes.** Comparing individual LEAs' measurement to the statewide average helps to identify those that are not adhering to the crash reporting threshold. The State can contact agencies with high or low ratios to discuss any issue that needs to be addressed.

Performance measures of internal uniformity are usually calculated based on aggregate data. For example, a State may wish to compare actual to expected levels of reporting on a quarterly basis to share with LEAs. These performance measures are automated in the crash records data management system as an analytic report, generated upon request. Measure 2 later in this chapter shows how this might work using the ratio measure from Chapter 4.

5.4 Developing Uniformity Performance Measurements

The *Model Performance Measures for State Traffic Records Systems* report (NHTSA, 2011, 2011) defines one performance measure for crash data uniformity:

> **C-U-1: The number of MMUCC-mappable data elements entered into the crash database or obtained via linkage to the other databases.** (p. 14)

This suggested performance measure is based on a detailed comparison of the State's data to the elements defined in MMUCC. The process for calculating this performance measurement is covered in full detail in Chapter 8 (Mapping to MMUCC). States are encouraged to go beyond the measure of MMUCC-mappable data elements suggested in the NHTSA (2011) report. States are also encouraged to track adherence to the SafetyNet reporting requirements using the crash data quality performance measurements reported by FMCSA.

Performance measures of internal uniformity are designed to compare data collected by law enforcement to the State's own standards. A lack of uniformity may point to agencies using an incorrect reporting threshold, an incorrect or out-of-date data definition, or any other non-standard practice that affects the data collected at the crash scene. Ideally uniformity performance measurements should point to actions that will improve data quality such as training on the State's standards, rejection of non-compliant data forms, and enforcement of the State's reporting threshold.

5.5 Examples of Uniformity Performance Measurements

Measure 1 (C-U-1): The Number of MMUCC-Mappable Data Elements Entered Into the Crash Database or Obtained Via Linkage to the Other Databases

The process for mapping to MMUCC is covered in complete detail in Chapter 8. This performance measure is included here in Chapter 5 because it is the single uniformity performance measure described in the *Model Performance Measures for State Traffic Records Systems* report (NHTSA, 2011). Since the time of that report, the review process has changed to become a more detailed mapping to MMUCC. The following is a summary of the steps required to complete an MMUCC mapping under the most recent guidance.

A) **Gather all necessary documentation relating to crash data.** The State will need to have the following items to compare the data elements from the crash data system to the MMUCC: the PCR, the officer's instruction manual, and the statewide crash system data dictionary. The data dictionary should include each data element in the database. For convenience, the data dictionary should be organized into the four areas data elements included in MMUCC: crash, vehicle, person, and roadway.

B) **Total the number of data elements that are listed in MMUCC versus the State-obtained data elements with the PCR.** Using the process and tools described in Chapter 8 (Mapping to MMUCC), the State records MMUCC elements and attributes present in its PCR and crash database.

C) **Execute a thorough review based on MMUCC mapping rules.** This is the actual process where the report and/or the system are evaluated. Using the latest MMUCC mapping rules as a guide, States determine if each element and associated attributes from both their PCR and crash database map to the elements and attributes detailed in the latest MMUCC Edition. PCR mapping is considered complete if the form can minimally map to all attributes designated in MMUCC. Database mapping is considered complete if the database contents can map to all attributes detailed within MMUCC. Calculations are based on total attributes mapped rather than by elements so that elements with many attributes would not diminish the value of each attribute. Thus, the MMUCC mapping percent mappable score counts all attributes as having equal value (of 1).

D) **Divide the number of State attributes that map to MMUCC by the total number of attributes in MMUCC.** States can use the free MMUCC mapping Excel spreadsheet to track the number of attributes that map, which will automatically calculate their percent mappable score. Alternatively, States can tally using their own method, dividing the total number of State attributes that mapped to MMUCC by the total number of attributes in MMUCC.

Figure 5.1 shows the calculations page from the NHTSA MMUCC 4th Edition mapping spreadsheet. This provides the State with the overall mapping for all MMUCC elements, and separately for the 77 elements collected at the scene and the 33 derived or linked elements.

Mapping for all 110 Elements		77 Elements Collected in the Field		33 Linked or Derived Elements	
46.3%	For 110 Elements	49.9%	For 77 Elements	38.0%	For 33 Elements
44.8%	Crash Elements	38.5%	Crash Elements	62.2%	Crash Elements
65.2%	Vehicle Elements	65.2%	Vehicle Elements	55.6%	Vehicle Elements
43.6%	Person Elements	40.6%	Person Elements	17.9%	Roadway Elements
17.9%	Roadway Elements				

Figure 5.1. MMUCC Mapping Results Example from NHTSA's Spreadsheet

Measure 2: Ratio Measure of Uniformity

A ratio measure of serious (fatal+injury) divided by total crashes can be used to measure internal uniformity. The statewide average of this ratio sets the expected value for all LEAs in the State. A lack of uniformity is shown when LEAs' ratios vary substantially from that statewide average. Specifically, a large difference can indicate when an agency is interpreting the reporting threshold incorrectly and either reporting crashes that do not meet the State threshold, or failing to report crashes that do meet the threshold. Both are evidence of a lack of uniformity. When an agency's ratio is high in comparison to the statewide, average this indicates potential under-reporting of PDO crashes—a circumstance that would also demonstrate a problem with completeness. In the event that an agency's ratio is unexpectedly low, one possible cause is that the agency is submitting crashes that are below the State's crash reporting threshold. Note that this would inflate the total number of crashes *and* lower the proportion of serious crashes within that total. This sometimes happens when the agency sets a policy of completing a full crash report for every crash event as a courtesy to their citizens to help with insurance claims. To show how this performance measure can be used to spot inconsistent reporting, Table 5.1 presents the same data as the ratio measure presented in Chapter 4.

Table 5.1. 2013 Ratio Measure of Uniformity

Agency	Injury Crashes	Fatal Crashes	Fatal + Injury Crashes	Total Crashes	Ratio of Fatal + Injury/Total
A	1,520	380	1,900	10,000	0.19
B	320	80	400	4,000	0.10
C	576	144	720	2,000	0.36
D	141	35	176	800	0.22
E	1,800	450	2,250	9,000	0.25
…					
W	1,560	390	1,950	3,000	0.65
X	32	8	40	400	0.10
Y	256	65	320	2,000	0.16
Z	36	9	45	900	0.05
State Total	**40,000**	**6,000**	**46,000**	**200,000**	**0.23**

The injury and fatal crashes are summed for each agency and divided by the total number of reported crashes in the year. This is the ratio of serious crashes. In this example the 2013 statewide total shows the ratio of 0.23. Agencies that diverge far from this value may require further review. Values far below a statewide average, like those of Agency Z, may have one of two common reasons for the disparity. One, Agency Z might contract with other LEAs to investigate serious crashes. Two, this could indicate they are not responding to or are submitting crash reports for less serious crashes; essentially having a much higher threshold than defined by the State. On the other hand, Agency W has a ratio of 0.65, indicating that 65 percent of its reported crashes involved an injury or fatality. Agency W is either under-reporting non-injury (PDO) crashes or is contracted by other LEAs to investigate their serious crashes, thereby changing Agency W's ratio.

When evaluating uniformity, agencies with a ratio well above or well below the statewide average should be reviewed. The State can use the agency-specific data to identify those which *might* have deliberately increased or decreased their reporting of PDO crashes and encourage them to adhere to the statewide reporting threshold. This data should prompt a conversation rather than an accusation. There may be legitimate reasons for the high ratios including chance fluctuations, increasing variability, and contractual agreements to assist other jurisdictions.

5.6 Final Considerations

In order to measure uniformity, States should identify data elements on the crash report and crash database and compare them to the MMUCC Guideline. Measure 1, the MMUCC-mapping measure, is of critical importance to safety and crash analyses at the national level. Knowing how well States' data maps to MMUCC can help analysts develop a National embodiment of a traffic safety issue. From the States' perspective, however, that performance measure is one that States will likely revisit only when they are considering revisions to the PCR or the crash database. Once those changes are in place, the value of the performance measure will remain unchanged, often for years.

The State can also measure internal uniformity by comparing trends in reporting over time, and by comparing each LEA's reporting patterns to the statewide average. There is clearly some overlap in how a State may wish to measure accuracy, completeness, and internal uniformity. Such multipurpose performance measurements are valuable in that they can illustrate the impact of a variety of problems upon data quality.

Comparisons among jurisdictions and populations are critical for State safety program management. If uniformity problems are making those comparisons less reliable, the State needs ways to identify the problem and ways to address it. Uniformity performance measures should be incorporated into feedback systems for LEAs and officers. The State should provide the information to data collectors and users to promote discussion of ways to improve completeness on a periodic basis. NHTSA encourages using these performance measures and integrating them into a formal data quality management program, as discussed in the *Traffic Records Program Assessment Advisory* (NHTSA, 2012).

5.7 Questionnaire

The following questions are intended to identify how the State measures the uniformity of the crash data. This section will evaluate how the State reports the measure and how it is calculated.

85. What output measures of uniformity are currently reported? (AQ-69) (SOE: Provide baseline current data and trend if available)

86. Has the State participated or conducted a mapping-to-MMUCC measurement of uniformity? If yes, please supply the analysis. If one is conducted as part of this CDIP, the TAT will have access to it.

87. 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)?

88. How are these output measures calculated? (SOE: provide sample reports and their frequency of issuance)

Chapter 6: Integration

6.1 Purpose of Measuring Integration

Highway safety is multidisciplinary and encompasses the 4 E's of highway safety: engineering, enforcement, education, and emergency medical services. Using data from resources related to all of these areas provides a more robust foundation for decision-making and resource allocation. While each of the six traffic records system datasets are valuable independently, together they provide a more comprehensive understanding of all components of highway safety.

Crash data integration performance measurement serves to:

- Support more robust analysis and effective decision-making.
- Identifying potential data quality issues within each of the linked datasets.
- Reduce redundant data collection.

6.2 Defining Integration

Model Performance Measures for State Traffic Records Systems (NHTSA, 2011) defines integration as, "…the ability of records in a database to be linked to a set of records in another of the six core databases – or components thereof – using common or unique identifiers." This report, as well as *Traffic Records Program Assessment Advisory* (NHTSA, 2012), describes these distinguishing features associated with data integration:

- **Integration always involves two or more traffic records subsystems (i.e., databases or files).** Unlike other performance measures that may be applied to each of the six core databases individually, data integration inherently requires at least two.
- **Integration is different from interface.** Integration is the linkage of databases to support in-depth analysis, usually done using a dataset that has been deemed final (at the close of a calendar year or other regularly scheduled period). Interface is more commonly a real-time connection between datasets to support business processes.

Types of Data Integration

Data linkage for the purpose of integration and subsequent analysis can be done in two ways: deterministically or probabilistically.

Deterministic Linkage

Deterministic linkage uses one or more unique identifiers—event identification code, EMS run report number, patient identifier, or crash report number—to link one or more datasets. In deterministic linkage, there is no uncertainty as to whether or not the linkage worked. When the unique identifiers match, so do the records.

A common example of deterministic linkage is matching crash data with roadway inventory data using geographic coordinates. Crash reports can often contain latitude/longitude for the crash location. The roadway inventory data is also often geolocated through a GIS platform. The latitude/longitude from the crash report can be linked to the roadway inventory information to determine which roadway characteristics are associated with that crash location.

Figure 6.1 is an example of a map produced by integration. Crashes are linked to the roadways through the location information and information from both the crash data (e.g., crash severity and first harmful event) and roadway data (e.g., Annual Average Daily Traffic) are available for the linked crashes. A more robust analysis is possible as a result of that linkage. For example, safety analysts can perform network screening to identify locations with higher-than-expected crash counts and look for patterns of crashes associated with particular roadway characteristics like narrow shoulders, medians, or curves.

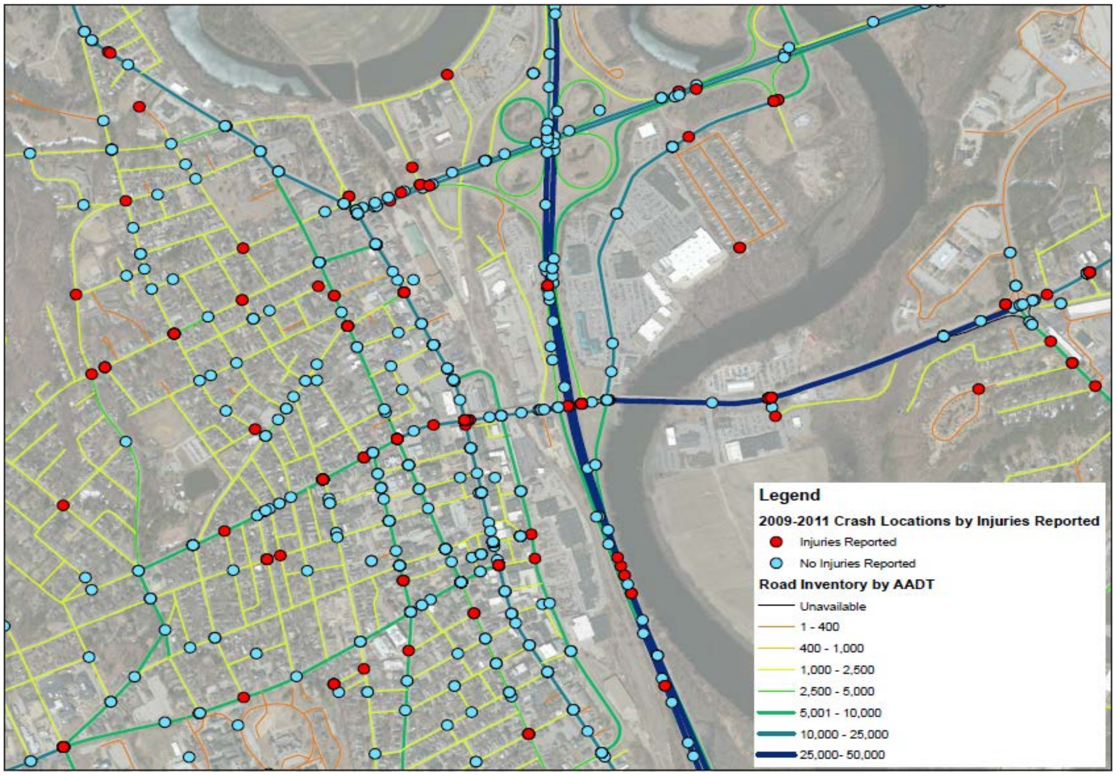

Figure 6.1. Example of Deterministic Linkage Between Crash and Roadway Data in a State

The crash to roadway linkage is an example of data integration used primarily for analysis purposes. There are also opportunities to use deterministic linkage to auto-populate databases and reduce the burden placed on data collectors. Two examples of this are the following:

1. Driver license number can be collected on the crash report form and later used to auto-populate driver related fields such as age, sex, and address as well as license status at the time of the crash.
2. VIN can be collected from the vehicle registration or on the vehicle itself and recorded on the crash report form. This can later be used to auto-populate vehicle-related fields such as make, model, and year. VIN decoding can also be used to identify what safety features a vehicle has such as airbags, backup cameras, and lane departure warning systems.

In deterministic linkage, if the unique identifiers do not match, neither do the records. A non-match does not mean that there *was* no match possible; it only means that the deterministic linkage failed to find the matching records. For example, the State knows that they only collect crash information for reportable events occurring on public roads and should therefore be able to match crash records to a record in the roadway database. The impact of data quality on data integration is important. Failure of deterministic linkage is often due to incorrect or missing data. For example, in an analysis of drivers' likelihood of a crash, a State might want to link each driver's crash history to their history of convictions for moving violations. This linkage will fail for out-of-State drivers currently since State motor vehicle agency databases are not yet interconnected. It will also fail when the driver-identifying data on the crash report form is inaccurate or missing. These are expected matches that do not happen because the linking variables do not match precisely across the two databases.

Probabilistic Linkage

Probabilistic linkage uses statistical techniques to determine the probability of a correct association between two records from different datasets that have corresponding information in a set of common data elements. Probabilistic linkage results in a value measuring the likelihood that the two records are associated with the same person, vehicle, or event, even if those common data elements are not unique or if they do not match precisely.

A State's EMS run report and crash report may each record the time, date, location, and identity of the injured person who was transported from the crash scene. It is very likely that the date of the crash event will match in the two records (although crashes occurring near midnight may have one date for the police response and the next day's date for the EMS response). The event time in the two records is likely to be similar, but not exactly the same. The location information may be very different even if the law enforcement and EMS responders are using GPS to obtain location coordinates. This is because location information is based on the operator's position when taking a reading. Even personally-identifying information may differ among two data sources. The crash report may record the person's name as it appears on their driver's license whereas the EMS report may record the name as reported to them by the injured person or someone else at the scene. There are limited opportunities for police and EMS to compare records at the scene of a crash, so two correct—but disparate—records of the same event may result.

Deterministic linkage in that situation would fail. Other than the crash date, none of the linking data elements match precisely, so there would be *no match.* Probabilistic linkage can address these issues by (1) using multiple variables simultaneously; (2) allowing for near matches; and, (3) allowing for a window of agreement to accommodate anticipated differences such as the crash time and the EMS arrival time.

Data Integration Variables

Linking variables are the data elements that are expected to match across two or more databases. The variables used for data integration will depend greatly on (1) the type of integration (deterministic or probabilistic) and (2) the datasets being linked.

Integration Variables by Type of Integration

If the linkage is deterministic, the linking variables will be a common unique identifier. The location coding example earlier in this chapter uses a unique location code assigned to the crash report and the roadway segment. An event identifier is another example of a data element that may be used for deterministic linkage. If all first responders use the same unique event identification number, the crash report and the EMS report may be linked using that identifier.

EMS and law enforcement may, for example, share the run report number or crash report number so that both databases contain at least one shared event/report number. Some States have implemented unique *patient* identifiers. These are numbers that can be assigned by EMS and shared with the law enforcement officer at the scene. The same number can also follow the person through ED treatment, outpatient observation, hospital admission, and even later rehabilitation. Unique patient identifiers make it possible to link crash and injury surveillance data without sharing restricted data such as names, social security numbers, or driver license numbers. When such personal identifiers *are* used for deterministic linkage, their release is strictly controlled under Federal and State laws.

If the linkage is probabilistic, it will rely on a selection of linking variables that are common to the datasets being linked. For example, if a crash report and an injury surveillance record (i.e., EMS, ED, hospital) report the same sex, date of birth, crash date, crash location and other common fields, the records are likely to describe the same person and event. None of the linking variables are unique to an event or person, but together they can be used to identify records that are likely matches across the datasets. This method can determine if various records are likely associated with the same crash with a high enough probability to deem the linkage sufficient for analysis purposes.

Table 6.1 provides example fields used to probabilistically link crash and hospital data using probabilistic linkage. Data linkage allows for the analysts to set tolerances that allow for the linkage of records that don't match all fields exactly. For example, location information such as ZIP code or town can be linked with a tolerance of a reasonable mile radius to allow for the likelihood that the crash does not occur in the same ZIP code or town as the hospital where they will be treated.

Table 6.1. Example Linkage Fields for Probabilistic Crash-Hospital Linkage

Crash Data		Hospital Data
Age	←→	Age
Sex	←→	Sex
Date of Birth	←→	Date of Birth
ZIP Code	←→	ZIP Code
Crash Date	←→	Admission Date
Crash Town	←→	Hospital Town
Role in Crash	←→	Role in Crash
Collided With	←→	Collided With
Vehicle Type	←→	Vehicle Type

Integration Variables by Datasets Being Linked

When linking datasets to support safety decision-making, crash data is often the core dataset used to link to the other sources of safety data. Crash data can be integrated with vehicle, driver, roadway, citation and adjudication, and injury surveillance data. However, non-crash datasets may be linked to each other without crash data. For example, driver data may be linked to citation and adjudication data to understand the potential relationships between driver history (e.g., suspensions, driver education completion) and citations.

Although any of the six traffic records data sets can be linked with any of the others, the CDIP focuses on the relationship between crash data and each of the five other data systems—vehicle, driver, roadway, citation and adjudication, and injury surveillance—as shown in Figure 6.2.

Figure 6.2. Crash Data Linkage Opportunities with Other Traffic Records Data Systems.

The datasets being linked will largely govern which fields are used for linkage. Table 6.2 provides example variables for the linkage of crash data with each of the other traffic records datasets. The data elements listed are described in the MMUCC 4th edition (GHSA & NHTSA, 2015).

Table 6.2. Linkage Variables by Datasets Linked and Type of Linkage

Crash Data Link To	Potential Deterministic Linkage Variable included in MMUCC	Potential Probabilistic Linkage Variables Included in MMUCC
Vehicle	VIN	Vehicle type, registration State, year, make, model, model year, body type category
Driver	Driver License Number	Sex, birthdate, age, person type, driver license jurisdiction, driver license class, driver name
Roadway	GPS Coordinates	Bridge/structure ID number, part of National Highway System, location by other means (e.g., distance from mile marker or landmark, address)
Citation/ Adjudication	Driver License Number	Driver information (e.g., sex, birthdate, age), and incident information (e.g., town, date, time)
Injury Surveillance	Common Event/Incident Number	Sex, birthdate, person type, injury status, seating position, restraint use, crash/hospital location, vehicle type, manner of collision

6.3 Assessing Integration

Measuring the effectiveness of data linkage may be a relatively simple process, especially since much of the information needed is produced automatically from the linkage process. It is important to remember, however, that the linkage itself can be a complex process. Even relatively 'simple' deterministic linkages may be more complicated than expected. Probabilistic linkages are inherently more complex and should only be undertaken by experienced statisticians or data professionals. Measuring integration is therefore even more important to ensure the linkages are appropriate, provide the expected resulting dataset, and are useful to safety professionals.

It is important to remember that data linkage relies heavily on the participation of various owner agencies. Data governance and coordination through an entity such as the State TRCC may be imperative to successfully balance the requirements and resources of each contributing owner agency. It is important to assess the success of data integration so users know the limitations of the data sub-set and what types of records are likely to be left out of the final database.

In general, data integration performance measures can be grouped into the following categories:

- **Existence and quality of data integration.** Did the linkage capture the majority of records expected to link? With what certainty were records linked (if probabilistically linked)?
- **Timeliness of the linkage.** Were the finalization calendars for each of the linked datasets sufficiently aligned to allow for a timely linkage?
- **Usefulness of linked data.** Are data owner agencies and other highway safety professionals able to use the linked data for program and policy development, implementation, and evaluation?

6.4 Developing Integration Performance Measurements

Measuring data integration provides information for data custodians, those performing the linkage, and data users. This should be done on a regular basis (such as annually) to measure changes over time. When determining which integration performance measures to use, two factors should be taken into consideration:

1. **At what stage of the data integration process is the State?** If this is a new endeavor for the State, they may choose to focus on successful integration with a reasonable level of quality. States that are more experienced with data linkage may wish to add performance measures aimed at improved timeliness and use of linked data.
2. **What are the current needs for data integration?** If the current focus is on integration for data quality control across datasets, timeliness may be more important than it is for linked data being used for long-term research.

The *Model Performance Measures for State Traffic Records Systems* report (NHTSA, 2011) defines one measure of crash data integration, which addresses the existence of data integration:

> **C-I-1: The percentage of appropriate records in the crash database that are linked to another system or file.**
>
> A key word in this measure is "appropriate." This may also be translated as "expected." For example, not every crash record is expected to link to a citation/adjudication record since there are some crashes where no citation is issued. Conversely, not every citation/adjudication record is expected to link to a crash record since citations may be issued without a crash occurring.

States may also wish to measure the number of systems that are integrated and the current status of inter-agency agreements formalizing those data sharing arrangements, as follows:

> **Number of data owner agencies that have current signed memorandums of understanding or data sharing contracts allowing their data to be used in linkage.**
>
> Institutionalizing data sharing agreements formalizes the process. This is especially helpful in addressing turnover in agency leadership as well as staff. Having a current, formal data sharing

agreement from the data owner agency can facilitate a more seamless transfer of data to the linkage team, even in the face of organizational instability.

Another performance measure for crash data integration focuses on the quality and usability of the linked data set. Some potential performance measures include the following:

Identify potential users of the linked database and query them about their satisfaction with the data as it applies to their specific use (e.g., data quality monitoring, in-depth analysis).

Since the linked dataset may reside outside the control of the crash data manager or custodian, it is important to consider it independently. The organization conducting the linkage, which may be another State agency or an external entity such as a university, is likely to be responsible for ensuring satisfaction with the usefulness of the linked data.

NHTSA measure C-I-1 is generally calculated automatically as part of the linkage process.

6.5 Examples of Integration Performance Measurements

The crash data integration performance measure defined in the *Model Performance Measures for State Traffic Records Systems Report* (NHTSA, 2011) is calculated as described in the following section.

Measure 1(C-I-1): The Percentage of Appropriate Records in the Crash Database That are Linked to Another System or File

A) **Identify data sets to be linked and measured.** The State may choose to measure the percentage of appropriate records in the crash database that are linked to one or more other traffic records datasets. The State should determine which of these traffic records data set linkages they will measure.

B) **Identify how many records in each dataset are expected to be linked.** Each dataset included in the linkage will have more records than are expected to link. For example, when considering crash data linkage to driver license data, the State should only expect to link records for drivers registered in that State.

C) **Determine what the linkage probability threshold is for considering records successfully matched.** If the linkage is deterministic, the State can skip this step since the record will either link or it will not; there is no probability associated with the linkage. If the linkage is probabilistic, the State should determine what threshold is considered sufficient for deeming the records matched a true match (e.g., 90 or 95% probability).

D) **Tally the total number of records expected to match in each dataset and the number of records that linked.** The State should have two numbers for each dataset included in the linkage: total records expected to match and total that matched (using the acceptable match probability threshold for probabilistic linkage).

E) **Calculate the percentage of expected records that linked.** Divide the number of linked records for each data set included in the linkage by the number of records expected to link for that dataset. Then, multiply by 100 to get the percentage of expected records that linked to another system or file.

Example data for this calculation for a linkage of crash records to driver license records in a State is shown in Table 6.3.

Table 6.3. Example Data for Integration Calculation C-I-1

Data	Total Records	Expected Link Records	Actual Link Records	Percent Linked
Crash	125,256	24,899	23,242	93%
Driver License	548,969	31,550	25,632	81%

Since linking data yields a new dataset, States may consider applying the performance measures they use for crash data to the linked dataset separately. Specifically, timeliness and accessibility have the greatest opportunity to be different than the performance measures for the crash data.

Measure 2: Number of Data Owner Agencies That Have Current, Signed MOUs or Data Sharing Contracts Allowing Their Data to Be Used in Linkage

The process of data integration requires agreement among the data owners (custodians) and the managers of the systems that will use the various data sources in combination. A set of MOUs or some other form of data-sharing contract among the agencies is helpful in formalizing the agreements and States would do well to track the number and current status of such agreements. The simplest measure would be one that tallies the number of current agreements and compares it to the number of agreements sought for data linkage. For example, if the linkage team asked for agreements from seven agencies and has current signed MOUs for four of the data systems, the measure would be reported as four of seven. If more detail is desired, the integration team can report information such as that shown in Table 6.4.

Table 6.4. Example of Tracking Data Sharing Agreements

System	MOU Status	Effective Date	End Date	Signed By
Vehicle	Current	01/01/2016	12/31/2017	Secretary DMV
Driver	Current	01/01/2016	12/31/2016	Secretary DMV
Roadway Inventory	Current	01/01/2016	12/31/2016	Secretary DOT
Traffic Volume	Current	01/01/2016	12/31/2016	Secretary DOT
Citation	No MOU	N/A	N/A	N/A
EMS	Lapsed	01/01/2015	12/31/2015	Secretary DHHS
Trauma Registry	Lapsed	01/01/2015	12/31/2015	Secretary DHHS

6.6 Final Considerations

Opportunities to improve data integration exist at any point in the data system lifecycle. However, as with many of the other traffic records performance measures, integration is most successfully considered during the design or update phases of each of the individual traffic records data systems. The following considerations are the most notable:

- Creating a traffic records system inventory provides an opportunity to identify potential linking variables.
- Identifying linking variables helps to highlight the need for specific data elements that may otherwise not be recognized as important for safety analysis.

- Designing a new system or system upgrade provides the ideal opportunity to address data integration as well as new performance measures aimed at assessing the success of data integration efforts.
- Sharing data governance standards among system design and update efforts can aid in data integration efforts. Data governance is a set of processes that ensure that important data assets are formally managed throughout the enterprise. In particular, determining when datasets are considered sufficiently complete for use in linkage or standardized definitions of variables to facilitate linkage. Data governance processes also promote shared data definitions and quality standards which, in turn, aid in data linkage.
- Gaining input from IT leadership and others who are outside the typical safety functions in a department may facilitate future linkage of data systems that are also under their purview. This is especially important in States with centralized IT functions for all State government.
- Automating portions of the data integration process can be included in new system designs. At that time, it is possible to also automate the data quality performance measures that assess linkage.

Data integration improvements can be successful at any time, even if a system design or update is not on the horizon. The NHTSA and CDC report *Assessment of Characteristics of State Data Linkage Systems* (Milani et al., 2015) identifies key facilitators and challenges in establishing data linkages that should be considered at all stages of linkage. In addition to factors discussed in this chapter, the report identifies the following facilitators:

- Housing overall responsibility for the data linkage program in one organization.
- Establishing memorandums of understanding or data sharing agreements that consider the institutional processes required by each data owner to address issues of privacy.
- Staffing with personnel that have the appropriate skills.
- Providing technical assistance and training in linking and analyzing linked data.
- Providing stable funding.
- Participating in a community of practice of data linkage practitioners.

A key facilitator identified in the report is establishing a coalition that includes all data owners. The State TRCC is an ideal organization to serve in this capacity and to facilitate data integration the following ways:

- Negotiating access to data across State agencies;
- Addressing privacy and security concerns;
- Agreeing upon a data governance approach that meets the needs of each data owner agency;
- Identifying the best party to conduct the data linkage, whether it is one of the data owner agencies or an external organization; and
- Establishing performance measures of the success of integration efforts.

The NHSTA/CDC report also identifies barriers to data linkage that should be considered by agencies in advance of data linkage to identify means for minimizing their impact. These barriers include the following:

- Insufficient funding;
- Staffing turnover;
- Lack of process documentation;
- Long lag times in obtaining source data for linkage;
- Complex linkage techniques;
- Marketing linked data so that others understand its value and utility; and
- Statutory requirements for obtaining and reporting data.

This last issue can be especially difficult when handling data that contains personally identifiable information (PII). Privacy restrictions for driver and citation data containing PII are generally mandated by State law while privacy restrictions for injury surveillance data is mandated by the Federal Health Insurance Portability and Accountability Act. Organizations responsible for collecting and managing these datasets generally have established practices for providing access to them for linkage that may include Institutional Review Board review.

The complexities of data linkage and the need to coordinate across potentially disparate owner agencies should not discourage a State from integrating their traffic records data. Resulting linked datasets support a more comprehensive, quantitative understanding of the many facets associated with crash characteristics and outcomes before, during, and after the crash. Linked data has been successfully used by States to guide programming, influence policy-making, and support evaluation.

The following examples are found in NHTSA's *The Crash Outcome Data Evaluation System (CODES) and Applications to Improve Traffic Safety Decision-Making* (NHTSA, 2010):

- Linked crash and injury surveillance data provided key information as State legislatures considered passing a primary seat belt law.
- Linked crash and injury surveillance data was used to measure the effect of safety device use on crash outcomes for children ages four through eight.
- Linked crash, injury surveillance and roadway data was used to examine the causes and consequences of crashes occurring on State roads with posted speed limits of 50 mph.
- Linked crash, injury surveillance and citation data was used to strengthen the State's graduated driver licensing laws.

The *Crash Outcome Data Evaluation System (CODES): Program Transition and Promising Practices* (Kindelberger & Milani, 2015) report presents ongoing use of probabilistic matching methods and documents its effects on safety program design and policy, problem identification, legislation, and education. Examples include the following:

- Child occupant protection, including seating position and device use analyses in support of multiple State's safety programs.
- Impaired driving analyses aimed at reducing the economic impact of alcohol- and drug-related crashes in Delaware.
- Motorcycle safety analyses of helmet use, motorcycle and route choice, and number of involved vehicles in Delaware, Minnesota, and Maine.

- Primary seatbelt legislative support in Minnesota and Utah.
- Seatbelt law effectiveness evaluation in South Carolina and Missouri.
- Young driver and distracted driver legislative support in Connecticut and Utah.

Crash data is not the only potential beneficiary of a successful linkage. Additional examples of the benefits of linkage to other data systems include the following:

- Linking crash to roadway data can provide an alert system for DOTs regarding roadway environment issues that should be addressed. Examples include missing or broken traffic control devices that may have contributed to the crash or the need for median barrier or guardrail repair/replacement if it has been compromised by the crash.
- Linkage of crash and EMS data can help track EMS response times and improve the EMS system overall.
- Linkage of crash, citation, and driver data greatly improves the driver file, keeping unsafe drivers off the road. If this linked data is available to law enforcement officers at the time they are issuing citations, it may be helpful to ensure the most appropriate violations are cited in cases where multiple offenses of the same violation are associated with stronger penalties. Ultimately, this may help keep the most unsafe drivers off the road.

Data linkage can be important for data quality efforts related to the individual datasets being linked. The inability to link, or a low percent of expected links matching, may point to a data quality issue in either or both of the individual datasets. In addition, when using probabilistic linkage, States can see where discrepancies in high probability matched records lie. In some cases, small discrepancies are expected (such as the difference in time for the crash time versus the EMS arrival time). Other discrepancies may be the result of data collection or coding errors (such as a difference in sex for two records that match on all other linkage fields for the individual).

Measuring how well the State is integrating their traffic data sets ensures that the highest quality linked data is being used for these safety efforts.

6.7 Questionnaire

The following question is intended to identify how the State measures the accuracy of the crash data. This question evaluates how the state reports the measure.

89. What output measures of integration are currently reported for your State? (AQ-70) (SOE: Provide baseline current data and trend if available)

Chapter 7: Accessibility

7.1 The Purpose of Measuring Accessibility

Accessibility is one of the most difficult quality attributes to understand and achieve, but it is also one of the most critical. All the efforts to improve a crash database's timeliness, accuracy, completeness, uniformity, and integration are a missed opportunity if the data is not accessed and used by the decision makers who need it. Data driven decision-making is the proven, best way safety improvements are made with limited resources. Data driven decision-making limits the risk of waste and misapplication of solutions compared to when decisions are based on perception or political will rather than quantifiable reality. Furthermore, providing access to the crash data for a variety of users results in multi-prong feedback on other data quality attributes, especially accuracy and completeness. If the users identify errors and missing data, the State can use that information to determine if and how the problems should be addressed.

Measuring crash data accessibility will allow data managers to:

- Quantify how well users are served by the crash data system.
- Demonstrate compliance with existing laws and policies regarding data access and security.
- Identify unmet user needs.
- Assess the need for new technologies and access methods to better serve users.
- Document the value of the crash data system to users and senior managers.
- Communicate information about crash data access to users and potential users.

7.2 Defining Accessibility

Accessibility in crash data systems means that users can obtain the data they need in easy-to-use formats and with appropriate controls in place to safeguard sensitive, personally identifying information (PII) and other data elements for which the State restricts release. Users employ crash data in different ways and for different purposes; States must decide for themselves which uses they will support. However, users' perceptions of their ability to access the data they need defines overall success.

According to the NHTSA (2011) *Model Performance Measures for State Traffic Records Systems:*

> Accessibility reflects the ability of legitimate users to successfully obtain data (p. 15). For every database and file in a traffic records system, there is a set of legitimate users who are entitled to request and receive data. The accessibility of the database or sub-file is determined by obtaining the users' perceptions of how well the system responds to their requests. Some users' perceptions may be more relevant to measurement of accessibility than others. Each database manager should decide which of the legitimate users of the database would be classified as principal users, whose satisfaction with the system's response to requests for data and other transactions will provide the basis for the measurement of accessibility (p. 4).

This definition focuses on the satisfaction of principal users and implies that accessibility is based on the documented needs and perceptions of these users. States may, however, go beyond this definition to take into account the needs of all users. For example, some States make their crash data available on the internet so that anyone with an interest can obtain the database. Typically, this is a public-release version of the database from which personally identifying information (PII) has been removed. Even though the goal may be full public access to the data, States may still wish to collect some information on who is using the data and what analyses are being performed. Voluntary requests for contact information and user surveys may be helpful in assessing accessibility for all users. This information can help identify important constituents and their safety concerns. User feedback may also highlight new uses for the data, identify unanticipated problems, and lead to new opportunities to improve data quality overall. One example is when safety advocacy groups (e.g., bicycle, pedestrian, etc.) use the data and provide feedback to the State requesting that more detail be added regarding non-motorized use of the roads.

7.3 Assessing Accessibility

NHTSA strongly encourages States to develop data quality performance measures, and build such measures into the crash data management software as automated reporting functions whenever possible. Accessibility measurement is different from the other five data quality measures because it cannot be assessed based on the contents of the crash database or a simple count of the number of users logging into a data system. In order to measure accessibility, States need to correctly identify and categorize users. Users' level of satisfaction is the most direct way to measure accessibility, by answering the question, "Did the users get the data they needed?"

Surveys can serve as helpful tools for gathering quantifiable data on the satisfaction levels for user groups. Automation may be helpful in conducting and reporting the results of surveys. A State could, for example, set up an electronic (e.g., web-based) survey of user satisfaction and ask users to complete it after each data or analytic request is closed out by the State crash data managers or analysts. The electronic survey results could be compiled in a database and reported periodically. If the State chooses to make data or reports available online, it may also wish to track users' page visits and downloads. This information can quantify the level of user access, but does not address user satisfaction. Another option is to prompt online users to complete a satisfaction rating upon exiting the site.

Each State must develop its own metrics for accessibility—setting targets for both the number and variety of uses as well as the overall level of user satisfaction. Ideally, the statistics would address principal users separately from all other user groups so that the State can readily provide the data quality measure described in NHTSA (2011) *Model Performance Measures for State Traffic Records Systems.*

7.4 Developing Accessibility Performance Measures

The goal of accessibility measurement is to provide useful information for data managers and users. In developing accessibility performance measures, system managers should consider what is most important to the various stakeholders and the managers of the system.

The *Model Performance Measures for State Traffic Records Systems* report (2011) defines one measure for crash data accessibility:

C-X-1: Crash Data Accessibility for Principal Users. (p.15)

This measure is intended to allow a State to quantify the use of the products from the state crash data system. States are encouraged to get feedback from the principal data users. Each database manager should decide which of the legitimate users would be classified as "principal." At a minimum, the list of principal users should include those who conduct State-sanctioned analyses and issue official reports. Ideally the Traffic Records Coordinating Committee would help decide which users should be included in this list.

7.5 Examples of Accessibility Measures

Measure 1: (C-X-1) - Measure of accessibility through user assessment:

It is intended for the State to assess the level of satisfaction the primary users have with the accessibility of the crash system. Working through the defined process, the State needs to perform the following items listed in the *Model Measures Report* (2011):

- 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.

Once the principal users are identified, the State can ask those users who request crash data or an analysis to periodically complete a survey. Figure 7.1 provides an example of a user satisfaction survey a State might use.

The survey asks principal users about their ability to access the data and satisfaction with the data received. To create the suggested measure of accessibility, the answers to Questions 4, 5, and 6 are needed. These questions are framed to gather more information on the users' ability to obtain the data, overall satisfaction with the data, and timeliness. As shown in this example, a State may also take the opportunity to ask users additional questions about data uses and experiences.

1. Agency Name: ____

2. What was the purpose for the requested data?
   - ○ Site Specific engineering analysis
   - ○ Population or site-specific behavioral analysis
   - ○ Aggregate engineering analysis (e.g., multiple locations/regional/statewide)
   - ○ Aggregate behavioral analysis (e.g., multiple groups or statewide)
   - ○ Other: Please explain

3. What do you plan to accomplish with the data (mark all that apply)?
   - ○ Calculate crash frequencies and rates
   - ○ Calculate trends over time
   - ○ Compare safety among locations, juridictions, or groups of people
   - ○ Integrate crash data with other safety data resources
   - ○ Other: Please explain

4. Were you satisfied with your ability to access the data?
   - ○ Fully Satisfied
   - ○ Satisfied
   - ○ Partially Satisfied
   - ○ Not Satisfied

5. Were you satisfied with the data that you received?
   - ○ Fully Satisfied
   - ○ Satisfied
   - ○ Partially Satisfied
   - ○ Not Satisfied

6. Were you satisfied with the response time for the data request?
   - ○ Fully Satisfied
   - ○ Satisfied
   - ○ Partially Satisfied
   - ○ Not Satisfied
   - ○ Not applicable: Please explain

7. How would you rate the data's accuracy?
   - Excellent
   - Acceptable
   - Unacceptable
   - Don't know

8. How would you rate the data's completeness?
   - Excellent
   - Acceptable
   - Unacceptable
   - Don't know

9. Are there any improvements that you would like to see in this system or the data you received?

   ____

   ____

   ____

   ____

Figure 7.1. Example User Satisfaction Survey

To create the accessibility measure, individual responses are aggregated to produce period summaries of user satisfaction. This is further illustrated in Table 7.1, which displays aggregate data collected in surveys of principal users. The table shows percentages for each score on questions 4 through 6, the average score over all surveys collected, and the total number of responses collected for each question.

Table 7.1. Agency Response Survey Table from Principal Users Satisfaction Survey

Response	Q4 – Access Satisfaction	Q5 – Data Satisfaction	Q6 – Timeliness Satisfaction
1-Fully Satisfied	10%	4%	60%
2-Satisfied	30%	20%	38%
3-Partially Satisfied	40%	70%	2%
4-Not Satisfied	20%	6%	0%
5-Not Applicable	-	-	0%
Average Response Score	**2.7**	**3.3**	**1.8**
Number of Responses	**97**	**89**	**77**

The average response score is calculated from the mean of the individual responses. On average, principal users were satisfied or partially satisfied with access and data quality, and were fully satisfied or satisfied with the timeliness with which requests was completed. This measurement could provide useful information for the State to evaluate the process of how data requests are handled and look for ways to make it more efficient.

Measure 2: Web Analytics

When States invest in making data, reports, and query tools available via the internet, they may wish to document how that investment has paid off in terms of increased user access to crash information and a corresponding decrease in requests requiring staff time to complete. Web analytics are available through a third-party subscription. As shown in *Table 7.*, these tools provide statistics on numbers of visitors, page hits, downloads, and other data that a State may use to monitor site usage. While this information does not measure user satisfaction, it can provide important indicators of the services provided to users and help to identify repeat users. States may wish to survey some or all of these users as well as the principal users discussed in Measure 1. In an ideal system, the survey would be presented online and the results aggregated automatically. States may also wish to quantify the impact that placing information and resources on the web has on the number and complexity of requests requiring staff involvement. Table 7.2 includes measures of the total time staff spent answering requests for data, analysis, or information.

Table 7.2. Example Web Site Usage Statistics for Use in Accessibility Measurement

YTD Web Statistics	Item Count
Number of visits	12,268 visits
Number of unique visitors	7,405 visitors
Number of data extracts created	130 data extracts
Number of queries completed	251 queries
Number of pre-defined reports downloaded	1072 downloads
Staff time spent on requests for assistance	76.5 hours
Number of requests for staff assistance	174

The State agency's IT division or the web site hosting this information can generate far more detailed reports than the example provided. Ideally, the State can create this report on a monthly basis and track accessibility measures such as the number of times that a page is accessed and the number of unique visitors. Trends over time will indicate if the user base is increasing, decreasing, or remaining stable. The concurrent tracking of requests for assistance can show, over time, if use of the web-based resource correlates with a reduction in the number of requests and the staff time required to meet those requests.

7.6 Final Considerations

In order to measure accessibility, States need to correctly identify and categorize users. Principal users' level of satisfaction is the most direct way to measure accessibility. Surveys are helpful tools for gathering quantifiable data on the satisfaction levels for user groups. States are encouraged to develop additional measures, including measure that use web analytics and surveys of a broader group of users. NHTSA encourages States to use such performance measures and integrating the results into a formal data quality management program, as discussed in the *Traffic Records Program Assessment Advisory* (NHTSA, 2012).

States have legitimate concerns regarding data release including the illegal and damaging release of PII and the potential for misuse of the data by users who don't fully understand the data structure, definitions, and limitations. Release of PII is generally mitigated by creating special, public-use data sets from which all PII has been redacted. If sensitive information is still included in the dataset (such as injury types), users are often blocked from accessing individual records and any results tables with small record counts are blocked or limited. These steps help to make it more difficult for a user with illegal or unsanctioned intentions to use the crash data as a way to zero in on the identity of any specific crash-involved person. States may also choose to limit the completeness or specificity of location information provided by public-use systems. This is usually not satisfactory for some user types (e.g., the State Department of Transportation, metropolitan planning agencies, and others who need the precise location information). Such limits on publicly accessible data may, however, help the State avoid external publication of location-specific analyses by users who are not well versed in the limitations and proper use of the data. Such restrictions on public-use datasets are not typically addressed in a State's performance measures of data accessibility. If, because of restrictions, users are having problems obtaining the data they need, however, it is expected that the user survey results will reflect some level of dissatisfaction. The State may wish to use the survey results as part of a periodic re-evaluation of its data release policies.

7.7 Questionnaire

The following question is intended to identify how the State measures accessibility of the crash data. This question evaluates how the state reports the measure.

90. What output measures of accessibility are currently reported? (AQ-71) (SOE: Provide baseline current data and trend if available)

Chapter 8: MMUCC Mapping Program Guide

8.1 Introduction

Background

The Model Minimum Uniform Crash Criteria Guideline (FHWA, FMCSA, &. NHTSA. 2012) referred to as the *MMUCC Guideline*, recommends a minimal set of standardized data elements and attributes that States should use for describing the characteristics of motor vehicle crashes, the vehicles, persons, and environment involved. Since the initial release of the *MMUCC Guideline 1st Edition* in 1998, this data set has been revised three times, to better support improved highway safety decision-making within States and nationally. The 4th edition is the most recent edition, containing 110 data elements. These are classified as "Data Elements Collected at Scene" and "Derived and Linked Data Elements." In each of these editions, the *MMUCC Guideline* has not provided States with guidance on implementation.

States have established their own data collection guidelines, resulting in substantial variation regarding the specific crash data collected. Some States use different formats and names for data elements and attributes, or combine (or split) elements and attributes in their PCRs and crash databases. Consequently, it is difficult to compare or share crash data among States, between State and Federal data sets, and—in some cases—between different agencies within a State. States are encouraged, but not required, to be more consistent with the *MMUCC Guideline, 4th Edition* (FHWA, FMCSA, &. NHTSA. 2012)*,* for their PCRs and crash databases, at both the element and attribute levels.

To assist States in evaluating their consistency with MMUCC, the Governors Highway Safety Association and NHTSA have developed guidance in their recent publication, *Mapping to MMUCC: A Process for Comparing Police Crash Reports and State Crash Databases to the Model Minimum Uniform Crash Criteria* (GHSA & NHTSA, 2015), hereafter called the *GHSA/NHTSA Mapping Process*. The process standardizes how States should compare both their PCRs and crash databases to MMUCC through establishing mapping rules.

For many of the MMUCC data elements throughout this chapter, there are sub-sections labeled "additional considerations" that present lessons learned from three pilot tests of the MMUCC mapping process. The notes clarify and expand upon the MMUCC mapping guidance and are consistent with NHTSA's and GHSA's presentation of the mapping process and rules.

Purpose

The purpose of this chapter is to build upon the *GHSA/NHTSA Mapping Process* to provide detailed guidance on mapping State PCRs and crash databases to MMUCC while following the rules outlined in the *GHSA/NHTSA Mapping Process*. The intention of this chapter is to serve as a self-assessment tool for a State to quantify the agreement between their PCR and the *MMUCC Guideline*—and their crash databases and the *MMUCC Guideline*—at both element and attribute levels. Throughout this chapter, the word "element" refers to the data fields on a PCR or in a database and the word "attributes" refers to the values that an element may take on.

Benefits

By conducting a MMUCC mapping assessment a State can determine and prioritize changes they could implement to increase their agreement with the *MMUCC Guideline*. States can use the information gleaned from the mapping process to plan updates or revisions to their PCR or to their crash database. This standardized process gives States vital information for implementing the *MMUCC Guideline*, which should also improve uniformity of crash data nationally.

Organization

This chapter is organized into the following five sections:

- **Introduction:** information on background, purpose, and benefits of this program guide;
- **Process for mapping a State PCR to MMUCC:** preparing for a PCR mapping assessment, general mapping rules and commentary element by element, and an example of how to apply the mapping rules;
- **Process for mapping State crash database to MMUCC:** preparing for a database mapping assessment, general mapping rules and commentary element by element, and an example of how to apply the mapping rules;
- **Computing compatibility scores and ratings:** methods for calculating compatibility scores for (a) each individual element, (b) overall PCR to MMUCC, and (c) overall crash database-to-MMUCC mappings, and the four-scale rating system;
- **Summary:** general discussion of mapping to the *MMUCC Guideline*.

8.2 Process for Mapping a State PCR to MMUCC

This section provides States with detailed guidance on how to map their PCRs to MMUCC. In general, the mapping process contains two primary steps: preparation and review following the mapping rules. These following sections provide additional information.

Preparation

Prior to conducting a thorough review of mapping State PCRs to MMUCC, the person conducting the review (the assessor) needs to obtain documentation for both the source and target data elements. The documentation includes, but may not be limited to:

- A PCR identifying all data elements and any attributes defined on the form.
- Any associated PCR overlay listing attributes for elements on the form.
- Any instruction manuals associated with the PCR, providing definitions for elements on the PCR as well as all attributes for those elements.

As a part of the preparation, States are encouraged to use the NHTSA Mapping worksheet to map each of the data elements and associated attributes. This spreadsheet is used for a direct comparison of the PCR elements and attributes to the elements and attributes in the *MMUCC Guidelines.* Elements in *MMUCC* are grouped into two categories: “collected at scene” and “linked or derived.” As indicated in the *GHSA/NHTSA Mapping Process,* the “collected” elements in MMUCC are separated by their MMUCC classification: Crash Data Elements, Vehicle Data Elements, and Person Data Elements.

Interested agencies can obtain the NHTSA-developed mapping spreadsheet, which follows this structure one of three ways.

- GHSA’s website at www.ghsa.org/html/issues/mmuccmappingdraft.html
- NHTSA’s website at www.nhtsa.gov/DOT/NHTSA/NVS/TrafficRecords/MMUCC_Mapping_v5.xlsx
- Contact the NHTSA Regional Office

The following elements are included within each worksheet:

- **Column A:** Lists the number of each MMUCC element in order, such as C1, C2.
- **Column B:** Lists the name of each MMUCC element.
- **Column C:** Lists all the MMUCC attributes associated with the MMUCC element. If a MMUCC element has subfields, Column C is the list of attributes for the first subfield.
- **Column D:** The space for recording whether the PCR has an element/attribute that can be mapped to the corresponding MMUCC element/attribute. Enter "1" if the PCR element/attribute matches to the MMUCC element/attribute, otherwise, enter "0."
- **Column E:** A field for recording the specific PCR element/attribute name and number that matches to the MMUCC element/attribute.
- **Columns G, H, and I (and subsequent column groups):** List of MMUCC attributes for additional subfields of a MMUCC element, if that MMUCC element has multiple subfields.
- **Column O:** Computes the percentage mappable for each MMUCC element. This is the mapping score for each MMUCC element.

Mapping Rules

Once the assessor has set up mapping tables and gathered the needed documentation, they can start conducting a thorough evaluation using the mapping rules and specific notes described in this section. There are both general rules and rules/notes for individual data elements.

General

The *GHSA/NHTSA Mapping Process* has established general rules for mapping a PCR to MMUCC. As indicated in the *GHSA/NHTSA Mapping Process,* States should adhere to the following rules for most of the elements when mapping State PCR to *MMUCC Guideline*:

1. The *MMUCC Guideline* assumes that States will collect data for all types of crashes (e.g., fatalities, serious injuries, non-motorist, and commercial motor vehicle). Exclude State data elements collected or coded onto a crash database for only a subset of all crash types (e.g., only crashes involving a fatality) from the MMUCC mapping process.
2. The State PCR element name does not need to match the MMUCC element name, but the definition should be essentially the same. The reverse is not true. If a PCR element/ attribute has the same name as a MMUCC element/attribute, the definitions must be the same for a match. Mapping by name alone will result in errors.
3. Similarly, a State PCR element/attribute may be mapped to a MMUCC element/attribute even if the same term (or name) is not used as long as the State term is synonymous and unambiguous, or has the same definition.
4. An element/attribute on a State PCR that is "close enough" should not be mapped because it will be difficult for others to understand and will corrupt data integrity.
5. If the MMUCC element has multiple reporting iterations (subfields), the matching State element must have opportunities to code as many times as the MMUCC element has Subfields. For example, the MMUCC Element Contributing Circumstances, Road (C15) has 3 Subfields: Road Circumstances 1, Road Circumstance 2, and Road Circumstances 3. If the State PCR only allows

for reporting one subfield, then the PCR would map only to the first Subfield for MMUCC Element C15 and would not map completely.

6. A single attribute of a State PCR element may be mapped only to one MMUCC element/attribute. Suppose a State element "Roadway Conditions" has an attribute of "Snow." It may not be mapped to both the attribute "Snow" in MMUCC element Weather Conditions (C11) and the MMUCC Element Roadway Surface Condition (C13) attribute "Snow." Because the State PCR element is "Roadway Conditions," mapping to MMUCC Element C13 may be more appropriate. The State must choose one and only one.

7. If a State PCR element has an attribute that combines several terms (i.e., it has a broad definition), it may not be mapped to MMUCC element/attributes that are included in that broad definition. For example, a State's attribute "Frozen Precipitation" may not be mapped any of the four MMUCC Element Weather Conditions (C11) attributes "Snow," "Blowing Snow," "Sleet or Hail," or "Freezing Rain or Freezing Drizzle" because it does not distinguish between the four possibilities.

8. Two or more elements on a PCR may map to a single MMUCC element. For example, the MMUCC Element Restraint Systems/Motorcycle Helmet Use (P8) may be listed as separate State elements "Restraint Systems" (or "Occupant Protection") and "Motorcycle Helmet" on the State PCR.

9. If an element on a State PCR has attributes that map to attributes included in separate MMUCC Elements, they are permitted to make a match so long as individual State PCR element attributes are not mapped to more than one MMUCC element/attribute.

10. A State PCR element that is reported as an open text field—the officer either writes or types in the information and is not limited to a specific set of possible values—may be used to map to a MMUCC Element only if the PCR instruction manual clearly indicates what should be written/typed in the field.

11. PCR-to-MMUCC mapping is done only at the element/attribute level. If a MMUCC element/attribute is present on the PCR (in some way, including as a "freeform" or "text" field), then the State element/attribute maps to the MMUCC element/attribute. The number of characters allowed should not matter.

12. For a State element attribute "Other" to map to a MMUCC element attribute "Other," the State element must possess all of the specific attributes for the MMUCC element in question. For example, Subfield 1 of MMUCC Element Transported to First Medical Facility by (P28) has the following attributes: Not Transported, EMS Air, EMS Ground, Law Enforcement, Other, and Unknown. If the State element being mapped has the first four MMUCC Attributes, then the State attribute "Other" would map to the MMUCC "Other." But if the State element does not have an attribute that mapped to one of the four (for example, EMS Air), then "Other" would also not map. The State "Other" treats EMS Air as a possible undefined value. A visual representation is shown in Figure 8.1 and Figure 8.2.

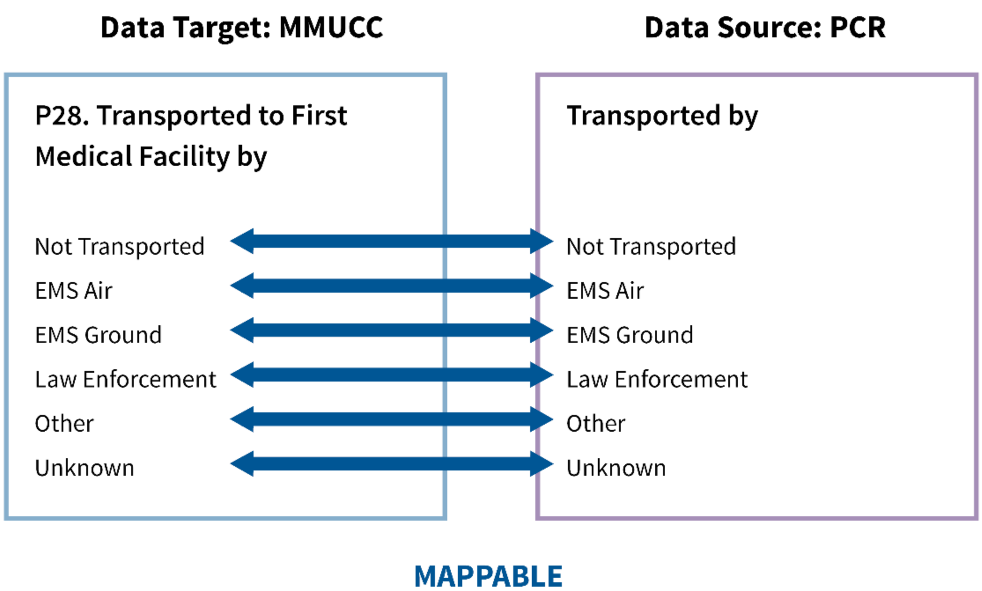

Figure 8.1. State PCR Attribute "Other" Can Be Mapped to the MMUCC

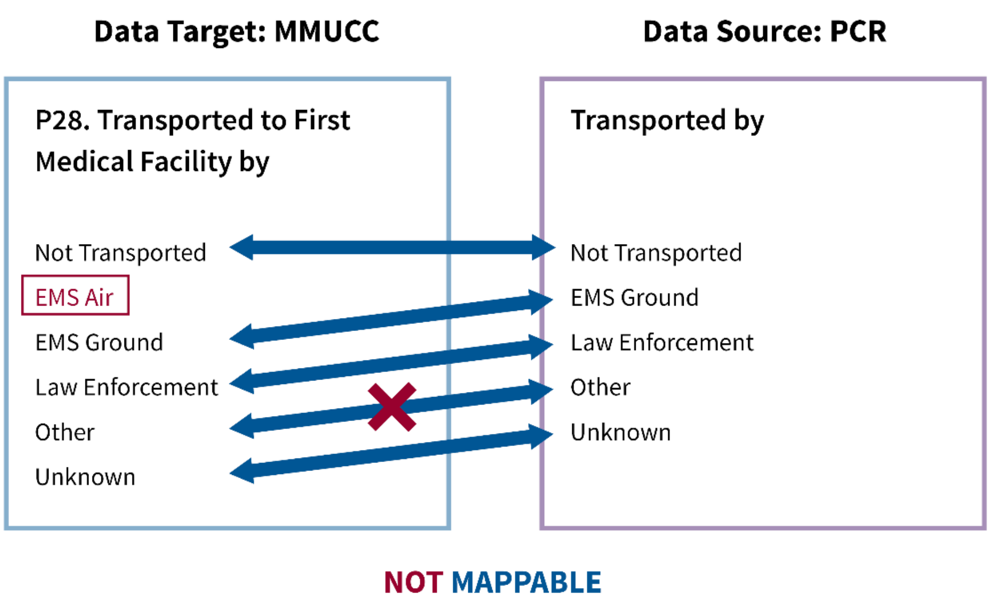

Figure 8.2. State PCR Attribute "Other" Cannot Be Mapped to the MMUCC

13. If a MMUCC element has both attributes "Other" and "Unknown," then the State attribute "Unknown" will map only if the State element also has the attribute "Other," regardless of whether or not "Other" mapped. For the previously cited example, it does not matter if the State element being mapped has all five MMUCC Attributes (including "Other"), only that the State element has an attribute "Other."

14. However, if a MMUCC element has the attribute "Unknown" but does not have an attribute "Other," then the situation is similar to that outlined in General Rule 12 for the attribute "Other." For example, MMUCC element Speeding Related (P13) has the attributes "Racing," "Exceeded

Speed Limit," "Too Fast for Conditions," "No," and "Unknown." If the PCR element being matched has an attribute "Unknown," it must have exact matches to the first four MMUCC attributes with no other attributes in order to have a match to "Unknown."

15. If the list of element attributes on a PCR does not include a value for "Other" or "Unknown," they can be mapped to a MMUCC element/attribute list. This can only occur if the instructions for completing the State PCR directs the officer to enter a code (e.g., "00," "99," "UNK") in the event the appropriate response is "Unknown" or falls into an "Other" category.

Individual Data Elements

The remainder of this section presents only those individual MMUCC elements for which the GHSA, NHTSA, and CDIP teams developed commentary derived from the pilot tests of State PCR-to-MMUCC mappings. These are considered as supplementary and explanatory information and do not replace the guidance and notes provided in the *GHSA/NHTSA Mapping Process.*

Crash Elements

This section provides the specific notes for each crash element collected at the scene.

C2. Crash Classification

Notes from the GHSA/NHTSA Mapping Process

State PCRs may have a "Private Property" check box. This can be mapped with the first subfield. (Not selecting "Private Property" would be equivalent to indicating that the crash occurred on public property.) Additionally, if the State does not have a separate element to identify public versus private property but has an element that classifies the location by the type of road on which the crash occurred (e.g., Interstate, Primary, Secondary), then that element can be used to match to this subfield if it includes the attribute "Private Road" or something similar.

Additional Considerations

Previous to mapping, the assessor should understand the definition of each individual attribute.

For mapping to Subfield 1, the State PCR could have a check box of "Private Property" or an element such as "Roadway System" that can distinguish private property from public property. In general, the State PCR does not need the exact same element; as long as the PCR element can be used to identify ownership of the land where the crash occurred, it is considered mappable.

For mapping to Subfield 2, the State PCR needs to have an element/attribute that can be used to identify the characteristic of the crash with respect to its location on or off a trafficway.

The *MMUCC Guideline* Appendix E provides illustrations of a trafficway. Note the Subfield 2 is different from MMUCC element C8. Location of First Harmful Event Relative to the Trafficway**.** The MMUCC element C8 only captures the location of the first harmful event, while Subfield 2 captures where the unstabilized situation originated and the harmful events occurred.

C7. First Harmful Event

Notes from the GHSA/NHTSA Mapping Process

To map to this MMUCC element, the State PCR must have a similar element at the crash level. This MMUCC element refers to the first harmful event occurring in the entire crash.

Additional Considerations

The *MMUCC Guideline* categorizes this element's attributes into three types, including non-collision; collision with person, motor vehicle, or non-fixed object; and, collision with fixed object. For successful mapping, attributes in the State PCR must have the same definitions as the *MMUCC Guideline*. However, the category in which the attribute is listed on the State PCR is irrelevant. For example, the *MMUCC Guideline* has the attribute "Thrown or Fallen Object" listed under the event type "Non-Collision." The State PCR may have this attribute listed under a different type of event, such as "Collision With Person, Vehicle, or Non-Fixed Object." In this case, the State should still get credit for mapping to this MMUCC attribute as long as the definitions are the same.

The assessor should also consider the following when mapping to the attribute "Other:"

- To map to the attribute "Other Non-Collision," the State PCR must possess all of the seven attributes for event type "Non-Collison."
- In order to map to "Other Non-Motorist," the State PCR must have attributes that can be mapped to both MMUCC attributes "Pedestrian" and "Pedalcycle."

  Note "Bicycle" cannot be mapped to the MMUCC attribute "Pedalcycle" since a pedalcycle includes tricycles, unicycles, pedal cars, and other types of cycles in addition to bicycles.
- To map to "Other Non-Fixed Object," the State PCR must have the exact same attributes for the list of "Collision With Person, Motor Vehicle, or Non-Fixed Object" as the *MMUCC Guideline*.
- To map to "Other Post, Pole, or Support" and "Other Fixed Object" (wall, building, tunnel, etc.), the State PCR must have the same attributes for the list of post, pole or support and for the list of fixed objects for the event type of "Collison With Fixed Object."

C9. Manner of Crash/Collision Impact

Notes from the GHSA/NHTSA Mapping Process

Diagrams of collision types are acceptable if the diagram unambiguously represents the same collision types as the corresponding MMUCC attributes and as explained in Appendix C of the *MMUCC Guideline*.

Additional Considerations

Some States may have these attributes under element "Crash Type." As long as the definitions are the same as the MMUCC attributes, they are mappable.

The *MMUCC Guideline* provides illustrations for each manner of collisions in Appendix F. Please refer to these illustrations and definitions before conducting mapping to ensure that the attributes in the PCR represent the exact same manners of collisions as in the *MMUCC Guideline*. For example, some State PCRs may use "Right Angle" crash rather than "Angle" crash and define the crash type as "two vehicles approaching from non-opposing angular directions collide." However, in the *MMUCC Guideline*, the angle crash can be caused by two vehicles approaching from any direction. In this case, the State cannot get credit for this mapping due to the definition differences. Some State PCRs may use the crash type "Backing," which is the combination of MMUCC attributes "Rear to Side" and "Rear to Rear." If so, the State PCR cannot map to either of the two MMUCC attributes.

C14. Contributing Circumstances, Environment

The *MMUCC Guideline* also defines some of the attributes as follows:

- **Glare**: A very harsh, bright, dazzling light that impairs vision.
- **Visual Obstruction**: An object that blocked the driver's sight, contributing to the crash (e.g., bush, tree).

Notes from the GHSA/NHTSA Mapping Process

Weather conditions reported in a separate element corresponding to the MMUCC element C11Weather Conditions should not be counted for the "Weather Conditions" attribute of MMUCC element C14 Contributing Circumstances, Environment.

Additional Considerations

First, to map to all the subfields, the State PCR must allow officers to code for environmental contributing circumstances at least three times.

If the State PCR has environmental contributing circumstances integrated with other types of contributing circumstances, such as roadway contributing circumstances and motor vehicle contributing circumstances, the assessor should not give credit for mapping to this MMUCC element because the attributes cannot be disambiguated for each type of contributing circumstances even though the attribute lists may match.

If the State PCR does not include this specific data element, but has some of the attributes included in other elements such as "Visual Contributing Circumstances." In this case, the attributes of "Visual Contributing Circumstances" can be mapped to the MMUCC attributes if the definitions match.

C15. Contributing Circumstances, Road

Notes from the GHSA/NHTSA Mapping Process

Attributes should not be combined in one field.

Additional Considerations

First, to map to the three subfields, the State PCR must allow officers to code at least three times for road contributing circumstances.

Some State PCRs have roadway contributing circumstances integrated with environmental or motor vehicle contributing circumstances, such as environmental contributing circumstances and motor vehicle contributing circumstances. If that is the case, the assessor should not give credit for mapping to this MMUCC element. This is because the attributes cannot be disambiguated for each type of contributing circumstance even though the attribute lists may match.

C16. Relation to Junction

Additional Considerations

Prior to mapping, the assessor should understand the definition of each individual attribute, as listed above. It is important to understand the differences between "Intersection" and "Intersection Related," "Driveway" and "Driveway Access Related," "Entrance/Exit Ramp" and "Entrance/Exit Ramp Related."

To successfully map to Subfield 1: Within Interchange Area, the State PCR must have a similar data element indicating whether the crash occurs within an interchange area or not.

Note that if the State PCR has a data element indicating whether a crash is an intersection crash, it cannot be used to map to Subfield 1 due to the inconsistency with the definition of intersection versus interchange. The *MMUCC Guideline* Appendix G provides illustrations of an interchange and Appendix H provides illustrations of an intersection.

In order to map to attributes of the Subfield 2, the assessor should carefully compare the PCR attribute to MMUCC attribute even though similar terms may be used on the State PCR. For example, a State PCR may have an attribute "Crossover in Median" or "Median Crossing," which cannot be mapped to the MMUCC attribute "Crossover-Related" because crossover-related crashes do not need to occur on a crossover while a crossover-in-median crash does.

The State will get credit if they use "On/Off Ramp" instead of "Entrance/Exit Ramp." The State can also get credit using "On Ramp Merge Area and Off Ramp Diverge Area" instead of "Acceleration/Deceleration Lane" so long as the PCR also has attributes for intersection crashes and crossover crashes.

C19. Work Zone Related

Notes from the GHSA/NHTSA Mapping Process

If the State PCR combines Subfield 4, Workers Present and Subfield 5, Law Enforcement Present, it must allow at least two entries since MMUCC is looking for both the presence of workers and of some type of law enforcement presence.

Additional Considerations

State PCRs may have a similar element, such as "Temporary Traffic Control Zone," which can be mapped with the first subfield. That is, selecting "No" would be equivalent to indicating that the crash is not within a work zone and selecting other types of temporary traffic control zone would be equivalent to selecting "Yes" for the first subfield of the MMUCC element. If the matching PCR element has the attribute "Unknown" or the instructions direct the Officer to enter a code for "Unknown," the MMUCC attribute "Unknown" is considered mappable.

Vehicle Elements

There are 30 vehicle elements the *MMUCC Guideline* suggests to be collected at the scene of each crash. This section contains MMUCC definitions, MMUCC attributes, specific notes, and commentary for each vehicle element.

V2. Motor Vehicle Unit Type and Number

Notes from the GHSA/NHTSA Mapping Process

Most States will have "Unit Number" separate from "Unit Type." States with PCRs that have the same attributes under separate elements can map these attributes to those in MMUCC element V2 Motor Vehicle Unit Type and Number.

Additional Considerations

If the State PCR does not have a separate element to identify motor vehicle unit type, but has an element at vehicle level to indicate the vehicle use (e.g., "Personal, Machinery in Use"), attributes of that element can be used to map to attributes of the subfield with the same definitions.

State PCRs may have a "Parked" check box at the vehicle level. This can be mapped with the MMUCC attribute "Parked Motor Vehicle" because selecting "Parked" would be equivalent to selecting the MMUCC attribute "Parked Motor Vehicle."

If the State PCR has attributes that can be mapped to the MMUCC attributes "Parked Motor Vehicle" and "Working Vehicle/Equipment," then the State would get credit for mapping to the MMUCC attribute "Motor Vehicle in Transport" regardless whether they have this specific attribute because it can be derived based on not selecting the other two attributes.

V3. Motor Vehicle Registration State and Year

Notes from the GHSA/NHTSA Mapping Process

State of registration and year of registration are often two separate fields on State PCRs. This is acceptable.

Additional Considerations

The *MMUCC Guideline* Appendix D provides a complete list of States.

In order to successfully map to the MMUCC attribute "State Identifier," the State PCR needs to have values related to State, foreign country, U.S. government, Indian Nation, and so forth.

V8. Motor Vehicle Body Type Category

Additional Considerations

Note that this MMUCC element describes body type, not vehicle use. So a State PCR that only lists vehicle uses—such as "School Bus" or" Transit Bus"—is not an acceptable mapping for those corresponding V8 Motor Vehicle Body Type attributes.

Some State PCRs do not have specific attributes equating to the MMUCC attributes "Other Light Trucks (10,000 lbs GVWR or less)" and "Medium/Heavy Trucks (more than 10,000 lbs GVWR)." In those cases, the State PCR attributes indicating truck type cannot be mapped to the MMUCC attributes in combination with GVWR/GCWR (Gross Vehicle Weight Rating/Gross Combined Weight Rating) attributes. For example, if the State PCR has an attribute "single unit truck (2 axles)" and a check box for "10,000 lbs or less," they cannot use the vehicle type in combination with GCWR/GVWR for this mapping.

V12. Motor Vehicle Posted/Statutory Speed Limit

Additional Considerations

Many States may have an open text field for speed limit. If instructions direct officers to enter a code for "Unknown" (e.g., "99," "00") or allow officers to leave it blank, this would be equivalent to having an "Unknown" attribute. The State would get credit for mapping to the MMUCC attribute "Unknown." Similarly, the State would get credit for mapping to the MMUCC attribute "Not Applicable" if instructions direct officers to enter codes for "Not Applicable."

V13. Direction of Travel Before Crash

Specific Notes from the GHSA/NHTSA Mapping Process

Arrow diagrams are sufficient if they clearly equate to the MMUCC attributes and follow the MMUCC definition of this element.

Additional Considerations

Many States may have a similar PCR element, such as "Direction of Travel of Vehicle." It would not be an acceptable mapping unless the instructions clarify that it is the State-designated direction of the road.

V14. Trafficway Description

Additional Considerations

The *MMUCC Guideline* Appendix E provide an illustration of a trafficway. States may have a PCR element "Roadway Description" or similar, rather than "Trafficway Description." Attributes of the State PCR element can be mapped to the corresponding MMUCC attributes as long as the definitions match.

V15. Total Lanes in Roadway

Additional Considerations

States may have an open text field for recording number of lanes. States will be given credit for this mapping if the instructions for completing the State PCR element match the MMUCC definition.

V18. Motor Vehicle Maneuver/Action

Additional Considerations

States may combine motor vehicle maneuver/action with non-motorist maneuver/action. They can get credit for this element if there is a way to unambiguously identify the actions of motor vehicles.

V19. Vehicle Damage

Specific Notes from the GHSA/NHTSA Mapping Process

A State diagram may be used to report both Subfield 1, Initial Contact Point on Vehicle, as well as for Subfield 2, Damaged Areas, if the former is unambiguously identified. A State diagram may contain more than the recommended 12 points (as long as those points can be mapped to the MMUCC 12-point diagram), but the State diagram may not contain fewer.

Additional Considerations

The *MMUCC Guideline* Appendix J provides a 12-point clock diagram.

State PCRs may have three separate fields for this MMUCC element. In order to map to the Subfield 1, the State PCR needs to have elements / attributes capturing the initial contact point. In order to map to the Subfield 2, the State PCR needs to have an element allowing officers to code as many times damaged areas as needed to record *all* damage to the vehicle. States may have an element recording the most damaged area, which is not an acceptable mapping for Subfield 2 because of the inconsistent definitions.

In order to map to the Subfield 3, the State PCR needs to have an element indicating the total damage to the motor vehicle. States may have a similar element, such as "Vehicle Deformity for Most Damaged Area" which cannot be mapped to Subfield 3 due to inconsistent definitions.

V20. Sequence of Events

Specific Notes from the GHSA/NHTSA Mapping Process

MMUCC element V20 includes non-harmful events as attributes, whereas State PCRs may include only harmful events. If this is the case, the PCR will not match to the non-harmful MMUCC attributes. Also note that V20 includes four subfields. Consequently, the State PCR must allow for four entries for its element to fully map to this MMUCC element.

Additional Considerations

The *MMUCC Guideline* Appendix L shows an example of how to encode sequences of events.

V21. Most Harmful Event for this Vehicle

Specific Notes from the GHSA/NHTSA Mapping Process

The State PCR must have a similar element for each vehicle involved for the PCR to match to this MMUCC element.

Additional Considerations

There are two "Unknown" attributes in this MMUCC element. To map to the first "Unknown" in category "Collision With Fixed Object" the State PCR must have an attribute "Unknown" or have the capability for officers to code (e.g., "99", "00") for this specific category. If the State PCR has the attribute "Unknown" or allows officers use codes for "Unknown," but does not tie to the category "Collision With Fixed Object" or other categories, this would be considered a match to the uncategorized, second MMUCC attribute "Unknown."

V23. Hit and Run

Additional Considerations

States may have a check box for a hit and run crash, which can be mapped to these two MMUCC attributes. This is because selecting the check box would be equivalent to the MMUCC attribute "Yes, Driver or Car and Driver Left Scene." Not selecting the check box would be equivalent to the MMUCC attribute "No, Did Not Leave Scene."

Some States may not have a specific element for hit and run crashes, but have a similar element such as "Disposition of Vehicle," with attributes "Hit and Run" and "Retained by Driver." The mapping would be acceptable if the definitions match.

V24. Towed Due to Disabling Damage

Specific Notes from the GHSA/NHTSA Mapping Process

A State PCR that has an element "Towed" as a checkbox or "Towed, Y/N" will map to this MMUCC element.

However, the attributes to map will depend on how the PCR instruction manual indicates that "Towed" is to be defined. If it means only "towed due to disabling damage" and not for other reasons, then it can only be mapped to the MMUCC attribute "Towed Due to Disabling Damage." The State PCR cannot map to the other two MMUCC attributes. If the PCR instruction manual is unclear as to whether being towed is due to damage, then the State PCR can only map to the State PCR attributes "unchecked box" or "N" to the MMUCC attribute "Not Towed."

Additional Considerations

The State PCR may lack this specific data element, but have an element indicating whether the vehicle was retained by the driver or was towed. If the State PCR does not include the attribute "Not Towed," whether or not selecting other attributes would be equivalent to "Not Towed" depends on how instructions direct officers to code. For example, the State PCR has attributes "Towed," "Retained by Driver," "Towed/Disabled," but does not have attribute "Not Towed." There are situations where the driver may engage a private company without law enforcement assistance, in which case the officer would code "Retained by Driver" rather than code "Towed." Failure to select any of the other attributes is insufficient proof that the vehicle was not towed.

V25. Contributing Circumstances, Motor Vehicle

Additional Considerations

In order to map to both subfields, the State PCR needs to provide opportunities to code this element at least twice.

A State PCR may combine "Contributing Circumstances, Motor Vehicle" with other contributing circumstances, such as roadway and environmental. In this case, the State would not be given credit for mapping to any of the contributing circumstances. The *MMUCC Guideline* explicitly separates the roadway, environmental, and vehicle contributing circumstance so that officers can choose independently within each of these categories. Combining all of them together will force officers to pick from all of the attributes across categories and thus does not meet the MMUCC mapping rules.

V26. Motor Carrier Identification

Additional Considerations

States may have separate fields for each of the MMUCC subfields. This mapping is acceptable as long as the definitions match.

V27. Gross Vehicle Weight Rating / Gross Combination Weight Rating

Specific Notes from the GHSA/NHTSA Mapping Process

For mapping purposes, a State PCR may either report the GVWR (the manufacturer's operational weight limit for a motor vehicle and any cargo) or the GCWR (the sum of all GVWRs for each unit in a combination unit motor vehicle, such as a truck tractor and trailer).

Additional Considerations

State PCRs may have a check box for each weight level, which is acceptable. If the State PCR does not have the MMUCC attribute "Not Applicable," it would be an adequate mapping if the instructions clarify that a blank in this field is interpreted as "Not Applicable."

V28. Vehicle Configuration

Additional Considerations

The *MMUCC Guideline* Appendix K displays types of truck configurations.

Some State PCRs may combine this MMUCC element with V8. Motor Vehicle Body Type Category. The attributes can be mapped to this MMUCC element as long as the definitions match. Note that if the attribute has been mapped to the previous MMUCC element V8. Motor Vehicle Body Type Category, the attribute should not be mapped here again since each attribute can only be mapped once.

V29. Cargo Body Type

Additional Considerations

In order to map to the MMUCC attribute "Other," the State PCR needs to have exact same attribute list for the corresponding PCR element.

In order to map to the MMUCC attribute "Unknown," the State PCR must have the attribute "Other" regardless of whether it can be mapped to the MMUCC attribute "Other." Additionally, the State PCR needs to have an attribute "unknown or instructions direct officers to enter a code for unknown, such as "99" or "00."

V30. Hazardous Materials (Cargo Only)

Additional Considerations

Many State PCRs may have separate fields for each of the Subfield, which is acceptable. If the State PCR does not have element match to the Subfield 1, but has elements for Subfield 2, the State would be given credit for mapping to the MMUCC attribute "Yes" of Subfield 1. Filling out HM ID number or Class number would be equivalent to "Yes." However, the State would not get credit for mapping to the MMUCC attributes "No" or "Not Applicable" of Subfield 1 because the values cannot be derived by not filling HM placard information. That is, it is difficult to distinguish between "No" and "Not Applicable" without HM placard information.

With respect to Subfield 3, many State PCRs may have checkboxes for "Yes" and "No," indicating whether or not hazardous material were released from the package. These two PCR attributes can be mapped to the corresponding attributes of Subfield 3. If the State PCR does not have an option for "Not Applicable" but the instructions clarify that not selecting would be equivalent to "Not Applicable," the State would be given credit for mapping to "Not Applicable" as well.

Person Elements

There are 28 person elements that the *MMUCC Guideline* describes to be collected at the scene of each crash. This section discusses the definitions, subfields, and attributes defined by the *MMUCC Guideline* as well as the specific notes and commentary for mapping.

P2. Date of Birth

Specific Notes from the GHSA/NHTSA Mapping Process

Note that the *MMUCC Guideline* definition for this element states that Subfield 2 "Age," is "to be used only if the date of birth cannot be obtained." Consequently, if the State PCR has Subfield 1, "Date of Birth," it need not have Subfield 2, "Age."

Additional Considerations

Note that this MMUCC element is to record age information for each person involved in the crash, including drivers, occupants, and non-motorists. A "Driver Age" or "Driver DOB" alone is insufficient. Age or DOB fields in separate sections of the PCR (i.e., driver section, occupant section) are acceptable.

If the State PCR does not have "Age" but has "Date of Birth," the State would still get credit for mapping to the MMUCC Subfield 2 since age can be derived from date of birth. However, if the State PCR only has an element for Subfield 2 "Age," they would not be given credit for mapping to Subfield 1 "Date of Birth" since the exact date of birth could not be derived based on knowing the age.

Many State PCRs have an open text field for "Date of Birth." States could get credit for mapping to the MMUCC attribute "Unknown" if the instructions demonstrate that blank is equivalent to unknown or if officers are instructed to indicate unknown values in that text field.

P3. Sex

Additional Considerations

Note that this MMUCC element is to record the sex of each person involved in the crash, including drivers, occupants, and non-motorists. A field "Driver Sex" alone is insufficient.

P4. Person Type

Specific Notes from the GHSA/NHTSA Mapping Process

If the State PCR has separate motorist and non-motorist sections, and in the non-motorist section it has separate attributes for pedestrians and pedalcyclists, then the State receives credit to mapping to this MMUCC element as Person Type can be derived.

Additional Considerations

States do not to have this specific data element as long as the values can be derived from other elements. For example, States may number the vehicle, pedestrian, and bicyclist differently, such as use number starting with "V" for vehicles (e.g., V1), number starting with "P" for pedestrians (e.g., P2), and number starting with "B" for bicyclist (e.g., B1). In this case, it is possible to derive person type from the numbering information.

Also note that pedalcyclist is different from bicyclist, as pedalcyclist includes other types of cyclists in addition to bicyclist. Thus, if the State PCR only has one attribute "Pedalcyclist," it would not be an acceptable mapping to "Bicyclist" or "Other Cyclist."

P5. Injury Status

Specific Notes from the GHSA/NHTSA Mapping Process

In accordance with the MAP-21 requirement that the U.S. DOT establish performance measures for reporting fatalities and serious injuries, the FHWA released a Notice of Proposed Rulemaking on March 10, 2014 (79 FR 13845), indicating that States will be required to use the definition of the MMUCC attribute, Suspected Serious Injury (A), of the MMUCC element P5 Injury Status to report serious injuries (to be effective 18 months after the final rulemaking).

For that reason, a State PCR serious injury attribute equivalent (for example, "Incapacitating Injury") must have the same definition as the MMUCC attribute, Suspected Serious Injury (A). Other Injury Status attributes may be mapped to MMUCC attributes using synonymous terms, such as "Dead" for "Fatal Injury (K)" or "Non-incapacitating Injury" for "Suspected Minor Injury (B)."

Additional Considerations

Many State PCRs may have the attribute "Incapacitating Injury (A)," which can only be mapped to the MMUCC attribute "Suspected Serious Injury (A)" if the definitions are exactly the same.

Other injury status attributes may be mapped to MMUCC attributes using synonymous terms, such as "Non-Incapacitating Injury" to "Suspected Minor Injury (B)," "No Injury" to "No Apparent Injury (O)," "Complaint of Pain" to "Possible Injury (C)," and "Killed" to "Fatal Injury."

P7. Seating Position

Specific Notes from the GHSA/NHTSA Mapping Process

A diagram is acceptable for mapping as long as all MMUCC position attributes are represented.

Additional Considerations

The *MMUCC Guideline* Appendix N includes diagram of common vehicle types. Please refer to this appendix for mapping and also include ambulance seating/positioning.

Some State PCRs may combine Subfield 1 and Subfield 2, such as having attribute "Front Left" and "Second Row Middle." It is an acceptable mapping if the State PCR has all of the combinations.

P8. Restraint Systems/Motorcycle Helmet Use

Additional Considerations

An attribute "Helmet" alone is insufficient to map to any of the helmet types in Subfield 2 except for the "No Helmet" attribute. The State would get credit for mapping to the MMUCC attribute "No Helmet" as this value can be derived by not selecting the attribute "Helmet."

If the State PCR element/attribute does not specify whether the helmet is compliant with DOT requirements but the instructions direct officers to code "Helmet" only when it is DOT-compliant, then the State would be given credit for mapping to the MMUCC attribute "DOT-Compliant Motorcycle Helmet."

P9. Air Bag Deployed

Additional Considerations

The *MMUCC Guideline* Appendix M provides a diagram of air bag types. States may not have attributes for "Not Applicable" or "Deployment Unknown." If the instructions specify that officers could use a code (e.g., "99," dash) for "Not Applicable" or "Deployment Unknown," the State would be given credit for mapping to these MMUCC attributes.

P10. Ejection

Specific Notes from the GHSA/NHTSA Mapping Process

An "Ejected" checkbox or "Ejected, Y/N" is sufficient to map to this MMUCC element. However, the MMUCC ejection attributes to which the State PCR maps depends upon how PCR instruction manual defines "Ejected." For example, if "Ejected" means completely ejected, then "Y" maps to the MMUCC attribute "Ejected Totally." The State PCR will not be able to map to any of the other MMUCC attributes for this element.

Additional Considerations

If the State PCR does not have attributes coding for "Not Applicable" or "Unknown" but the instructions direct officers to enter codes for these attributes—such as "99" for "Unknown" and "NA" for "Not Applicable"—the State would get credit for mapping to these MMUCC attributes.

P11. Driver License Jurisdiction

Specific Notes from the GHSA/NHTSA Mapping Process

An open text field is acceptable for mapping for the attribute "State." For an open text field to be acceptable for mapping to the other MMUCC attributes, the PCR instruction manual must indicate that a reporting officer may report the other specific jurisdiction types.

Additional Considerations

If the State PCR only has an open text field for "State," the State would not get credit for mapping to "Not Applicable," "Not Licensed," and "Unknown" unless the instructions direct officers to enter codes representing these attribute values. Additionally, if the instructions do not specify that "State" includes Indian Nation, U. S. Government, and other jurisdiction types listed in the *MMUCC Guideline*, the open text field can only be mapped to the MMUCC attribute "State."

P14. Driver Action at Time of Crash

Specific Notes from the GHSA/NHTSA Mapping Process

Driver Action attributes are sometimes included under Contributing Circumstances or other more general State elements. In such cases, State PCR attributes may be mapped to P14. Driver Action at Time of Crash attributes, but will fully map only if at least four circumstances/factors/actions may be reported. It will map partially if fewer than four are reported.

Additional Considerations

The assessor needs to pay attention to the following situations.

- Many States may not have the MMUCC attribute "Ran Red Light" but have a closely related attribute such as "Disregard Traffic Signals." Such closely related attributes are not equivalent to "Ran Red Light" given that "Disregard Traffic Signals" may include other situations such as ignoring flashing signals.
- An attribute "Disregard Traffic Signs" cannot be mapped to the MMUCC attributes "Ran Stop Signs" or "Disregarded Other Traffic Sign."

P16. Driver Distracted By

Specific Notes from the GHSA/NHTSA Mapping Process

State attributes equivalent to "No Apparent Distraction" or "None" may be mapped to the MMUCC attribute for this element "Not Distracted." "Cell Phone" by itself cannot be mapped to any MMUCC attribute. "Driver Inattention" or "Inattentive" will not be mapped to any P16. Driver Distracted By attributes.

Additional Considerations

A State PCR attribute coding for the presence or use of a "Cell Phone" cannot be mapped to any of the MMUCC attributes as "Cell Phone" contains many uses such as texting, dialing, talking, and listening to music. With "Cell Phone" alone, it is impossible to distinguish among all these possibilities.

Similarly, the State PCR attribute "Texting" by itself cannot be mapped to the MMUCC attribute "Manually Operating an Electronic Communication Device (texting, typing, dialing)" given the inconsistent definitions.

In order to map to the MMUCC attribute "Other Activity, Electronic Device," the State PCR must have the previous three MMUCC attributes regarding electronic devices, including:

- Manually Operating an Electronic Communication Device (texting, typing, dialing),
- Talking on Hands-Free Electronic Device, and
- Talking on Hand-Held Electronic Device.

To map to the MMUCC attribute "Other Inside the Vehicle (eating, personal hygiene, etc.)," the State PCR must have the previous five MMUCC attributes, including:

- Manually Operating an Electronic Communication Device (texting, typing, dialing),
- Talking on Hands-Free Electronic Device,
- Talking on Hand-Held Electronic Device,
- Other Activity, Electronic Device, and
- Passenger.

P18. Law Enforcement Suspects Alcohol Use

Specific Notes from the GHSA/NHTSA Mapping Process

Law Enforcement Suspects Drug Use will not be mapped unless the reporting officer may unambiguously indicate whether alcohol or drug use or both is involved.

Additional Considerations

The State PCR may have a data element combining alcohol use and drug use together, such as "Alcohol/Drug Use Suspected," which cannot be mapped to this MMUCC element unless it is possible to indicate whether alcohol or drug use or both is involved.

For example, a State PCR has a data element "Alcohol/Drug Use Suspected" with the following attributes.

- None
- Both
- Alcohol
- Drug
- Unknown

In this case, a combination of "None" and "Drug" can be mapped to the MMUCC attribute "No" of P18. Law Enforcement Suspects Alcohol Use, as both of these State PCR attributes indicating that alcohol was not involved. The State PCR attribute "Both" in combination with "Alcohol" can be mapped to the MMUCC attribute "Yes" given that both the State PCR attributes are an indicator of alcohol involvement. "Unknown" can be mapped to the MMUCC attribute "Unknown."

P19. Alcohol Test

Specific Notes from the GHSA/NHTSA Mapping Process

The subfields and associated attributes must be on the PCR in order to get credit for mapping. Test results are often obtained after completing and filing the PCR, in which case (until result is obtained), Subfield 3 on the PCR, BAC Test Result would be coded "Pending."

Additional Considerations

The State PCR may have a separate data element for each of the subfields, which is acceptable.

For Subfield 1, many States may combine P19 Alcohol Test and P21 Drug Test as one element, which cannot be mapped to the either of the MMUCC element unless it is possible to unambiguously indicate whether the test was performed for alcohol use or drug use.

For example, a State PCR has an element "Alcohol/Drug Test" with the following attributes.

- Not Given
- Refused
- Alcohol
- Drug
- Both
- Unknown

Then, the State PCR attributes "Not Given" and "Drug" can be mapped to the MMUCC attribute "Test Not Given" for this variable which focuses *only* on alcohol testing. The State PCR attribute "Refused" can be mapped to the MMUCC attribute "Test Refused." The State PCR attribute "Alcohol" in combination with "Both" can be a complete mapping to the MMUCC attribute "Test Given." Last, "Unknown" can be mapped to the MMUCC attribute "Unknown if Tested."

For Subfield 2, if the State PCR combines alcohol and drug test into one element, the attributes cannot be mapped to the MMUCC attributes unless there is a way to distinguish between alcohol tests and drug tests. One exception is that State PCR attribute "Breath" can be mapped to the MMUCC attribute since currently there is no breath test for drug use.

State PCRs may have an open text field for BAC Test Result, which can be mapped to both of the MMUCC attributes "Value" and "Pending." If instructions direct officer to enter a code representing "Unknown," the State would be given credit for mapping to the MMUCC attribute "Unknown" as well.

P20. Law Enforcement Suspects Drug Use

Specific Notes from the GHSA/NHTSA Mapping Process

A State PCR element that combines P18. Law Enforcement Suspects Alcohol Use and P20. Law Enforcement Suspects Drug Use will not be mapped unless the reporting officer may unambiguously indicate whether alcohol or drug use or both is involved.

Additional Considerations

The State PCR may have a data element combining alcohol use and drug use together, such as "Alcohol/Drug Use Suspected," which cannot be mapped to this MMUCC element unless it is possible to indicate whether alcohol or drug use or both is involved.

For example, a State PCR has a data element "Alcohol/Drug Use Suspected" with the following attributes.

- None
- Both
- Alcohol
- Drug
- Unknown

In this case, a combination of "None" and "Alcohol" can be mapped to the MMUCC attribute "No" of P20. Law Enforcement Suspects Drug Use as both of these State PCR attributes indicate that drug was not involved. The State PCR attribute "Both" in combination with "Drug" can be mapped to the MMUCC attribute "Yes" given that both the State PCR attributes are indicators of drug involvement. "Unknown" can be mapped to the MMUCC attribute "Unknown."

P21. Drug Test

Specific Notes from the GHSA/NHTSA Mapping Process

The subfields and associated attributes must be on the PCR in order to get credit for mapping. Test results are often obtained after completing and filing the PCR, in which case (until result is obtained), Subfield 3 on the PCR, Drug Test Result, would be coded "Pending."

Additional Considerations

State PCRs may have a separate data element for each of the subfields, which is acceptable.

For Subfield 1, many States may combine P19. Alcohol Test and P21. Drug Test as one element, which cannot be mapped to the any of the MMUCC attributes unless it is possible to unambiguously indicate whether the test was performed for alcohol use or drug use.

For example, a State PCR has an element "Alcohol/Drug Test" with the following attributes.

- Not Given
- Refused
- Alcohol
- Drug
- Both
- Unknown

Then, the State PCR attributes "Not Given" and "Alcohol" can be mapped to the MMUCC attribute "Test Not Given." The State PCR attribute "Refused" can be mapped to the MMUCC attribute "Test Refused."

The State PCR attribute "Drug" in combination with "Both" can be a complete mapping to the MMUCC attribute "Test Given." "Unknown" can be mapped to the MMUCC attribute "Unknown if Tested."

For Subfield 2, combining alcohol and drug test into one element cannot be mapped to the attributes in Subfield 2 unless there is a way to determine that the test is for drug use.

For Subfield 3, combining alcohol and drug testing cannot be mapped to the MMUCC attributes unless it is possible to determine whether the result is for an alcohol test or a drug test. For example, if a State PCR element "Alcohol/Drug Test Result" with the following attributes.

- No Test
- Alcohol-Positive
- Drug-Positive
- Both-Positive
- Negative
- Unknown

Then a combination of State PCR attributes "Drug-Positive" and "Both-Positive" can be mapped to the MMUCC attribute "Positive." The State PCR attribute "Negative" in combination with "Alcohol-Positive" can be mapped to the MMUCC attribute "Negative."

P23. Non-Motorist Action/Circumstance Prior to Crash

Specific Notes from the GHSA/NHTSA Mapping Process

If the State PCR combines P23. Non-Motorist Action/Circumstance Prior to Crash and P24. Non-Motorist Actions/Circumstances at Time of Crash as one State PCR element, it will partially map to both elements only if the PCR permits coding of at least three actions/circumstances (because of the subfields in these MMUCC elements). To map completely the State PCR needs a separate field for "Going to or From School (K12)."

Additional Considerations

The State PCR attributes must have the identical definitions. "Entering or Crossing Roadway" would not be equivalent to "Crossing Roadway" because entering a roadway is merging into traffic and proceeding in the appropriate direction along the roadway.

To map to Subfield 2, the State PCR needs to have a separate field for "Going to or from School (K12)." States PCR may have an attribute "Walking to/From School" on the attribute list which cannot be mapped to any of the attributes in Subfield 2.

P24. Non-Motorist Actions/Circumstances at Time of Crash

Specific Notes from the GHSA/NHTSA Mapping Process

If the State PCR combines P23. Non-Motorist Action/Circumstance Prior to Crash and P24. Non-Motorist Actions/Circumstances at Time of Crash as one State PCR element, it will partially map to both elements only if the PCR permits coding of at least three actions/circumstances (because of the subfields in these MMUCC elements).

Additional Considerations

To map to the MMUCC attribute, the State PCR does not need to use the same terms. States will be given credit as long as the definitions match. For example, a State PCR attribute "Getting On or Off Vehicle" can be mapped to "Entering/Exiting Parked/Standing Vehicle."

P25. Non-Motorist Location at Time of Crash

Specific Notes from the GHSA/NHTSA Mapping Process

P25. Non-Motorist Location should not be mapped based on a State PCR "Non-Motorist Action or Circumstance" element. For mapping purposes, the State PCR must have a specific "Non-Motorist Location" element ("At Time of Crash" is not necessary).

Additional Considerations

Assessors should refer to Appendix E in the *MMUCC Guideline* for a clarification of what a trafficway contains.

P26. Non-Motorist Safety Equipment

Specific Notes from the GHSA/NHTSA Mapping Process

Some or all of the attributes of this MMUCC element may be listed under a more general State PCR element that combines motorist and non-motorist equipment. This is acceptable as long as the Person Type for the reported individual is unambiguously a non-motorist.

Additional Considerations

Assessors should check the definition of each attribute to ensure that the State PCR attribute has the exact same definition even though the attribute names may be very similar to the MMUCC attribute names. For example, a State PCR attribute "Safety Vest (Ped Only)" cannot be mapped to the MMUCC attribute "Reflective Clothing (jacket, backpack, etc.)" for the reason that: (1) it is for pedestrian only and, (2) in addition to safety vest, reflective clothing also includes jacket, backpack, or similar. Given these, it would not be a complete mapping.

Example

This section provides an example demonstrating how to map a State PCR to MMUCC by following the general rules and the specific notes. The example here is from the *GHSA/NHTSA Mapping Process*.

Step 1: Collect both source and target documents. The relevant excerpt of each is shown in Figure 8.3.

Data Target: MMUCC

C11. Weather Conditions

Defintion

The prevailing atmospheric conditions that existed at the time of the crash

Element Attributes

Subfield 1: Weather Conditions

- Clear
- Cloudy
- Fog, Smog, Smoke
- Rain
- Sleet or Hail
- Freezing Rain or Freezing Drizzle
- Snow
- Blowing Snow
- Severe Crosswinds
- Blowing Sand, Soil, Dirt
- Other
- Unknown

Subfield 2: Weather Conditions

See attributes in Subfield 1

Data Source: PCR

Weather Condition **C2**

- ⬭ 1. No Adverse Condition (Clear/Cloudy)
- ⬭ 3. Fog
- ⬭ 4. Mist
- ⬭ 5. Rain
- ⬭ 6. Snow
- ⬭ 7. Sleet/Hail
- ⬭ 8. Smoke/Dust
- ⬭ 9. Other
- ⬭ 10. Blowing Sand, Soil, Dirt, or Snow
- ⬭ 11. Severe Crosswinds

Figure 8.3. Data Target (Left) and Data Source (Right)

Step 2: Set up a mapping table (or use the NHTSA-supplied spreadsheet) so that data elements and attributes from the *MMUCC Guideline* and the PCR are arranged for a direct comparison, as shown in Figure 8.4. Since the *MMUCC Guideline* has two subfields for C11 Weather Condition, the table should be set up to include both subfields.

Crash Data Elements

A	B	C	D	E	F	G	H	I
Target Data: MMUCC			Ability to Map?	Source: State PCR		Target Data: MMUCC	Ability to Map?	Source: State PCR
C11	Data Element	Data Attribute (Subfield 1)	1 = Yes 0 = No	Data Element*/ Data Attribute		Data Attribute (Subfield 2)	1 = Yes 0 = No	Data Element / Data Attribute
	Weather Conditions	Clear	0	(1)No Adverse Condition (Clear, Cloudy)		Clear	0	
		Cloudy	0	(1)No Adverse Condition (Clear, Cloudy)		Cloudy	0	
		Fog, Smog, Smoke	1	(3) Fog, (8) Smoke/Dust		Fog, Smog, Smoke	0	
		Rain	1	(5) Rain		Rain	0	
		Sleet or Hail	1	(7) Sleet/Hail		Sleet or Hail	0	
		Freezing Rain or Freezing Drizzle	0	N/A		Freezing Rain or Freezing Drizzle	0	
		Snow	1	(6) Snow		Snow	0	
		Blowing Snow	0	(10) Blowing Sand, Soil, Dirt, or Snow		Blowing Snow	0	
		Severe Crosswinds	1	(11) Severe Crosswinds		Severe Crosswinds	0	
		Blowing Sand, Soil, Dirt	0	(10) Blowing Sand, Soil, Dirt, or Snow		Blowing Sand, Soil, Dirt	0	
		Other	0	(9) Other			0	
		Unknown	0	N/A			0	

*To save space, the name and number of the State PCR element, Weather Condition, C2, are not listed in the Column E of the table.

Figure 8.4. Example of a Mapping Table

Step 3: Map the data attributes from the PCR to the MMUCC attributes.

The PCR attribute "No Adverse Condition (Clear, Cloudy) (1)" cannot be mapped to the MMUCC attributes "Clear" or "Cloudy" because the PCR combines these MMUCC attributes, referring to General Rule 7.

Likewise, the PCR attribute "Blowing Sand, Soil, Dirt, or Snow (10)" from the PCR cannot be mapped to the MMUCC attributes "Blowing Snow" or "Blowing Sand, Soil, Dirt."

However, the PCR attributes "Fog (3)," and "Smoke/Dust (8)" can be mapped to the MMUCC attribute "Fog, Smog, Smoke," without a loss in data integrity.

Four attributes from the State PCR were mapped "one-to-one" to a MMUCC attribute, including:

- Rain.
- Sleet or Hail.
- Snow.
- Severe Crosswinds.

For the MMUCC attribute "Other," according to general rule 12, the State PCR attribute "Other" cannot be mapped to it since the PCR did not possess all of the specific attributes in the *MMUCC Guideline.*

The PCR did not have an attribute to map to the MMUCC attribute "Unknown." Based upon general rule 15, the State would get credit for this if the instructions for completing the State PCR directs the Officer to enter a code ("00," "99,""UNK") in the event the appropriate response is unknown.

The *MMUCC Guideline* included two subfields for Weather Conditions (C11), reporting the same 12 attributes in each subfield. However, the PCR collected only one value. Thus, according to general rule 5, the PCR did not match to any of the Subfield 2 attributes and "0" scores were entered.

As a result, of the 24 total MMUCC attributes (12 for each subfield), the State would be given credit for mapping to just 6.

8.3 Process for Mapping State Crash Database to MMUCC

The overall purpose of the *MMUCC Guideline* is to provide a minimal set of data elements that can be used to explain and investigate the status of highway safety within a State, and enable comparisons of the results among different States. In order to do this evaluation, data collected (and derived or linked) must be entered into to a crash database, which is typically some form of computerized or electronic file.

Typically a State's crash database is comprised of the corresponding data collected on PCRs, derived from data collected on PCRs, and obtained from other data sources (e,g., a roadway database). The elements and attributes contained in a State's crash database should be mapped to all 110 MMUCC elements and their attributes, including the 77 elements discussed in previous section and the additional 33 derived and linked elements described in this section.

> *"Since elements collected on a PCR may not be captured or coded into a State's crash database, especially those that are text based, the State crash database is to be mapped to MMUCC independently of the State's PCR to MMUCC Mapping." (**Error! Reference source not found.**)*

Preparation

Similar to mapping a State PCR to MMUCC, States need to have materials ready prior to conducting a thorough mapping of the crash database to MMUCC. For mapping crash data to MMUCC, State should have the following documentation:

- State Crash Data Dictionary; and
- State PCR and Police Instruction Manual (optional, but may be needed if the data dictionary is incomplete).

The data dictionary for the State Crash Database should list all data elements and element attributes used in the crash database. The police crash report form and police instruction manual should not be needed for mapping if the data dictionary contains all relevant terms and definitions.

The next step is to set up mapping tables. States should follow the procedure outlined in the preceding section. States are also encouraged to use the NHTSA-supplied Mapping Spreadsheet.

Mapping Rules

General

The process for mapping a State crash database to MMUCC follows the same process used for mapping a State PCR to MMUCC. Consequently, all the general rules, specific notes, and commentary in previous section apply to mapping a crash database to MMUCC. The major difference between the two processes is that the State crash data dictionary is to be used primarily (or exclusively) instead of the PCR for the mapping compatibility comparison. Note that the general rules and the mapping process for the first 77 MMUCC elements are not repeated here. This section deals exclusively with the 33 derived or linked data elements. A full database-to-MMUCC mapping must include a review all 110 elements because the database may not capture all of the data from the PCR.

The *GHSA/NHTSA Mapping Process* presents the following general rules particularly for mapping a crash database to *MMUCC Guideline*:

> How the database is populated would not affect the mapping result.

1. The means by which a data field is populated is irrelevant for mapping purposes when mapping a crash database to MMUCC. The presence of a State data element and its associated attributes that match to a MMUCC element and its attributes is sufficient to enable its mapping. Examples of possible acceptable non-conventional State crash database sources include: 1) Obtaining PL4. Injury Area from the State PCR (collected at the scene); (2) Deriving P28. Transported to First Medical Facility By through linkage to an EMS or hospital record; and (3) Obtaining RL17. Total Volume of Entering Vehicles through a combination of RL6. Annual Average Daily Traffic and C6. Crash Location.
2. If a State crash database does not carry a MMUCC derived element, but carries the MMUCC element that would allow the value to be determined, then the State can receive credit for mapping to this derived element. The State must demonstrate through documentation that the element used to derive is present for each case on the database. For example, if a State does not calculate CD9. Day of Week, it can derive that information through CD3. Crash Date and Time, if carried on the crash database for all cases.

3. If the State does not carry linked elements on its crash database, but can demonstrate through documentation that appropriate linkage elements are and would permit linkage for each case on the database, then the State may be given credit for mapping to those linked elements. For example, if the State can access data for all roadways (State and local) so that it can be linked to the crash database, then the State should get credit for mapping under the 17 Roadway Linked Elements.

Individual Data Elements

The remainder of this section presents the remaining individual MMUCC elements, any applicable notes from the *GHSA/NHTSA Mapping Process*, element definitions and attributes from the *MMUCC Guideline*, the expected source of the data, and commentary derived from initial pilot tests of State PCR-to-MMUCC mappings.

Crash Elements

There are nine crash elements not usually collected by law enforcement at the scene. However, the *MMUCC Guideline* recommends these elements be derived from computerized crash scene information. Depending on the system used, they could be derived automatically by electronic data collection systems or generated when the data are computerized and merged at the local, regional or State level.

CD1. Crash Severity

Specific Notes from the GHSA/NHTSA Mapping Process

This MMUCC element is expected to be derived from the maximum value of P5. Injury Status, or the State's equivalent to P5, for each person involved in the crash. However, it could be obtained through linkage to a record in a non-crash database of all injuries occurring in a given crash (e.g., emergency department database).

Additional Considerations

The State crash database does not need to carry this specific element for mapping. A State will be given credit as long as they can demonstrate that the attributes can be derived from other existing data elements or obtained through linkage. For example, if the State crash database does not carry this specific data element, but does map well to MMUCC Injury Status (P5), the value of this element can be derived by identifying the most severe injury (or lack of any injuries) in the crash. Additionally, the State must demonstrate through its documentation that MMUCC element Injury Status (P5) is present for each case on the database.

CD2. Number of Motor Vehicles Involved

Additional Considerations

The State crash database does not need to carry this specific element for mapping; they would be given credit as long as they can approve that the value can be derived from other existing data elements. For example, if the State crash database does not carry this specific data element but carries this specific MMUCC element V2 Motor Vehicle Unit Type and Number (which would allow the value to be derived), then the State gets credit for mapping to this MMUCC element. Additionally, the State must demonstrate through its documentation that MMUCC element V2 Motor Vehicle Unit Type and Number is present for each case on the database.

CD3. Number of Motorists

Additional Considerations

The State crash database does not need to carry this specific element for mapping. To map to this element, the State mush show that the value can be derived from other existing data elements. For example, if the State crash database carries the MMUCC elements P6. Occupant's Motor Vehicle Unit Number, P7. Seating Position, and P4. Person Type, the number of motorists could be derived. In that case, the State would get credit for mapping to this MMUCC element. Additionally, the State must demonstrate through its documentation that the data elements used to derive the MMUCC element are present for each case in the database.

CD4. Number of Non-Motorists

Additional Considerations

The State crash database does not need to carry this specific element for mapping. To get credit for mapping, the State must show that the value can be derived from other existing data elements. For example, if the State crash database carries the MMUCC element P22. Non-Motorist Number, the number of non-motorists could be calculated. In that case, the State would get credit for mapping to this MMUCC element. The State must also demonstrate the source variables used to derive this MMUCC element are present for each case in the database.

CD5. Number of Non-Fatal Injured Persons

Specific Notes from the GHSA/NHTSA Mapping Process

This MMUCC element is intended to include not only persons who are coded as having sustained suspected serious (A) or suspected minor (B) injuries (or State equivalent terms), but also persons coded as having sustained possible injuries (C) as indicated in the MMUCC element P5. Injury Status.

Additional Considerations

The State crash database does not need to carry this specific element for mapping. To get credit for mapping, the State must demonstrate that the value can be derived from other existing data elements. For example, if the State crash database carries the MMUCC element P5. Injury Status, the number of non-motorists could be calculated. In that case, the State would get credit for mapping to this MMUCC element. The State must also demonstrate the source variables used to derive this MMUCC element are present for each case in the database.

CD6. Number of Fatalities

Additional Considerations

The State crash database does not need to carry this specific element for mapping. To get credit for mapping, the State must demonstrate that the value can be derived from other existing data elements. For example, if the State crash database carries the MMUCC element P5. Injury Status, the number of non-motorists could be calculated. In that case, the State would get credit for mapping to this MMUCC element. The State must demonstrate the source variables used to derive this MMUCC element are present for each case in the database.

CD7. Alcohol Involvement

Specific Notes from the GHSA/NHTSA Mapping Process

A State data element that is mapped to MMUCC element P18. Law Enforcement Suspects Alcohol Use or P19. Alcohol Test cannot also be used to map to CD7. Alcohol Involvement. MMUCC intends CD7. Alcohol Involvement to be a separate element on the State crash database derived from P18 and P19.

Additional Considerations

The State crash database needs to carry a separate element for mapping to Alcohol Involvement CD7. The MMUCC elements P18. Law Enforcement Suspects Alcohol Use and P19. Alcohol Test are insufficient.

CD8. Drug Involvement

Specific Notes from the GHSA/NHTSA Mapping Process

Similar to CD7. Alcohol Involvement, State elements that map to either P20. Law Enforcement Suspects Drug Use or P21. Drug Test cannot also map to CD8. Drug Involvement. There must be a separate element on the State crash database equivalent to CD8. Drug Involvement.

Additional Considerations

A State crash database needs to carry a separate element for mapping to CD8. Drug Involvement. The MMUCC elements P20. Law Enforcement Suspects Drug Use and P21. Drug Test are insufficient.

CD9. Day of Week

Additional Considerations

The State crash database does not need to carry this specific element for mapping. To get credit for mapping, the State must demonstrate that the value can be derived from other existing data elements. For example, if the State crash database carries the MMUCC element C3. Crash Date and Time, the number of non-motorists could be calculated. In that case, the State would get credit for mapping to this MMUCC element. The State must also demonstrate the source variables used to derive this MMUCC element are present for each case in the database.

Person Elements

There are seven derived or linked person elements. These elements can be obtained based on values in other crash data elements, or by linking to driver history, injury data, or other State data sources. When a State does not have the capability to link to other State data, they should collect as many of the person "linked" data elements as possible at the scene.

PD1. Age

Specific Notes from the GHSA/NHTSA Mapping Process

If the State collects age only, and not date of birth, this may be used to map to both Subfield 2 of P2. Date of Birth and to PD1. Age.

Additional Considerations

If the State crash database does not have PD1. Age but has P2. Date of Birth and C3. Crash Date and Time, the age value can be derived. In this case, the State would be given credit for mapping to MMUCC element PD1. Age. The State must also demonstrate through its documentation that P2. Date of Birth and C3. Crash Date and Time are present for each case on the database.

PL1. Driver License Restrictions

Additional Considerations

The State crash database does not need to carry this specific element for mapping. They would be given credit as long as they can demonstrate through documentation that the appropriate linkage elements are accessible to the crash database. If linkage is available for each in-State driver in the crash database, the State may be given credit for mapping to this element.

PL2. Driver License Status

Additional Considerations

The State crash database does not need to carry this specific element for mapping, they would be given credit as long as they can demonstrate through documentation that the appropriate linkage elements are accessible. If the linkage is available for each in-State driver in the crash database the State may be given credit for mapping to PL2. Driver License Status.

PL4. Injury Area

Additional Considerations

States may collect this information at the scene and include this element in the crash database. In this case, the State would get credit for mapping to this MMUCC element even though the *MMUCC Guideline* suggests they be obtained through linkage to medical data.

If the State can demonstrate through its documentation that the appropriate elements are accessible to the crash database for each case on the database to support linkage to obtain injury area information, then the State may be given credit for mapping to this element.

PL5. Injury Diagnosis

Additional Considerations

The State crash database does not need to carry this specific element for mapping. A State will be given credit as long as they can demonstrate that the attributes can be obtained through linkage. For example, if the State crash database does not carry this specific data element but can demonstrate through documentation that the appropriate linkage elements are accessible to the crash database to support linkage for obtaining injury diagnosis information, then the State may be given credit for mapping to PL5. Injury Diagnosis. Note that how the data is linked is relevant to the mapping. Some States may obtain this information through automatic linkage and some may have law enforcement officers update the information manually from information supplied by EMS providers. In both cases, the State would be given credit for mapping.

PL6. Injury Severity

Specific Notes from the GHSA/NHTSA Mapping Process

Note that this MMUCC element is explicitly intended to be obtained through linkage to clinical health records. A State data element that maps to P5. Injury Status cannot also be used to map to PL6. Injury Severity. The values used here must be derived from the clinical scale used in the State's linked injury database.

Additional Considerations

If the State obtains the values through linkage to medical or injury surveillance records, they would be given credit for mapping to PL6. Injury Severity regardless of whether the linage is automated or manual.

Roadway Elements

There are 17 linked roadway elements. There are no notes or additional considerations for any of these data elements. Computing Compatibility Scores and Ratings

If the State chooses to use the NHTSA MMUCC Mapping Spreadsheet, each of the scores and ratings described in this section will be calculated automatically.

The information presented here is adapted from the *NHTSA/GHSA MMUCC Mapping Process*. Once completed, a State PCR-to-MMUCC mapping and Crash Database-to-MMUCC mapping will yield a series of mapping tables (or worksheets in the NHTSA-supplied spreadsheet) that will show which MMUCC elements and attributes were successfully mapped by State crash data elements and attributes and which were not.

This information can be used to compute a MMUCC Element Mapping Compatibility Rating for each MMUCC element and, then, the overall compatibility ratings for PCR-to-MMUCC and Crash Database-to-MMUCC.

MMUCC Element Mapping Score and Compatibility Rating

Definition

For each of the 110 MMUCC elements, the MMUCC Element Mapping Score is defined as:

$$\textbf{MMUCC Element Mapping Score} = \frac{\textbf{Number of Attributes for State Element that Map to MMUCC Element}}{\textbf{Total Number of MMUCC Attributes}^{*}\ \textbf{for Element}} \times \textbf{100}$$

*Note: If a MMUCC element has subfields, the value for "Total Number of MMUCC Attributes for Element" is the sum of the number of attributes across all off the subfields. Table 8.1 provides a suggested Compatibility Rating Scale to be applied to each MMUCC element based on the MMUCC Element Mapping Score to provide a measure of how well the State PCR maps to individual MMUCC elements.

Table 8.1. MMUCC Element Compatibility Rating Scale

PCR to MMUCC Mapping Score	Rating
100	Full
70 – 99	High
40 – 69	Moderate
1 - 39	Low

Example

There are 24 attributes for MMUCC element C11. Weather Conditions, with 12 attributes for each subfield. Of these 24 attributes, 6 attributes of Subfield 1 can be mapped to the State PCR. So, the calculation should be:

$$\textbf{MMUCC Element Mapping Score} = \frac{6}{12 \times 2} \times 100 = 25$$

Referring to *Table 8.*, the State Compatibility Rating for this specific element would be "Low."

Overall PCR to MMUCC Mapping Compatibility Rating

Definition

A State's Overall PCR to MMUCC Mapping Compatibility Rating is based on the number of MMUCC element attributes that were mapped by element attributes from the State's PCR. To calculate this rating, a MMUCC element mapping score must be computed for each of the 77 MMUCC elements designated to be collected at the crash scene. Since there are 77 such scores, the average, overall mapping compatibility rating is just the average (the sum of all the individual scores, divided by 77). Because the individual element scores are percentages, this is the average percentage mapped. This rating provides the State with a generalized score as to how well its PCR maps to MMUCC.

An Overall PCR to MMUCC Mapping Compatibility Rating is calculated by adding the MMUCC Element Mapping Scores for all 77 MMUCC Elements to be collected on a PCR and divide by 77:

$$\textbf{Overall PCR to MMUCC Mapping Score} = \frac{\sum \textit{MMUCC Element Mapping Score}}{77}$$

Table 8.2 provides a suggested Compatibility Rating Scale that can be applied to the Overall PCR to MMUCC Mapping Score to obtain a measure of how well the State's PCR mapped to MMUCC.

Table 8.2. PCR to MMUCC Mapping Compatibility Rating Scale

PCR to MMUCC Mapping Score	Rating
100	Full
70 – 99	High
40 – 69	Moderate
1 - 39	Low

Example

The total score for the 77 MMUCC elements mapping results is 5,010. So, the calculation should be:

$$\textbf{Overall PCR to MMUCC Mapping Score} = \frac{\mathbf{5010}}{\mathbf{77}} = \mathbf{65}$$

Referring to Table 8., the State Compatibility Rating for the overall PCR to MMUCC mapping would be "Moderate."

Overall Crash Database to MMUCC Mapping Compatibility Rating

Definition

A State's Overall Crash Database to MMUCC Mapping Compatibility Rating is based on the number of MMUCC element attributes that were mapped by element attributes from the State's crash database. To calculate this rating, a score must be computed for each MMUCC element. Keep in mind that the crash database must be mapped independently to the full set of 110 MMUCC elements (including the 33 linked or derived elements). This Rating provides the State with a generalized score as to how well its crash database maps to MMUCC.

An Overall Crash Database to MMUCC Mapping Compatibility Rating is calculated by adding the MMUCC Element Mapping Scores for all 110 MMUCC elements and dividing by the total number of MMUCC elements (i.e., 110). The result is the percentage of MMUCC element attributes mapped by the State crash database:

$$\textbf{Overall Crash Database to MMUCC Mapping Score} = \frac{\sum \textbf{\textit{MMUCC Element Mapping Score}}}{\mathbf{110}}$$

Table 8.3 provides a suggested Compatibility Rating Scale that can be applied to the Overall Crash Database to MMUCC Mapping Score to obtain a measure of how well the State's Crash Database mapped to MMUCC.

Table 8.3. Crash Database to MMUCC Mapping Compatibility Rating Scale

Crash Database to MMUCC Mapping Score	Rating
100	Full
70 – 99	High
40 – 69	Moderate
1 - 39	Low

Example

The total score for the 110 MMUCC elements mapping results is 8111. So, the calculation should be:

$$\textbf{Overall PCR to MMUCC Mapping Score} = \frac{\mathbf{8111}}{\mathbf{110}} = \mathbf{74}$$

Referring to Table 8.3, the State Compatibility Rating for mapping crash database to MMUCC would be "High."

Final Considerations

Once a State has calculated the three scores, they can use the information to develop an action plan for updating their PCR and crash database. NHTSA recognizes that it may not be possible or desirable for States to update everything. States are free to prioritize the elements that most need to be revised.

It would be helpful for States that are considering a PCR or database revision to establish a template for the development of the action plan and include the following items for each element for which they are considering a change:

- Element.
- Priority for Change.
- Rationale.
- Deadline.
- Person Responsible.

8.4 Summary

MMUCC is typically updated every five years. The next update (the 5th Edition) is scheduled to be published in 2017. By following the *GHSA/NHTSA Mapping Process,* States can identify how closely their State PCR and crash database follow the *MMUCC Guideline* to establish a baseline of conformance. Over time, improvements in conformance can be used to show measurable progress. States can benefit from the knowledge gained through the MMUCC mapping exercise by (1) knowing how their data elements and attributes compare to the set developed by a national team of experts, and (2) identifying the elements and attributes that could be shared and compared in national crash databases.

References

American Association of State Highway Transportation Officials. (2010). *Highway safety manual.* (Web document). Washington DC: Author. Available at http://www.highwaysafetymanual.org/Pages/default.aspx

Blincoe, L. J., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. (2015, May). *The Economic and societal impact of motor vehicle crashes, 2010* (Report No. DOT HS 812 013). Washington, DC: National Highway Traffic Safety Administration. Available at www-nrd.nhtsa.dot.gov/Pubs/812013.pdf

Federal Highway Administration, Federal Motor Carrier Safety Administration, &. National Highway Traffic Safety Administration. (2012, July). *MMUCC Guideline: Model minimum uniform crash criteria, 4th Edition* (Report No. DOT HS 811 631). Washington, DC: Authors. Available at https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811631

FHWA. (2012, December 27). *Guidance on State Safety Data Systems* (FHWA web page headed "Map-21 -- Moving ahead for progress in the 21st Century). Washington, DC: Author. Available at www.fhwa.dot.gov/map21/guidance/guidesafetydata.cfm

FHWA. (n.d.a). *Crash modification clearinghouse*. Washington, DC: Author. Available at c

FHWA. (n.d.b). *Notice of Proposed Rulemakings: Highway Safety Improvement Program & Safety Transportation Performance Management* (Flyer for 23 U.S.C. 148 NPRM. No. FHWA-SA-14-027). Washington, DC: Author. Available at http://safety.fhwa.dot.gov/hsip/nprm/outreach_flyer.pdf

FHWA. (2013, September 12). *Strategic Highway Safety Plan (SHSP) Interim Guidance* (FHWA web page headed "Map-21 -- Moving ahead for progress in the 21st Century). Washington, DC: Author. Available at www.fhwa.dot.gov/map21/guidance/guideshsp.cfm

FHWA, & Federal Transit Administration. (n.d.). *2013 Status of the Nation's highways, bridges, and transit: conditions & performance* (Report to Congress). Washington, DC: Authors. Available at www.fhwa.dot.gov/policy/2013cpr/pdfs/cp2013.pdf

Goodwin, A., Kirley, B., Sandt, L., Hall, W., Thomas, L., O'Brien, N., & Summerlin, D. (2013, April). *Countermeasures that work: A highway safety countermeasure guide for State Highway Safety Offices.* Seventh Edition (Report No. DOT HS 811 727). Washington, DC: National Highway Traffic Safety Administration. Available at www.nhtsa.gov/staticfiles/nti/pdf/811727.pdf

Governors Highway Safety Association & National Highway Traffic Safety Administration. (2015, July). Mapping to MMUCC: A process for comparing police crash reports and state crash databases to the Model Minimum Uniform Crash Criteria (Report No. DOT HS 812 184). Washington, DC: National Highway Traffic Safety Administration.

Kindelberger, J., & Milani, J. A. (2015, July). *Crash Outcome Data Evaluation System (CODES): Program transition and promising practices* (Report No. DOT HS 812 178). Washington, DC: National Highway Traffic Safety Administration. Available at https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812178

Moving Ahead for Progress in the 21st Century Act (MAP-21). 23 U.S.C. 101(a)(2) (2012)

Milani, J., Kindelberger, J., Bergen, G. Novicki, E. J., Burch, C., Ho, S. M., & West, B. A. (2015, August). Assessment of characteristics of state data linkage systems. (Report No. DOT HS 812 180). Washington, DC: National Highway Traffic Safety Administration, and Atlanta: Centers for Disease Control and Prevention. Available at www.cdc.gov/motorvehiclesafety/pdf/linkage/assessment_of_characteristics_of_state_data_linkage_systems-a.pdf

National Center for Injury Prevention and Control. (2014, October 7). *Motor Vehicle Crash Injuries: Costly but Preventable* (Web page). Atlanta: Centers for Disease Control and Prevention. Available at www.cdc.gov/vitalsigns/crash-injuries/index.html

NHTSA. (2010, April). *The Crash Outcome Data Evaluation System (CODES) and applications to improve traffic safety decision-making*. (Report No. DOT HS 811 181). Washington, DC: National Highway Traffic Safety Administration. Available at https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811181

National Highway Traffic Safety Administration. (2011, February). *Model performance measures for state traffic records systems* (Report No. DOT HS 811 441). Washington, DC: National Highway Traffic Safety Administration. Available at https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811441

NHTSA. (2012, July). *Traffic Records Program Assessment Advisory* (Report No. DOT HS 811 644). Washington, DC: Author. Available at www-nrd.nhtsa.dot.gov/Pubs/811644.pdf

NHTSA. (2015, January 15). *2013 Crash Overview: The Fatality Analysis Reporting System (FARS) and National Automative Sampling System (NASS) General Estimates System (GES)* (PowerPoint presentation). Washington, DC: Author. Available at www-nrd.nhtsa.dot.gov/Pubs/13WPPP.pdf

Scopatz, R. A., Benac, J., & Lefler, N. (2014, January 2). *Crash Data Improvement Program (CDIP) Final Report* (Report No. FHWA-SA-14-001). Washington, DC: Federal Highway Administration. Available at http://safety.fhwa.dot.gov/rsdp/downloads/fhwasa14001cdip_final.pdf

Transportation Research Board. (2010). *Target-setting methods and data management to support performance-based resource allocation by transportation agencies* (NCHRP Report 666). Washington, DC: Author. Available at http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_666.pdf

Uniform Procedures for State Highway Grant Programs Final Rule, 79 Register 15, vol. 78, pg. 4985. (2013).

Appendix A: Pre-Site Visit Information Collection

Crash Data Improvement Program
Pre-Site Visit Information Collection

The following lists documents that will help the CDIP project team develop a clear understanding of the State's crash data practices, policies, and procedures. Providing this information in advance can help the CDIP team be well informed going into the CDIP to get more meaningful results during the site visit.

Document Upload for the CDIP Module

Please upload the following documents to the Document Library:

- The most recent Traffic Records Assessment (from State/NHTSA).
- The most current Section 405(c) (formerly Section 408) grant application (from State/NHTSA). This should also include the most current version of the State's Traffic Records Strategic Plan. If it does not, please provide a copy of the Strategic Plan as well.
- Any separate or expanded written plans (beyond what is included in the Strategic Plan) for increasing electronic data collection and electronic submission of crash reports.
- Any separate or expanded written plans (beyond what is included in the Strategic Plan) for projects aimed at improving crash data timeliness, accuracy, completeness, uniformity, integration, or accessibility.
- The completed CDIP questionnaire (the questionnaire is attached to this document or was sent under separate cover).
- A current version of the crash production process flow chart.
- The data dictionary for the crash system.
- The current crash report form, and copies of any older report forms that are still in use.
- The crash data users guide.
- The State crash reporting manual used in law enforcement officer training.
- A list of edit checks for field data collection systems.
- A list of edit checks for acceptance of records into the State crash database.
- Any section of the Strategic Highway Safety Plan dealing with crash data improvement or safety data integration (from State DOT). Only provide if SHSP's content differs from the content of the Traffic Records Strategic Plan; if the SHSP adopts the traffic records strategic plan by reference, please let us know.
- The most recent Highway Safety Plan (HSP) sections dealing with crash data or safety data integration.
- Any other materials the State feels may be relevant, e.g., research studies that highlight difficulties using State crash data, or any previous documentation on measuring crash data quality.

State Crash Data Upload

A crash data sample, or results of the State's analysis, will give the TAT a better understanding of the crash data itself. The benefit is two-fold: it allows the project team to determine what performance measures can be developed based on the information in the crash data, and also allows the CDIP project team to run some initial analysis to determine the quality of the data itself. This enables the CDIP project team to make more detailed recommendations specific to the issues in the State. If the State can provide results of the State's analysis of the crash data, upload of the crash data file is not needed. Notes on data file extracts:

- Data extracts are required if the TAT is conducting the analysis.
- If you choose to provide a data extract, it is useful if the data element coding for law enforcement agency identifier is included.
- Please do not provide personal identifiers (name, address, license number).
- Please provide data files as comma-separated values with one file per database table, and a header row with data element labels at the top of the file. Alternatively, please check with the TAT to make sure they have the capability to process the data files in the format you propose to use.

1. Who will be conducting the crash data analysis for the development of performance measures?

The following questions are to be used if the State performs the analysis or works with a third party to do so. For each of the following questions the State answers yes, please provide the analysis described by the question.

Timeliness Analysis

2. Can the state provide a measurement for the overall timeliness for a crash to be available in the Statewide database after the event date?
3. Which of the following timeliness component measurements can you provide:
    - Law Enforcement submission: Time from event to receipt of crash reports—annual averages for the state and for each law enforcement agency.
    - 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.
    - Rejected Report Correction: Time from report rejection (or returned to the law enforcement agency) to receipt of the corrected report—annual averages for the state and for each law enforcement agency.
    - Other: Any other processes (Quality Control, data cleansing, financial responsibility, other?) required *prior to data availability for use/analysis* - annual averages.

Accuracy Analysis

4. Can the State provide the percentage of reported crash locations that automatically code (without human intervention) to the appropriate State locating coding system?

5. Can the State provide the percentages of successful location coding for State-maintained (on-system) and local (off-system) roadways calculated?

6. Can the State provide the initial automated location coding and a final percent coded successfully (after human intervention/correction)?

7. Can the State provide annual statewide averages and, if available, annual averages for each law enforcement agency?

8. Can the State provide the percentage of crash reports rejected (i.e., failing to pass mandatory—fatal error—edit checks) upon initial submission organized by annual averages statewide and for each law enforcement agency? NOTE: this may only be relevant to those reports submitted on paper (if any)

9. Can the State provide examples of the following 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—annual averages.
    - Agreement between number of vehicles/units & number of drivers recorded (adjusting count for hit & run or recorded phantom vehicles).
    - 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 (people) and the crash-level injury severity (this may also be a calculated field, in which case this analysis is moot).
    - Results of any other data cleansing analysis/accuracy measurements that are run routinely.
    - VIN (if recorded)—the percent that decode to a valid vehicle year, make & model.

Completeness Analysis

10. Can the State provide the percentage of reports received lacking the narrative, the diagram, or both shown by annual average statewide and for each law enforcement agency?

11. Can the State provide the percentage 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) by annual statewide average?

12. Can the State provide the percentage of other key fields left blank? Examples include making allowances for hit-and- run or phantom vehicles (if recorded), identify the percentage of records with unexplained blanks, missing data, and null codes like "NA," "unknown," and "other" in select fields such as:

    - Driver name/address;

- VIN;
- Harmful Event;
- Environment, Vehicle and Driver contributing factors; and
- EMS data (if recorded).

13. Can the State provide evidence for any law enforcement agencies that have seen a significant drop in crash report submissions that is larger than the drop experienced by the State as a whole compared across the most recent three complete, consecutive years of data?
14. Can the State provide evidence for any medium-to-large agencies (in terms of annual crash report submissions historically) that have turned in either zero or an uncharacteristically low number of crash reports for the most recent YTD in comparison to the average for prior years?

Uniformity Analysis

15. Can the State provide the following ratios to compare across years?
    - Percentage of PDO crashes of total crashes by law enforcement agency
    - Percentage of Injury crashes of total crashes by law enforcement agency
    - Percentage of Fatal crashes of total crashes by law enforcement agency
16. Can the state provide the ratio [(Fatal Crashes + Injury Crashes)/Total Crashes] to compare across years overall for the State and individually for each law enforcement agency?

    NOTE: A short cut method to calculate this ratio is [1-(PDO crashes/Total Crashes)].

Integration Analysis

17. Can the State provide the information from any relevant data linkage project (proportion of records linked and the strength of the matching achieved) for *integration*?
18. Can the State provide the percentage of linkage/matching you have when linking the crash database with roadway inventory, citation and adjudication, driver, vehicle, or injury surveillance databases?

Accessibility Analysis

19. Is there a record of number of requests, analyses run, web-enabled data analysis page hits, or any other indicator you might have?
20. Is there a measure of user satisfaction or unmet user needs for data?

Appendix B: Assessment Questionnaire

Crash Data Improvement Program Questionnaire

These questions should be assigned to people 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 Advisory 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.***

This document presents the Crash Data Improvement Program Questionnaire. This questionnaire is intended for States to complete well in advance of the CDIP Technical Advisory Team 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.
- 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.

1. Which department/agency is the custodian of the State's crash database of record? (AQ-37)
2. Which section/office within that department has the principal responsibility for managing/maintaining the crash database? (AQ-37)
3. 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, web page links are acceptable.)
4. 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)

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.)? (AQ-36) (SOE: provide a description of statewide database and specify how the data is consolidated into a statewide system)
7. 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. For example, an interface between crash data collection systems and the driver licensing system may be used to A) autocomplete portions of the crash report; B) verify current address and contact information; and C) validate information during quality control post-processing. 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.

8. 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)

9. 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)? If there are people who are assigned less than full time on tasks related to the crash data system, please estimate staffing based on full-time-equivalents (FTEs).

10. 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?

11. What type of coordination is there among the offices involved in crash report processing? For example, crash involved driver information may be needed for ascertaining financial responsibility, including assessment of any damages to public property—in such cases, multiple offices or agencies may be involved and information must flow between them. Location coding, in some States, is another example of inter-office or inter-agency coordination as the crash data system record is enhanced with location-specific information provided by another office. FARS and SafetyNet are also potential examples of coordinated processes.

12. Are any changes to this process being considered? If so, what are they and why are they being considered?
13. 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 MMUCC provides a suggested minimum set of crash-related data elements that enable valid statistical analysis. 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 the applicable standard published by the American National Standards Institute: ANSI-D16, the *Manual on Classification of Motor Vehicle Traffic Accidents*, 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.

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 interfaces with the State's centralized crash records database.

14. 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, web links are acceptable – this may have been provided in answer to question 3 earlier, but please point to the appropriate section of the law)

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

16. Are there any reasons that a law enforcement agency might have an incentive for not providing crash reports in a timely manner (for example, if a municipality charges fees to the public for providing a crash report copy they may hold onto it for a time before making it available through another source such as the statewide crash reporting system)?

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

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

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

20. If multiple (out-of-date or specific to a municipality) versions of the PCR are used, how many different versions of the PCR form are being used? (SOE: provide a copy of all PCRs currently in use; and under what circumstances/ jurisdictions they are used)
21. Is there a short form used for PDO crashes? (SOE: provide the form.)
22. Does the State require crash-involved drivers to submit an operator report? If yes, does the State include operator-submitted data in the statewide crash database (i.e., blended with data from crash reports supplied by law enforcement officers)? (SOE: provide the operator form if the data are used in the statewide crash database.)
23. 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.)
24. Are there any plans to modify/update the crash or supplemental forms in the State? If 'yes', which ones, when and who will be involved?
25. 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) Statewide, what percentage of crash reports is collected electronically? (AQ-53) Does the electronically collected data identically match the data elements and attributes on the paper PCR elements?
26. Does the State have a single electronic crash software product for use by all law enforcement?
27. If not, is there a standard for data collection systems to be attained by all electronic reporting systems? Does the State approve vendors and/or law enforcement agencies to validate their electronic data collection software? If yes, please describe the approval process.
28. By what processes are PCRs from law enforcement agencies submitted to the crash database (e.g., mail, internet, secure electronic upload, CDs)?
29. Does the State crash database accept any crash data electronically? If so, what percentage is accepted electronically? Is there a standard for electronic data submissions? Does the State approve vendors and/or law enforcement agencies to validate their electronic data submissions? If yes, please describe the approval process.
30. Are there law enforcement agencies collecting crash data electronically that do not submit their data to the statewide crash database electronically (i.e., they collect the data electronically using software, but print the report and submit it on paper or send a static image file)?
31. 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)
32. What technologies are used for the verification?
33. What are the data verified against (e.g., driver license file, vehicle registration file)?
34. Are there edit checks/validations run by the crash database on the PCRs submitted electronically, prior to uploading the data? (AQ-54)

35. 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)

36. What feedback has your State received regarding problems with either the paper or electronic PCR 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 key 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.

37. Does the State have a process to locate each crash onto a base map? (AQ-60, 167, 168)
38. Is the base map inclusive of all public roads in the State? (AQ-60, 167, 168)
39. 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)
40. 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.)
41. Does the crash database custodial office perform the 'crash locating' function or is this effort managed by another entity within the State? (AQ-56)
42. What percentage of reportable crashes is locatable on the statewide base map? (AQ-385)
43. What percentage of reportable crashes is successfully located automatically? (AQ-385)
44. Does the State crash database include the vehicle identification numbers of crash involved vehicles? (AQ-59)
45. Does the State crash database include the Blood Alcohol Concentration (BAC) test results of crash involved drivers? (AQ-50)
46. 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)
47. How are crash-involved pedestrians and bicyclists recorded in the statewide crash database? Are they recorded as "involved units" (i.e., at the vehicle-driver level) or as "persons" (i.e., at the person level), or in some other manner?

48. Does the State crash database include data elements related to distracted driving? Please provide the data element definitions and attribute list.
49. What is the injury severity scale used on the crash report?
50. How does the State define "serious injury" on the crash report?
51. Are a narrative and diagram required on all crash reports? (AQ-50)
52. Are the narrative and diagram for all crashes retained in the crash database? If yes, are these searchable or otherwise available for use in case selection, data aggregation, or analysis?
53. 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.

54. 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? (AQ-63) If so, please describe. (SOE: or provide a listing of the edit checks and validations used by the State.)
55. 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? (AQ-64) If so, please describe.
56. Does the State have processes in place for transmitting and tracking rejected reports between the statewide crash system and the originating officer? (AQ-65) If so, please describe.
57. Does the State establish numeric goals—performance metrics—for each performance measure? (AQ-72) If so, please describe.
58. Does the State provide performance reporting to law enforcement agencies regarding the timeliness, accuracy, and completeness of their submissions to the statewide crash database? (AQ-73) If so, please describe.
59. Are high-frequency errors tracked and used to generate new training content and training manuals, updated validation rules, and prompt form revisions? (AQ-74) If so, please describe.
60. Does the State's data acceptance process include quality control reviews that compare the narrative, diagram, and coded report contents? (AQ-75) If so, please describe.
61. Does the State periodically conduct independent, sample-based audits of crash reports and related database contents? (AQ-76) If so, please describe.
62. Does the State periodically conduct comparative and trend analyses to identify unexplained differences in the data across years and jurisdictions? (AQ-77) If so, please describe.
63. Does the State regularly collect data quality feedback from key users and share it with data collectors and managers? (AQ-78) If so, please describe.
64. Does the State regularly provide the TRCC data quality management reports? (AQ-79) 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

Crash Data System Data Quality Performance Measures	
Timeliness	• The median or mean number of days from the crash date to the date the crash report is entered into the database. • The percentage of crash reports entered into the database within XX days after the crash (e.g., 30, 60, or 90 days).
Accuracy	• The percentage of crash records with no errors in critical data elements (e.g., 95% of reports with no critical data error). • The percentage of in-State registered vehicles on the State crash file with VINs matched to the State vehicle registration file (e.g., 95% of in-State VINs match).
Completeness	• A decrease in the percentage or number of missing data elements from the crash database for a given time period (e.g., 1 year). • The percentage of crash records with no missing data elements (e.g., 90% of reports with no missing data elements). • The percentage of unknowns or blanks in critical data elements for which unknown is not an acceptable value (e.g., 99% of reports with no unknowns or blanks in critical data elements).
Uniformity	• An increase in the number of crash data elements collected as part of a State's PCR that are mappable to MMUCC (e.g., 95% of data elements and attributes map successfully to MMUCC 4th Edition).
Integration	• The percentage of appropriate records in the crash database that are linked to another system or file. Examples: crash with in-State driver linked to driver file, crash with EMS response linked to EMS file (e.g., 85% of crash-involved persons coded as "transported" matched to EMS run report data).
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 (e.g., 90% of respondents rate their satisfaction as "satisfied" or "extremely satisfied").

65. Does the State currently have a processes for assessing the quality of the information within the crash database in terms the data quality attributes listed in Table 1? For each of the six data quality attributes, please indicate if the State has the listed data quality performance measures and list any other relevant data quality performance measures that are calculated on a statewide basis.

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

Timeliness

The CDIP Guide defines 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 the following.

- 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.

67. What output measures of timeliness are currently reported for your State? (AQ-66) (SOE: Provide baseline current data and trend if available)
68. 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.)
69. Does the State measure the time it takes to get the PCRs from each individual law enforcement agency? (SOE: provide the measurement)
70. Has the State developed a 'year-to-date' measure that calculates the timeliness of PCR submissions compared to the same time-period during the prior year for each law enforcement agency?
71. 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.)
72. Does the State produce an official annual file for use in analyses and data extract releases? If so, does it track the timeliness of the availability of that file in a performance measure? (SOE: provide the overall data file release timeliness performance measure, preferably with multiple years to show the trend).

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 persons, vehicles 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.

A. With internal validation, 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.

B. 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.

73. What output measures of accuracy are currently reported? (SOE: Provide baseline current data and trend if available, and specify frequency of issuance) (AQ-67)
74. How does the State assess the accuracy of content (attributes) for each of the data elements within the crash database (e.g., a descriptive analysis reporting the frequencies or percentages of responses)?
75. 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)
76. How does the State assess the relational content of information? (e.g., business edits and validations to assess the logical relations between date elements)
77. 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)
78. 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, internal and external.

A. Internal Completeness: For individual crash reports, completeness addresses whether or not the report is missing any required information.

B. External Completeness: 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.

Changes over time in 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. Another measure of completeness of report submissions compares the count of serious crashes (injury + fatal) to the overall number of reports submitted. 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. Under-reporting agencies will typically have a high ratio compared to the statewide average because they report fewer PDO crashes than expected.

79. What output measures of completeness are currently reported for your State? (SOE: Provide baseline current data and trend if available, and provide sample reports) (AQ-68)
80. How does the State assess missing data for each of the data elements within the crash database?

81. 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)

82. How does the State assess the frequency with which 'unknown' is the reported value for each of the data elements within the crash database? (e.g., a descriptive analysis listing the frequency and/or percentage of 'unknowns' per data element)

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

84. How does the State report to law enforcement agencies on the completeness of their crash report submissions in terms of numbers of reports submitted? (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 MMUCC Guideline, FARS and Motor Carrier Management Information System (MCMIS)/SafetyNet data system reporting requirements, and the ANSI D-16.1 crash reporting standard. 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.

85. What output measures 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 PCR)? (AQ-69) (SOE: Provide baseline current data and trend if available)

86. Has the State participated or conducted a mapping-to-MMUCC measurement of uniformity? If yes, please supply the analysis. If one is conducted as part of this CDIP, the TAT will have access to it.

87. 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)?

88. How are these output measures calculated? (SOE: provide sample reports and their frequency of issuance)

Integration

Integration is "the ability of records in a database to be linked to a set of records in another of the six core databases – or components thereof – using common or unique identifiers." Integration is typically measured based on the existence and quality of data integrating, timeliness of the linkage, and usefulness of linked data. 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.

Integration can be measured by tallying the number of systems that are integrated and the current status of inter-agency agreements formalizing those data sharing arrangements. Integration can also be measured by calculating the percentage of appropriate records in the crash database that are linked to another system or file. 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.

89. What output measures of integration are currently reported for your State? (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, by counting the total number of analyses performed on behalf of users, and by measuring user satisfaction with their level of access. It typically includes three steps: (1) identify the principal users; (2) query the principal users to assess their ability to obtain the data or their satisfaction with the timeliness of the response to their request; and (3) document the method of data collection and the principal users' response. 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.

90. What output measures of accessibility are currently reported? (AQ-71) (SOE: Provide baseline current data and trend if available)

Page intentionally left blank

Page intentionally left blank

DOT HS 812 419
December 2017

U.S. Department of Transportation
National Highway Traffic Safety Administration

NHTSA

12712-122017-v5