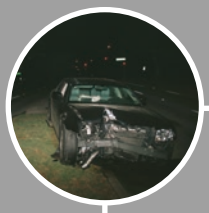


U.S. Department of Transportation
National Highway Traffic Safety Administration

NHTSA

DATA INTEGRATION:
LINKING IT ALL TOGETHER

Integrating Data: A Business Case

A State's traffic records system includes six core data sources—crash, driver, vehicle, roadway, citation and adjudication, and injury surveillance—and within each of these, there may be multiple databases. Taken in isolation, each of these more than two dozen data sets fit the defined business needs of the agencies that collect and manage them. States have been combining or **integrating** some of these core data sources for decades, but what if more of this data could be integrated so that multiple agencies could use this information to make better decisions that prevent crashes and save lives?

Collectively, the agencies responsible for the six core data sources strive to improve safety, but no one agency can do it alone. Data integration has the potential to promote resource and expertise-sharing, and when that happens everyone wins. Expanding data integration can help your agency solve complicated problems as well as foster widespread collaboration with other traffic safety and health care organizations. Linked data becomes valuable to decision-makers in the organizations, and by extension the legislators, the public, and private community groups. High-quality, detailed, and on-topic data and analyses move all stakeholders toward evidence-based decision-making.

The value of data-driven decision-making is reflected in national guidelines such as the *Highway Safety Manual* for engineering safety, *Countermeasures That Work* for addressing behavioral safety, and the various data-driven safety planning efforts that States engage in every year. This guide provides a brief overview of data integration to help bolster traffic records data systems and data-driven decision-making.

This guide is intended for data users, data analysts, and decision-makers who work with the traffic records system. If you represent or work closely with any of the core traffic records system components—crash, vehicle, driver, roadway, adjudication, or emergency services—this guide will assist your data integration efforts.

What Is Data Integration and Why Do It?

Combining or merging data sources for analysis purposes is typically referred to as data integration. Often performed annually, integration merges the data sets into a more comprehensive file containing data elements from each source. The two most common types of integration are deterministic and probabilistic. **Deterministic integration** relies on common elements shared among the data sets to make exact matches. For example, when a State inputs all crashes into the same linear referencing system as they put their roadway inventory and traffic volume data, they use location code or coordinates to exactly match records belonging to the same point in the roadway. In contrast, **probabilistic integration** relies on similar elements and values shared among the data sets to make matches, such as slight differences in date and time ranges on EMS records and crash reports.

Combining two or more data files to check data quality or auto-complete data collection forms is referred to as a **data interface**. A data interface is typically performed in near-real time and would not add new elements to either source.

Linking data from across the four E's—Engineering, Education, Enforcement, and Emergency Services—can result in a more complete understanding of traffic safety issues. Researchers, highway safety office program managers, and other safety stakeholders can use the data to make better decisions and develop solutions.

Linking data systems can also reduce the number of data elements required in each source. For example, a police crash report based on the *Model Minimum Uniform Crash Criteria (MMUCC) 5th Edition* has a roadway section that can be populated based on linkage to the roadway inventory data by using the location information as linking variables. This dramatically reduces the amount of time officers spend collecting data at the scene of the crash, increases data completeness and accuracy, and supports more detailed analyses of roadway-related contributing factors in crashes.

Linking data can result in a more complete understanding of traffic safety issues so that better decisions can be made about how to address them. Linking data systems can also reduce the number of data elements required in each source.

Example A police crash report based on the *MMUCC 5th Edition* has a roadway section that can be populated based on linkage to the roadway inventory data by using the location information as linking variables. This dramatically reduces the amount of time officers spend collecting data in the field, increases data completeness and accuracy, and supports more detailed analyses of roadway-related contributing factors in crashes.

Are You a Safety Stakeholder?

All safety stakeholders—data professionals (IT), educators, enforcement officials, legislators, engineers, and planners—play a role in data linkage and integration. State Traffic Records Coordinating Committees (TRCCs) are responsible for prioritizing, promoting, and coordinating quality control and data improvement programs that affect any portion of the traffic records system. State TRCCs include members that have different roles and responsibilities for data. Data collectors directly impact data timeliness, completeness, and accuracy, which helps make linkage possible and effective; data program managers monitor and maintain overall performance of a data system; and data users rely on high-performing data systems to investigate problems and identify solutions.

Data linkage and integration benefits all safety stakeholders by:

	**Enhancing decision-making through more advanced, evidence-based approaches to safety**
Providing broader access to the data, including to the public	**Improving support for safety program grant requests and justification for all resource allocations**
Directing analytic support for detailed countermeasure deployment, even spot location and individual driver levels	**Identifying trends in safety performance and outcomes at the State level that can drive programs and raise public awareness**

Identify Your Business Needs

Robust data systems allow safety professionals to develop countermeasures, allocate resources, evaluate program effectiveness, and advocate for improved data. Data linkage efforts require cooperation among data collectors, managers, and users. Each of these stakeholders should consider what they need from the data in order to do their jobs.

- What data do you need?
  - What is the job and what data does the job require?
  - What do you do *with* data that cannot be done *without* data?
- What do you need to enhance your services?
  - Are there requests for information that you cannot meet now?
    - What additional data would meet those needs?
  - What additional data would allow you to be more efficient and provide more relevant support for decision-making?

The Role of Data Governance

Data integration is not a quick or simple process. Successful integration and linkage require communication and collaboration. Experienced States note that formal data governance processes are essential for ensuring success.

Data governance is an approach to managing data by a committee that creates policies, documents data quality standards, and enforces those standards across all systems and partner agencies. A Memorandum of Understanding or Agreement (MOU/MOA) can formalize interagency collaboration and cooperation. Through these mechanisms, agencies work together to decide the steps needed to integrate data—such as agreed-upon identifiers like a naming mechanism or number system—to make the linkage process as efficient as possible. The National Highway Traffic Safety Administration (NHTSA) encourages agencies to use the National Information Exchange Model (NIEM) and other national data models and guides so that States' data can be integrated at a national level. The standards and guides save States time and labor when designing data systems because the data definitions and data exchange methods are already worked out.

Data governance groups establish the documentation defining each data element, as well as how it is collected and evaluated against State and national standards. They also help prioritize data and system changes based on partner organizations' business needs. When all traffic records system components are managed through a formal data governance process, State personnel interested in data linkage can easily identify the contact persons and procedures for gaining access to the required data sets.

Collectively, all stakeholders will have a better understanding of the value of each system because of data governance practices. Individually, each stakeholder group may experience benefits specific to their business practices. For example, TRCCs can review detailed summaries of collected data. Data managers can identify data element locations, increasing linkages and creating streamlined traffic records systems and eliminating redundancies.

An MOU/MOA typically includes the following components:

Purpose: Describes the collaboration that two or more agencies will maintain.
Sample Text: *The purpose of this MOA is to establish a citation data exchange between the State Department of Public Safety (DPS) and State Courts Administration for sharing initial charges and final dispositions of all traffic cases managed in the State, county, and appellate courts.*

Roles and Responsibilities: Details the tasks each agency is responsible to perform.
Sample Text: *State DPS will be responsible for providing the XML schema for the State traffic records data repository; wherever possible, the NIEM will be used as the schema. State Courts Administration will adhere to the State DPS schema as published.*

Reporting Requirements: Describes the data that needs to be reported, responsible parties, and the reporting interval.
Sample Text: *The State Courts Administration will provide an electronic record for every traffic citation and every charge entered into the State Court Administration's Case Management System on a nightly basis. A summary report of total citations and charges will be sent separately for use as a completeness check.*

Timeframe: Defines the period in time that the document will be valid.
Sample Text: *This MOA shall be effective on the date of the last signature of the parties to the agreement. This will be in effect for 3 years from the date of execution.*

Successful State Data Linkage

Injury Surveillance Data Linkage

States are linking crash, EMS, and emergency department data to determine outcome and economic data for crash types as well as the impact of educational programs addressing risky behaviors. **Utah's Crash Data Initiative** is linking data to understand injury severity, seat belt use, and distraction. **Maryland's Crash Outcome Data Evaluation System (CODES)** project links crash, hospital, driver license, citation, and vehicle registration data to support program evaluation and policy decision-making. Recently, the CODES data was used to support a major update in the State's ignition interlock policy.

Data Visualization Linkages

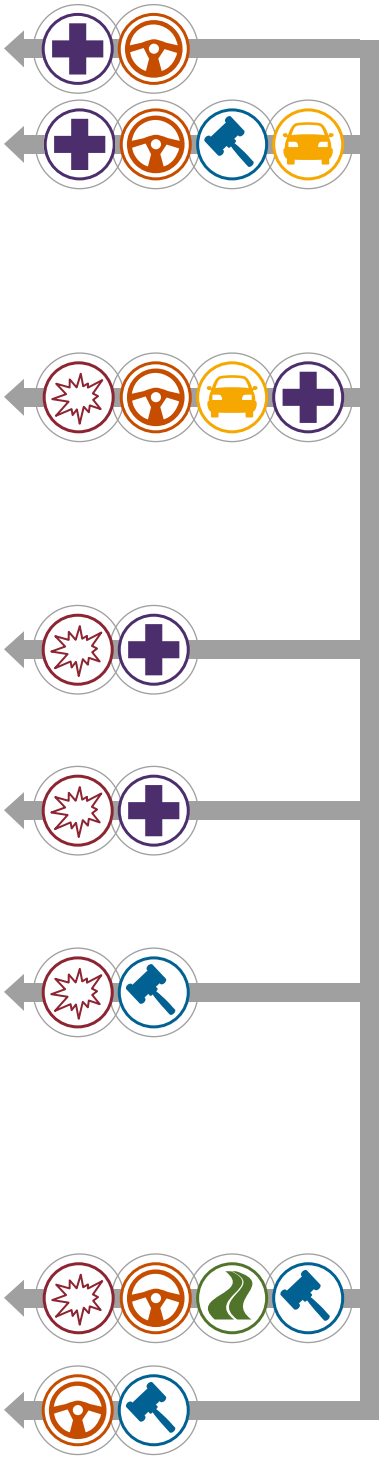
States are developing new ways to engage stakeholders with novel data presentations. Using crash, driver, vehicle, and injury data, **Louisiana** created a website with dashboards that show how the State is addressing the performance measures outlined in its Strategic Highway Safety Plan. The website allows for the dynamic display of traffic safety problems such as driver distraction and contributing factors for pedestrian-involved crashes, which helps with emphasis area problem identification, resource allocation, and countermeasure selection.

Nevada produces a quarterly Traffic Research and Education newsletter with safety infographics that is disseminated to the public via a website and social media platforms. Within the newsletter, the State provides summaries of linked data from crash, EMS, and trauma centers that focus on injuries, outcomes, and the economic consequences of crashes for a variety of road user types. **Michigan** is developing data visualization tools, customizable widgets, fatal crash analytics, and reports tailored to its multiple safety program areas.

Citation and Crash Data Linkages

States are also linking crash and citation data to generate driver alerts. In **North Dakota**, the parents of teen drivers receive an Early Warning Letter (EWL), when their novice driver commits a traffic violation or is involved in a crash within 9 months of receiving his/her license. The EWL details how their teen is at greater risk for being involved in a fatal crash because the incident occurred early in his/her driving career. The initiative, which is the result of linking crash, citation, and adjudication data, is helping to reduce teen driver violations and fatal crashes.

Ohio and Virginia link crash and citation information to manage behavioral programs, support law enforcement planning and resource allocation, and identify roadway infrastructure improvement locations. **Virginia** uses their Traffic Records Electronic Data System to perform linkages giving the State the ability to analyze high-crash locations and unsafe behaviors at a level of detail that neither system alone could support. **Ohio's** Department of Public Safety Driver's Records Retrieval System links driver and citation data to reduce errors and produce more complete records.



Where do You Start?

Equipped with the basic data integration tools, you are ready to start. The first step is to identify the greatest needs and their short- or long-term payoff. To be most effective, focus your efforts on the business and data needs most important to your agency.

Learn more about filling an important role in data integration by:

- **Engaging with your State's TRCC** to learn what is being done to promote data integration (or linkage) and explore the Traffic Records Inventory, which documents data definitions and responsible agencies for each traffic records system component.
- **Sharing your subject matter expertise** with the Data Governance Group and the TRCC members who are responsible for the data sources you need.
- **Visiting the NHTSA traffic records website** (www.nhtsa.gov/research-data/traffic-records) and take advantage of the agency's technical assistance, which includes access to subject matter experts (GO Teams) who can provide guidance to help improve your State's traffic records system.
- **Tap into existing data integration resources such as:**

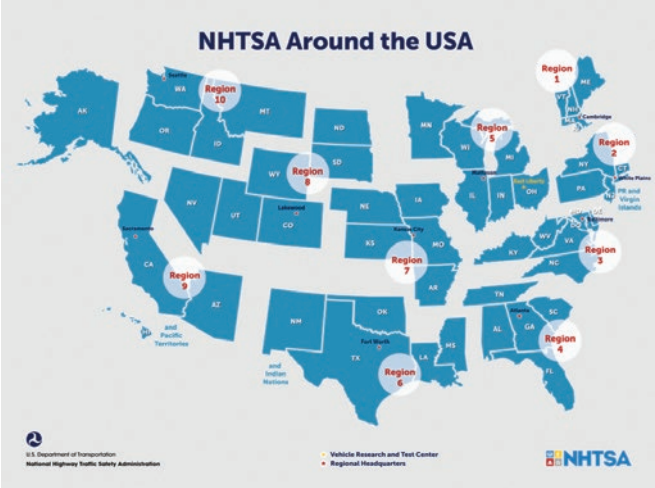
  National Information Exchange Model (www.niem.gov)

  NHTSA's State Traffic Records Coordinating Committee Noteworthy Practices (www.transportation.gov/government/traffic-records/state-traffic-records-coordinating-committee-noteworthy-practices)

  NHTSA's TRCC Traffic Records Strategic Planning Guide (Upcoming)

  Crash Data Improvement Program Guide (www.nhtsa.gov/research-data/traffic-records#cdip-38851)

  NHTSA's Crash Outcome Data Evaluation Services (CODES) and Applications to Improve Traffic Safety Decision-Making (https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811181)



Need help?

Contact your NHTSA Regional Office for technical assistance and more information on data integration resources.

DOT HS 812 784
July 2019

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

NHTSA

14304-072419-v3