Crash Data Improvement Program Guide

Safe Roads for a Safer Future

Investment in roadway safety saves lives

April 2010

FHWA Program Manager

Robert Pollack

Performing Organization

KLS Engineering

(L. Boodlal and J. Emery)

And

Iowa State University

(R. Souleyrette)

U.S.Department of Transportation

Federal Highway Administration

Table of Contents

List of Tables

List of Figures

1.0 Introduction

1.1. Purpose of the Crash Data Improvement Program Guide

The purpose of this Crash Data Improvement Program Guide (Guide) is to assist state crash database managers and other traffic safety professionals in identifying, defining and measuring the characteristics of the data quality within the state crash database. The quality characteristics consist of the timeliness, accuracy, completeness, consistency, integration and accessibility of the crash data.

The purpose of the Guide is to assist State database managers and other traffic safety professionals in identifying, defining and measuring the quality characteristics of the data within the State crash database.

The Guide will assist States to establish baseline measures that reflect the current status of the quality characteristics and to 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 the quality measures.

The CDIP Guide is intended to address the following issues relating to the Crash Database:

1. What are the data quality characteristics?
2. Why is each quality characteristic important?
3. What is the definition of each data quality characteristic?
4. What metrics can be used to measure the quality characteristics?
5. How are the metrics calculated or derived?
6. How is the performance of the quality metric assessed?
7. What is the importance of establishing business practices for working with agencies that are not currently submitting quality data?

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

Intended Audience

The CDIP Guide is intended for a target audience of State crash database administrators and managers, State Traffic Records Coordinating Committee (TRCC) members, State highway safety office and State Department of Transportation (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. However, the information should benefit anyone who has a chance to use the state crash data.

Reference States Used in the Guide

Most of the examples of good practices cited in the Guide are drawn from practices adopted by the states of Michigan, Iowa and Kentucky. All three states are among the leaders in developing, maintaining and managing crash databases that are capable of providing good quality crash data to users. Their ability to provide this quality data is in large part due to the constant vigilance with which they scrutinize the information (data) being provided to the database and the actions they take to maintain and improve the quality of their crash data.

1.2. Background

Despite significant gains in reducing traffic related fatalities since the enactment of Federal motor vehicle and highway safety legislation in the mid-1960s, the annual toll of traffic crashes remains tragically high. In the United States, over 35,000 people are killed annually in traffic crashes on the nation's highways and an additional three million people suffer serious injuries. Motor vehicle crashes are the leading cause of death and disability in the United States for two year olds and people of every age from 4 to 33. Furthermore, traffic crashes are not only a grave public health problem for our nation, but are also a significant economic burden. In 2005, traffic crashes cost the U.S. economy approximately $250 billion, or more than two percent of the Gross Domestic Product.

Congress, in passing the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU, 2005), a transportation reauthorization bill, identified reductions in the number of crashes and the associated fatalities and injuries as the basis for judging the effectiveness of highway safety programs under the Highway Safety Improvement Program (HSIP). The determination of effectiveness is contingent on each State having good quality traffic safety data and using that data to determine the location, severity, and changes in their traffic safety problems.

> *...data which are timely, accurate, complete, consistent, integrated and accessible.*

Traffic safety data is the primary source of our knowledge about the traffic safety environment, human behavior, and vehicle performance. Therefore, in order to address these safety problems, we require good traffic safety data, meaning data which is timely, accurate, complete, consistent, integrated, and accessible.

1.3. Safety Decisions Are Improved With Good Quality Data

Good quality data has the potential to improve problem identification, the prioritization of different safety problems and the evaluation of the effectiveness of countermeasures. Safety problems may have several potential solutions, such as:

> *Good quality data has the potential to improve problem identification, analysis, and prioritization of a specific safety problem.*

- Engineering the infrastructure to remove or minimize the hazard (e.g., sharp curve, inconsistency in geometrics, and visibility issues).
- Enforcing existing laws to ensure driver/vehicle compliance (i.e., graduated driving licensing, alcohol, and speeding).
- Educating the public on safety issues (i.e., seat belt use, aggressive driving, and speeding).
- Improving Emergency Medical Services through processes such as training EMS personnel or the deployment of EMS units.

The availability of good safety data allows us to accurately identify the problems, assess the potential effectiveness of the selected countermeasures and to actually evaluate the effectiveness of those countermeasures.

1.4. Traffic Safety Databases

Data for determining the severity and extent of traffic safety problems include information on the crashes, drivers, vehicles (including commercial motor vehicles), the roadway environment (including traffic volumes), injuries, and traffic violations. Therefore, the databases that are considered to comprise a traffic safety information system are the:

- Crash
- Roadway/Traffic
- Driver Licensing
- Vehicle Registration
- Emergency Medical Services/Injury
- Citation/Adjudication

Figure 1 depicts the relationships among these systems as an interlocking set of related information. For traffic safety purposes, crash data is at the center of this *honeycomb* of information; while all of the other databases are integrated with or linked to the crash data.

Figure 1: The Traffic Records System

Data collected at the State and local levels serve as the foundation for any traffic records system. Whenever a vehicle is registered, a driver's license is issued, a traffic counter clicks to record traffic volumes, or a car crashes; data is generated. The collection, storage, and use of this data support the primary business process of the collecting agency (e.g. driver licensing, vehicle registration etc.) but this data also provides a useful source of information for traffic safety. Thus, the traffic safety information system has the capability to serve as an information resource for traffic safety professionals to identify traffic safety problems, select countermeasures, manage countermeasure programs, and evaluate the performance of safety programs.

1.5. The Crash Data Component

The focus of the Guide is on the crash data component of the traffic safety information system, which is the driving force for most traffic safety programs. The crash database will contain information from the law enforcement officers investigating and reporting on traffic crashes. It may also contain operator crash reports where there were no fatalities or injuries and the property damage may have been at or below the state's reporting threshold.   The property damage reporting thresholds will vary from state to state.

The data collected in the crash report will contain information on person(s), vehicle(s), circumstance(s), location and environment in which the actual crash event occurred. The crash data will be obtained from direct collection by law enforcement officers or possibly derived through linkages from other state databases. Table 1 presents examples of the type of information (data) that may or should be available in a state's crash database.

Table 1: Sample Data Elements of a Crash Database

Components	Examples
Crash	• Weather condition and pavement surface condition • Illumination • Time of Day, Day of Week • Avoidance maneuvers • Violation of traffic law (speed, turns, failure to obey, reckless driving) • Number of fatal, severe injury or property damage only crashes • Number of fatalities and severe injuries • Number of vehicles involved • Manner of collision and speed • Object struck • Person type (driver, occupant, pedestrians) • Substance abuse • Safety device use
Roadway	• Location referencing system • Roadway character (jurisdiction, classification, surface, geometrics) • Structures (bridges, tunnels) • Traffic control devices, signs, delineations, and markings • Roadside features (hardware, conditions, bike lanes, sidewalks, land use) • Rail grade crossings • Traffic volume and characteristics
Vehicle—All	• Type and configuration • VIN • Age/model year • Weight • Registration information/Plates • Defects • Owner information • Safety devices (type and condition)
Vehicle—Commercial	• Carrier information (including DOT number) • Hazardous materials/Placards • Inspection/Out of Service Records

Components	Examples
Driver	• Age/DOB • Gender and Ethnicity • Experience, driver education • License status • Conviction history
Injury Surveillance System	• EMS response time for driver/pedestrian/pedal-cyclist • Hospital assessment of injury severity • Hospital length of stay and cost • Rehabilitation time and cost

1.6. General Observations About State Crash Data

There are many potential reasons for inadequate crash data. The problems associated with this data may be related to three broad categories: people, processes and/or technology.

People

There may be many reasons that the quality of crash investigation and reporting varies from agency to agency and, in fact, from officer to officer. For example, training requirements vary among states. The level of training provided to officers may vary among law enforcement agencies. Officers may lack an understanding of the crash data element definitions or how to measure or interpret some of the information they are asked to report on. They may lack an understanding of the importance of crash data collection because they do not know the multitude of uses for this information. Some police agencies may not view the timely and accurate completion of the crash form as a "mission-critical" item, and thus data has the potential to be delayed, error-ridden or incomplete. One refrain commonly heard from police is that "crash forms are being completed just for insurance companies." One method to help police agencies to perform better crash investigation and reporting may be to provide them with feedback on the quality of reports that they are submitting.

Processes

Inaccurate data can result from numerous causes relating to the processing of crash data. The accuracy of the submitted crash data can suffer from cumbersome edit checks with paper crash report forms or inadequate edits with electronic collection. The timeliness of data can be affected by the number of times the forms are handled by the custodians of the crash database. For instance, delays could result if the forms shipped to another office outside the custodial office for location coding. Accuracy errors can result from errant "keystrokes" by data input personnel. Again, if the individuals responsible for processing the data are provided information and feedback on their processing, they are in a better position to improve their data handling and performance.

Technologies

To the degree that states can afford it, adoption of new and innovative technologies can help improve the quality of the crash data. Electronic data collection, whether through laptop computers or "on-line" entry of crash reports, can help improve the timeliness, accuracy, and completeness of the crash data. The use of global positioning system (GPS) units or GIS-based "smart-maps" can more precisely determine the location of crashes. And the creation of "data warehouses" can assist in making crash data available to users and also assist in integrating crash data with other traffic safety information system databases.

These situations and more affect the quality of crash database. However, if a state has established a mechanism to assess the quality of its data, it is in a much better position to detect deficiencies and take steps to correct them.

1.7. Ensuring Data Quality

Purpose of Measuring Crash Database Quality

One of the first actions in improving the traffic records system in any state is to identify its strengths, weaknesses, and areas of potential efficiency improvements. The ability to recognize the type and scope of deficiencies is necessary to begin to take steps to correct the deficiencies. As a first step, states should determine the quality measure(s), or metric(s), that will provide the greatest utility to the database administrators in determining the quality of the data within the database. A benchmark or baseline measure of this characteristic should be determined for each data quality characteristic. Through benchmarking, a State is able to obtain a quantitative, performance-based, measure to determine the current functional status and to gauge future progress in terms of that data quality characteristic.

> *... evaluated in terms of standardized, quantitative performance-based measures and matrices for measuring progress, including its own benchmarks.*

Data Quality Characteristics

The six data quality characteristics this Guide is concerned with are:

1. ***Timeliness:*** Information should be available within a specific timeframe to allow for meaningful analysis of the current status of the issue under investigation (e.g., the number of injury crashes at a specific location within a limited timeframe).
2. ***Accuracy:*** Information within the database should be correct and reliable in describing the data element it purports to describe. Accuracy is typically enhanced through the practice of conducting consistency checks and validations on the data being entered into the database.
3. ***Completeness:*** Information within the database should be complete in terms of all reportable instances of the event/characteristic being reported and available within the database, and all required data elements within the record should be completed with appropriate responses.
4. ***Consistency/Uniformity:*** Information collected should be consistent among all reporting jurisdictions with all reporting jurisdictions using the same reporting threshold and reporting the same information on a standard data collection form(s). Ideally, information will be reported using nationally accepted and published guidelines and standards (Model Minimum Uniform Crash Criteria (MMUCC), ANSI D.16, ANSI D.20).

5. ***Integration:*** By using common data elements, information in one database should be capable of being linked with information from other databases. An example of integration is the linkage of crash data with roadway inventory data by having a common location element in each database.
6. ***Accessibility:*** Information within the database should be readily available to all eligible users of the information**.**

These performance areas are also referred to as the "*6 pack.* "

Performance Measures of Data Quality

Performance Measures should include a unit of measurement in assessing data quality, such as:

- Days (is one measure of timeliness)
- Errors (is one measure of accuracy)
- Empty Fields (is one measure of completeness)
- Number of different forms used to collect information (is one measure of consistency)
- Number of different databases that can be linked (is one measure of integration)
- Number of individuals querying the database (or number of times the database is queried) for information (is one measure of accessibility)

There are various ways to measure the performance of the quality characteristics. Table 2 presents some examples of performance measures that can be determined to obtain a quantifiable measure of the quality characteristic. The performance-based measures are judged in terms of an increase or decrease from a baseline in terms of the unit of measurement.

Table 2: Example of Performance-Based Measures for the Crash Database

Characteristic	Performance Measure	Sample Benchmark
Timeliness	• # days from crash event until data is available for analysis in crash database	30
	• % of crash reports entered into the system within 30 days of crash	>95
Accuracy	• % of crashes *locatable* using roadway location coding method	>95
	• % VINs that are valid (i.e., match to vehicle records that are validated with VIN checking software)	>90
	• % of interstate motor carriers *matched* in MCMIS	>95
	• % crash reports with 1 or more uncorrected "fatal" errors	<1
Completeness	• % Law Enforcement Agencies with unexplained drop in crash reporting one year to the next	<5
	• # of reportable data elements with no value	<98
Consistency	• % of time *unknown* code is used in crash fields with that possible value	<2

Characteristic	Performance Measure	Sample Benchmark
	• % logical error checks on crashes that fail	<5
	• Number of elements missing per MMUCC guidelines	X of 111
Accessibility	• Number using on-line crash data system for data retrieval and statistical reports	
Integration	• % of crashes posted to Driver history file for drivers	>99
	• % of injury traffic crashes with indicated EMS response linked to associated EMS run	>90

1.8. Paper Versus Electronic Reporting of Crashes

Many states and local agencies have begun to collect data electronically. The use of electronic field data collection software, along with electronic transfer of data, typically has major advantages over a system that relies on paper reports. The most important advantages generally are:

1. More accurate data
2. More timely data
3. More complete data
4. Faster retrieval and easier access
5. More effective use of resources
6. Better opportunity for quality control monitoring
7. Better opportunity for electronic integration with other databases (e.g. driver and citation/adjudication)

While measures of implementation of electronic reporting (e.g. percent of law enforcement agencies using electronic data collection/transmission) are not of themselves measures of data quality, it is clear that the data quality of the crash database, as well as other databases, can be significantly enhanced by electronic field reporting. Most states are pursuing efforts to increase the amount of crash information being collected and transmitted electronically. When assessing the quality characteristics of the crash database it would be advisable to conduct two separate analyses. One set of quality analyses should be conducted for the data submitted electronically and one for crash data submitted via paper crash forms. The quality distinctions noted between the electronic and paper methods, in addition to highlighting the expected advantages of electronic collection/transmission, can potentially be used for marketing the benefits of electronic collection to agencies not yet committed to this method.

Adoption of Practices Advocated by the CDIP Guide

One of the principal sources of federal funding for data and traffic records system improvements is National Highway Traffic Safety Administration's (NHTSA) Section 408, State Traffic Safety Information System Improvement Grants. These grants are provided to assist states to develop and implement improvements to their traffic safety databases. As states adopt and implement the practices contained in the CDIP Guide they stand to benefit by enhancing their ability to demonstrate their compliance with portions of the qualifying criteria for Section 408 funding.

2.0 Timeliness

The call for more timely crash data is a common theme in traffic records. If a 2005 statewide crash file still is not released to users until 2007, the only knowledge gained is its unacceptable lateness. On the other hand, calling for timeliness to be *measured* is still rare. How can determining the actual measurement of data timeliness contribute to improving the timelines of crash data?

The six safety data quality performance areas are timeliness, accuracy, completeness, uniformity, integration, and accessibility. The monitoring and measurement of timeliness, is examined in the following section.

The acquisition of crash data at the state level is a multi-step process. It involves the collection, transmission, processing and management of the crash data. The collection of crash data is generally done by law enforcement officers who work within their jurisdictional boundaries and then transmit the data to the state custodial agency. Once the state custodial agency receives the data, depending on how the information is received (e.g. on paper forms or electronically), there are different processes for entering the data into the state's crash database.

2.1. Purpose Of Measuring Timeliness

Routinely monitoring one or more time intervals within the crash-reporting and processing system serves the following purposes:

- Early detection of slow or slowing processes.
- Identification of where the slowing is occurring (in which measured time interval it is occurring) so corrections and improvements can be made.
- Facilitation of management projects aimed at improving crash data timeliness.
- Quantification of the effectiveness of completed data improvement projects aimed at improving crash data timeliness.
- Detection of unintended consequences that other crash system changes and improvements may have on timeliness (such as changing from one type of database to another, changing the content of crash report forms or adding an edit program to the processing sequence).

Taken together, these advantages of measurement help a State achieve and maintain a level of timeliness required for data-driven decision making by all State safety disciplines and safety data users.

2.2. Defining Timeliness

Timely crash data is defined as: *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, and when the needs of data users are met.*

Crash data is timely when ... the needs of data users are met.

This means that good data management policies and practices keep delays to a minimum. It may also require modernizing antiquated, time-consuming methods and assuring that more efficient methods are working as intended. The implementation of timeliness changes as the state-of-practice in traffic records changes and as the users needs and expectations change. Some crash data users feel that *timely* ultimately means approaching real-time data. While this is not typically the state-of-practice today, there are technologies available that come close to real-time reporting.

> *In Iowa, police agencies using Traffic and Criminal Software (TraCS) for electronic crash data capture can create a near real-time local crash database provided they enter all reported crashes into TraCS. The upload to the local database takes place simultaneously with the upload to the State custodial office, which gives State authorities immediate access to those crash reports as well.*

It is important to distinguish between two different aspects of crash data timeliness:

1. **For an individual crash report**: Timeliness for an individual crash report is indicated by the time between a traffic crash occurrence and the time when that case information is placed into an accessible database. The crash case can then be retrieved from the electronic database, studied as a single event, and related to other crashes and events in the same location and time frame.

2. **For a crash dataset**: Timeliness for a complete and fully edited dataset is indicated by the time between the occurrence date of the most recent crash contained within the dataset and the time when the dataset is available for analysis. The dataset (whether for 1 month, 1 year or some other time period) must be comparable in completeness, accuracy, and uniformity to previous datasets released for statistical use. (The term "dataset" may be used to represent a complete statewide crash database or some component of the database, for instance a specific city or county, for some fixed period of time.)

2.3. Assessing Performance For Timeliness

As indicated in Table 2, there are many different measures for assessing timeliness. Generally speaking, the more precisely the measure for timeliness is determined, the greater the value of the measure in determining deficiencies and implementing corrective processes.

In measuring the timeliness of a state's crash database, the most basic measure will be the number of days (on average) between the occurrence of the crash and when the data is available in an electronic database. This is a value that should be calculated and not estimated.

Because assessing the performance of crash data timeliness is a multi-step process, it is helpful to separate out the police crash reporting phase for separate evaluation from other, possibly distinct, aspects of the data entry process.

A. Assess timely performance of police submission of reports.

Steps in police crash reporting include the investigating officer at the scene, collecting crash information at the scene (and also afterward), entering crash information into a report, approval of the report by a supervising officer, possible additional processing or filing by the police agency, and delivering the report to the State custodial agency. Delays in police report of crashes may be due to blood alcohol concentration (BAC) tests that have been taken, but await laboratory results before entry into the crash report. There are many other reasons for the process to be delayed at numerous points in this sequence.

When crash reports are not completed and reported by police in a timely manner, the delay in this first step of the system will cascade to all subsequent stages. It is impossible to measure the timeliness using only anecdotal evidence. Having a metric (a specific measure) for a process is not sufficient unless someone is taking time to read its output reports, interpret the information, and takes necessary action. Effective use of a metric requires its integration into core business practices of the assessing entity. Taken together, these steps assess performance of timely report submission. Every state should institutionalize the process of assessing the quality of their data.

B. Assess timely performance of the crash database system.

Police reporting is not the only part of the system affecting timeliness. Once reports (electronic or paper) reach the custodial office, the business practices and automated processes of the custodial office come into play. Some delays may be caused by assessing whether or not involved drivers have proper insurance coverage or converting the crash location into the State's integrated crash location methodology. Sometimes more than one State agency or office is involved in these processes that ultimately delay entry into the State crash data file. If a component function, like location coding, gets backlogged, then the completion of the state crash file will be delayed.

> *When Iowa implemented an electronic field-based method of capturing location information into a Geographic Information System (GIS) via a "smart" map approach, the custodial office simultaneously switched to the "smart" map as its method of crash location. Data clerks previously used paper maps; the new method made it not only more efficient but also eliminated coding backlogs. While electronic location capture in the field is preferable, the application of the smart map to paper reports still being received produced a timeliness benefit to the crash file as a whole.*

2.4. Developing a Metric For Timeliness

A metric is a specific measure that can be taken to assess where a database is in terms of a specific unit of measure (e.g. number of days, percent of cases reported within a specific timeframe, etc.). A metric provides a "snapshot" of where the database stands at a specific point in time, in terms of the unit of measure. The steps required in creating a metric for timeliness include:

1. Decide what aspect or component process of timeliness is to be measured. The State TRCC can help in making a preliminary selection. The agency most affected should develop the final metric. For example, these may include:
    a. Overall time from crash to data available for analysis.
    b. Time from police submission of reports to custodial office.
    c. Time from the custodial office processing from receipt to inclusion in database.
    d. Time from the locating of the crash to a base map process.

2. Choose a time interval to be measured that fits the process being monitored. Ensure it has a start date and an end date available. If using a unit of measure other than *days elapsed*, such as the percent of a process completed by a certain time, those numbers

should be readily obtainable as well. Dates useful in developing a metric for timeliness include:

a. Date of crash.
b. Date of receipt by crash database management office.
c. Date entered into crash database.
d. Date validated and posted.
e. Date the data is available for analysis by traffic and safety personnel.

3. Determine how the time interval data (or other criteria) are channeled into its own output report for use by data managers, and by what formula or algorithm. For example, if *days elapsed* for a crash report to arrive at the State custodial office is the metric to be used; will the data for all crashes be averaged by each reporting agency by month? (It would not make sense to measure the time it takes police to file a report and add it to the State's crash database unless the performance of specific police agencies can be identified and measured.)

Various criteria may be used for metrics, such as:

- ***Days elapsed*** *per crash case. (from crash occurrence to custodial office).*
- ***Average number of days statewide from crash occurrence to availability for analysis****.*
- ***Percentage of crash reports processed*** *by a specific time.*
- ***Actual versus expected number of crash reports from a reporting agency*** *in a given time interval, and so on.*

4. Develop or establish a benchmark for timeliness based on the criteria agreed to above. A benchmark is a "starting point" measure of the timeliness of the overall crash database or a component process. It is important for assessing progress to determine the starting point against which future progress will be measured.

5. Produce the report. At this point, a means of measurement (metric) has been created, but it is not useful for data improvement until someone with the designated authority examines the metric's output (or outcome) and takes action as needed. Metrics must be part of the business practices of the agencies involved in order to provide meaningful information.

2.5. Examples Of Metrics For Timeliness

Metrics for timeliness and their implications must be part of the State agencies business practices if it is to have any benefit. Michigan generates several routine reports of timeliness. Two examples of the Michigan timeliness metrics are presented below.

The metrics in the first example report on timeliness in terms of the average number of days between crash occurrence and when the information from the report is entered into the state's electronic database by the Michigan State Police (the custodial agency for the Michigan Crash Database). The timeliness metric in the second example assesses the expected reporting level by agency and the actual number of reports submitted by that agency on a year-to-date basis.

For each example, additional information is presented on:

- An explanation of the metric—how it is obtained, calculated, and reported.
- Sample output of the metric.
- Assessing performance from the metric.
- Using the metric in a performance measure.

Example 1. Using *Average Reporting Days* to Assess Police Reporting Timeliness by Agency and Statewide

Metric—How it is Obtained, Calculated, and Reported

The metric in this example measures the time between the crash occurrence and when the data on the crash report are entered into the crash database and are available for analysis. **This metric provides the most basic measure every state should be calculating.** The metric is calculated by counting the number of days between when the crash occurred (this information is listed on the crash report form) and when the crash information is entered into the electronic crash database. The date of entry into the crash database is information that the state maintains in a management database. To determine the statewide average number of days between crash occurrence and electronic database availability, simply sum the number of days between the crash and data entry for each individual crash report and divide the sum by the total number of crashes that have been reported. This will provide a calculation of the average number of days it takes in the state of Michigan for a crash to be available for analysis. Michigan also uses this process to identify the average number of reporting days for each individual law enforcement agency. This calculation is the same as for the state as a whole, but it is done for each individual reporting agency.

Figure 2 shows Michigan's output report using this metric, which lists the average reporting days for a group of law enforcement agencies from January 1, through June 19, 2006. Included in the report is the number of crashes reported to date for a sample of agencies and the calculated average number of days between the crash and entry of data into the electronic database. Based on the number of cases reported, one is able to discern, as an indirect indicator, the relative size of each reporting agency. At the bottom of Figure 2 is the summation for all agencies and cases reported. This is the statewide average number of reporting days for Michigan from January 1, 2006 through June 19, 2006.

2006 Crash Reporting Stats
Status Date 06/19/2006

Agency Name	2006 Crashes	Average Reporting Days
Huntington Woods Police Dept	42	12.58
Keego Harbor Police Dept	28	29.36
Kensington Metro Park Police Dept	5	10.99
Lake Angelus Police Dept	0	0
Lake Orion Police Dept	39	43.25
Lathrup Village Police Dept	147	19.24
Madison Heights Police Dept	444	11.38
Milford Police Dept	159	10.1
Novi Police Dept	515	11.93
Oak Park Police Dept	265	13.85
Orchard Lake Police Dept	77	11.5
Oxford Police Dept	28	26.11
Pleasant Ridge Police Dept	56	24.9
Pontiac Police Dept	765	26.72
Rochester Police Dept	107	15.98
Royal Oak Police Dept	481	15.11
Royal Oak Twp Police Dept	0	0
South Lyon Police Dept	48	10.23
Southfield Police Dept	1083	25.18

Summary of Output

Statewide Total Crashes	**121,787**	
Statewide Average Reporting Days (Database Load - Crash Date)		**19.65**

Figure 2: Sample Output of Average Reporting Days by Agency and Statewide

In assessing the Average Reporting Days, a reviewer would determine which agencies are slower in reporting crashes to the state by examining the Average Reporting Days column for police agencies having the highest averages. These agencies should be evaluated based on the relative size of the agency size from the crashes column and compared to the statewide average days. By examining the **average reporting days report** for each agency each month, agencies with the highest average reporting days may be identified. Often a simple phone call to determine why reports are delayed can spur an agency to improve its reporting.

Figure 3 is an example of a useful metric. It represents the actual average number of reporting days that Michigan has experienced between 2003 and 2006 (as of June 19, 2006).

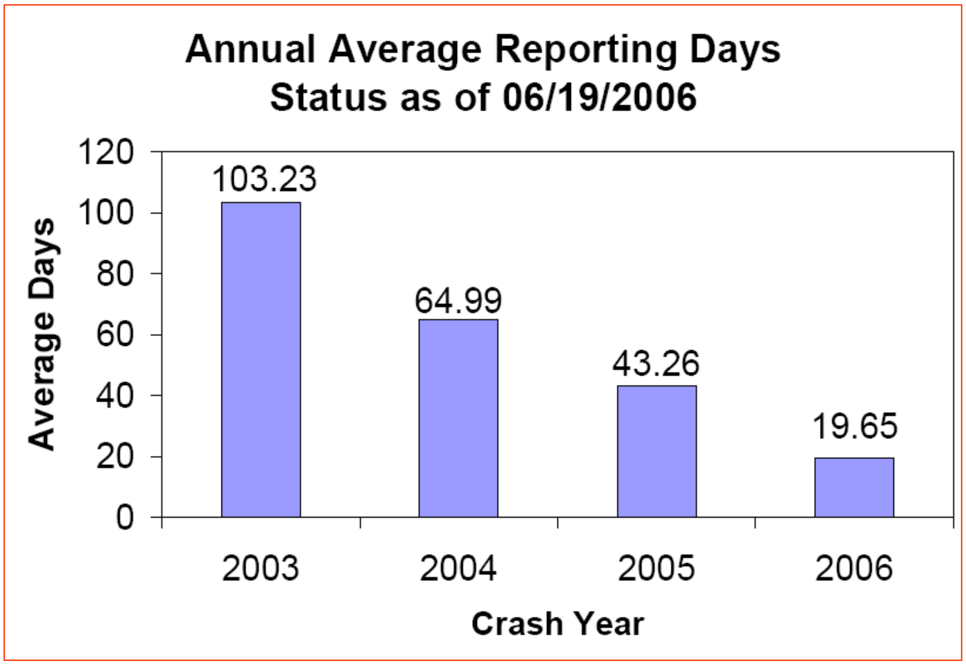

Figure 3: Comparison of Annual Average Reporting Days

This report demonstrates the improvement in timeliness of crash data that Michigan has realized over the last few years. It is easy to generate such a multi-year report once this metric has been implemented and continued for a number of years.

Michigan creates monthly reports on this metric, but they could also be generated for any interval without reducing accuracy. Michigan creates it monthly report to fit its business practices. It is worth noting that there are at least two components to this measure. One is the time taken by the reporting agency; the other is the time taken by the State custodial agency from receipt of the report (whether paper or electronic) and its addition to the database. It may be advantageous for some States to measure each of these components separately.

Using the Metric in a Performance Measure

Below is a hypothetical sample performance measure for using *average reporting days*:

Sample Performance Measure

The State will improve the **timeliness** of the **crash data base** by demonstrating a measured decrease in **average reporting days** (crash occurrence to database upload) where the **baseline level was 63 days** (hypothetical example) in 2006, and the **goal levels** for the future are:

- 50 days by the end of 2007
- 40 days by the end of 2008
- 30 days by the end of 2009
- 20 days by the end of 2010

Example 2. Assessing Police Reporting Timeliness Using Expected Reporting Level.

Metric—How it is Obtained, Calculated, And Reported

At the State or individual law enforcement agency level, crash frequency rarely varies much from year to year, despite efforts to reduce the toll. This reliability allows for a comparison of expected crashes (the number of crashes reported by this date a year ago) with actual crashes (the number of crashes reported by this date this year). This comparison allows crash database managers to monitor underreporting as the year progresses and to detect agencies that either fall behind in investigating crashes (this may be a function of a reduction in crashes) or in submitting crash reports to the state. To calculate this metric a state needs to maintain a record of the number of crashes reported by individual agencies on a year-to-date basis, thus permitting a comparison of the number of cases an agency has submitted this year compared to the number of cases they had submitted to the same point in time last year. Through a simple summation procedure this comparison can be extended to the whole state.

Figure 4 shows Michigan's output report for this metric, which is used primarily as a way to flag agencies that are significantly behind in reporting this year for follow up. An observed crash reduction from previous years may result from effective programs or to chance alone, but, at least the possibility that the crash reduction results from incomplete or untimely reporting should be considered, and examined if found.

CRASH Reporting Status - June 19, 2006

Expected Level of Reporting 41%

Agency Name	Penalty Total	Data Edit Error Total	2006 Crash Total	2005 Crash Total	Per Cent of 2005
Alcona Co Sheriff's Office	0	10	184	560	32.86%
MSP Munising	0	114	98	306	32.03%
Alger Co Sheriff's Office	3	30	10	49	20.41%
Munising Police Dept	0	36	17	62	27.42%
MSP Wayland	0	177	200	581	34.42%
Allegan Co Sheriff's Office	0	375	702	2000	35.10%
Allegan Police Dept	0	73	34	81	41.98%
SAUGATUCK/DOUGLAS P	0	26	8	34	23.53%
Hopkins Police Dept	0	6	2	0	0.00%
Otsego Police Dept	0	79	25	80	31.25%
Plainwell DPS	0	77	42	119	35.29%
Wayland Police Dept	0	48	19	42	45.24%
MSP Alpena	1	135	184	541	34.01%
Alpena Co Sheriff's Office	0	10	99	274	36.13%
Alpena Police Dept	1	134	104	240	43.33%
Antrim Co Sheriff's Office	1	324	286	891	32.10%
Bellaire Police Dept	2	16	3	14	21.43%
Central Lake Police Dept	0	3	3	8	37.50%
Elk Rapids Police Dept	0	39	16	49	32.65%
Mancelona Police Dept	0	18	2	25	8.00%
Arenac Co Sheriff's Office	0	500	245	688	35.61%
Augres-Sims Police Dept	0	4	8	32	25.00%
MSP L'Anse	0	65	110	331	33.23%
Baraga Co Sheriff's Office	1	54	42	111	37.84%
Baraga Police Dept	0	0	0	15	0.00%
L'anse Police Dept	0	0	0	17	0.00%
MSP Hastings	0	196	271	806	33.62%
Total Crashes Reported			**121,787**	**352,592**	**34.54%**

Figure 4: Sample Output of the Expected Frequency of Reporting Metric

The Crash Reporting Status Table shows the expected reporting level for individual agencies based on the number of 2005 UD-10's (the Michigan State Uniform Crash Report form) that had been submitted in 2005 and compares that to the actual number of 2006 UD-10s that have been submitted through 6/19/2006. The statewide expected rate for 6/19/2006 is about 35 percent. The "Percent of 2005" column is the key for determining agencies that may be delinquent in reporting 2006 cases. Based on the Table, most agencies appear to be submitting at a normal rate. The Baraga and L'anse Police Departments towards the bottom of the Table may be delinquent in reporting and could be flagged for follow up contact.

This measure will be a better monitor of agency performance as the number of crashes becomes larger. To interpret the data for a single agency, it is better to know if it has few crashes or many. The expected reporting level is less a measurement than an indicator or flag for looking for timeliness problems.

Michigan generates this report monthly, but reports do not show separate months. Rather, the table shows accumulated crash totals. Early in the year, the random variation inherent in crash frequencies is exacerbated by small sample sizes. As the year progresses and accumulated crash totals grow, the method becomes better able to identify agencies that are potentially late. This measure lacks the precision and sensitivity of the prior example, but works well to flag potential problems, particularly toward the end of the year.

Using the Metric in a Performance Measure

Below is a hypothetical sample performance measure using the metric *expected reporting days*:

Sample Performance Measure

The state will improve the **timeliness** of the **crash data system** by identifying **the 15 percent of agencies with the smallest actual/expected percentages, averaged over 2005-2006,** and increasing the actual/expected percentages of this group by 10 percentage points each year from 2007 through 2010 (approaching 100 percent is considered ideal).

The baseline is calculated to be 43 percent over 2005-2006. The goals for this subset of agencies for subsequent years are:

- 53 percent in 2007
- 63 percent in 2008
- 73 percent in 2009
- 83 percent in 2010

There are many ways to write performance measures for this metric by changing the baseline definition, changing the method of computation, and setting goals. A State should carefully consider what can work best to meet its own needs.

Table 3: Summary of Examples for Timeliness

Metric Example	Description	Measure	Business Practice
1. Average Number of Reporting Days (by agency and statewide)	Michigan statewide average number of reporting days has been reduced from 103 in 2003, to 20 in 2006.	Measured from crash date to data entry in the database date.	Statistical report generated at regular intervals. Law enforcement agencies (LEAs) are monitored.
2. Expected Reporting Level	If at a given time of year, 83 percent of the year's crash reports are expected (based on the previous year), and only 74 percent were actually turned in statewide, then agencies reporting a percentage significantly less than 74 percent could be potentially late with their reports and targeted for follow up.	Comparison of expected to actual number of crashes reported by each agency.	Report generated at regular intervals for review. LEAs not in compliance are identified for follow-up.

3.0 Accuracy

The six safety data quality performance areas are timeliness, accuracy, completeness, uniformity, integration, and accessibility. The monitoring and measurement of accuracy, is examined in the following section.

Accuracy is a characteristic of data quality. On an individual crash report basis it is a measure of how precisely the information about the person(s), vehicle(s) environment and circumstances of that crash are reported. On an aggregate basis, it is a measure of how well the information in the crash database reflects correct information about the who, what, when, where and why of all the crashes reported for a particular jurisdiction or the whole state. Yet, measuring accuracy (validity or degree of truthfulness) can be more difficult than measuring timeliness.

3.1. External And Internal Validity

There are really two measures of accuracy. One measure of accuracy requires validation information from outside the crash reporting system itself. This is referred to as "external validity." Examples of this type of measure would be, when the "Vehicle Make" reported to have been involved in a crash as a "Chevrolet" is confirmed to be a Chevrolet rather than a Ford. Another example is when the crash location is reported as "the intersection of First Street and Main Avenue" when in-fact the actual location was Second Street and Main Avenue. In the first case, we have valid or accurate information while in the second case we have invalid or inaccurate information. External validity measures the accuracy of "does the reported information" reflect what actually happened in the crash.

Assessing external validity may require its own special study or data collection process. Sometimes external validation can be performed by checking specific crash data elements against a separate data file containing the allowable codes, which provides for continual monitoring—the ideal situation. For example, by checking the license plate number of in-state vehicles against the state's vehicle registration file, we can confirm that crash involved vehicle reported as being a "2002 model year Chevrolet" was in fact a 2002 model year Chevrolet. However, it may be more difficult to confirm the actual location of the crash as being Second Street and Main Avenue in the absence of any information source, such as a roadway inventory database, other than the crash report itself.

Internal validity is assessed by examining the values or attributes that are coded for the data elements. Are the reported attributes within a range of acceptable responses? For example, if an officer is to report on the actual level of intoxication (e.g. Blood Alcohol Concentration - BAC) of a driver subjected to a BAC test, the acceptable values would range between 0.00 and 0.99 (depending on state coding conventions). A response of "the driver was very drunk" would be an invalid response. Another means of assessing internal validity is to check the responses to one data element reported on in a crash form against other data elements being reported on in a crash form. An example of this type of internal check for accuracy would be to identify when the coded value for the "Light Condition" element is coded as "Daylight" but the "Time of Crash" is coded as occurring at "11:30 PM," there is likely to be an error in reporting on one these elements. An internal consistency check would highlight the fact that one of these elements is probably being 'misreported' and should be further examined to determine which element is incorrect.

One of the benefits of the electronic collection of crash data is that the software used to collect the information will often have these internal validation edits built into the software. So if the investigating officer tried to submit or upload a crash report with one or more of these internal consistency errors, this “front–end” edit would alert the officer to the problem data elements and the officer would have to make a correction(s) to one or more elements to remove the error(s), before the case could be uploaded. On the other hand if the agency responsible for processing the reports into the crash database discovers errors through its accuracy assessment edit process (e.g. back-end edit), they may have to return the report or check with the investigating officer to get the accurate response for the data element.

For the majority of data elements, it is often easier to check for errors through internal consistency edits and to use that approach for developing a metric for accuracy. This is not meant to minimize the importance of external validation checks and studies. Data managers should consider doing both, but the emphasis in this Guide will be on internal validation to assess crash data accuracy.

One final consideration for accuracy is “how accurate is the submitted data?” States should be cognizant of what the data is being used for and attempt to have the information precise enough to meet those needs. One example of this may be found in the precision of “locating” rural crashes. With the lack of convenient fixed reference points in rural locations, officers may often estimate the distance from a fixed reference point in approximate measures, such as “a quarter-mile west of the intersection of Route 1 and County Road A.” While this description will generally locate the crash to the road, it may overlook the fact the crash actually occurred at a sharp curve in the road that is 1,980′ (3/8 of a mile) “west of the intersection of Route 1 and County Road A” and that the combination of vehicle speed and the curvature of the roadway contributed to the crash. Though not a “cure all”, the use of GPS units or “smart maps” may assist in more accurately locating or depicting the roadway to assist in more precisely locating rural crashes.

3.2. The Purpose Of Measuring Accuracy

Routinely monitoring one or more aspects of accuracy within the crash reporting and processing system will serve the following purposes:

- Detect inaccurate data.
- Identify where the problems are occurring and the types of problems that are occurring so early intervention is possible.
- Facilitate management of projects aimed at improving crash-data accuracy.
- Quantify the effectiveness of completed data improvement projects aimed at improving crash data accuracy.
- Detect unintended consequences that other crash system changes and improvements may have on crash data accuracy (such as changing from one type of database to another, or adding an edit program to the processing sequence).

Taken together, these advantages of measurement help a State achieve and maintain a level of accuracy required for data-driven decision making by all safety disciplines and State safety data users.

3.3. Defining Accuracy

Accuracy is the condition or quality of being "true." Other terms sometimes used for accuracy are validity and reliability, with reliability conveying that the degree of accuracy is consistent in repeated measurements.

The accuracy of a database must be sufficient for the purposes for which it is used. This is a matter of judgment for the data users, whose opinion regarding which data elements are most crucial and the level of sufficient accuracy will differ.

3.4. Measuring Accuracy

The two approaches to measuring accuracy are, again, through internal (within crash report) validation edits and through external validation checks.

The two approaches to measuring accuracy are through internal (within crash report) validation edits and through external validation checks.

With internal edits, data is assessed to assure that the values reported for the data elements are within a range of acceptable responses (e.g. BAC is reported as a value between 0.00 and 0.99) or one data element is cross-checked with another to see if the pair makes sense. If the crash vehicle action is *ran traffic signal*, and the crash location is reported to be on a non-intersection segment of rural highway, then at least one of the two data elements is likely incorrect. Without more information, one cannot tell which it is. If the crash report locates the crash at an intersection, then the roadway file can be consulted for whether that intersection is really signalized. This would be an external validation. If one assumed the roadway file was more likely to be accurate, one would then conclude the information on the crash report was incorrect. Either the intersection was not signalized or the vehicle action was incorrect.

While accuracy throughout the report is desired, accuracy is more critical for some data elements than for others. One of the most critical accuracy issues for safety is the ability to identify where crashes occur. When a number of crashes occur within a particular section of roadway or intersection, this may indicate a safety problem. Therefore, one of the most critical accuracy issues is the ability to identify the location of crashes on all public roadways within a state. States use a variety of methods to locate crashes onto maps where the frequency and severity of crashes can be ascertained. Some of the methods include: link-node, linear referencing system or a global positioning system to identify the location of crashes. However, if the officer investigating the crash does not identify the location on the report in terms the state is able to use, the state will be unable to locate the crash. The ability to locate crashes is a particular aspect of accuracy that states must be measuring. Similarly, the accurate identification of data elements (VINs, driver license number, CDL licensing state, etc.) identifying commercial motor vehicle crashes are critical elements for other groups of crash data users.

Places Where Accuracy Errors Can Be Made In the Crash Reporting and Processing System

Aside from the internal/external validation dichotomy, there are other factors associated with measuring accuracy of the crash reporting system. Opportunities for error occur throughout the collection and processing of the data.

Below is a list of some steps in the data management sequence where accuracy errors can occur:

- Police reporting.
- The custodial office.
- The locating process.
- The editing/validation process.

Various criteria may be used to measure accuracy, such as:

- *Percentage of crash locations that have coordinates located on the roadway network/in the correct county, and so on.*
- *Average number of errors per crash report as found by internal edits.*
- *Percentage of crash reports that contain no errors or that meet a certain threshold for errors allowed.*
- *Percentage of valid Vehicle Identification Numbers (VINs) match to vehicle records that are validated with VIN checking software*
- *Number of police agencies that have invalid construction zone information among their top 10 most common errors.*

In order to choose one or more ways to measure crash data accuracy, a State must assess what its key challenges in accuracy are, and then devise metrics that allow for monitoring those key areas. If the internal edits are checking 150 cross validations, that does not necessarily mean all of them should become metrics. Editing results for total errors, error type and critical errors may be sufficient for ongoing monitoring.

3.5. Developing A Metric For Accuracy

General Steps in Creating a Metric

- Decide what aspect or component process of accuracy is to be measured. The State TRCC can help in making a preliminary selection; with the agency most affected developing the final metric.
- Select one or more criteria for the metric, such as, *total errors per data element, average number of errors per report*, *percentage of reports meeting a certain threshold of accuracy*, and so on.
- Determine what information is needed to calculate the metric.
- Determine how the metric's data are channeled into its own output report for use by data managers, and by what formula or algorithm.
- Develop or establish a benchmark for data accuracy based on the criteria agreed to above. Along with benchmarking, also establish goals for improving the metrics over time (with intermediate milestones).

- Produce the report. At this point, a means of measurement (metric) has been created, but it is not useful for data improvement until someone with the designated authority looks at the metric and takes action as needed. To have any benefit, metrics must be part of the business practices of the agencies involved.

3.6. Assessing Performance For Accuracy

A common situation with crash data is that accuracy is often left to those performing the edits on the data, individuals doing their best to achieve accuracy but who may not always be fully aware of the varying needs of crash data users. Users, in turn, are often unaware of what edit checks (if any) were performed on the data they are using. When users find a discrepancy in the data, or a glaring error, they do not know if they have stumbled upon a fluke or if such errors are found throughout the database. Communication and teamwork are essential in preventing this kind of disconnects between data processing managers and data users.

When the crash data edit process required an overhaul in Iowa, the Office of Driver Services (custodial office) and the Office of Traffic and Safety (a data user office) partnered to redevelop the edit process following major changes in the crash database.

The accuracy of crash data almost always needs better monitoring because of the challenge presented by the large number of data collectors. Even in small States, law enforcement officers often have minimal or inconsistent training in crash reporting. The State should employ quality control methods to ensure accurate and reliable information for each crash report by methods such as:

- Performing validity and consistency checks in the data capture and data entry processes.
- Providing feedback to jurisdictions regarding the accuracy of the reports they submit.

Managers should apply quality control methods to the aggregate crash database by using methods such as performing:

- Edit checks to detect over- or underreporting of specific data elements or categories.
- Edit checks to detect errors in the data processing software or methods (part of the debugging process when new or revised software is applied to the crash file, and also to detect unintended consequences of changes in crash processing strategies within the custodial office).
- Tests of data processes to identify what range of accuracy they provide, such as benchmarking of crash location accuracy.

3.7. Examples Of Metrics For Accuracy

Three examples are presented below for measuring Accuracy: Most Frequent Error Types by Agency, Average Number of Errors per Crash by individual agencies; and Classification of Error Type. Table 4, at the end of this section, summarizes the examples.

Example 1. Assessing the Accuracy of Reporting Agencies via Most Frequently Occurring Types of Errors on Crash Reports

Explanation of Metric - How it is Obtained, Calculated, and Reported

In Michigan, a Data Performance Report generates a report for each reporting agency showing data errors by type, listed in order of frequency of the error. The report is generated by running an internal validation procedure on the cases submitted by each agency to identify the data elements with the most errors and the type of errors that the reports contain. This permits the state to identify those elements that are misreported (accuracy errors) or not reported (completeness errors).

Figure 5 is specific to one agency. The Description column contains a ranking of the most frequently occurring errors for that agency by data element and the type of error that is being identified. The frequency with which the error is occurring is reported in the Count column. At the bottom of the Data Edit Errors report, the agency can view the number of crashes it has reported per month compared to the same month a year ago.

Data Edit Errors

Description	Count	Class
Missing Airbag Deployed	329	P
Invalid Construction Zone Info	317	C
Invalid Direction for Crash Type - Angle	258	U
Missing or invalid Alcohol Suspected	165	P
Missing Driver Condition	132	P
Missing or invalid Position for Driver	127	P
Missing Injury	123	P
Invalid Direction for Crash Type - Head On Left	104	U
Missing Location of Greatest Damage	94	U
Missing Access Control	87	C
Invalid Direction for Crash Type - Rear End	82	U
Missing Hazard Action	81	P
Missing Vehicle Registration Number	80	U
Invalid Most Harmful Event > 1 most Harmful	57	U

Error Class: C = by Crash, U = by Unit, P = by Person or Party. *** Multiple Data Edit Errors can be assigned for a single crash UD-10 form.

Crash Comparison Report
Totals As Of: 07/23/2006

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2006	747	537	476	488	582	505	172*						3507
2005	723	646	544	584	583	651	555	594	570	655	801	816	7722

* = The monthly total reflects at least a 30% difference from the previous year's monthly total.

Figure 5: Data Performance Report Identifying Crash Report Data Edit Errors with Month-To-Month Comparison of 2006 to 2005

This report addresses both accuracy and completeness errors (completeness errors will be discussed in the next chapter). Based on the findings from this report, steps can be taken to improve the accuracy of the reports being submitted by this particular agency. For example, a training session could be developed and presented to the officers to train them on the information being sought to complete the “construction zone” data element. Following the presentation of this training, monitoring the accuracy of data edit error assessments should find fewer “construction zone” errors.

Agency Name	Penalty Total	Data Edit Error Total	2006 Crash Total	Errors/Crash
Alcona Co Sheriff's Office	0	10	184	0.05
MSP Munising	0	114	98	1.16
Alger Co Sheriff's Office	3	30	10	3.30
Munising Police Dept	0	36	17	2.12
MSP Wayland	0	177	200	0.89
Allegan Co Sheriff's Office	0	375	702	0.53
Allegan Police Dept	0	73	34	2.15
SAUGATUCK/DOUGLAS POLICE	0	26	8	3.25
Hopkins Police Dept	0	6	2	3.00
Otsego Police Dept	0	79	25	3.16
Plainwell DPS	0	77	42	1.83
Wayland Police Dept	0	48	19	2.53
MSP Alpena	1	135	184	0.74
Alpena Co Sheriff's Office	0	10	99	0.10
Alpena Police Dept	1	134	104	1.30
Antrim Co Sheriff's Office	1	324	286	1.14
Bellaire Police Dept	2	16	3	6.00
Central Lake Police Dept	0	3	3	1.00
Elk Rapids Police Dept	0	39	16	2.44
Mancelona Police Dept	0	18	2	9.00
Arenac Co Sheriff's Office	0	500	245	2.04
Augres-Sims Police Dept	0	4	8	0.50
MSP L'Anse	0	65	110	0.59
Baraga Co Sheriff's Office	1	54	42	1.31
Baraga Police Dept	0	0	0	0.00
L'anse Police Dept	0	0	0	0.00
MSP Hastings	0	196	271	0.72

Summary Report for State:

Total Crashes Reported	**518**	**181,427**	**121,787**	**1.49**

Figure 6: Sample Output of the Metric: Average Number of Errors per Crash Report by Reporting Agency for 2006, Year-to-Date

Using this Metric in a Performance Measure

These agency-specific reports are best used by the agencies themselves to reduce errors. For example, each agency could use its last annual report as a baseline, and set goals for error reduction similar to the sample performance measure shown below.

If this information is summed for all reporting agencies, a statewide total of the most error-prone data elements will be generated. The error information contained in Figure 6 can be summed for each reporting law enforcement agency in the state to identify those data elements containing the most errors as well as the types of errors on a statewide basis. This would provide a statewide data accuracy report by data element and error type.

Sample Performance Measure

The State will improve the accuracy of the crash file data by reducing the frequency of each agency's five top-ranked errors from the level of their 2006 data edit report to:

- 80 percent of 2006 level by the end of 2007
- 60 percent of 2007 level by the end of 2008
- 40 percent of 2008 level by the end of 2009
- 20 percent of 2009 level by the end of 2010.

Example 2. Assessing Police Reporting Accuracy by Average Number of Errors per Crash Report

Explanation of Metric—How it is Obtained, Calculated, and Reported

Another measure for data accuracy would be to develop a metric to assess the average number of errors per crash report. Figure 6 presents sample output from a metric that Michigan uses to assess the accuracy of reports submitted by reporting agencies. This metric determines the average number of errors per crash report submitted for each reporting agency.

This is an example using internal (within crash report) validation edits. Software operating on the Michigan statewide crash file finds and tallies data errors. These are reported by reporting agency in the column Data Edit Error Total. (The Penalty Total column does not pertain to this metric. It indicates the number of paper reports that cannot be processed and therefore cannot be included in subsequent steps of the edit process). The software producing this report also prints out the Crash Total for the most recent complete year, by agency. The final column, Error/Crash, is calculated by dividing the error total for each agency by the crash total for that agency.

For example, using the Alcona County Sheriffs Office (top-listed agency), the error total of 10 is divided by the crash total for 2006 of 184 to get errors per crash equal to 0.05. Note that "crash" in this example means "crash report."

As shown in the summary report the Alcona County Sheriff's errors per crash (0.05) are much less than the statewide average of 1.49 errors per crash.

Using this Metric in a Performance Measure

Sample Performance Measure (Hypothetical)

The State will improve the accuracy of crash data by reducing the average errors per crash statewide from 5.2 errors per crash in 2007 to:

- Four errors in 2008
- Three errors in 2009
- Two errors in 2010

Example 3. Development of Error Classification Codes

Explanation of Metric – How It is Obtained, Calculated, and Reported

In order to determine the types of errors that are contained in the crash reports received by the Michigan State Police, the custodial agency of the crash database for Michigan has developed what amounts to an "Error Dictionary." This document attributes a unique numeric code to each type of error for which the state monitors. The "classification" column identifies how the errors are identified (e.g. through an edit process), the "severity" column identifies critical errors ("severe," which prohibit the information from being entered into the database, differ from "informational" errors, which are less critical but do not prohibit the information from being entered into the database). The "MC Error" column identifies errors pertinent to crashes involving commercial motor carriers. "Certification Errors" indicate that the element must be reported as a standardized form. "Performance Errors" indicate that the error can be tracked back to the total errors for an individual agency. Finally, the "Short Description" column identifies the type of error that occurred.

While this document is not a measurement of accuracy, it is necessary to be able define and identify the type of errors the state is identifying and tracking.

Error #	Classification	Severity	MC Error	Certification Error	Performance Error	Short Description
185	EDIT	Informational	No	Yes	Yes	Invalid Public Property
190	EDIT	Severe	No	Yes		Missing MSP County
195	EDIT	Informational	No	Yes	Yes	Missing City/Township
[illegible]	EDIT	Informational	No	Yes	Yes	Missing Road Name
205	EDIT	Informational	No	Yes	Yes	Missing Intersecting Road
210	EDIT	Informational	No	Yes	Yes	Missing Distance
211	EDIT	Informational	No	Yes	Yes	E-Source: Initial Framework Version not provi
212	EDIT	Informational	No	Yes	Yes	E-Source: Mapping Tool Version not providec
213	EDIT	Informational	No	Yes	No	Minimum Linear Referencing data missing
215	EDIT	Informational	No	Yes		Invalid Roadway Area - Must be 1, 4, 5, or 7
217	EDIT	Informational	No	Yes		Invalid Distance - Exceeds 10 miles

Figure 7: Sample Portion of an Error Code Table that Identifies Error Types the State is Monitoring

The State can use the information in Figure 8 to assess common errors. This information will be helpful in determining what corrective action needs to be undertaken. Further, when the information is broken down by reporting agency as shown in Figure 8 for MSP Wayland, it is clear where the deficiencies may lie (e.g. most errors occur for this agency relate to Construction Work Zones).

Law Enforcement Agency	Error Type	Errors	Crashes	Errors/Crash
0305600 - MSP Wayland	Invalid Construction Zone Info	51	235	0.22
0305600 - MSP Wayland	Missing VIN	45	235	0.19
0305600 - MSP Wayland	Missing Access Control	22	235	0.09
0305600 - MSP Wayland	Missing or invalid Restraint for Vehicles: PA, VA, PU, ST, GC, Oth	22	235	0.09
0305600 - MSP Wayland	Missing Vehicle Registration State	21	235	0.09
0305600 - MSP Wayland	Missing Traffic Way	20	235	0.09
0305600 - MSP Wayland	Missing Airbag Deployed	18	235	0.08
0305600 - MSP Wayland	Missing Age	16	235	0.07
0305600 - MSP Wayland	Missing or invalid Passenger Position Code	16	235	0.07
0305600 - MSP Wayland	Missing or invalid Drugs Suspected	14	235	0.06
0305600 - MSP Wayland	Carrier name does not match USDOT Legal name or DBA name	11	235	0.05
0305600 - MSP Wayland	Invalid Direction - Ramp Related	11	235	0.05
0305600 - MSP Wayland	Missing or invalid Alcohol Suspected	11	235	0.05
0305600 - MSP Wayland	Missing or invalid Gender	10	235	0.04
0305600 - MSP Wayland	Missing Age - Driver is present	9	235	0.04
0305600 - MSP Wayland	Missing or invalid Position for Driver	9	235	0.04
0305600 - MSP Wayland	Missing Driver Condition	8	235	0.03
0305600 - MSP Wayland	Missing or invalid Interlock	8	235	0.03
0305600 - MSP Wayland	Missing Injury	7	235	0.03
0305600 - MSP Wayland	Missing Location of Greatest Damage	7	235	0.03
0305600 - MSP Wayland	Invalid Investigated at Scene	6	235	0.03
0305600 - MSP Wayland	Missing Vehicle Use for Motor Vehicle	6	235	0.03
0305600 - MSP Wayland	Cargo Body Type is required	5	235	0.02
0305600 - MSP Wayland	Invalid Total Occupants	5	235	0.02
0305600 - MSP Wayland	Missing Action Prior	5	235	0.02
0305600 - MSP Wayland	Missing Special Circumstances	5	235	0.02
0305600 - MSP Wayland	Missing Vehicle Direction	5	235	0.02
0305600 - MSP Wayland	Missing or invalid Ejected	5	235	0.02
0305600 - MSP Wayland	Missing or invalid Trapped	5	235	0.02
0305600 - MSP Wayland	Invalid Direction for Crash Type - Sideswipe Opp	4	235	0.02
0305600 - MSP Wayland	Invalid Posted Speed	4	235	0.02
0305600 - MSP Wayland	Missing Drivers License Number	4	235	0.02
0305600 - MSP Wayland	Missing Hazard Action	4	235	0.02

Figure 8: Error Types for MSP Wayland

Table 4: Summary of Examples for Accuracy

Metric Example	Descripton	Measures	Business Practices
1. Data Performance Report (by agency)	Generates data errors in ranked list by agency	Errors most common per agency.	Provides agency feedback on performance and any follow up needed
2. Average number of errors per crash	Formula is total errors in database divided by total crashes in database	Average number of errors for each crash report received	Generated both statewide and by agency to provide agency feedback on performance
3. Classification of error type	See figure 7	*Error classification—Severe* (not allow crash to be loaded) versus *informational errors*. (passenger, unit or crash errors). Crash upload to database depends on error classifications.	Provides agency feedback (may include training) on key error types

4.0 Completeness

Incomplete data makes it difficult, misleading or even impossible to conduct factual analyses of crash data. Yet, incomplete crash reports are still prevalent in State crash files.

The six safety data quality characteristics are timeliness, accuracy, completeness, uniformity, integration, and accessibility. The monitoring and measurement of completeness is examined in the following section.

Another aspect of completeness that may not be readily discernable from review of available data are crashes which should be reported to the custodial agency, but are not reported. This situation occurs most frequently when no deaths or injuries are involved in a crash but the damage to vehicles or property exceed the state's property damage reporting threshold. Crash reports are often not submitted on these events. Assessing the extent of these "unreported-but-reportable" crashes is often difficult.

4.1. Purpose Of Measuring Completeness

Measuring completeness helps determine if there is a pattern to the data elements that are most often not reported by investigating officers, and whether law enforcement agencies are investigating and reporting on all crashes that should be reported. While completeness is primarily considered a quality characteristic for crash reporting, it can also pertain to the post submission processing of the reports. For example, if the State custodial office is charged with locating the crash to a base map but is unable to identify the precise location of the crash, the database will not be complete with regard to where crashes are occurring in the state.

When one or more aspects of the crash reporting and processing system are routinely monitored for completeness, the following purposes are served:

- Early detection of changes in completeness.
- Identification of where the problem is occurring so interventions are possible early on.
- Better management of projects aimed at crash data completeness.
- Quantification of effectiveness of completed data improvement projects aimed at crash data completeness.
- Detection of unintended consequences that other crash system changes and improvements may have on completeness (such as changing from one type of database to another or adding an edit program to the processing sequence).
- Avoidance of statistical bias resulting in faulty analysis and presentation, leading to improper actions taken in engineering, enforcement, education or emergency response planning.

Taken in total, these advantages of measurement can help data managers achieve and maintain a level of completeness required for data-driven decision making by all safety disciplines and State safety data users.

4.2. Defining Completeness

A Complete Crash Dataset is one in which:

- All reportable crashes throughout the State are available for analysis.
- All reportable crashes throughout the State are located and available for site-specific analysis.
- All required data elements on individual crash records are completed, as appropriate.

4.3. Measuring Completeness

The completeness of a crash report is a relatively straightforward characteristic of the data to measure. Either the data are there or they are not. Blank fields, where fields should not be blank are easy to detect and count. Computer programs can easily assess the crash data file to ascertain what is not complete. It is equally easy to identify and count crashes that, for whatever reason, cannot be located on a statewide crash location map. On the other hand, it can be challenging to determine if all crashes that meet the State's reporting criteria are actually being reported to the custodial office.

Data managers should identify specific aspects of the crash data in which to measure the completeness of the data. This will typically be based around areas of responsibility for that data, such as:

- Police reporting of crashes and required data element fields
- The custodial office editing and validating process
- The locating process

Various criteria can be used for measurement, such as:

- *Data elements that are most frequently missing*
- *Percentage of report forms containing missing data elements.*
- *Percentage of crashes located to a uniform location system.*
- *Number of missing values per crash.*
- *Percentage of crash reports having more than 5 missing values*

4.4. Developing A Metric For Completeness

General Steps in Creating a Metric

- Decide what aspect of completeness (e.g., most incomplete data element(s), crash location, and BAC level) is to be measured. The State TRCC can help in making a preliminary selection; the agency most affected should develop the final metric.
- Select one or more criteria for the metric, such as which data elements are most often incomplete, *average number of missing values* or *percentage of reports with no missing values*, as discussed above.
- Determine how the metric's data are channeled into its own output report for use by data managers, and by what formula or algorithm is relevant (averaging, percentages, relative vs. absolute measures, etc.).

- Decide on the level of aggregation and timing combination — monitor larger agencies more frequently than smaller agencies where fluctuation would tend to be higher.
- Develop or establish a benchmark for completeness of the crash dataset based on the criteria agreed to above. Along with benchmarking, also establish goals for improving the metrics over time (with intermediate milestones).
- Produce the report and market it to all interested parties. At this point, a means of measurement (metric) has been created, but it is not useful for data improvement until someone with the designated authority looks at the metric and takes action as needed. Metrics must be part of the business practices of the agencies involved in order to have any benefit. Not all data users maybe aware of the ways in which their agency can benefit from the report.

4.5. Assessing Performance For Completeness And Accuracy

The process for assessing crash data elements for completeness is virtually identical to assessing the data for accuracy in terms of internal validity. The assessment for both data quality characteristics entails examining the data in the crash database or on the crash form before entry into the database for validity (accuracy) of the information reported and for the presence (completeness) of data for each element. One computer generated content analysis review of the data for an element can perform reviews for accuracy and completeness simultaneously. In Example 3 in Chapter 3 (Accuracy), the Development of Error Classification Codes identifies errors based on invalid information and on incomplete or missing information (See Figure 7, for sample listings of both types of error). Both data quality characteristics are being measured through the same assessment process. The determination of the accuracy and completeness levels of the crash database are a matter of distinguishing what type of error (accuracy or completeness) is occurring within each data element. Algorithms can be developed to sort the type of errors being identified in the data and develop output reports that differentiate accuracy errors from completeness errors. The findings from this assessment process can be used to produce performance measure reports for accuracy and completeness.

4.6. Examples Of Metrics For Completeness

Three examples of Completeness metrics are: Ratio of Injury Crashes to Total Crashes, Status of Crash Locating by County, and Monthly Progress Report. Table 5, at the end of this chapter, summarizes Completeness examples.

Example 1. Assessing Police Reporting Completeness Using Ratio of Injury Crashes to Total Crashes

Explanation of Metric - How It is Obtained, Calculated, and Reported

As mentioned earlier in this chapter, it is difficult to assess whether all reportable crashes are being submitted to the state crash database. One possible way to measure the completeness of the database for the entry of all reportable crashes entails developing ratios for combinations of fatal, injury, property damage only crashes (PDO) and total crashes.

Figure 9 presents an example of using ratios to assess the completeness of the crash database. All of the columns of information in this table are generated from the crash file itself. The totals are cumulative from the beginning of the calendar year up to the month the report is generated. The columns for fatal crashes, injury crashes, PDOs, and total crashes are listed first. The next three columns consist of three different ratios: fatal crashes to total crashes, injury crashes to total crashes, and fatal-plus-injury crashes to total crashes. While the full report lists all crash reporting agencies, only a sample page from the report is shown. The statewide totals for the first four columns plus statewide averages for the last three columns appear at the end of the report. This line is excerpted and shown inside a red line at the bottom of Figure 9.

2004 Crash Reporting Stats
Status Date 12/09/2005

Agency Name	Fatal Crashes	Injury Crashes	PDO Crashes	Total Crashes	Ratio Fatal/Tot	Ratio Inj/Tot	Ratio Fat-Inj/Tot
Niles Twp Police Dept	1	50	205	256	0.39	19.53	19.92
Coloma Twp Police Dept	2	41	109	152	1.32	26.97	28.29
Chikaming Twp Police Dept	0	22	105	127	0	17.32	17.32
Saint Joseph Twp Police De	0	64	232	296	0	21.62	21.62
Branch Co Sheriff's Office	7	129	809	945	0.74	13.65	14.39
Bronson Police Dept	0	3	29	32	0	9.38	9.38
Coldwater Police Dept	0	92	328	420	0	21.9	21.9
Union City Police Dept	0	2	16	18	0	11.11	11.11
MSP Coldwater	1	78	682	761	0.13	10.25	10.38
MSP Battle Creek	5	191	1221	1417	0.35	13.48	13.83
Calhoun Co Sheriff's Office	1	159	1562	1722	0.06	9.23	9.29
Albion Police Dept	1	18	162	181	0.55	9.94	10.5
Battle Creek Police Dept	5	440	1953	2398	0.21	18.35	18.56
Bedford Twp Police Dept	0	1	7	8	0	12.5	12.5
Emmett Twp DPS	1	75	384	460	0.22	16.3	16.52
Homer Police Dept	0	3	8	11	0	27.27	27.27
Marshall Police Dept	0	39	236	275	0	14.18	14.18
Pennfield Twp Police Dept	0	2	5	7	0	28.57	28.57
Springfield DPS	0	40	131	171	0	23.39	23.39
Cass Co Sheriff's Office	10	219	1157	1386	0.72	15.8	16.52
Cassopolis Police Dept	1	7	28	36	2.78	19.44	22.22
Dowagiac Police Dept	0	30	126	156	0	19.23	19.23
Ontwa Twp - Edwardsburg F	0	45	139	184	0	24.46	24.46
Charlevoix Co Sheriff's Offic	0	145	712	857	0	16.92	16.92
Boyne City Police Dept	0	10	76	86	0	11.63	11.63
Charlevoix Police Dept	0	25	145	170	0	14.71	14.71
East Jordan Police Dept	0	8	42	50	0	16	16
MSP Cheboygan	3	107	423	533	0.56	20.08	20.64
TOTALS	**1082**	**74435**	**298995**	**374512**	**0.29%**	**19.88%**	**20.16%**

Figure 9: Developing Ratios of Fatal, Injury and PDO Crashes to Assess

How the last three columns are calculated: Using the top-listed agency in Figure 9 as an example, (Niles Township Police Department), the ratios in the last three columns are obtained as follows.

- *Ratio Fatal/Tot* is found by dividing the number of Fatal Crashes (1) by Total Crashes (256), equaling 0.0039 or 0.39 percent.
- *Ratio Inj. /Tot* is found by dividing the number of Injury Crashes (50) by Total Crashes (256), equaling 0.1953 or 19.53 percent
- *Ratio Fat-Inj/Tot* is found by adding the number of Fatal Crashes (1) to Injury Crashes (50) and then dividing by Total Crashes (256). The result is 19.92 percent.

Although the report gives three calculated crash ratios for each agency, an inspection shows that the last two ratios are based on larger samples because injury crashes happen more frequently than fatal crashes. Adding the fatal crashes to the injury crashes does not change the ratio very much: the last two column's ratios are almost identical for each agency. Both of the last two columns have the advantage of a larger sample size and either could be used as a metric. It is not necessary to use more than one.

Assume that the second to the last column, *Ratio of injury crashes to total crashes*, is chosen as the metric. Why is this ratio considered indicative of completeness of an agency's crash reporting? Normally the ratio of injury crashes to all crashes is fairly consistent year to year and across agencies. If an agency fails to deliver reports on reportable crashes to the state custodial agency, it is usually the property damage crashes that go unreported. That in turn changes the ratio. Suppose Niles Township Police had reported only 150 property damage crashes of the 205 that occurred, 50 injury crashes divided by (150 + 1 + 50) x 100 = 24.88. This value is larger than the actual value of 19.53, as well as the statewide average of 19.58 for the same ratio. Ratios larger than that expected for an agency, and greater than the statewide average, might indicate that not all property damage crashes are being reported or there is a sudden increase in injury crashes that may warrant immediate attention.

This report showing the three ratio columns can be generated monthly from the custodial agency's statewide crash database or at any other chosen intervals. By nature, there is much randomness in the ratios generated, which makes it only an approximate indicator of potential incomplete reporting by police agencies. Because of sample size, the ratios generated for larger or generally more urban agencies are usually more indicative than for smaller or rural agencies. This metric becomes a more sensitive indicator of potential incomplete reporting as sample size grows. This metric may be used to *flag* agencies for follow up, and usually toward the end of the year, this metric provides a better indicator of a potential problem.

Using this Metric in a Performance Measure

The state will monitor the completeness (reportable crashes that are in fact submitted to the custodial office) of crash reporting from the large urban law enforcement agencies by examining the injury-crash-to-total-crash ratios of these agencies in July and October for high ratios that suggest underreporting of total crashes for the current year. Ratios much higher than the statewide average ratio or that are much larger than the same agency showed in previous year-to-same-date reports, should be flagged. The flagged law enforcement agencies can be contacted so that any potential underreporting or incomplete data that may exist can be reduced before the close of the data year.

Example 2. Assessing Completeness Using Status of Crash Locating by County

Explanation of Metric - How It is Obtained, Calculated, and Reported

One of the biggest problems confronting many states is their inability to identify the precise location where a crash occurs. In Michigan, 80 percent of crash locations are determined automatically based on the literal description provided by the reporting officer. Other crashes are located using Microsoft Maps, Google Maps, or other map tools: "Smart maps" are used to reconcile otherwise un-locatable crashes. As a result of these various methods, the location data provided by the reporting officer is successfully interpreted for the great majority of crashes, to place of the crash event at a specific point on the road system (i.e. "locating" the crash). When crashes in the database are "located" it means they can be queried "by location" using the State's crash analysis software.

A summary page from Michigan Crash Locating report is shown in Figure 10. This figure shows the completion rate for locating crashes for each county. In the case of Oakland County, the percent of located crashes (99.77) is derived simply by dividing the number of located crashes (15,603) by total crashes (15,639). This ratio is important to that county when compared with the equivalent statewide ratio of 99.41 percent and serves as an excellent metric for the completeness of "crash locating."

Status of Crash Locating - 2006

County	Total Crashes	Total Located Crashes	Percent Located	Unlocated Crashe
Oakland County	15,639	15,603	99.77%	36
Oceana County	436	434	99.54%	2
Ogemaw County	305	304	99.67%	1
Ontonagon County	124	119	95.97%	5
Osceola County	421	420	99.76%	1
Oscoda County	126	124	98.41%	2
Otsego County	389	385	98.97%	4
Ottawa County	2,658	2,652	99.77%	6
Presque Isle County	208	208	100.00%	0
Roscommon County	326	324	99.39%	2
Saginaw County	2,168	2,149	99.12%	19
St. Clair County	1,640	1,636	99.76%	4
St. Joseph County	680	678	99.71%	2
Sanilac County	546	546	100.00%	0
Schoolcraft County	166	162	97.59%	4
Shiawassee County	971	970	99.90%	1
Tuscola County	683	683	100.00%	0
Van Buren County	906	906	100.00%	0
Washtenaw County	3,936	3,926	99.75%	10
Wayne County	22,377	22,002	98.32%	375
Wexford County	503	497	98.81%	6
Total	**121,161**	**120,441**	**99.41%**	**720**

Status As Of 06/19/2006
Excludes Non Traffic Crashes

Figure 10: Sample Output of the Status of Crash Locating Metric

Inability to assign a location may be due to inadequate literal description, or improper conversion of valid literal description by the automated system. The metrics shown (three different ones, all measuring locating completeness) could be improved by replacing the county column with a local enforcement agency column. This would take the metric to a lower level of measurement. For example, in a county with many law enforcement agencies investigating crashes, if one or two agencies produce the majority of crashes that cannot be located, that will bring the metric of "percent located" down for the entire county, without identifying the particular agencies that may be less precise. Therefore, it is still a very helpful tool for monitoring location completeness in the general sense. Michigan further intends to investigate the accuracy of successfully located crashes to ensure they are correct as well as complete.

Example 3. Assessing Completeness Using a Monthly Progress Report

Explanation of Metric—How it is Obtained, Calculated, and Reported

This report is not a single metric but rather an ongoing weekly summary report that lists numerous statistics like number of fatal crashes, injury, and property damage only crashes, total crashes, number of crashes located, etc. as shown in Figure 11. They are not compared to previous years or converted to percentages. Their usefulness is as a way to rapidly screen different statistics by data managers. It requires interpretation by someone experienced in what to look for. This summary report is very useful in many ways, in that the data manager can review:

- The number of total crashes and compare with previous year
- Time of year when crashes are concentrated and where
- The percent of crashes located by month and compare with previous year average or goal set by department

Statistic Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Fatal Crashes	70	61	60	56	39	1	0	0	0	0	0	0	287
Injury Crashes	5,050	4,362	4,165	4,237	4,509	816	0	0	0	0	0	0	23,139
Property Damage	22,602	20,354	17,637	16,436	17,282	3,411	0	0	0	0	0	0	97,722
Traffic Crashes	27,722	24,777	21,862	20,729	21,830	4,228	0	0	0	0	0	0	121,148
Unresolved Traffic	0	0	0	0	0	1	0	0	0	0	0	0	1
Non Traffic	130	203	97	92	80	22	0	0	0	0	0	0	624
Total Crashes	27,852	24,980	21,959	20,821	21,910	4,251	0	0	0	0	0	0	121,773
Detroit Crashes	1,676	1,728	1,846	1,790	1,849	228	0	0	0	0	0	0	9,117
Crashes Located	27,661	24,719	21,806	20,664	21,310	3,847	0	0	0	0	0	0	120,007
% Located	99.8	99.8	99.7	99.7	97.6	91.0	0.0	0.0	0.0	0.0	0.0	0.0	99.1
Not Located (2)	5	10	15	20	305	269	0	0	0	0	0	0	624
Could not Locate (4)	53	48	38	38	35	6	0	0	0	0	0	0	218
Manually Updated (6)	0	0	0	0	0	0	0	0	0	0	0	0	0
Missing Data (9)	3	0	3	7	180	107	0	0	0	0	0	0	300
MDOS Drivers	45,312	41,483	37,259	35,894	37,790	7,085	0	0	0	0	0	0	204,823
MDOS Pending	86	56	53	109	1,672	2,596	0	0	0	0	0	0	4,572
MDOS Updated	39,357	35,888	32,264	30,792	30,974	3,502	0	0	0	0	0	0	172,777
MDOS Rejected	0	0	5	4	134	149	0	0	0	0	0	0	292
MDOS Drivers Bypass	5,740	5,423	4,797	4,778	4,885	834	0	0	0	0	0	0	26,457
Trunkline Crashes	10,952	9,509	8,337	8,078	8,789	1,610	0	0	0	0	0	0	47,275

Figure 11: Sample Output of the Monthly Progress Report Metric

Using the Metric in a Performance Measure

Below is a hypothetical sample performance measure using the metric *incomplete data elements per crash report*:

Sample Performance Measure

The state will improve the **completeness** of the **crash data system** by reducing average number of missing data elements per report from five in 2007 to one by 2010. .

The baseline average, as calculated using the 2005 and 2006 annual crash databases, is five incomplete data elements per crash report. The goal for reducing these incomplete data elements in subsequent years are:

- Four incomplete data elements per case in 2007
- Three incomplete data elements per case in 2008
- Two incomplete data elements per case in 2009
- One incomplete data elements per case in 2010

Table 5 provides a snapshot of other metrics that can be used to determine the completeness of the crash database. The content is useful to the Crash Unit manager in gauging overall dataset performance, and state assessment.

Table 5: Summary of Examples for Completeness

Metric Example	Description	Measure	Business Practices
1. Ratio of Injury crashes to total crashes.	All injury crashes divided by total crashes reported by agency	Ratios higher than two standard deviations above the statewide average ratio should be flagged	Use to provide feedback mainly to the larger local enforcement agencies regarding potential under-reporting
2. Status of Crash Locating by County	Total located crashes divided by all reported crashes by county	Table showing county and statewide, total crashes that can be located to public roads within the state	Use to provide internal feedback to managers and potential corrective action/s for agency location reporting
3. CRASH STATISTICS monthly progress report.	This report assembles tallies for 20 crash categories/crash-processing milestones	Various statistics updated weekly and also serves as a good comparison to previous years data	Provide an overview of the entire crash-reporting process to management to gauge performance – present and past and potential corrective action/s

5.0 Consistency

Data consistency is essential for making sense of aggregated data. The information collected about each crash and the reporting thresholds for crashes should be identical from one jurisdiction (within a state) to the next jurisdiction. The guidelines and standards for reporting crashes must be consistent among all reporting agencies. Without this it would not be possible to effectively compare results across jurisdictions and time regarding the application of mitigating treatments in engineering, enforcement, education, and emergency response planning. Without consistent information safety efforts could be misguided, wasteful or ineffective.

The six safety data quality characteristics are timeliness, accuracy, completeness, uniformity, integration, and accessibility. The monitoring and measurement of consistency is examined in the following section.

5.1. Purpose Of Measuring Consistency

To promote consistency across jurisdictions, the state needs a single crash report form. Variations created by individual jurisdictions and municipalities can create significant problems in data consistency as well as data uniformity. If an individual jurisdiction wishes to collect additional information on crashes for its own purposes, this information should be maintained in a local database or maintained at the state level in such a way as to not confound the uniformity of the state crash database. Crash reports should be completed in a consistent manner across reporting jurisdictions. Since many law enforcement agencies are involved in any state crash reporting system, achieving consistency in how they interpret and complete the standard crash form is often a significant challenge. Only through training and close monitoring of the data can this challenge be overcome.

After reports arrive at the custodial agency, consistency in processing is another area of concern. However, far fewer individuals are involved at this level and are likely to be in a single work unit, making consistency in their work tasks easier to attain.

The federal government and many state governments have sought a greater degree of consistency from crash data to help establish national policies regarding highway safety and to do state-to-state comparisons of traffic safety issues. The development of the Minimum Model Uniform Crash Criteria (MMUCC) initiative was a joint effort of federal, state and local agencies as well as university researchers and private sector consultants. The MMUCC effort established a set of uniform crash data elements, definitions and attributes that every state should be collecting. The third edition of the MMUCC has 107 data elements (including derived and linked elements) that are a *guideline* to states regarding crash data elements and their attributes on which they should be collecting crash information.

The Minimum Model Uniform Crash Criteria (MMUCC) initiative was developed to help States adopt uniform definitions for data elements and to collect a minimum set of data elements from every State.

5.2. Defining Consistency

A consistent crash data file is one in which all jurisdictions are using standard reporting form(s) (i.e. the same data elements and attributes), interpret and report the data elements uniformly and use the same reporting thresholds. Further, any procedures in processing the dataset such as, locating crashes are handled identically at all times. To ensure "consistency" of data the following guidelines and standards exist:

- Minimum Model Uniform Crash Criteria (MMUCC), 3rd Edition, 2008.
- Manual on Classification of Motor Vehicle Traffic Accidents, 6th Edition, ANSI D16.1-1996.
- Data Element Dictionary for Traffic Records Systems, ANSI D20.1, 2003.

5.3. Measuring Consistency

Measuring consistency usually involves comparing the uniformity across time and across agencies in the completion of the crash form. It can also assess the performance of an element of the crash-data system with some external standard or guideline (e.g. MMUCC). How does the State compare to the standard, and is there positive change (i.e. greater compliance with standards/guidelines) over time in response to data-improvement efforts?

Some of the most important requirements for consistency or uniformity are:

- *All reporting agencies using a standard crash report form*
- *Completing the crash form in a consistent manner by all law enforcement reporting agencies*
- *The crash report form is in maximum practicable conformance with MMUCC data elements.*

Assuming a standard statewide crash form exists in state law or administrative rule, one could count the number of police agencies that fail to use it for crash reporting. The metric would be number or percent of agencies using (or not using) the standard crash form.

Some States report on a single crash report form, but there still maybe concerns about consistency in crash reporting, such as investigating officers understanding of the terminology, definitions and their interpretation of the crash report data elements. This aspect of consistency is closely associated with measurements of data accuracy and completeness. Where accuracy and completeness errors are being identified in the crash data, these errors may be due (at least in part) to the type and extent of training the officer received in conducting crash investigations and reporting their findings on the crash form. One way to mitigate some types of consistency errors is through the use of software to collect crash data. Electronic crash collection, with pull-down menus of attributes, a GPS locator function, business edits and validations limit the selections an officer can make on the crash report form thereby assuring greater consistency with standard report formats, as well as increasing the accuracy and completeness of the report.

For MMUCC compliance, the number of MMUCC elements/attributes included (or not included) in the statewide crash report/database could be counted. The criterion of MMUCC compliance would be determined by how many of the MMUCC elements on which the state collects information and maintains this data in its crash database. The NHTSA has an assessment of how compliant each state is with the MMUCC guidelines.

5.4. Developing A Metric For Consistency

General Steps in Creating a Metric

- Decide what aspect or component process of consistency is to be measured. The State TRCC can help in making a preliminary selection; the agency most affected should develop the final metric.
- Select a criterion for the metric, such as *number of law enforcement agencies not submitting crash reports using the standard state report form,* or *number of data elements/attributes missing from crash report per MMUCC guidelines*.
- Determine how the metric's data will be channeled into its own output report for use by data managers, and by what formula or algorithm.
- Develop or establish a benchmark for consistency of the crash data based on the criteria agreed to above. Along with benchmarking, also establish goals for improving the metrics over time (with intermediate milestones).
- Produce the report and market it to all interested parties. At this point, a means of measurement (metric) has been created, but it is not useful for data improvement until someone with the designated authority looks at the metric and takes action as needed. Metrics must be part of the business practices of the agencies involved in order to have any benefit. Not all data users maybe aware of the ways in which their agency can benefit from the report.

5.5. Assessing Performance For Consistency

Assessing performance for data consistency in some ways is rather simple, as shown below:

- Determine if all law enforcement agencies are using a standard crash report form
- Assess increases/decreases in MMUCC compliance
- Identify types of training law enforcement officers and law enforcement students receive in investigating crashes and completing the state's crash report form. If several agencies provide this training an evaluation should be conducted to assess proficiency levels.

Generally, all reporting jurisdictions should use the same reporting threshold and the same set of core data elements. Should it become necessary to change or modify a data element, **changes to data elements should be fully documented**, including the date the change became effective. For example, data values expanded to provide greater detail on truck involvement in crashes (e.g., trucks involved in crashes were previously coded as light or heavy; the new values are changed to *under 10,000 pounds, 10,001 to 20,000 pounds, greater than 20,000 pounds*). Manual crash reporting makes this task laborious and time consuming. However, with electronic reporting, it is simply uploading the new files and alerting the respective officers. The training in the latter case is minimal compared to the requirements for manual data entry.

5.6. Examples Of Metrics For Consistency

Two examples are presented for metrics involving consistency: *assessing statewide usage of a standard crash form* and a*ssessing MMUCC compliance*. These examples are for a hypothetical State that does not yet have good compliance with its crash report form or with MMUCC.

Example 1. Assessing Consistency of Use of the Statewide Standard Crash Report Form

Paper Report Forms

In assessing the consistent usage of a standard crash report form, one would only need to identify and count the number of law enforcement agencies not using the standard reporting form. Once identified, these agencies should be contacted about changing to the state standard report form.

Electronic Report Forms

Assessing consistency in electronic software is not quite as simple as identifying and counting the number of agencies not using a standard report form. However, it is generally not a difficult process. In order to assess consistency between different electronic data collection software packages, it is necessary to verify that each software package collects data on the required data elements, uses required attributes for those data elements and is uploaded to the State crash database. Additionally, to the extent that the state uses business edits and validations for the crash database, these same edits and validations should be incorporated into the electronic collection software being used by individual agencies, thus assuring consistency in reporting.

Using this Metric in a Performance Measure

Sample Performance Measure

The State will improve the consistency of crash data by reducing noncompliant local enforcement agencies not using the Statewide Standard Crash Report from an average of 12 (30 percent of total crashes) in 2006 to:

- No more than eight (24 percent) in 2007.
- No more than four (15 percent) in 2008.
- No more than one (< 2 percent) in 2009 and thereafter.

Example 2. Assessing Consistency with MMUCC

The second example of a consistency check is to assess the state's level of compliance with the MMUCC. This is one of the processes that can assist a state in qualifying for a subsequent year Section 408, State Traffic Safety Information System Improvement Grant. In this process a state reviews the data elements and attributes they maintain within their crash database relative to the elements and attributes specified in the MMUCC Guideline. For each data element (and its attributes) that the state crash database contains that duplicates the MMUCC element, the state would be deemed to be consistent with the MMUCC. Where the MMUCC contains elements not collected or maintained in the state's crash database, the state would not be compliant (consistent) with the MMUCC, even if the element is collected on the crash report form.

NHTSA has developed an assessment of each states level of MMUCC compliance. This assessment is used in determining whether or not a state has improved its MMUCC compliance for purposes of awarding Section 408 grants. Each state is encouraged to review NHTSA's assessment of their MMUCC compliance to assure that the information within this rating criterion is current and correct. If a state feels it is not being given credit for a particular element, there are procedures to resolve the disputed element.

Example 3. Assessing Consistency with Measures of Completeness

The information presented earlier in Figure 9 and Figure 11 for completeness can also be used to assess consistency. The ratios developed in Figure 9, particularly the ratio of fatal and injury crashes to total crashes because of the larger number of crashes used in this measure, can help assess the consistency of crash reporting within the state and by agencies. By examining the ratio values for the current year relative to the ratio value for the previous year, for a comparable time period, the consistency of reporting can be assessed. This same technique of comparing the current year's reporting of various types of crashes and other information presented in Figure 11, can be used to assess consistency by agency or for the state when compared to a prior comparable time period. When the current year's values are substantially below the prior year's values, this may indicate actual reductions in different types of crashes or it may indicate problems with agency reporting.

Using this Metric in a Performance Measure

Sample Performance Measure

The state will improve the consistency of the crash data to full MMUCC compliance (baseline for 2005-2006) in the next report form revision scheduled for 2009.

The goals might then look like this:

- No more than ten missing elements in 2007
- No more than ten missing elements in 2008
- No more than two missing elements in 2009

Table 6: Summary of Examples for Consistency

Metric Example	Description	Measure	Business Practice
1. Statewide Crash Form	Crashes reported using statewide crash form divided by all crashes reported on any other form(s).	In Sample State, an estimated 95 percent of reportable crashes are actually reported using the uniform report (base year 2006)	Sample State Motor Vehicle Code requires a statewide crash form, and specifies that the State will develop the crash form (uniform) and that all agencies will send this to the crash repository. The goal is to approach 100 percent (in theory) by 2010.
2. MMUCC Compliance	Assessment of State crash form against the MMUCC criteria/elements	Review found 70 of the 77 elements matched (base year, 2006)	To be fully compliant (as soon as practical) by 2010

6.0 Integration

The ability to integrate diverse databases has long been a goal of highway safety analysts. By combining databases, safety researchers create a richer source of data and are better able to understand the factors that may affect the occurrence and severity of crashes. Many States are currently combining or linking their crash database with other traffic safety databases such as their driver licensing database, vehicle registration database, and injury database.

> *The six safety data quality characteristics are timeliness, accuracy, completeness, uniformity, integration, and accessibility. The monitoring and measurement of integration is examined in the following section.*

6.1. Purpose Of Measuring Integration

As with other quality characteristics of safety data, assessing integrated databases can provide the following benefits:

- 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.

6.2. Defining Integration

Data integration combines multiple sources and types of data to create a broader array of information to be analyzed for better safety decisions. Data integration as applied to crash data means the data elements in the crash database are linked with data elements in other safety databases through the use of common data elements or "linking variables" to create a database that is a combination of all the data elements in two (or more) databases. Examples of linking variables may include elements such as a unique dispatch number, driver's name, date of birth, driver's license number, specific location and date/time of incident, etc. The more common elements databases share increases the reliability of the linked data bases.

Where common data elements are not available, another means of file linkage is based on using probabilistic linkage methods. This methodology examines two or more databases for approximate or similar data elements and similar responses to the data elements. For example if a law enforcement officer investigates a crash that occurs at First and Main Street on January 1 at 11:00 AM and an ambulance run report was submitted to the EMS database for a motor vehicle crash at 10:51 AM on the same date at the same location, there is a reasonable probability that the crash report and the EMS run report are being generated in response to the same event. If these two reports are judged to be in response to the same event, the reports can be linked based on the likelihood that they are in response to the same event..

This is the method commonly used by states to generate Crash Outcome Data Evaluation System (CODES) databases. They link crash data with available injury data (e.g. EMS run reports, emergency department admissions, and/or hospital discharge reports). Approximately 30 states have established a CODES linked database.

Another method of integration is to use physical location of the crash as the common linking element. Global Positioning Systems (GPS) coordinates provide one means of locating different crash related elements such as the crash itself with the roadway inventory characteristics at the spatial location of the crash (e.g. rural 2 lane asphalt road, 2 foot shoulders, no guardrail etc.). GIS software can be used for the manipulation of all database information contained within it and linked to a specific location. Crash and roadway information are more commonly linked in GIS, but citations and EMS run reports can also be identified by location and integrated for analysis.

There are many challenges to crash data integration, including:

- Inconsistent/incomplete data.
- Incomplete or nonexistent metadata (metadata is information about the database).
- Insufficient training and/or lack of skilled technicians/programmers to manipulate the integrated databases.
- Unclear partnership agreements that fail to define responsibilities, data transfer, quality checking, metadata, and other data stewardship issue.
- Institutional and funding issues.

6.3. Measuring Integration

There are several possible ways to approach measuring integration. At the *macro* level, data managers may simply count the number of independent databases (roadway, driver, vehicle, injury etc.) to which the crash database is linked. At the *micro* level, as with probabilistic linkages, what can be measured is the **degree of linkage** achieved. If each specific record on one file is linkable to a specific record on another file, then the degree of linkage is 100 percent. Often, file linkages do not accomplish this, and then it makes sense to monitor the degree of linkage as part of the improvement effort.

> *As with other data quality characteristics, various criteria can be used to measure integration, such as number or percentage of crashes linked to another driver/vehicle data file or the percent of EMS run reports that can be matched to hospital discharge records.*

Setting up useful metrics for data integration should be tied to the State's goals and needs but at a minimum the state should identify and have a count of the other databases to which the crash database is linked.

6.4. Developing A Metric For Integration

General Steps in Creating a Metric

- Decide what aspect or component process of integration is to be measured at the macro or micro level. The State TRCC can help in making a preliminary selection; with the agency most affected developing the final metric.
- Select a criterion for the metric, such as *ability to link crash databases with injury outcome data, driver and vehicle information.*
- If applicable, determine how the metric's data are channeled into its own output report for use by data managers, and by what formula or algorithm.
- Develop or establish a benchmark for integration of the crash data based on the criteria agreed to above. In addition to benchmarking, also establish goals for improving the metrics over time (with intermediate milestones).
- Produce the report. At this point, a means of measurement (metric) has been created, but it is not useful for data improvement until someone with the designated authority looks at the metric and takes action, as needed. Metrics must be part of the business practices of the agencies involved in order to have any benefit.

6.5. Assessing Performance For Integration

As with other metrics, performance of the integration metrics must be monitored over time. Typically, the simple assessment of integration involves identifying and counting the number of other databases which the crash file is linked. This simple count should identify which databases are linked to the crash file. The answer to this question basically addresses the question of to what extent the crash database is integrated. There are additional questions that a state may wish to consider. These questions, including:

- What linkages are most important within the safety context?
- Which data files are or should be linked?
- By what mechanism are they linked?
- How complete are the linkages?
- What kinds of analyses is possible using the integrated data?
- Who has the access and ability to use the integrated file?
- What safeguards are in place to prevent unauthorized access?
- What report outputs exist for the integrated file that were not possible previously? (Have the integrated files been used to solve highway safety problems?)

6.6. Examples of Metrics for Integration

Three examples of Data Integration are used as settings for looking at how metrics can fit into these systems: Iowa's GIS for Highway Safety, Michigan's linkages to crash file, and Iowa CODES.

The Guide does not attempt to describe each example of integration exhaustively, but rather to show where some opportunities for the use of metrics may exist. While the integrated systems in the examples are real, the metrics suggested are hypothetical. States with this level of integration may already have the information in their databases necessary for creating the metrics as discussed in this Guide.

Example 1. Assessing the Integration Capabilities of Iowa's GIS for Highway Safety

A pivotal development in Iowa traffic records improvements was the "Smart Map" Location Tool. This tool enabled law enforcement officers in the field to capture the crash location, as well as locating the crashes at the State custodial agency level. The coverage of the Location Tool is statewide, and it has the capability to capture coordinates for crashes, citations, and other spatial events. Thus, all data being captured electronically by TraCS's agencies could be immediately located for local agency use and state use. An analysis tool, the Incident Mapping Analysis Tool (IMAT), was created to make GIS data within the TraCS local dataset easy for law enforcement to use.

Thus, all data being captured electronically by TraCS agencies could be immediately located for local agency use.

By locating all crashes statewide with the tool, the Office of Driver Services (the Iowa custodial office for crash data) insured that all crashes reported to the State, whether electronically or via paper, became accessible within the Statewide Highway Safety GIS. The Iowa DOT Office of Traffic and Safety created two analysis packages to provide all crash data users with location-specific query capability of the statewide crash file. The Safety Data Visualization and Exploration Resource (SAVER) was created for users needing the most rigor, while the Crash Mapping Tool (CMaT) was created for users needing a versatile data "look-up" tool that could be mastered quickly. Crash data from the Statewide Crash File has the advantage of combining all crash reports from a geographical area, regardless of which police agency submitted the report, and includes driver reports when appropriate.

The Location Tool insures linkage of the crash file with the roadway file and provides a means of spatial linking of other files to the crash/roadway GIS as well. These additional opportunities are only partially exploited at the present time, and therefore the State is pursuing broader local implementation of this technology.

Business practices in Iowa have led to the creation of a large variety of output reports, both standard and user-specified, that have utilized the integrated data and provided safety managers with useful information for decisions and problem solving.

Example 2. Assessing the Integration Capabilities of Michigan's File Linkages.

It is typical for States to be able to link Crash and Roadway files in some manner. By having the capability to also link Driver and Vehicle files to the Crash and Roadway files, Michigan is well positioned to develop an even broader system of linked files. Michigan is currently working to improve and expand their linkage to other traffic record system components, including the injury component. When this comes onboard, additional capability will be created to validate the reported injury on the crash forms, as well as on the EMS run files.

The business practices used by Michigan relating to integration include:

1. **Safety databases linked—fostering data integration—**Michigan formed the Crash Data Users Group (CDUG), a subgroup of Traffic Records Coordinating Committee.
2. **Plans to link crash data with additional datasets—**There is a plan to have a database of citations (Judicial Data Warehouse) and sentencing (such as for tying DUI back to crashes). The "cradle to grave" citation project intends to collect all electronic citations.
3. **Processes for crash/roadway linkage** – Most of Michigan's 83 counties use the Michigan Tech Transportation Institute software RoadSoft. There is a vision in the strategic plan for a roadway inventory project that would bring representatives together to identify standards, like Model Inventory of Roadway Elements (MIRE), that might be brought into RoadSoft or use beyond the local level.

Example 3. Use of the Crash Outcome Evaluation System (CODES) by Iowa

CODES is a collaborative approach to obtain medical and financial outcome information related to motor vehicle crashes for highway safety and injury control decision making.

How CODES Works:

Crash data are collected at the crash scene and then linked to injury outcome data also collected at the scene, en route to the emergency department, at the hospital or trauma center, and after discharge.

The type of injuries, their severity, and the costs incurred by persons injured in motor vehicle crashes are described and computerized.

The linked crash outcome data are also linked to other traffic records such as vehicle registration, driver licensing, citation and roadway inventory data in order to generate more comprehensive information to evaluate highway safety.

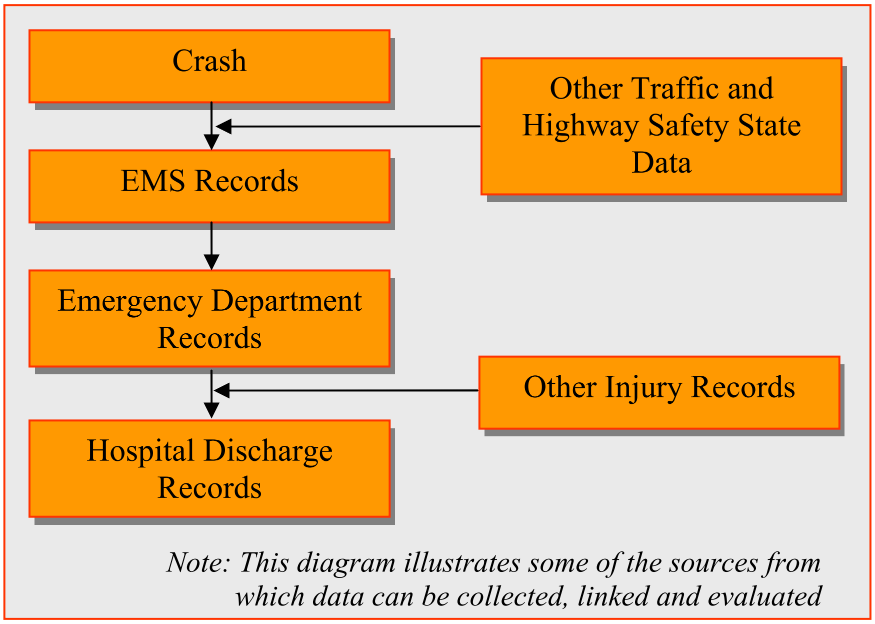

Figure 12: Iowa CODES Process

Iowa CODES

Iowa combined highway crash and medical records to evaluate National and State medical cost-component estimates. The use of CODES data resulted in estimates of crash cost which were significantly different from those used by FHWA and the Iowa DOT. It also demonstrated that the identification and ranking process currently used by the Iowa DOT is relatively insensitive to medical costs. That is, even though medical cost represent a significant and previously difficult-to-quantify portion of total crash cost, changing them results in only minor differences in location rank. Although mitigation prioritization is not found to receive a great benefit from the use of CODES, the system does show promise for other applications, e.g., determining appropriate societal crash costs resulting from various safety improvements.

This validated the existing business practice in use by the Iowa DOT in terms of identification and ranking of high crash locations.

Using this Metric in a Performance Measure

Sample Performance Measure

At the macro level, the state will improve the integration of their crash data by linking the crash database with other traffic safety databases. These linkages will be established over the next couple of years. The following presents a timeline for crash database integration with other traffic safety databases. Each file linkage provides a value of one to the state's integration performance measure

The goals might then look like this:

- In 2007 integrate the crash database with driver license database (value 1)
- In 2008 integrate the crash database with vehicle registration database (value 1)
- In 2009 integrate the crash database with roadway inventory database (value 1).

Table 7: Summary of Examples of Integration

Example	Description	Measure	Business Practice
1. Iowa's Highway Safety GIS	Crash and roadway data currently in the GIS statewide. Has the capability to include all TraCS databases with a location component.	The number of TraCS agencies locating additional safety events (besides crashes, such as citations) using the Location Tool.	The Location Tool used for crash locating is also available for locating all other TraCS datasets into the Highway Safety GIS.
2. Michigan's linked Safety databases	Analysis capability with linked files. Files currently linked to the crash file: Driver, vehicle, and roadway environment files. Linkages are achieved by DL number, plate or VIN, linear referencing, or coordinate systems, respectively.	The number of EMS services files added.	The baseline for 2006 is zero local EMS services files added, with an addition of 100 services per year to the integrated database starting in 2008.

Example	Description	Measure	Business Practice
3. Iowa CODES	Combine highway crash and medical records to evaluate cost-component estimates.	Comparison with National and State cost estimates.	Present ranking process for high crash locations is not significant affected by medical cost (existing business practiced validated).

7.0 Accessibility

Accessibility is the final characteristic of crash data quality. Unless users have access to the data, the benefits of all the work and expense to collect the data and maintain the crash system are of little use. While this would seem logical enough, often the planning and funding given to this "back end" of the crash data system is an afterthought when compared to the front end of data capture and storage.

The six safety data quality characteristics are timeliness, accuracy, completeness, uniformity, integration, and accessibility. The monitoring and measurement of accessibility is examined in the following section.

7.1. Purpose Of Measuring Accessibility

Measuring the extent of data accessibility and progress in increasing data accessibility serves the following purposes:

- Identifies ways that accessibility can be improved.
- Promotes marketing of data to additional users.
- Supports data driven decisions (political, institutional, and technical).
- Garners additional resources for further improvement and maintenance of crash data.
- Encourages better management of projects that increase data accessibility.
- Evaluates effectiveness of completed data accessibility projects.

Examples of accessibility include Kentucky's traffic records system which features a web portal with GIS functionality. The GIS allows web users to connect directly to the enterprise database system to query locations for analyses, including high-crash locations and alcohol-related incidents. The system allows the retrieval of individual incident reports or customized summary reports for data elements. Extracting data is a core function that allows users to bring raw data into their own systems for further analysis. A user can also use the online statistical analysis package to analyze data directly from the open portal.

Kentucky's system features a Web portal with GIS functionality. GIS allows Web users to connect directly to the enterprise database system to query locations for analyses, including high-accident locations and alcohol-related incidents.

The Iowa Department of Transportation has created a custom desktop interface to link users to its wide array of system databases. The interface contains limited GIS mapping of incidents and a data-export module to help users bring incident reports into their own applications for further analysis. Iowa's system comes with advanced statistical analysis and charting capabilities as well as a portal to Intersection Magic for diagramming of incidents at specific intersections.

Iowa's system comes with advanced statistical analysis and charting capabilities as well as a portal to Intersection Magic for diagramming of incidents at specific intersections.

7.2. Defining Accessibility

Accessible crash data is data that is readily obtainable by the eligible users for both direct (automated) access and periodic outputs (standard reports) from the system. Compact disks, the internet, and special data websites have opened up new avenues of accessibility to an ever increasing range of crash data customers. However, on-line access is not the only way to achieve electronic access for hands-on querying or to distribute data and reports in electronic formats. Other strategies include enabling local agencies to create near-real-time crash files of their own, making data and query tools available on compact disks, and using e-mail both to submit data requests and receive results.

7.3. Measuring Accessibility

States have the responsibility to make crash data available and easily accessible to eligible users. State laws may dictate what data can be made available to the public and what can be released to professional analysts and to local jurisdictions only. The restrictions placed on the availability of data are usually related to personal identification type information.

Developing a quantified measure(s) of accessibility can be difficult. There are several potential metrics that may provide measures of accessibility. One measure of accessibility may be derived from when the crash data is made available to users in an electronic database. An example would be the date when the state updates its crash database with the most current year of crash data. This measure identifies when the crash data is available (accessible) however, it does not provide a measure on the use of the data. Another measure of accessibility may be obtained by counting the eligible users of the crash data, especially if a state limits or restricts access to the crash database to “registered or subscribed” users of the crash data. This measure would provide an indication on the potential accessibility of the data, but would not provide an actual indication of how often the data is being accessed. To determine the actual usage of the crash data, it would be necessary to monitor and count the number of times the data is actually requested and if the requested information is successfully provided. By counting ‘hits’ on a state’s crash data website or counting the number of requests for information that the state receives and the number of requests that are successfully fulfilled, the state may obtain a measure of the accessibility of it’s crash data.

As a practical matter, the degree to which users access the data may be the easiest to measure. Marketing the availability of the data to key users is critical to measuring accessibility. Examples of possible metrics are:

- *A count of the number of users making data requests to a centralized resource*
- *The percentage of all counties that request crash data at least yearly*
- *The number of data users able to perform their own queries directly from the data*
- *A count of the number of weekly hits on a website*
- *A count of the number and distribution of standardized reports distributed*

As with all the other quality characteristics, the ways states measure accessibility should relate to the state’s own needs and goals.

7.4. Developing a Metric for Accessibility

General Steps in Creating a Metric

- Decide which aspect of accessibility is to be measured. The State TRCC can help in making a preliminary selection; with the agency most affected developing the final metric.
- Select a criterion for the metric, such as "number of data requests filled monthly," "percentage of local agencies that utilize State-provided data analysis tools," or "number of hits to the crash facts web pages per week."
- Determine who is going to collect the criterion data and how it is channeled into a report for data managers.
- Determine the baseline for the criterion along with goals for improvement. The baseline may be zero when a new program is being set up, such as with a new website. If a website is being revised to be more useful, then the baseline might be average number of weekly hits for the year prior to the revision.
- Produce the output report and market it to all interested parties. At this point, a means of measurement (metric) has been created, but it is not useful for data improvement until someone with the designated authority looks at the metric and takes action as needed. Metrics must be part of the business practices of the agencies involved in order to have any benefit. Not all data users may be aware of the ways in which their agency can benefit from the report.

7.5. Assessing Performance for Accessibility

There are numerous aspects to accessibility. The more ways a State provides accessibility, the more options they have for assessing performance such as:

- Accessibility of all crash databases a State has: If there are local or regional crash databases in addition to the statewide file, performance may be determined for each one of them. Maintaining multiple crash files does not necessarily mean duplication of effort, as they may be a result of the practice "capture and enter once, create datasets simultaneously" such as in States using TraCS.
- The timeliness aspect of accessibility: If access to data is only via one path of limited capacity, a bottleneck may exist that is a source of delay to users. This is an aspect of performance important in some States.
- Accessibility over time as progress is made in completing the file: User accessibility may differ for the crash file "in progress" and the crash file when released for statistical work. One or the other may be easy to access but both are not necessarily equally accessible to all users.
- User accessibility may vary by safety disciplines within the State: Access by law enforcement agencies may differ from access possible for engineering office's, and different yet for injury prevention purposes.
- Accessibility to integrated files: Accessibility may differ for the crash/roadway file and the crash/medical outcome file. For example, if accessibility differs between various integrated files, the state may wish to conduct a separate assessment of accessibility for the different integrated databases.

- Accessibility through standardized reports versus hands-on querying and analysis. Again, some States have developed one or the other approach while other States have attained some access by each method.
- Accessibility in relation to privacy laws: Privacy laws in a given State may govern specific information from crash reports, dictating what portions may be accessed and by whom.

Over 120 Kentucky agencies within the state and FHWA have direct access to these tools, including 90 predefined management and statistical reports.

Kentucky is a leader in data and tool accessibility. Its web-based system allows users 24/7 access to its information resources. The Web solution (Figure 13) enables GIS mapping, summarized database query and export, and individual incident-report lookup. Over 120 agencies within the State have direct access to these tools, including 90 predefined management and statistical reports.

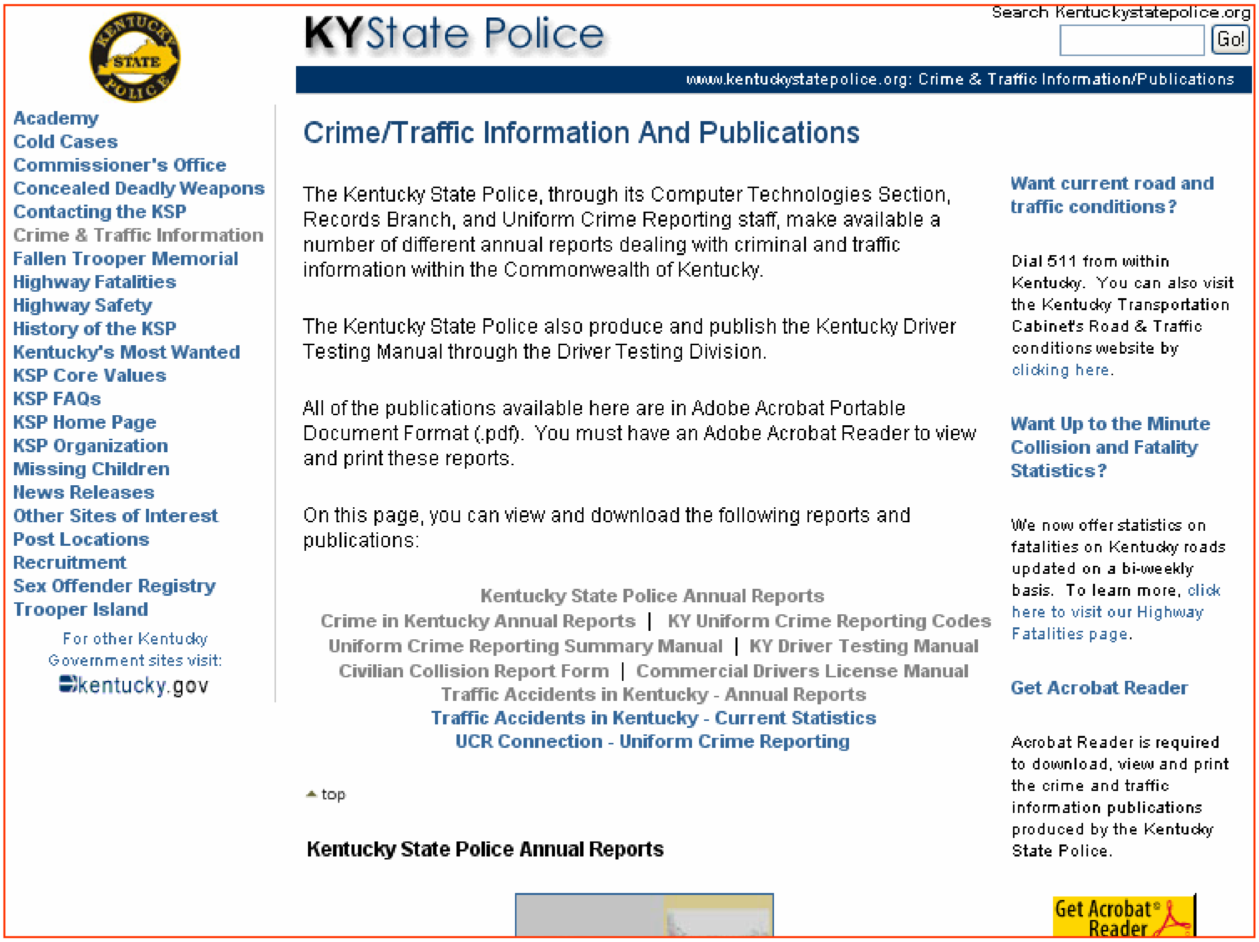

Figure 13: Kentucky State Police Web Site

Michigan also has extensive tools for data accessibility; including public access as well as password protected access for local enforcement agencies, and approved safety personnel. Figure 14 illustrates the Michigan web site.

Figure 14: MSP Traffic Crash Reporting Web Site

Iowa also has a progressive accessibility innovation, but it does require users to have licensed GIS software to fully utilize the capabilities. The GIS allows users direct access to the enterprise system from their desktops for data export, spatial query, and mapping. The Iowa system's customized interface allows all users access to the custom analytical tools and database queries directly from their desktops.

The Iowa system's custom interface allows all users access to the custom analytical tools and database queries directly from their desktops.

7.6. Business Practices For Accessibility

States vary widely in their business practices related to accessibility. In some States, the custodial office may control all access, while in others, access is obtained through a second office or is a shared responsibility among a set of offices or agencies. Accessibility to crash data can be facilitated through a wide array of agencies extending their distribution networks far beyond the State custodial office. The State police may provide access to local law enforcement; the State engineering office may provide a different sort of access to city and county engineers, and so forth. In the end, what matters is not how a State has chosen to organize access, but rather if the data users have access and the capability (or assistance) to get what they need when they need it.

Electronic Data Entry and Transfer Technologies Impact Data Accessibility

Kentucky and Iowa are examples of states where automation and technology at the data entry stage positively impacts business practices for data accessibility by users.

Through the utilization of its systems, Kentucky has eliminated its entire backlog of incident reports. Its databases are now day-current by virtue of electronic data entry and electronic data transfer for reports that are entered by hand in the office. Kentucky uses a single, shared repository for data storage and a web-deployed interface to minimize staff involvement in the data-entry process and ensure data consistency for end-users. The Web portal also gives managers and analysts quick and seamless access to the necessary resources to achieve decreased response time for critical safety issues.

> *Kentucky crash databases are now day-current through the utilization of electronic data entry and electronic data transfer*.

Iowa's TraCS system maximizes readily available data in the data-collection stage, reducing data-entry time and duplicate data entry. The associated business rules mitigate the need for additional staff intervention to validate and provide quality control on incident records. Automatic electronic data-transfer processes distribute incident records to all necessary agencies, eliminating the need for end-users to export data from the enterprise system for analysis and reporting. The desktop interface puts custom tools and reporting capabilities at managers' and analysts' fingertips.

> *Automatic electronic data-transfer processes distribute incident records to all necessary agencies, eliminating the need for end-users to export data from the enterprise system for analysis and reporting*.

The use of electronic, field-deployed data entry and electronic data transfer are associated with dramatic improvements in data accessibility. Enterprise database systems at all three States - Michigan, Iowa, and Kentucky provide quick and seamless data access to those who need the information. Specialized tools on the desktop, or the Internet, are providing managers and analysts' quick access to tools and to reports that previously required significant amounts of time to generate.

7.7. Examples Of Metrics For Accessibility

Michigan, Iowa, and Kentucky all have outstanding provisions for providing accessibility. However, none of them have metrics in this area but probably have the capacity to generate them. As with Integration, hypothetical ways of adding metrics to these States' programs of data offerings can be envisioned.

Example 1. Assessing Crash Data Access in Michigan

Michigan aspires to provide data customers what they desire – "data on demand." However, the State acknowledges the difficulty in knowing how to create that demand.

Nevertheless, the custodial agency of the crash and other safety data managers continually tries to respond to the needs of data users.

In its tool kit for access, Michigan includes:

- Crash Facts (at www.michigantrafficcrashfacts.org)
- Limited on-line queries and standard reports.
- Traffic Crash Reporting System (TCRS) Web – At the present time, the numbers of "hits" on the websites (number of users accessing it) or requests made are not being tallied.
- Southeast Michigan Council of Governments (SEMCOG) web site – This site offers a wide variety of information including access to data and to safety planning (www.semcog.org)
- Links to TCRS, Traffic Crash Purchasing System (TCPS), advisory to law enforcement agencies, crash form updates, etc.

Michigan currently uses the criterion "appropriate access to data" as measured by the number of agencies and groups appropriate for that level of access. Their goal is to provide access to all law enforcement agencies and selected road commissions and researchers between 2006 and 2009. By the end of 2010, access is anticipated to extend to all safety partners.

Although not presently used, Michigan could easily measure access by tracking hits to its website, numbers of persons performing on-line queries and/or requesting standard reports.

Example 2. Assessing Crash Data Access in Iowa

Iowa provides numerous ways of accessing the crash data to meet user needs, from the simple to the complex. When desktop computing became common, Iowa released crash data and analysis software first on floppy disks and later on compact disks to crash data users statewide. The software generated several types of standardized reports and had limited but simple query capabilities.

When the State converted to a GIS format for safety data, it became necessary to do location-specific queries using a map interface. Most users were unfamiliar with GIS software and thus the SAVER and CMaT tools were created and distributed as needed. Later, the Incident Mapping and Analysis Tool (IMAT) was created, originally as a spin-off of CMAT, to operate on TraCS local datasets for law enforcement applications. Both SAVER and IMAT are capable of analyzing multiple kinds of safety data (not just crashes). All three are distributed on compact disks that include multiple years of the statewide crash file.

For data customers who are not capable of doing the analysis themselves, or do not have the staff-time available to use the analysis tools, the Center for Transportation Research and Education has a program that can fill data requests—the Iowa Traffic Safety Data Service (ITSDS). ITSDS is funded by the Iowa Governor's Traffic Safety Bureau and also by the Iowa Department of Transportation.

Recently, web pages for crash data were upgraded and the new Iowa DOT web-based "County Crash Profiles" and "City Crash Profiles" expanded the offerings of standardized reports by an order of magnitude. Iowa can document numbers of users of each of its three analysis tools, including who and where they are. It has not tracked changes in these numbers over time, however. This would be problematic to accomplish retrospectively but easy to set up for the future.

The ITSDS has kept records of data services provided over time, the best existing resource for creating a metric for accessibility. Like Michigan, Iowa could start tracking website hits, particularly for the new County or City Crash Profiles.

Example 3. Assessing Crash Data Access in Kentucky

Presently approximately 25 percent of law enforcement agencies (LEAs) are using the on-line crash data system for data retrieval and statistical reports. Table 8 illustrates Kentucky's calendar year (2007) summary requests by program area. This metric can be generated monthly, quarterly or yearly. It is an excellent tool to measure how well a State's accessibility initiatives are working as well as whether the LEA's and other safety-related agencies are monitoring themselves. If the percentage remains stagnant, then it is fairly likely that the LEA's are not accessing the on-line crash data for data retrieval and statistical reporting. This points to an even larger problem in that the LEAs may not be reviewing their own performance, or using all the data available for decision making. If the percentage of LEA's not using the system is high then the potential for this to affect the timeliness, accuracy, and completeness of other parts of the Traffic Records system is increased.

Table 8: Kentucky Calendar Year Totals

Monthly Programs	Jan-07	Feb-07	Mar-07	Apr-07	May-07	Jun-07	Yearly Totals
Research	37	36	50	40			
Corridor	3	18	30	11			
Education	1	9	19	23			
Community	7	2	15	19			
DIP	0	0	0	2			
CPS	10	41	23	17			
Monthly Totals	***58***	***106***	***137***	***112***			

Table 9: Summary of Examples for Accessibility

Examples	Types of Data Access
1. Michigan	Crash facts website; on-line queries; standard reports
2. Iowa	Crash facts website; electronic personal queries via SAVER, CMaT, or IMAT GIS analysis software; Web pages with County crash profiles (standardized reports) and results of numerous studies
3. Kentucky	KY on-line system

Note: Each type of access affords opportunities for creating metrics.

8.0 Summary and Conclusion

Six data quality characteristics have been identified by the USDOT and its partners as the means by which States' safety data systems may be evaluated and improvement projects measured. The six quality characteristics, often called "the six-pack," are Timeliness, Accuracy, Completeness, Consistency, Integration, and Accessibility.

Four of the six quality characteristics – timeliness, accuracy, completeness, and consistency – define the quality of the data within the crash database. The last two quality characteristics, Integration and Accessibility, are mechanisms that enable the crash database to be utilized for solving highway safety problems**.** *The purpose of this Guide is to assist States in improving the quality of their crash data systems through a systematic approach of defining, monitoring, and measuring change over time for these six characteristics of crash data quality.*

Data quality improvements are often linked to technological advancements and always linked to sound business practices. Sweeping changes are usually expensive. The measurement of data quality to detect and verify improvements can also be facilitated through technology, but often requires only simple, low-cost methods.

The three States referenced in this report - Michigan, Iowa and Kentucky – have made extensive changes both in technology and business practices to improve crash data quality. More germane to this Guide, these States have made varying degrees of progress in the actual measurement of their improvements using the "six-pack." These examples may serve as "Good Practices" to states seeking to improve their crash data systems.

Kentucky and Iowa have shown that electronic, field-based data entry and electronic data transfer can expedite data entry and increase efficiency by reducing staff involvement in these processes. This also increases data consistency and accuracy though the use of data element standards and business rules for data validation and quality control. The use of an open relational database management system (RDBMS) for data storage allows great flexibility in data accessibility and analysis. Michigan's integration of outside databases (including direct links to roadway, driver, and vehicle databases) as well as their development of documented and effective business practices are also models.

Measurement of data quality characteristics enables States to know the true status of their crash data system as it currently performs. Improvement projects can be more appropriately chosen and implementation of improvements can be more accurately monitored, both for desired and for unintended consequences. Because a baseline was created by initial data quality measurements, the ultimate impact of the improvement can be determined. The effective use of public funds is thus assured. Quality of data is improved, providing a basis for effective safety countermeasures and reductions in the loss of life and property in highway crashes.