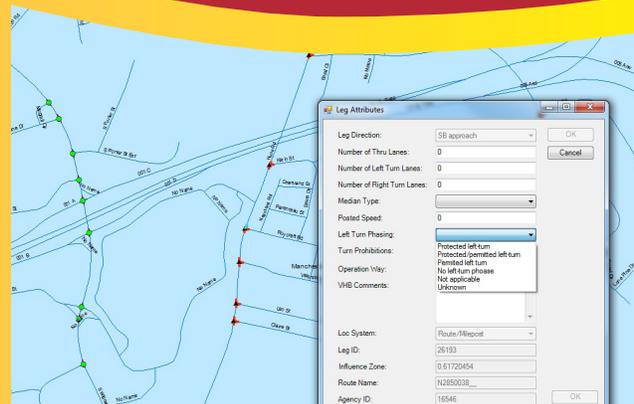


Informational Guide for State, Tribal, and Local Safety Data Integration



FHWA Safety Program

FHWA-SA-16-118
November 2016



U.S. Department of Transportation
Federal Highway Administration



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

<http://safety.fhwa.dot.gov>

NOTICE

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

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

QUALITY ASSURANCE STATEMENT

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



TECHNICAL DOCUMENTATION PAGE

1. Report No. FHWA-SA-16-118		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Informational Guide for State, Tribal, and Local Safety Data Integration				5. Report Date October 31, 2016	
				6. Performing Organization Code	
7. Author(s) Scopatz, R.A., Goughnour, E., Abbott, D., Tang, E., Carter, D., Smith, S., and Salzer, T.				8. Performing Organization Report No.	
9. Performing Organization Name and Address VHB 8300 Boone Blvd., Suite 700 Vienna, VA 22182-2626				10. Work Unit No.	
				11. Contract or Grant No. DTFH61-10-D-00022 (VHB) Task 13-006	
12. Sponsoring Agency Name and Address Federal Highway Administration: Office of Safety 1200 New Jersey Ave., SE Washington, DC 20590				13. Type of Report and Period Informational Guide August 2013 – August 2016	
				14. Sponsoring Agency Code FHWA	
15. Supplementary Notes The contract manager for this report was Stuart Thompson.					
16. Abstract This is an Informational Guide for State, Tribal, and local safety data integration. This Informational Guide provides guidance to States, Tribes, and local agencies on the steps and effective methods of safety data integration. This Informational Guide documents practices within a nine-step process for data integration. Detailed descriptions of the nine data integration steps are provide as are recommendations to achieve integration. The steps in this Informational Guide help agencies meet their needs for a flexible approach to data integration. Readers are encouraged to take those portions most relevant to their needs in order to develop and implement their own safety data integration projects. A State, Tribe, or local agency does not have to adopt all the activities described in the Informational Guide.					
17. Key Words: safety data, data integration, roadway inventory data, traffic data, crash data			18. Distribution Statement No restrictions.		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 169	22. Price

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

SI* (MODERN METRIC) CONVERSION FACTORS				
APPROXIMATE CONVERSIONS TO SI UNITS				
Symbol	When You Know	Multiply By	To Find	Symbol
LENGTH				
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
AREA				
in ²	square inches	645.2	square millimeters	mm ²
ft ²	square feet	0.093	square meters	m ²
yd ²	square yard	0.836	square meters	m ²
ac	acres	0.405	hectares	ha
mi ²	square miles	2.59	square kilometers	km ²
VOLUME				
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft ³	cubic feet	0.028	cubic meters	m ³
yd ³	cubic yards	0.765	cubic meters	m ³
NOTE: volumes greater than 1000 L shall be shown in m ³				
MASS				
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
TEMPERATURE (exact degrees)				
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
ILLUMINATION				
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m ²	cd/m ²
FORCE and PRESSURE or STRESS				
lbf	poundforce	4.45	newtons	N
lbf/in ²	poundforce per square inch	6.89	kilopascals	kPa
APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	When You Know	Multiply By	To Find	Symbol
LENGTH				
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
AREA				
mm ²	square millimeters	0.0016	square inches	in ²
m ²	square meters	10.764	square feet	ft ²
m ²	square meters	1.195	square yards	yd ²
ha	hectares	2.47	acres	ac
km ²	square kilometers	0.386	square miles	mi ²
VOLUME				
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m ³	cubic meters	35.314	cubic feet	ft ³
m ³	cubic meters	1.307	cubic yards	yd ³
MASS				
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
TEMPERATURE (exact degrees)				
°C	Celsius	1.8C+32	Fahrenheit	°F
ILLUMINATION				
lx	lux	0.0929	foot-candles	fc
cd/m ²	candela/m ²	0.2919	foot-Lamberts	fl
FORCE and PRESSURE or STRESS				
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in ²

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

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	VIII
ACRONYMS	X
EXECUTIVE SUMMARY	XI
OVERVIEW	I
WHAT IS DATA INTEGRATION?	2
WHY SHOULD AGENCIES CONSIDER DATA INTEGRATION?	3
AUDIENCE FOR THE INFORMATIONAL GUIDE	5
ORGANIZATION OF THE INFORMATIONAL GUIDE	6
Case Studies	7
Pilot Studies and Additional Work Plans.....	8
DATA INTEGRATION CONCEPTS	8
Data Program Concepts	8
METADATA	11
Data Dictionary.....	11
Data Flows	14
Data Access Rights	15
Data Quality and Standards	17
Data Retention.....	18
Multiple Users Benefit from Integration.....	19
User Support.....	20
SUMMARY	20
STEP I – LAY THE FOUNDATION	21
OBJECTIVES	21
CHALLENGES	23
RECOMMENDED SOLUTIONS	24

LESSONS LEARNED.....	25
SUMMARY	26
CHECKLIST.....	26
STEP 2 – CONDUCT GAP ANALYSIS	28
OBJECTIVES	28
CHALLENGES	28
RECOMMENDED SOLUTIONS	30
LESSONS LEARNED.....	31
CHECKLIST.....	32
SUMMARY	32
STEP 3 – ESTABLISH DATA GOVERNANCE PROCESS.....	33
OBJECTIVES	33
CHALLENGES	35
RECOMMENDED SOLUTIONS	37
LESSONS LEARNED.....	40
CHECKLIST.....	41
SUMMARY	41
STEP 4 – DEVELOP DATA COLLECTION AND INTEGRATION PLAN	43
OBJECTIVE.....	43
CHALLENGES	44
RECOMMENDED SOLUTIONS	45
LESSONS LEARNED.....	47
SUMMARY	48
CHECKLIST.....	48
STEP 5 – IDENTIFY TRAINING NEEDS FOR DATA COLLECTION, INTEGRATION, AND ANALYSIS.....	50
OBJECTIVES	50

CHALLENGES	51
RECOMMENDED SOLUTIONS	52
LESSONS LEARNED.....	53
SUMMARY	54
CHECKLIST.....	54
STEP 6 – PERFORM DATA INTEGRATION	56
OBJECTIVES	56
CHALLENGES	56
RECOMMENDED SOLUTIONS	58
LESSONS LEARNED.....	59
SUMMARY	60
CHECKLIST.....	60
STEP 7 – DEVELOP AND DEPLOY THE EXTRACT, TRANSFORM, AND LOAD PROCESS	62
OBJECTIVES	62
Data Extraction Process	62
Data Transformation.....	64
Data Load	64
GIS-Based Data Integration	65
Spatial Data Analysis, Filters, and Extraction	66
CHALLENGES	69
RECOMMENDED SOLUTIONS	70
LESSONS LEARNED.....	70
CHECKLIST.....	71
SUMMARY	71
STEP 8 – CONDUCT ANALYSES.....	74
OBJECTIVES	74
Types of Safety Analyses.....	75
How to Choose the Right Tool?	76

CHALLENGES	76
RECOMMENDED SOLUTIONS	77
LESSONS LEARNED.....	79
CHECKLIST	79
SUMMARY	79
STEP 9 – PERFORM EFFECTIVENESS EVALUATION.....	81
OBJECTIVES	81
CHALLENGES	81
RECOMMENDED SOLUTIONS	82
LESSONS LEARNED.....	82
CHECKLIST.....	83
SUMMARY	83
CONCLUSION	84
BENEFITS OF USING THIS INFORMATIONAL GUIDE	85
REFERENCES	86
APPENDIX A – PILOT STUDY REPORTS	88
APPENDIX B – ADDITIONAL WORK PLANS.....	114
APPENDIX C – ROADWAY SAFETY DATA AND ANALYSIS TOOLBOX.....	124
APPENDIX D – DATA SHARING AGREEMENT AND MEMORANDUM OF UNDERSTANDING	127
APPENDIX E – SAMPLE SURVEY	139
ACKNOWLEDGEMENTS.....	156

LIST OF TABLES

Table 1. Those who may be involved in data integration efforts.	6
Table 2. MIRE FDEs for non-local (based on functional classification) paved roads.	12
Table 3. MIRE FDEs for local paved roads.	13
Table 4. MIRE FDEs for unpaved roads.	14
Table 5. Data source summary.	89

LIST OF FIGURES

Figure 1. Chart. Nine-step process for safety data integration.....	xi
Figure 2. Chart. Safety Management Process.....	2
Figure 3. Chart. Data driver safety analysis process.....	4
Figure 4. Chart. Generalized crash data flow diagram.	15
Figure 5. Chart. Step 1: Lay the Foundation.	21
Figure 6. Chart. Step 2: Conduct Gap Analysis.....	28
Figure 7. Chart. Step 3: Establish Data Governance Process.....	33
Figure 8. Chart. Step 4: Develop Data Collection and Integration Plan.....	43
Figure 9. Chart. Step 5: Identify Training Needs.	50
Figure 10. Chart. Step 6: Perform Data Integration.	56
Figure 11. Chart. Step 7: Develop and Deploy the ETL Process.	62
Figure 12. Map. Sample GIS map showing horizontal curve locations.	65
Figure 13. Map. Sample GIS map intersection types and crashes.	67
Figure 14. Screenshot. Sample crash application filters for performing analysis.....	68
Figure 15. Chart. Step 8: Conduct Analyses.....	74
Figure 16. Chart. Integrating the roadway safety management and project development process.	75
Figure 17. Chart. Step 9: Perform Effectiveness Evaluation.....	81
Figure 18. Chart. Interactions among the steps in data integration.	84
Figure 19. Chart. Process overview.....	90
Figure 20. Screenshot. Data management tool.	92
Figure 21. Chart. Recommended LRS/MIRE road inventory work flow.	111

Figure 22. Screenshot. Roadway Safety Data and Analysis Toolbox..... 124

Figure 23. Screenshot. Advanced search feature..... 125

Figure 24. Screenshot. Filter options from advanced search page..... 125

ACRONYMS

AADT	Annual Average Daily Traffic
AASHTO	American Association of State Transportation Officials
ARNOLD	All Roads Network of Linear Referenced Data
ATIS	Arizona Transportation Information System
BIA	Bureau of Indian Affairs
CMF	Crash Modification Factor
ETL	Extract, Transform, Load
FDE	Fundamental Data Elements
GIS	Geographic Information System
HSIP	Highway Safety Improvement Program
HSM	Highway Safety Manual
IHSDM	Interactive Highway Safety Design Model
IRR	Indian Reservation Roads
IT	Information Technology
LBRS	Location Based Response System (Ohio)
LGIM	Local Government Information Model
LiDAR	Light Detection and Ranging
LTAP	Local Technical Assistance Program
LRS	Linear Referencing System
MAG	Maricopa Association of Governments
MIS	Management Information System
MIRE	Model Inventory of Roadway Elements
MOU	Memorandum of Understanding
MPO	Metropolitan Planning Organization
NHS	National Highway System
PAG	Pima Association of Governments
RIGIS	Rhode Island Geographic Information System
RMS	Records Management System
RPC	Regional Planning Commission
RSDP	Roadway Safety Data Program
SHSP	Strategic Highway Safety Plan
TRCC	Traffic Records Coordinating Committee
TRIMS	Tennessee Roadway Information Management System
TTAP	Tribal Technical Assistance Program
TTIP	Tribal Transportation Improvement Program
WISLR	Wisconsin Information System for Local Roads

EXECUTIVE SUMMARY

The Federal Highway Administration (FHWA) Office of Safety sponsored a project titled *Integration of State and Local Safety Data in 2013-2016* in response to a State identified gap in knowledge about how to integrate safety data across agencies and data types. This Informational Guide is intended to supply States, Tribal governments, and local agencies with a roadmap to implement safety data integration projects.

The audience for this Informational Guide includes State Departments of Transportation (DOTs), Tribal transportation agencies, county and municipal governments, metropolitan planning organizations (MPOs), regional planning commissions (RPCs), and researchers who support any of these agencies. This Informational Guide documents practices within a nine-step process for data integration as shown in Figure 1.

This Informational Guide is designed to help the reader understand safety data integration concepts by providing detailed descriptions of the nine data integration steps, and recommendations to achieve integration. The steps in this Informational Guide help agencies meet their needs for a flexible approach to data integration. Examples of multiple paths to successful integration are discussed throughout. A State, Tribe, or local agency does not have to adopt all the activities described in the Informational Guide. Depending on circumstances, not all steps will be necessary and the order of the need not be sequential. Agencies are encouraged to take those portions most relevant to their needs in order to develop and implement their own safety data integration projects.

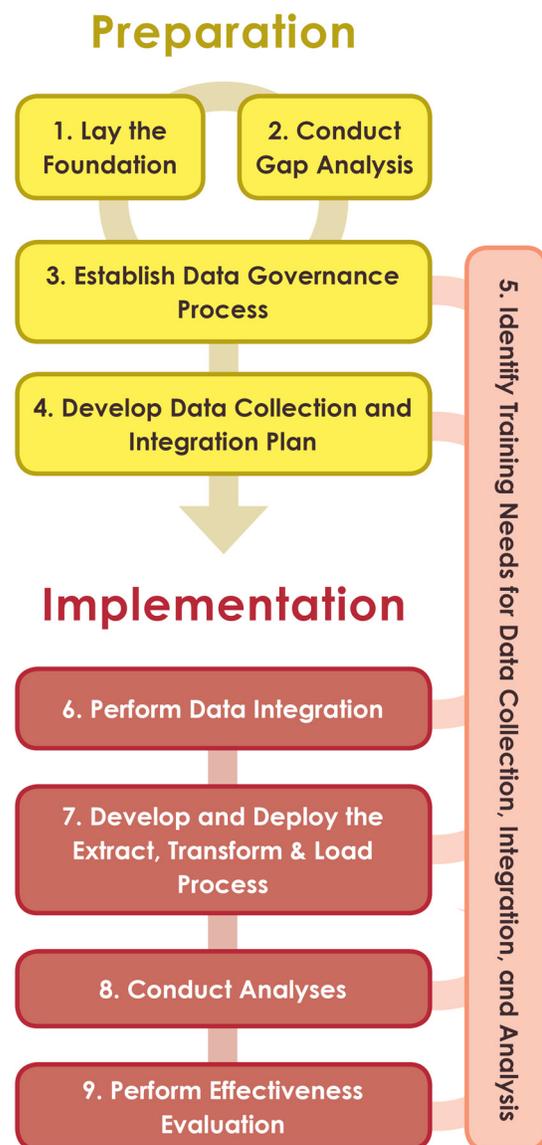


Figure 1. Chart. Nine-step process for safety data integration.

OVERVIEW

This document is an *Informational Guide for State, Tribal, and Local Safety Data Integration*. Its purpose is to give practical examples and guidance to agencies planning to integrate safety data. Throughout this document, “safety data” consists of three core data sources: crash records, roadway inventory, and traffic volume. The term “target agencies” is used to refer to State DOTs, Tribal transportation agencies, and local agencies working in traffic safety and roadway asset management. Two key views of data integration are discussed throughout: integration across jurisdictions and integration among safety databases:

- **Integration across jurisdictions** is when the target agencies agree to share data to support safety analyses, asset management, and other purposes. The examples and guidance in this document do not prescribe any single way to accomplish this cross-agency integration. The goal here is to describe the processes by which agencies can work together to improve data, conduct more complete analyses, arrive at more efficient solutions, and ultimately reduce the frequency and severity of crashes. When the need for data on all public roadways is met, then it does not matter if there’s a single centralized data resource managed by one custodial agency, or if the data are managed by individual agencies and combined at the time of analysis. The multiple agencies involved in partnership should decide how best to meet their needs.
- **Integration among safety databases** is when agencies bring together their crash, roadway inventory, and traffic volume data to create an analytic data resource. This resource may reside in one file or multiple files (just as with cross-jurisdictional integration). Target agencies can work together to implement a solution so that they can analyze safety on all public roads.

This Informational Guide focuses on integrating safety data. There are many other data sources that the target agencies may wish to include. The most prevalent example is asset management information on condition and location of pavement, signs, markings, culverts, and other physical resources and attributes of roadways. This information is not treated as a key component of safety data; however, target agencies may decide that integrating asset information with safety data is economically feasible and benefits an expanded list of users. Other data such as citations, adjudication, emergency response, and injury surveillance may also be a part of the data integration process. Throughout this document we provide examples of successful data integration that meets multiple needs. Successful data integration by target agencies will engage a broad audience for the data. The needs of all potential users of the integrated data will determine which information sources will be included beyond the core safety data described in this report.

WHAT IS DATA INTEGRATION?

Data integration is the linking of multiple data sources to meet users' needs. State-of-the-practice analysis methods supporting safety considerations in infrastructure investment decision-making require integrated safety data. Those analyses support decision making throughout the safety management process, as shown in in Figure 2, as well as planning, programming, maintenance, and project development processes. Data integration requires a common, or linking, variable in each of the to-be-integrated data files. For safety data integration, the linking variable is location. Linking based on location is facilitated by an all public roads linear referencing system (LRS) and basemap. By knowing the location of crashes, data integrators can associate those crashes with a specific spot on the roadway network. Using the same LRS, roadway inventory data links to crash data so that crashes can be associated not only with the specific locations where they occurred, but also with the roadway characteristics at those locations. Crash location descriptions are further enhanced with traffic volume data to provide a way to normalize the counts of crashes based on the amount of traffic experienced at each site. The resulting integrated data supports network screening, problem diagnosis, countermeasure selection, economic appraisal, project prioritization, and effectiveness evaluation.

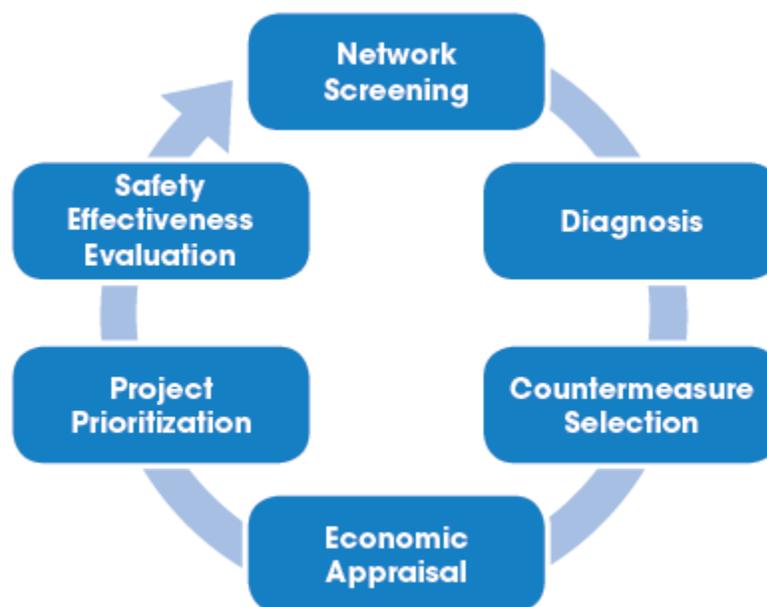


Figure 2. Chart. Safety Management Process.

WHY SHOULD AGENCIES CONSIDER DATA INTEGRATION?

The FHWA Data Integration Primer describes the benefits of integrating asset management data among transportation agencies. These benefits also apply to safety data integration:⁽¹⁾

- **Availability/Accessibility**—Data that is easily retrieved, viewed, queried, and analyzed by anyone within an agency encourages the integration of such data into every area of an agency that can benefit from it, spurring both innovation and better decision-making.
- **Timeliness**—Well-organized data can be quickly updated; one input will often apply the data across a variety of linked systems, and the information can be time-stamped to reflect its currency.
- **Accuracy and Integrity**—Errors are greatly reduced because the integration environment drives a higher quality of input and can include automatic or convenient error checking and verification.
- **Consistency and Clarity**—Integration requires clear and unique definition of various types of data, avoiding confusion or conflict in the meaning of terms and usage.
- **Completeness**—All available information, including both historical and recent data, is accessible in an integrated database, with any missing records or fields identified and flagged via the integration process.
- **Reduced Duplication**—Identical data are eliminated reducing the need for multiple updates and ensuring everyone is working from the exact same information.
- **Faster Processing and Turnaround Time**—Less time is spent on consolidating and transmitting data to various users in the agency. The integrated data environment saves time by eliminating consolidation and transmittal to disparate users and allows many users to conduct separate analyses concurrently.
- **Lower Data Acquisition and Storage Cost**—Data are collected or processed only once, and the information is consolidated and stored at locations supporting optimal convenience and ease of maintenance.
- **Informed and Defensible Decisions**—Highly organized, comprehensive databases allow users to drill down through successive levels of detail for an asset, supplying more information to support decisions and supporting different types of analysis using various data combinations.
- **Enhanced Program Development**—Comprehensive and coordinated system information advances program development by providing timely data for high-priority actions, promoting efficient distribution of funding among competing programs, and

improving consistency in programs from year to year and across departments, among other benefits.

- **Greater Accountability**—Data integration allows rapid and more accurate reporting of costs and accomplishments, including full attribution of results to relevant agency units and functions.

In short, data integration supports decision-making in ways that the individual data systems cannot. This brings us back to the opening discussion of integration across jurisdictions and among the core safety data sources. When target agencies bring together their safety data, the resulting integrated resource supports scientifically valid, rigorous, and complete analyses. The information developed from those analyses helps decision makers better understand the human, vehicle, and environmental factors that contribute to safety. As illustrated in Figure 3, improved data means that decision-making can address all public roads, resources are allocated where they will have the largest safety impact, lives are saved, and injuries are avoided.



Figure 3. Chart. Data driver safety analysis process.

Finally, safety data integration, and use of integrated safety data for decision-making are required under the Fixing America’s Surface Transportation Act as noted in the FHWA’s Guidance on State Safety Data Systems.⁽²⁾ Specific relevant guidance includes:

As part of its State highway safety improvement program, a State shall have in place a safety data system that can be used to perform analyses supporting the strategic and performance-based goals in the SHSP and HSIP. [23 U.S.C. 148 (c)(2)]. This section provides guidance on the capabilities a State’s safety data system should have in order to support analyses and evaluations in 23 U.S.C. 148, including: (1) types of roadways, (2) types of data, (3) geolocation of safety data to a common highway basemap, (4) analysis and evaluation capabilities, and (5) the subset of Model Inventory of Roadway Elements (MIRE) to be collected.

A State's crash, roadway, and traffic data must be able to be linked or combined by virtue of having common data elements. [23 U.S.C. 148 (c)(2)(A)(iii)]. These data should also be able to be linked to a State's other core safety databases including licensing, vehicle, citation/ adjudication, and emergency medical services or injury surveillance system. [23 U.S.C. 405 (c)(1)(C) and (c)(3)(C)]. Additionally, commercial motor vehicle data could also be linked based upon involvement in crashes and inspections. A State shall also improve the compatibility and interoperability of safety data with other State transportation-related data systems and with other States and national data systems, e.g., Fatality Analysis Reporting System. [23 U.S.C. 148 (c)(2)(A)(iv)].

Consistent with the purpose and scope of the HSIP, a State shall have in place a safety data system to perform safety problem identification and countermeasure analysis. [23 U.S.C. 148 (c)(2)(A)]. The statute also specifies that a State shall advance the capabilities of the State for data collection, analysis, and integration in a manner that includes all public roads, including non-State-owned public roads and roads on tribal land in the State. [23 U.S.C. 148 (c)(2)(D) and (D)(ii)]. Public road means "any road under the jurisdiction of and maintained by a public authority and open to public travel." [23 CFR 460.2(a)].

AUDIENCE FOR THE INFORMATIONAL GUIDE

The purpose of this Informational Guide is to provide target agencies with easy-to-use information and techniques to support their safety data integration efforts. The primary audience is the people leading the data integration at the target agencies.

Further, this document describes activities that will require others to get involved when the target agencies work to identify gaps and training needs, establish data collection priorities and data governance processes, and produce analyses that support decision making. That larger group includes practitioners throughout many business units in the target agencies. It also includes federal government agencies, researchers, consultants, and safety advocates.

Table I shows the types of people who might be involved in the activities described in this Informational Guide. While they aren't all part of the primary audience, this document describes roles that these individuals might play as part of a safety data integration effort.

Table 1. Those who may be involved in data integration efforts.

State and Tribal Personnel	Local Personnel	Others
<ul style="list-style-type: none"> • Executive staff in DOT and information technology (IT) roles • Safety engineers • Safety program managers • Geographic Information System (GIS) managers and staff • Enterprise data stewards • Roadway inventory data stewards • Crash data stewards • Traffic volume data stewards • Planners • Design engineers/managers • Maintenance engineers/managers • Asset managers • State law enforcement agencies 	<ul style="list-style-type: none"> • Executive leaders • GIS managers • IT staff • Public works managers • MPO and RPC staff • Designers/planners • Traffic engineers • Maintenance engineers • Asset managers • E911 managers • County appraisers/auditors • Local law enforcement agencies 	<ul style="list-style-type: none"> • FHWA division offices • FHWA headquarters staff • NHTSA and FMCSA regional offices • NHTSA and FMCSA headquarters staff • US Parks Service • Bureau of Land Management • Bureau of Indian Affairs • Engineering consultants • Safety advocates • Safety educators • News media • General public

ORGANIZATION OF THE INFORMATIONAL GUIDE

This Informational Guide has three main parts:

- I. **Overview and Data Integration Concepts.** This part presents definitions and examples of data management and data integration concepts so that readers may better understand the context and processes involved in safety data integration. These concepts apply generally to data management and integration. The discussion and examples presented in these chapters are specifically relevant to highway safety data sources.
- II. **The Nine Steps for Data Integration.** This part presents each of the nine steps of data integration. The steps and their purposes are:

- 1) Lay the Foundation – Identify appropriate target agencies to participate in safety data integration programs.
- 2) Conduct Gap Analysis – Compare existing and desired conditions to identify needed improvements.
- 3) Establish Data Governance Process – Institute a formal process to address data system and management issues.
- 4) Develop Data Collection and Integration Plan – Establish a project work plan to guide data integration efforts.
- 5) Identify Training Needs for Data Collection, Integration, and Analysis – Conduct an assessment of knowledge, skills, and abilities of existing personnel in comparison to identified needs for data integration tasks.
- 6) Perform Data Integration – Merge two or more data sources into a combined resource.
- 7) Develop and Deploy the Extract, Transform, and Load Process – Move the integrated data into an analytic tool.
- 8) Conduct Analyses – Implement the desired safety analyses using selected tools and techniques.
- 9) Perform Effectiveness Evaluation – Determine the impact of investments.

III. **Conclusions.** This part provides a summary of the project and the Guide.

Case Studies

In the first phase of the project, the staff researched noteworthy practices by States that already integrate local data into the statewide safety data. Four States—Michigan, Ohio, Tennessee, and Wisconsin—participated by helping to develop detailed case studies of their data integration practices. The case studies can be found in the FHWA Roadway Safety Data Program web page at http://safety.fhwa.dot.gov/rsdp/data_activities_state.aspx. Examples from these four States are used throughout this report as descriptions of methods that States have used in successfully collecting, integrating, and updating local safety data. The examples also describe ways that the States have developed data integration solutions that meet the combined needs for asset management and safety data.

Pilot Studies and Additional Work Plans

FHWA selected four agencies—Arizona DOT, Indiana Local Technical Assistance Program, (LTAP), Navajo DOT, and Rhode Island DOT—for full pilot studies including work plan development, proof of concept data integration implementation, and final report. Summaries of the pilot study reports appear as Appendix A in this report. This final Guide includes references to the pilot studies as sources of examples and lessons learned.

FHWA selected two additional agencies—Fort Belknap Reservation Transportation Department and North Carolina DOT—for work-plan-only implementations in which the agency was provided with a detailed work plan for local integration which they could choose to implement on their own. Summaries of these work plans appear as Appendix B and serve as further examples of how a State, Tribe, or Local/Tribal Technical Assistance Program might structure a safety data integration project.

DATA INTEGRATION CONCEPTS

Safety data integration involves multiple agencies, jurisdictions, software systems, and data types. Ultimately the goal is to provide decision makers with the information they need to be effective and efficient, making the best possible decisions with solid justification based on known facts. This introductory section presents concepts and techniques of data management from an IT and transportation perspective. This section covers:

- Data Program Concepts.
- Metadata.
 - Data Dictionary.
 - Data Flows.
 - Data Access Rights.
 - Data Quality and Standards.
 - Data Retention.
 - User Support.

Data Program Concepts

NCHRP Report Number 666: *Target-Setting Methods and Data Management to Support Performance-Based Resource Allocation by Transportation Agencies* (Section I: Chapter 4, pages I-58) provides definitions of key data program concepts.⁽³⁾ The following appear throughout this document and are crucial to the understanding of agencies' responsibilities over safety and other transportation-related data resources.

Data Management is the development, execution, and oversight of architectures, policies, practices, and procedures to manage the information lifecycle needs of an enterprise in an effective manner as it pertains to data collection, storage, security, data inventory, analysis, quality control, reporting, and visualization.

Data Custodians are the information technology professionals responsible for the maintenance and security of databases, hardware, and software used to support application systems.

Data Owner is a role or group who is empowered to make decisions about how a data entity can be structured, manipulated, or used.

Data Stewards are individuals accountable for the accuracy, integrity, and timeliness of the data as well as for informing users of the appropriate use of data.

Data Stewardship is the formalization of accountability for the management of data resources.

Data Users are the individuals who have authorized access to retrieve data from information systems to conduct business.

In safety data integration efforts these concepts take on another layer of meaning and importance. Data management includes the processes for obtaining specific data elements from existing sources and merging them into a combined resource. The definitions of data custodians, data owners, and data stewards now include layers of ownership—one for the source databases and another for the integrated database. Data users for integrated databases may span a broader audience than the users of the original sources.

Managing data integration adds the following conceptual definitions:

Data Interoperability is the compatibility between multiple data sources (usually with different owners) enabling sharing of database contents to create a single combined resource. Appendix D contains an examples of a Memorandum of Understanding (MOU), and a data sharing agreement.

Data Linkage is the merging of two or more data sources containing related, but dissimilar, information based on linking variable(s) present in each of the data sources.

In 2012, the FHWA Office of Safety assessed the status of the safety data systems in all fifty States plus Washington, D.C., and Puerto Rico as part of the Roadway Safety Data Program (RSDP). The *State Safety Data Capabilities Assessment*, used a Capability Maturity Model to score each State's safety data program on multiple dimensions, including data management and

interoperability—the two important areas of concern for this Informational Guide.⁽⁴⁾ The Capability Maturity Model is a five-point scale for each item rated. The five levels are:

1. **Initial:** The organization does not possess a stable implementation environment and the safety data process is “ad hoc” with no interconnection within the organization. There is no plan for interoperability or expandability.
2. **Repeatable:** The results of previous projects and the demands of the current project drive activities and actions. Individual managers decide what to do on a case-by-case basis during individual projects.
3. **Defined:** The organization documents the process rather than on a per-project basis. The organization’s standards tie to an adopted strategy and this guidance determines project outcomes.
4. **Managed:** Process management initializes and supervises individual projects. Through performance management, processes are predictable and the organization is able to develop rules and conditions regarding the quality of the products and processes.
5. **Optimizing:** The whole organization focuses on continuous improvement. The organization possesses the means to detect weaknesses and to strengthen areas of concern proactively.

At the time of the capabilities assessment, fewer than half of US States were managing their safety data systems with a formal process applied across all programs and projects. The implication is that State DOTs did not regularly apply data governance policies and procedures to address system maintenance or design needs. One concern arising from this is that States might defer valuable data integration efforts because they cannot formally manage their existing data systems (and necessary upgrades) well enough to ensure success from the outset. Another concern is that States may move forward with data integration efforts, but do so without the benefit of the formal IT processes that would result in a product that is cost-effective and meets the needs of the broadest possible set of users. The danger is that without formal procedures, the result will be a series of stand-alone integration efforts purpose-built to meet a narrow set of needs. The integration works, but without taking advantage of potential benefits of a broader look at available resources and the needs of a larger community of potential users. This does not imply failure of the individual integration projects. It does imply, however, that target agencies may have missed opportunities to build a more comprehensive solution at only marginally increased costs and project development times.

METADATA

Modern data management and data governance processes define a set of *metadata* (literally “data about data”). Metadata describes the data in a system: what data elements are in the database, how each element is collected and processed, the controls over data access and data quality, and data retention policies. The following sections provide a brief summary of the types of metadata that are most important for safety data integration.

Data Dictionary

A data dictionary is a list of all data elements present in a database. It provides the data element names, a descriptive title, the data type, the range of values each can take on, the data table in which the data element resides, and applicable validation rules. The Model Inventory of Roadway Elements (MIRE) provides an excellent example of a data dictionary—in this case defining data elements in an ideal roadway inventory system.⁽⁵⁾ FHWA developed the MIRE as a recommended listing of roadway and traffic elements critical to safety management. The MIRE data dictionary serves as a guide to help transportation agencies improve their roadway and traffic data inventories. MIRE was developed to enhance a State's ability to use advanced safety analyses such as those presented in the *Highway Safety Manual*.⁽⁶⁾

FHWA established a subset of the MIRE as part of the HSIP Final Rule changes to 23 CFR Part 924, effective April 14, 2016. This subset is referred to as the fundamental data elements (FDEs).⁽⁷⁾ States are to incorporate specific quantifiable and measurable anticipated improvements for collection of MIRE FDEs into their State Traffic Records Strategic Plan update by July 1, 2017. States will ultimately have access to the FDEs on all public roads by September 30, 2026.

FHWA's Guidance on State Safety Data Systems references the MIRE FDEs categorized by roadway functional classification and surface type and include three lists, one each for non-local paved roads, local paved roads, and unpaved roads.⁽²⁾ They are further refined into subcategories of data elements for road segments, intersections and interchanges for non-local paved roads.

MIRE FDEs for Non-Local Paved Roads

MIRE FDE requirements for non-local paved roads are presented in Table 2. These FDEs provide information on roadway segments, intersections, and ramps and are based upon data needed to conduct a sufficient review of a highway network using existing safety analysis methods. It should be noted that additional data to what is required in the FDEs would be needed to diagnose conditions at individual sites, to select countermeasures, and to prioritize

projects. States should consider all roadway and traffic elements needed to satisfy the full range of safety analyses they perform.

Table 2. MIRE FDEs for non-local (based on functional classification) paved roads.

MIRE Name (MIRE Number)^(2,5)	
Roadway Segment	Intersection
Segment Identifier (12)	Unique Junction Identifier (120)
Route Number (8)*	Location Identifier for Road 1 Crossing Point (122)
Route/street Name (9)*	Location Identifier for Road 2 Crossing Point (123)
Federal Aid/ Route Type (21)*	Intersection/Junction Geometry (126)
Rural/Urban Designation (20)*	Intersection/Junction Traffic Control (131)
Surface Type (23)*	AADT (79) [for Each Intersecting Road]
Begin Point Segment Descriptor (10)*	AADT Year (80) [for Each Intersecting Road]
End Point Segment Descriptor (11)*	Unique Approach Identifier (139)
Segment Length (13)*	
Direction of Inventory (18)	Interchange/Ramp
Functional Class (19)*	Unique Interchange Identifier (178)
Median Type (54)	Location Identifier for Roadway at Beginning Ramp Terminal (197)
Access Control (22)*	Location Identifier for Roadway at Ending Ramp Terminal (201)
One/Two-Way Operations (91)*	Ramp Length (187)
Number of Through Lanes (31)*	Roadway Type at Beginning Ramp Terminal (195)
Average Annual Daily Traffic (79)*	Roadway Type at Ending Ramp Terminal (199)
AADT Year (80)*	Interchange Type (182)
Type of Governmental Ownership (4)*	Ramp AADT (191)*
	Year of Ramp AADT (192)*
	Functional Class (19)*
	Type of Governmental Ownership (4)*

*Highway Performance Monitoring System (HPMS) full extent elements are required on all Federal-aid highways and ramps located within grade-separated interchanges, i.e., National Highway System (NHS) and all functional systems excluding local roads and rural minor collectors.

MIRE FDEs for Local Paved Roads

For paved public roads with a functional classification of local, States may collect a reduced set of FDEs, because these public roads routinely have no more than two through lanes and partial or no access control. This reduced set of FDEs is limited to the category of roadway segment elements. States shall collect, at a minimum, the reduced set of FDEs listed in Table 3 on local paved public roads. There is no requirement for collecting intersection data on local roads. Network screening for these low traffic volume roads can be performed using system-wide or corridor level analyses that combine (but do not distinguish) roadway segment, intersection, and ramp crashes. Corridor-level network screening would identify “intersection” hot spots, as well, and then an agency could collect specific roadway data relative to that location as needed.

Table 3. MIRE FDEs for local paved roads.

MIRE Name (MIRE Number)^(2,5)
Roadway Segment
Segment Identifier (12)
Functional Class (19)*
Surface Type (23)*
Type of Governmental Ownership (4)*
Number of Through Lanes (31)*
Average Annual Daily Traffic (79)*
Begin Point Segment Descriptor (10)*
End Point Segment Descriptor (11)*
Rural/Urban Designation (20)*

*HPMS full extent elements are required on all Federal-aid highways and ramps located within grade-separated interchanges, i.e., NHS and all functional systems excluding local roads and rural minor collectors.

MIRE FDEs for Unpaved Roads

A limited set of data elements must be collected on unpaved public roads in order to use HSIP funds on these roads. Table 4 provides a listing of the FDEs to be collected on unpaved roads regardless of their functional classification. Three of these elements support the Highway Performance Monitoring System (HPMS) all-public-roads LRS requirement (segment identifier, begin point segment descriptor, and end point segment descriptor) and enable States to locate all crashes on all public roads with their LRS. These elements, combined with the remaining two elements (functional class and type of government ownership), support States in fulfilling their Strategic Highway Safety Plan (SHSP) and HSIP requirements.

Table 4. MIRE FDEs for unpaved roads.

Name (MIRE Number)^(2,5)
Roadway Segment
Segment Identifier (12)
Functional Class (19)*
Type of Governmental Ownership (4)*
Begin Point Segment Descriptor (10)*
End Point Segment Descriptor (11)*

*HPMS full extent elements are required on all Federal-aid highways and ramps located within grade-separated interchanges, i.e., NHS and all functional systems excluding local roads and rural minor collectors.

Data Flows

A data flow is a text or graphic description of data management steps from initial collection, through to final storage and sharing. Figure 4 shows a generalized diagram for a statewide crash data system. The example in Figure 4 shows the flow from crash event through to creation of the statewide crash database and post-processing. Similar data flow diagrams should exist for roadway inventory and traffic volume data. Each of the data flows should show the steps in the process where the integration takes place. In the example in Figure 4, crash reports from State and local law enforcement merge to create the single statewide crash database in the pending and final crash databases.

A data flow is important for data integration efforts because it shows the integrator the complete record creation process. If the integration project relies on accurate location coding in an LRS or GIS (Steps 10 and 11 in this example), the integrator would need to wait for the DOT to complete its post-processing. State data custodians can monitor the duration of each processing step. An integrator would use this information to gauge how long to wait after a crash event before attempting to integrate crash data with the other safety data sources.

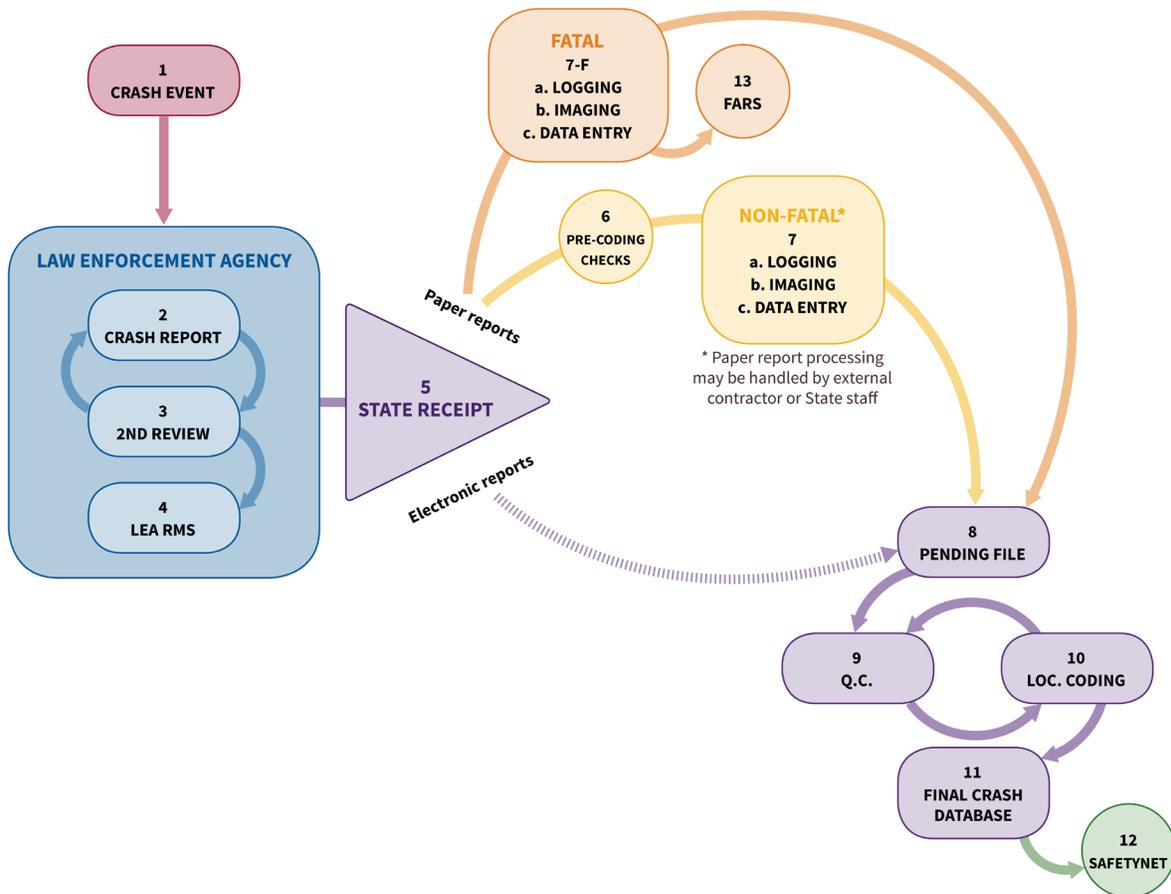


Figure 4. Chart. Generalized crash data flow diagram.

(Source: VHB)

Data Access Rights

Most government databases are access controlled, meaning that only authorized users may view, add, or edit records. A public-use mapping utility on the internet, for example, allows users to select and filter data, but there is usually no provision for them to add or alter records (e.g., they can't rename a street, add traffic counts, or erase a crash). Data custodians in the responsible agencies manage official records such as crash data, traffic volumes, roadway inventory, and asset management data. Their role is to monitor data integrity and maintain the systems (hardware, software, and databases). Data stewards manage day-to-day operations of the government's processes including assessing quality, approving user access requests, and assigning resources. The data custodian and data stewards work together to manage database access. In a typical system, designers define multiple levels of authority from top-level system

administration with access to all system functions down to general users who may be able to view only certain data elements. The data steward approves access requests made by the users and the data custodian implements the user access by creating the appropriate user credentials (ID, password, and user access rights).

Access rights are important in data integration because the merger of data between two or more systems creates a new resource that a custodian and steward must also manage. Data integration projects must define stewardship and custodial responsibilities as part of the data management plan. The data management plan addresses access rights based on input from the custodians and stewards of the contributing databases.

Creating merged datasets may require further access controls in addition to any in the contributing systems. In creating a single statewide crash database, the State data custodian interacts with all of the law enforcement agencies that also have custodial and stewardship responsibility over their own data. Similarly, a statewide roadway inventory file typically has a State custodian as well as local agency custodians. For example, in the Wisconsin Information System for Local Roads (WISLR) system, individual local agencies own the data, maintain it, and provide updates to the State upon request and to meet annual reporting requirements. The same is true in Ohio, where the local agencies are the final authority for non-State roads in their jurisdiction and provide annual updates to the State. The State pays for the system and controls user access. Other States take a different approach. In Tennessee, the State owns the data and the system. In Michigan, the local agencies own the data and data reside locally at each agency with no centralized integration except on an as-needed basis. The differences among these systems give rise to differences in data access management. In Michigan, no centralized, integrated file is maintained so access to that resource is only available to those who create it to meet a specific need. In Tennessee, the State DOT manages all of the data. In Ohio and Wisconsin, State and local data managers must confer on access.

The situation is further complicated when multiple data sources are involved. Integrating crash, traffic volume, and roadway inventory data may increase the number of custodial agencies and data stewards as well as the need for coordination to resolve access rights to any statewide-integrated system. Data sharing agreements and MOUs address concerns over who owns the data and who may access it. Appendix D provides examples of data sharing agreements and MOUs. In some States there are additional concerns over the threat of legal claims against the State or local governments in disclosing detailed highway safety data. For this reason, some States restrict access to linked crash and roadway data to approved users who must first sign a release. In other situations, keeping data secured may be incompatible with State open-records laws.

Even in States with open access policies, there must be controls over who can add, edit, or delete records. Different systems address this in different ways. In Michigan, the Roadsoft

system integrates local asset, crash, traffic, and roadway inventory data. Each local agency maintains its own copy of the software and its own data locally. Any sharing among agencies requires local action and approval. Local users are responsible for all data entry and management in their local system. In the WISLR system, there is one central integrated database available through the internet and can be used for statewide safety screening. Wisconsin DOT (WisDOT) can access all of the data, and creates accounts for local users upon request. Local agencies' approved users can access their own data from neighboring areas. WisDOT staff can edit all of the data whereas local users must have special approval and training before they can edit some portions of the data (e.g., line work). Though very different, both Roadsoft and WISLR work well and have good support among their users.

Data Quality and Standards

Data quality management is a central feature of a formal data governance process. The data dictionary documents the data quality standards for each element in the database. Setting standards imposes requirements on the data collectors as well as the data stewards and custodians of each agency that submits data to the centralized system. For example, in WISLR local agencies certify their centerline mileage annually. State law establishes the requirements for this certification, and WisDOT is ultimately responsible for its accuracy and use for funding decisions. WisDOT conducts a one percent sample validation as a data accuracy check.

Data quality is vitally important in data integration efforts. Integrating target agencies' data into a new resource requires that all data submissions adhere to the same standards. This implies that the quality attributes of each record are known. This may be considered another type of metadata—an indicator in the record of the level of confidence in the data's accuracy. For example, the WISLR system includes a confidence indication for pavement condition based on the number of years since the last actual field data collected for each site. As the data age, the confidence rating drops.

When merging multiple datasets, the final integrated database is generally of no better quality than the lowest quality level achieved in any of the contributing databases. This is certainly true for completeness—the final database will have incomplete records for any case that is missing data in *any* of the source databases. An integrated database of crash, roadway, and traffic data will only have as many complete records as there are jointly locatable records in *all three* databases. Accuracy is similarly limited by the lowest quality source file—when inaccurate data appear in any one source, the resulting integrated database is also inaccurate.

Each of the four pilot studies produced lessons learned related to data quality. All of the integration efforts used GIS and spatial location to support data integration. The pilot studies each sought to integrate data from multiple agencies into a single roadway inventory file, a single crash file, and/or a single traffic volume file. The data quality issues identified in the pilots

involved spatial and non-spatial aspects of the data. The spatial data concerns centered mostly on the ability to match locations across the multiple files and to align maps from multiple sources so that each location appears only once in the spatial database and there are no discontinuities introduced when roads cross jurisdiction (and GIS source file) boundaries. The spatial matching techniques are well understood, are common across GIS implementations, and can be anticipated as part of the project timeline and budget. The other types of data quality problems are often unique to a single situation and may require unique and sometimes unanticipated effort before a solution is found. For example, in one pilot study the State had segmented roadways in a way that allowed for overlap among adjacent sections of roadway. This segmentation method supported the analytic processes the State had been using prior to the pilot; however, the State wanted to implement AASHTOWare Safety Analyst™. That software requires unambiguous segmentation and, upon first attempt, the software rejected all but a small portion of the statewide data. Programming resolved most of the issues without much additional labor; however, there were some records that required manual correction by DOT staff. The pilot ultimately generated a sufficiently large integrated database for testing, but did not include all of the data originally planned for the pilot because of unresolved segmentation and completeness issues.

Data Retention

Data integration efforts in highway safety require information on the vintage of each record—the date collected and the normal cycle time between rounds of data collection (i.e., the data refresh frequency). This matters a great deal when integrating State and local data, especially traffic count and roadway data, because the value of data for decision-making decreases as it ages. In addition, custodians generally create an annual file as a snapshot of conditions to establish a permanent record for items such as roadway alignments, asset locations and conditions, and to support trend analysis.

Data retention refers to the number of years of data available for analysis. States typically set a standard number of years beyond which data are either archived (and thus require special effort to retrieve for use) or become inaccessible because the records have been purged. Local agencies' records retention policies are typically less well defined than are the State's policies and procedures. In an integrated dataset, the data custodians and stewards must agree on a record retention policy. In creating the integrated database, the data custodian has the responsibility to set a data retention policy and, together with the data steward, should ensure compliance. Formal data governance addresses these data retention issues.

The system metadata and the records in the database should include two pieces of information: the date of collection and the annual snapshot to which the data apply. This is used in various ways. For example, in Wisconsin's WISLR, pavement condition ratings feed into the budget

analysis tool. That tool provides a confidence indicator along with each result and that indicator is decremented as the pavement condition ratings age (condition ratings older than two years get a low confidence level). WISLR and the Michigan Roadsoft system both include historic alignment information so that event, asset, and feature location information remains accurate for safety analysis. For example, in a before and after analyses for safety effectiveness evaluation, analysts must know the conditions that applied in the “before” condition, when and what changes were made, in order to assess the impact of the changes.

Multiple Users Benefit from Integration

The data sources for safety data integration serve multiple purposes, and may exist primarily for some purpose other than supporting safety analysis. An important example is the annual snapshot of roadway data. This information includes the length and location of each roadway segment. States use this information to calculate roadway mileage in local agencies which, in turn, is used to calculate the amount of funding each local agency receives. The same resource is useful in safety data integration because it contains descriptive information about each location. Multi-year data will show important changes to the roadway network, including, for example, alignment changes that may affect safety. By integrating crash and roadway data, analysts can verify that the crashes are located on the year-current roadway alignments and generate a true picture of how the alignment changes affected safety. Thus, a data resource that originally served a financial purpose is critically important for safety analysis.

Data quality benefits from integration. Data integration tends to uncover errors and inconsistencies in the source data. One example is the overlapping segmentation discovered in one State’s roadway data. The State ran the risk of double-counting roadway miles based on the segmentation that had been in place prior to the data integration pilot study. The data integration project made it clear that overlaps needed to be resolved—because the integration simply failed when locations could not be uniquely matched among the data sources, and the analysis software tool rejected the records. As a result of the data integration effort, the State has resolved the overlaps and now has a much more reliable basis for calculating mileage for its funding formula. Each of the pilot studies identified data quality issues that needed to be resolved before data integration could succeed.

Data integration results in a new resource capable of supporting analyses that are not possible when the individual data sources stand alone. Integrating a single data type among all agencies supports cross-jurisdiction comparisons. Integrating the safety data sources supports analyses of the contributions to safety of people, the roadway environment, and traffic simultaneously. The result of having integrated data is the ability to answer more questions in greater detail, while considering the interactions among the various factors that contribute to crashes.

User Support

Because integrated safety data are a robust data resource, the user community for integrated safety data systems is quite broad—including target agencies’ engineers, planners, and law enforcement, but also often extending to other stakeholders such as State epidemiologists, safety advocacy groups, the media, and even members of the public. This user community may be much larger and more diverse than the user communities for the individual component data sources. The data custodians, data stewards, and data managers generally share responsibilities for ensuring that support is available to meet users’ needs. That support comes from the State data custodian or from contractors who maintain the system on behalf of the sponsoring agency. For example, the Michigan Technological University Center for Technology and Training provides user support for 400+ local agencies collecting and managing location, asset condition, and safety data. They allocate approximately three full time equivalent staff to training, technical assistance, and software user support.

SUMMARY

This section provided information and examples on data management concepts and their application to target agencies’ data integration. Concepts included integrating similar datasets, for example crash data, into a single statewide dataset, and integrating related but dissimilar datasets, for example roadway inventory and crash data, into a merged data resource.

This information is placed at the beginning of the report because the remaining sections require a basic understanding of data management, data governance, and data stewardship, as well as the roles of data managers, custodians, stewards, owners, collectors, and users. The next nine chapters go into greater detail on the steps of data integration.

STEP I – LAY THE FOUNDATION

OBJECTIVES

Laying the foundation for effective partnerships and cooperation is the first step in the safety data integration process (shown in Figure 5). This step brings together the people who will manage the data integration project and establishes the roles and responsibilities for the project. Target agencies work as partners to lead safety data integration programs. The necessary partnerships are developed through purposeful action and are expected to evolve over the life of a data integration project.

There are a variety of partners who can participate in safety data integration:

- State safety engineers, planners, data managers, data analysts, and other business units of the DOT.** The personnel responsible for safety plans such as the SHSP and for coordinating State and regional Transportation Improvement Programs have responsibility for decision-making related to safety performance, performance measurement, and analysis. The data managers are responsible for creating and maintaining statewide datasets with information on all public roads. Data analysts need complete, accurate, and timely data in order to provide services to decision-makers throughout the State. Other business units may contribute to the integration effort including IT, GIS, Asset Management, Planning, and Design.
- Tribal safety engineers, planners, data managers, infrastructure managers, and data analysts.** Tribes are listed separately from both the State and local agencies because they share characteristics and needs of both groups. Tribes, as sovereign

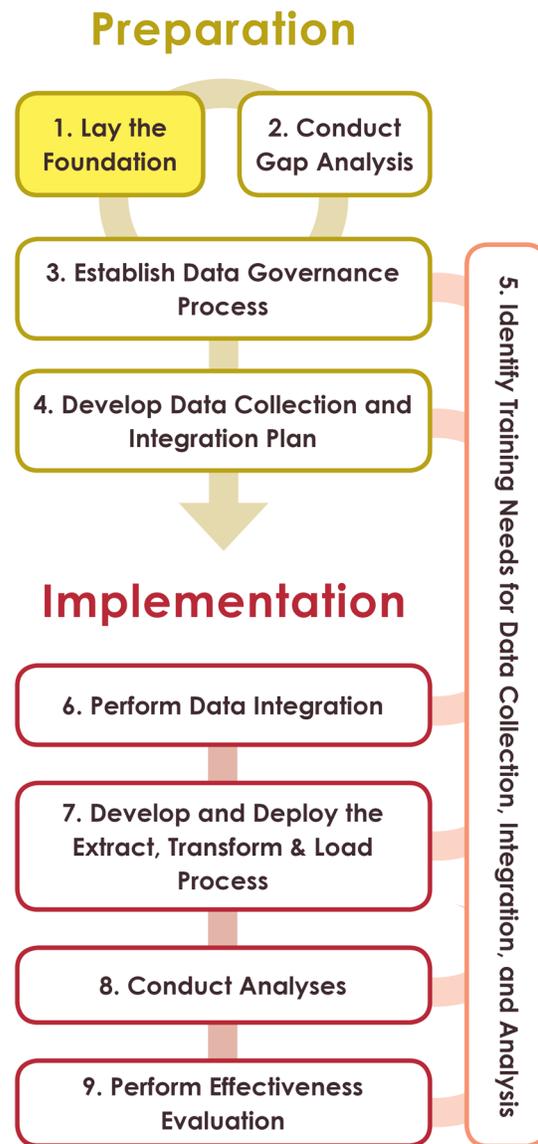


Figure 5. Chart. Step I: Lay the Foundation.

nations, control jurisdictions and set practices for all agencies to follow within their domains. They collect data from multiple sources and may integrate the data into a centralized database to support decision-making. Tribal areas also collect and provide data that States need.

- **City, county, and regional safety practitioners and infrastructure managers.** The local agencies (cities and counties), MPOs, and RPCs all need detailed data describing transportation safety in their jurisdictions. In their role as asset managers, they also benefit from having local data systems that can handle all spatial data in a single suite of software tools. This translates to a need for GIS and spatial data tools. These tools can assist local agencies in meeting their obligations to manage assets, plan, and participate in safety management with their State partners.
- **LTAPs/TTAPs.** The LTAP and TTAP centers throughout the US provide a variety of services to their local agency and Tribal government clients. In the safety focus area, LTAPs/TTAPs provide training and technical assistance in roadway safety. They also work closely with Tribal and local agencies to help manage roadway assets. This includes data collection and analysis that can be useful in safety data integration as well. Additionally, LTAP/TTAP professional development efforts can help improve the effectiveness and efficiency of safety funding. Because of those varied services, the LTAPs and TTAPs also have multiple roles to play in successful State and local safety data integration. They can represent Tribal and local agencies' interests in the integration effort, help target agencies identify gaps, and serve as a training resource.
- **Federal lands Land Management Agencies and other federal partners.** The National Parks Service and Bureau of Land Management directly manage roadways, and may also report crashes and traffic volumes on federal lands. States and the federal lands management agencies (FMLAs) within the States have compatible needs for data and can share information. The Bureau of Indian Affairs collects data on behalf of many Tribal governments and can supply information to States. These federal partners should be included in the inter-agency coordination as part of the safety data integration project. In the Arizona and Navajo DOT pilot projects, examples of data gaps related to federally managed roadways and sources of data from multiple Tribal governments were addressed through cooperation with these federal partners.
- **Other stakeholders.** States, Tribes, and local agencies work with a variety of stakeholders, advocacy groups, consultants, and researchers with a safety focus. These groups need to understand the sources of data and proper use of data for safety analysis and decision-making. In some cases, they are advocates for improved or expanded datasets addressing particular issues (e.g., motorcycle, pedestrian, and bicycle safety). In other cases, this group includes practitioners who conduct analyses and make

recommendations to State, Tribal, or local agencies, and thus must understand the data requirements and data integration methods.

The leaders of a data integration effort need to determine which partners should be part of the process. The list of partners above is a starting point and may expand as participants identify key stakeholders who are knowledgeable and can contribute to the data integration process.

This Informational Guide is purposefully not prescriptive about which agency or agencies must lead a safety data integration effort. The pilot studies each differed in the type of lead agency. Two were State DOTs (Arizona and Rhode Island) with local agency partners. One was a Tribal DOT (Navajo Nation) with State and local partners. Indiana's pilot project was led by the LTAP with State and county partners. This mirrors the diversity of lead agencies in the case studies as well. The primary point of this chapter is that none of these projects would exist or succeed without partnerships. As we will see in the next chapter, the data has many sources and there are many users from across all types of partner.

CHALLENGES

The pilot studies demonstrated challenges that closely parallel those of other cooperative efforts such as the Traffic Records Coordinating Committee (TRCC) and various State and local planning efforts such as the SHSP and safety performance measures target setting. The joint FHWA and NHTSA report *State Traffic Records Coordinating Committee Noteworthy Practices* (FHWA-SA-15-083) provides relevant advice on how to develop effective multi-agency coordination.⁽⁸⁾ The partners do well by articulating clear goals, establishing formal data sharing agreements and MOUs, and by fostering productive communication by enforcing rules for respectful dialogue during group meetings.

The FHWA Data Integration Primer describes the following challenges when building partnerships.⁽¹⁾

User groups within an agency may have developed databases on their own, sometimes independently from information systems staff, that are highly responsive to the users' particular needs. It is natural that owners of these functioning standalone units might be skeptical that a new system with integrated data would support their needs as effectively.

Other proprietary interests may come into play. For example, division staff may not want the data they collect and track to be at all times transparently visible to headquarters staff without the opportunity to address the nuances of what the data appear to show. Owners or users may fear that higher ups without appreciation of the peculiarities of a given method of operation will gain more control over how data is collected and accessed organization-wide.

In some agencies, the level of personnel, consultants, and financial support emanating from the highest echelons of management may be insufficient to dispel these fears and gain cooperation. Top management must be fully invested in the project. Otherwise, the likelihood is smaller that the strategic data integration plan and the resources associated with it will be approved. The additional support required to engage and convey to everyone in the agency the need for and benefits of data integration is unlikely to flow from leaders who lack awareness of or commitment to the benefits of data integration.

In short, agency executives' active support and participation is key to successful safety data integration. They should be informed on the plans being developed, the benefits of safety data integration, gaps and their resolution. The agency executives also need to commit to the formal processes involved in data integration, including data governance, and participate by committing resources and supplying direction as needed.

RECOMMENDED SOLUTIONS

In the pilot studies, external consultants facilitated large group meetings to which all the partner agencies were invited. One kickoff meeting included more than 60 attendees from a dozen agencies. The purpose of the meetings was to arrive at a shared vision for the project and to secure buy-in from all those who would supply or manage the data. For the pilots, it was crucial to have decision-makers who could commit their agencies' resources either present at the meetings or briefed by the staff who did attend. All of the pilot studies required some commitment of time and effort from the partners. After the project was underway, frequent communication among partners was critical. Because the pilots were used as proof of concept for the vision of more comprehensive data integration, the agencies had to learn how to work together to resolve data problems at the source. Ultimately, the pilot studies could reach a successful conclusion with integration of a sample data set. State, Tribal, and local agencies must work together to complete the data integration project.

The FHWA Data Integration Primer offers the following recommendations: ⁽¹⁾

- Any large-scale data integration project, regardless of model, demands that executive management be fully on board. Without the support of senior staff, the initiative is likely to fail.
- Informing and involving the diversity of players during the crucial requirements analysis stage, and then in each subsequent phase and step, is probably the single most effective way to gain buy-in, trust, and cooperation. Collecting and addressing each user's concerns may be a daunting proposition, particularly for professionals who prefer to "cut to the chase." However, without a personal stake in the process and a sense of ownership of the final product, the long-term viability of this major investment is likely to be

compromised by users who feel that change has been forced upon them rather than designed to advance their interests.

- Incremental education, another benefit of stakeholder involvement, is easier to impart than after-the-fact training, particularly since it addresses both the capabilities and limitations of the system, helping to calibrate appropriate expectations along the way.
- Since so much of the project's success is dependent upon understanding and conveying both human and technical issues, skilled communicators are a logical component of any data integration team. Whether staff or consultants, professional communications personnel are most effective as core participants, rather than occasional or outside contributors. They are trained to recognize and ameliorate gaps in understanding and motivation. Their skills also help maximize the conditions for cooperation and enthusiastic adoption. In many transportation agencies, public information personnel actually focus a significant amount of their time and budget on internal audiences rather than external customers. This makes them well attuned to the operational realities of a variety of internal stakeholders.

LESSONS LEARNED

The lessons learned about partnerships include the following:

- **There are relevant agencies who are easy to miss when extending invitations and soliciting partners.** In one pilot, the housing authority had the most prior experience integrating spatial data. They were a valuable voice in planning the project and this example encouraged the other teams to look beyond typical DOT issues to find partners with something to add.
- **Not all partners are equal to the tasks at the beginning.** One pilot project included promises of data from several agencies including large RPCs, Tribal, and federal partners. As the project unfolded, the project team and the State concluded that only one RPC could supply the data required for the test phase of the study. The other partners remained interested in the outcome, but the time and resources needed to make use of their data were too great. In a full data integration project, those problems would eventually be addressed.

It may also be beneficial to establish a data sharing agreement, MOU, or other foundational document that describes the powers and duties of each stakeholder involved with data integration. For example, Washington State's Traffic Records Coordinating Committee updates its MOU annually and publishes it online. The MOU for Washington State lists the nine participating agencies, explains the importance of

their shared responsibilities for traffic records data, describes the purpose of the committee, sets the committee's mission, goals and operational authority, and establishes a two-tiered organizational structure (executive leadership and working level TRCCs). The responsibilities of the executive and technical TRCC committees are listed individually. Member agencies are listed along with the process for appointing a representative to the TRCC. The MOU includes a formal process for replacing members who fail to attend meetings on a regular basis, and for adding members from newly identified stakeholder agencies. The director of each participating agency signs the updated MOU each year. Washington State uses the MOU update process to refresh the agency directors' understanding of the importance of the committee's work and to reinforce the ideal of inter-agency cooperation. A copy of the MOU is provided in Appendix D.

SUMMARY

Step I: Lay the Foundation addresses the need for pre-planning, securing buy-in and cooperation, and establishing partnerships among various agencies. The important items in this step are:

- Secure the commitment of executive management.** The executive staff have multiple roles to play. They identify funding, approve projects and resource allocations, and communicate with other agency's executives to promote the data integration program and goals. Data integration partners should seek formal approval of each agencies' leadership team to proceed with the data integration project. As the project progresses further contact with the executive leadership may be required to gain access to resources, signal agency commitment to data business plans, and approve data sharing agreements.
- Forge partnerships.** This is a major focus of Step I—identifying and securing participation from a variety of stakeholder agencies. It goes beyond participation to develop interagency cooperation and sharing through clear lines of authority, effective meetings, formal agreements, and building mutual trust. This is an ongoing process as stakeholders change and new needs arise. A deliberative body (such as a TRCC) can help to foster effective partnerships for data integration.

CHECKLIST

- Secure the commitment of executive management.
- Forge partnerships.
- Establish needed MOUs and Data Sharing Agreements to establish expectations.
- Establish communication processes to inform and involve all stakeholders.

The consistent and direct involvement of local agencies, and their representatives (LTAPs, MPOs, RPCs) is crucial to success. The local agencies have to see what they get out of participating in data integration—how their decisions are supported and spending is more effective—so that they join the process as eager contributors.

- **Establish needed MOUs and data sharing agreements to establish expectations.** Formal agreements can codify the understandings and relationships built through the committee structure. MOUs can demonstrate executive commitment, resource allocations, roles, and responsibilities. Data sharing agreements can set specific expectations for timeliness, access, and release of potentially sensitive data. Examples are presented in Appendix D.
- **Establish communication processes to inform and involve all stakeholders.** Establishing efficient and effective communication among all the partnering agencies will help to reinforce the commitments and retain goodwill during the course of the project. Group meetings (see also Step 3: Data Governance), project tracking, and frequent communication are needed so that partners feel engaged and know that there is activity leading to a specific goal.

STEP 2 – CONDUCT GAP ANALYSIS

OBJECTIVES

A gap analysis compares the current conditions to the desired conditions, and points to needed improvements. Conducting a gap analysis is the second step in the safety data integration process, as illustrated in Figure 6. For safety data integration, a gap analysis is a formal investigation comparing the available data sources to what is needed. Needs are defined by required data to support the intended type of safety analyses and the desired safety analytic tools.

Each of the pilot studies included a formal data gap analysis. These began with a survey of users' needs and available data sources. The surveys were distributed widely; however, the nature of the pilot study also dictated how much information was sought from partners. In Rhode Island, for example, the majority of the effort was aimed at State-led collection of MIRE data using a field data collection contractor. The contract for that effort was already in place before the pilot started and targeted the collection of the MIRE dataset. In contrast, the project in Indiana required a much broader-based understanding of the local agencies' capabilities and needs. There, the LTAP distributed a GIS gap analysis to all 92 counties and 3 State agencies. The gap analysis results told the technical assistance team what data were available, what the partnering agencies' analytic needs were, and who among the partners required training.

CHALLENGES

The gap analysis can also identify barriers to data integration. Thus it serves not only to show where needed data are missing, but what other needs must be met. Based on the pilot projects,

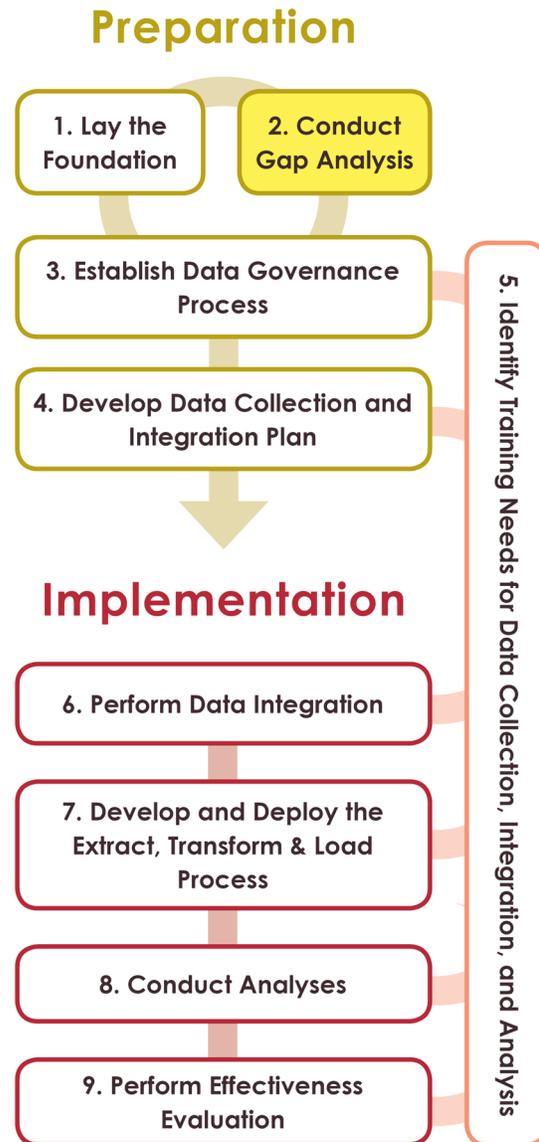


Figure 6. Chart. Step 2: Conduct Gap Analysis

those other needs may include training and new staffing, new technology or software tools, and new processes. The four case studies researched in this project (Michigan, Ohio, Tennessee, and Wisconsin) provide examples of barriers, including data gaps, and how they were addressed:

- **Funding and development schedule trade-offs:** Two of the case study projects started long ago with different management teams, so the original funding source information is not available. However, in all cases the cost to develop the integrated systems was higher than initially budgeted with the end result that development spread over more years than originally planned. Recent technology changes make it likely that cost information from early adopters such as Michigan, Wisconsin, and Ohio would be less relevant for modern new project launches; however, the maintenance costs for each of the case studies is meaningful for any new efforts. In all four States' data integration projects, local agencies contribute staff time but do not contribute dollars toward system maintenance. Wisconsin and Michigan, both reported that they fund the local data systems as one way to demonstrate the State-level commitment to their local partners and encourage adoption of the systems by local agencies.
- **Partnerships among agencies:** The data gaps analysis relies on inputs from data users, collectors, and managers. Data gaps may be overlooked if sufficient input from partners is not received. The gap analysis step and the forge partnerships steps are entwined. As users' needs are defined, the data providers must be polled to see if they have the required data. That is an iterative process that will go on even as later integration steps are being implemented.
- **Data gaps analysis as part of a data governance process:** One of the activities of a data governance group is to analyze data needs and decide how to meet those needs. Thus, the data gaps step and the data governance step are also inextricably tied together. In fact, the first three steps in data integration (lay the foundation, conduct a gap analysis, and establish a data governance process) form a three-part cycle that continues throughout the life of the data integration project. Without the continuous cycle, the data governance group may not be equipped or informed to adequately react to data integration issues.
- **System life-cycle management/sustainability:** Systems require maintenance and user support. This is an issue for gap analysis because it is far easier to examine how well a stable, multi-agency system meets a set of defined needs rather than the alternative of having to look into the capabilities of a large number of stand-alone systems that may change from year to year.

RECOMMENDED SOLUTIONS

The following describe solutions found in the Case Studies:

- **Funding and development schedule trade-offs:** There is no obvious single best way to fund a data integration effort among multiple partners; however, the partners are advised to consider more than just start-up costs and to plan the project roll-out schedule to deliver the most needed capabilities first so that the user base grows quickly. As described in the Overview section of this Informational Guide, States are required to create an all-public-roads base map and LRS. The next priority is to use those spatial resources to map crashes to the corresponding location information. For safety data integration, the same LRS and base map should be used for all safety data resources.

Additionally, building in some level of contingency funding may be ideal. For example, in Louisiana a multi-year, all-public-roads data collection contract includes a clause to hold 20 percent of the funding as a contingency against last minute changes and the need to collect additional data on some segments.

- **Partnerships among agencies:** The partnering agencies need to recognize that the gap analysis will likely identify even more potential partners. The likelihood is that as long as there is an integration effort, there will be new users and evolving user's needs. The partners will need to revisit the gap analysis periodically and involve new partners as they are identified.

Local agency participation is key. They need to get some benefit from the partnership. The benefits may include better data, more efficient reporting, improved tools for managing their assets, and analytic support. Training and user support are important factors in local agencies' acceptance and use of the tools and analyses.

- **Data gaps analysis is part of a data governance process:** The data governance process is led by a data governance group drawn from the same list of partners described in Step 1. As the list of partners changes and expands, so do the data needs and thus the data gaps analysis will need updating. All of this must be coordinated with the data governance group who have responsibility for setting data standards against which the gaps are defined.
- **System life-cycle management/sustainability:** Maintenance and user support for data systems require a reliable funding source over many years to meet the needs of users and to address future demands.

LESSONS LEARNED

The pilot studies provided lessons learned about data gaps:

- **Every integration project should include a data gaps analysis.** In some situations, it is tempting to just “do what needs to be done.” Rhode Island, for example, could have completed their entire data integration project without consulting the many business units within the DOT or the local agencies that are the stakeholders (and potential partners) in data integration. There were limited funds for the data collection effort and the State DOT was paying for MIRE data collection on all public roads. Even in that situation, a gap analysis proved important because it identified additional needs that might be addressed at very low cost as part of the data collection plan. In fact, Rhode Island’s data collection effort benefited from lessons learned in an earlier systemwide data collection effort by Utah—an effort that is also written up as a case study on the RSDP noteworthy practices website.
- **Data gaps are part of a larger package of needs.** The data gaps analyses conducted in the pilot studies were focused primarily on GIS capabilities and needs. The results uncovered multiple needs related to data in general, training, and in some cases, the need to adjust the project implementation plan. Giving partners an opportunity to weigh in early, and often, is a tangible benefit of surveying to identify data gaps.

It is also important to identify an end goal for the data integration. This means that the partners should decide together what they hope to get from the effort. The end goal may be a list of desired outcomes including a set of software tools and analyses; a specific level of data timeliness, accuracy, and completeness; and a decision-making process that is efficient and effective.

- **Data gaps are important to other steps in the data integration process.** Partnerships (Step 1) need to evolve as users’ needs change. Data governance (Step 3) needs to address the gaps. Integration implementation planning (Step 4) must include how the gaps will be addressed, how long it will take, and what in what sequence. Training (Step 5) can help to address data gaps by communicating data standards and helping agencies take on the role of a data provider.

SUMMARY

Step 2: Conduct Gap Analysis addresses the need for partners to assess their needs for analytic results in comparison to the data available from all sources, and the skills required to make the best use of the data. The important items in this step are:

- **Establish a team to perform the gap analysis.** The recommended team might be drawn from members of the State TRCC, the SHSP group, or the data governance group (Step 3). The team approach is important because the effort requires multiple perspectives including collectors, managers, and users as well as representatives from the various jurisdictions of target agencies.
- **Perform a survey of user needs and available data sources.** The survey should address common (known) needs for analytic results and specific data resources. It should be vetted by representatives from multiple perspectives to make sure that the survey questions make sense to the varied members of a broad audience. Lastly, consider splitting portions of the survey so that each section is addressed by people with true subject matter expertise and first-hand knowledge of the systems. Appendix E is an example survey used to collect relevant user needs and data sources.
- **Plan to repeat the gap analysis periodically.** Needs change. So do systems. The gap analysis is not a one-time process but one that can be repeated in order to understand how needs and capabilities have changed over time. Updated gap analyses can help with data business planning and promote discussion among the partners.
- **Recognize the interactions between laying the foundation, gap analysis, and data governance efforts.** This Informational Guide describes several interactions among the various steps in the data integration process. The first three steps are closely tied together and really could be treated as part of one larger process. They are broken out in this guide for clarity of explanation. By recognizing that the steps are tied together, the partnering agencies can be efficient in how they set up the various group processes and work together to establish agreements and plans.

CHECKLIST

- Establish a team to perform the gap analysis.
- Perform a survey of user needs and available data sources.
- Plan to repeat the gap analysis periodically.
- Recognize the interactions between forging partnerships, gap analysis, and data governance efforts.

STEP 3 – ESTABLISH DATA GOVERNANCE PROCESS

OBJECTIVES

Data governance gives the target agencies the assurance that their plans to address data gaps are sound and will result in the desired improvements. Establishing a data governance process is the third step in the safety data integration process shown in Figure 7. Data governance is a formal process that addresses data systems and management issues including data definitions, data standards, lines of authority and ownership, access rights and usage, and data lifecycles. Establishing data governance processes requires input from key stakeholders. This includes the agencies providing the data, those managing the data, and those planning to use the data for safety analysis. It also requires involvement by IT professionals familiar with system design and database management. A data governance group may already exist at the State or Tribal level. If that group is able to lead safety data integration data governance, then they may have already met some of the objectives described in this chapter. However, for States and Tribes without a formal data governance process already in place, tasking a select group of subject matter experts and IT professionals to fill this need is important. The data governance group reviews data needs, sets data standards, and establishes data quality measurements.

The FHWA Data Governance Plan describes six strategic goals which in turn are challenges that impact the establishment of good data governance.⁽⁹⁾

- **Leadership:** Identify champions to ensure accountability and to increase the value of data assets. The champions may be the official data owners or custodians; however, it is not uncommon for users to serve as the most vocal champions for a particular dataset.

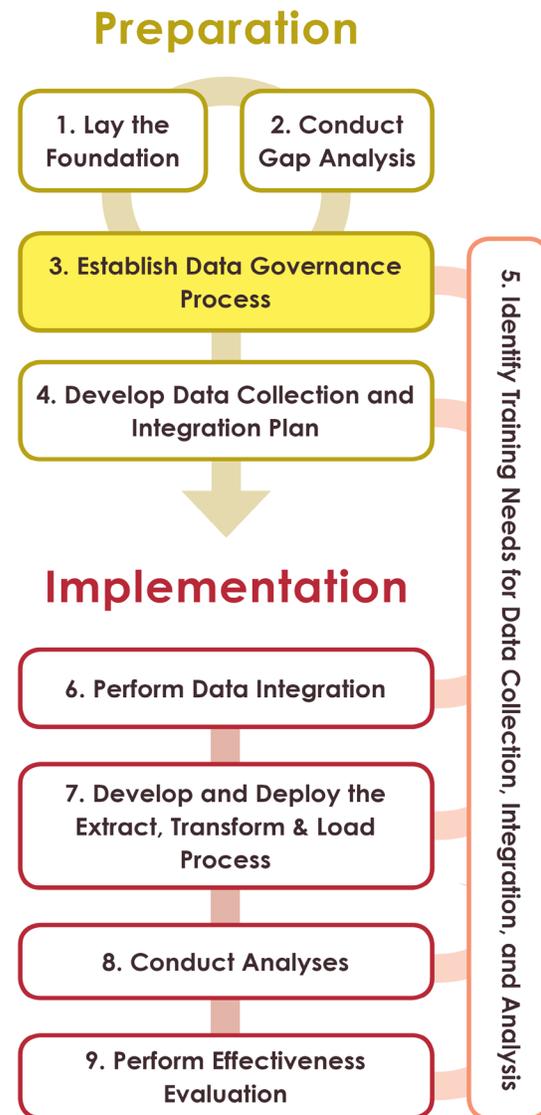


Figure 7. Chart. Step 3: Establish Data Governance Process.

For integrated safety data, analysts and decision-makers make strong data champions because the robust data resources help them do a better job.

- **Quality:** Establish procedures to ensure data are sufficient for the intended uses. This includes setting data standards (definitions, and operational parameters for collection and reporting). It also includes setting expectations for data quality attributes such as timeliness, accuracy, completeness, uniformity, interoperability (integration), and accessibility.
- **Prioritization:** Prioritize efforts to address data gaps and needs. The data governance group's efforts are fed by the data gaps analysis. The gap analysis may only define unmet needs in concrete terms. It is up to the data governance group to determine what the outcomes are when a data gap is not addressed. The group can then help agencies decide which gaps to address first, and which to defer or ignore.
- **Cooperation:** Facilitate cross-organizational collaboration, data sharing, and integration. A data governance group functions much like a TRCC. It is a deliberative body charged with, in this case, a set of specific tasks aimed at obtaining and improving data.
- **Flexibility:** Encourage creative and innovative solutions to data needs. The multi-disciplinary nature of the data governance group helps it to see beyond the typical solutions encountered in any single domain. Group members are expected to learn from each other and generate solutions that are better than if only one stakeholder type was in charge.
- **Utilization:** Improve data utilization and ease of access. It is axiomatic that better data will be used more, and more effectively. The work of a data governance group operationalizes what it means to have better data, including better access to that improved resource.

By establishing data governance practices, partner agencies will have an effective and standard approach when updating current data and developing new data resources. Data governance practices result in shared understanding of the data. That means that collectors, managers, and users all know how the information is gathered, managed, and judged for sufficiency. In a data integration effort, data governance makes absolutely certain that the data will meet analytic requirements once the various source files are combined. This is where data standards are especially important. The same standards must apply across all jurisdictions so that the data can be used in comparisons among areas of a State or Tribal lands.

Data governance also establishes metadata. Metadata describes the data and, importantly, includes detailed measures of its quality. Data quality performance measures give data collectors important feedback on their efforts. They also provide users with the information

they need to become confident in making decisions based on the data. Data governance makes the data more reliable and shows just how reliable the data are.

Data governance also promotes efficiency. By knowing the details of every potential source of data, partners can eliminate duplication of effort—the partner agencies do not all need to collect the same data in the same locations. Instead, the partners can decide which agency should collect and maintain which data elements based on available resources, technical capabilities, and long-term interest in maintaining a useful resource.

CHALLENGES

Based on the pilot studies and case studies, the following are important challenges for data governance:

- **Data security, ownership, and sharing:** The target agencies need each other's data. For some, data sharing is a new experience fraught with exposure to risk of tort liability, legislative backlash, or difficult public interactions. The data owners and custodians do not change when data are shared, but their control over the use and distribution of the data may be compromised. The solutions to these issues are not “one size fits all”—each integration effort must work within applicable laws and records management policies to build a positive data sharing relationship among the target agencies. That relationship must work both ways—the local agencies must work with the State to meet legal obligations and safeguard information from abuse. This issue gains even more importance when data integration efforts begin to link files with sensitive personal information (e.g., medical records, driver and vehicle records, citation and court records, and some portions of the crash database). Not only does the system have to safeguard information it contains, the system also has to have reasonable barriers against misusing the data to “back door” attempts to identify a specific person in violation of applicable State or Federal privacy laws.
- **System development ownership and custodial responsibilities:** There is a balance between ownership and control in the relationship among the target agencies. Participants need to be fully engaged in the data integration process. As individuals representing an agency they need to understand why they are there, and what the benefits are to sharing their agency's plans and its resources in an attempt to improve all of the State's data systems, not just the ones that they or their agency manage.
- **Data management and maintenance:** Each of the data sources, whether safety-related, asset management-related, or other relevant data, requires refreshing with new information as roadways change, physical conditions change, traffic volumes change, and crash experience changes. States set their own requirements for data refresh cycles

depending on statutory requirements, Federal regulations, and engineering needs. The gap analysis can help determine the ideal refresh cycle for each type of data, and each type of facility.

- **LRS maintenance for all public roads:** An LRS requires maintenance. Roadways change, new roads are added, and some roads may close or become private access only. With the inclusion of all public roads, the statewide LRS maintenance task expands. State and local agencies must agree on responsibilities for notifying the LRS custodian of changes, updating line work, reflecting those changes in the GIS, and the location data update cycle (immediate, annual, and other). Finally, the custodian must schedule periodic update releases of the LRS and basemap so that users have the most recent versions.
- **LRS temporality and how to handle historical changes/queries:** This issue is tied to LRS maintenance. Changes affecting roadways' representations in the LRS and basemap affect data integration and data analysis. It is important to know if a crash happened before or after a roadway change, and if traffic volume data were collected before or after the change.
 - **Querying data:** Users need the ability to query linked safety data that temporally synchronizes the physical roadway changes and the LRS, basemap, crash records, and traffic volumes. Knowing the effective date of the changes to the physical roadways allows users to identify the before- and after-periods and thus analyze the impact of the changes. The analysis tools must give users access to the dates as well as the historic and current information on the roadway's features, including geometry.
 - **Displaying historical map layers:** In a GIS environment, users also need the ability to view older versions of the map. A before-/after-study of roadway realignment, for example, could require generating maps of the two conditions along with traffic volume and crash information. The ability to roll back to prior maps makes it possible to create the required visual representation at any time.
- **Query Performance of Distributed Layers/Data:** Integrated data and systems refer to the link between disparate datasets and systems, but don't always assume they are physically in the same data structure or hardware. Query performance between integrated datasets/systems that are physically in different locations (i.e., distributed data) is a concern of an end user application or analysis toolset. There are pros and cons for taking data into a master physical data system, but data maintenance and custodianship issues discussed elsewhere in this report make the possibility of distributed data storage appealing.

RECOMMENDED SOLUTIONS

Among the case studies for this project, Wisconsin and Michigan are the most instructive with respect to data governance. In Wisconsin, the State DOT sets the data standards; however, there is extensive consultation with local agencies including cities, counties, MPOs, and RPCs. The local agencies are most immediately concerned with maintaining assets. They took the lead in selecting and implementing a pavement management process which the partners have refined over several years. The result is a well-honed tool that produces the kinds of predictive analysis needed to project pavement conditions in the future based on alternative spending plans today. From a data governance perspective, the partners set the data quality standards and established a data collection protocol that applies to all pavement condition ratings in the State, no matter whether the municipality or the RPC conducts the data collection. In Michigan, the Roadsoft system was designed specifically to meet the local agencies' needs. Part of that process required the local agencies to come into agreement about those needs. As a result, every Roadsoft user has the same set of data elements in their database and the same types of location definitions. This required some agencies to change their procedures to match the ones defined through joint discussion among the partners. The shared data standards mean that Roadsoft allows comparisons across jurisdictions much more efficiently than in the past when each local agency defined and managed their own data without consultation.

The pilot studies also demonstrated data governance solutions in safety data integration projects. A key aspect of data governance is establishing a way to share data. For the Navajo DOT pilot study, sharing data involves several agencies; three States whose boundaries include Navajo Nation lands, local law enforcement throughout the reservation territory, and several county departments. Available data include the Navajo DOT's and three State DOT's LRSs, crash data from seven law enforcement jurisdictions, roadway inventory, route information, archeological and environmental information, administrative boundaries, traffic volume data, and asset management information. With data being spread across a variety of different places, having a standard sharing system between all the parties is especially difficult. To give just two examples, police crash report forms may differ between agencies and LRSs are not always compatible. The GIS adopted by the Navajo DOT offers a convenient platform for integrating the shared data. The key to success is to share data definitions and quality standards so that each agency's data can flow seamlessly into the shared resource maintained by the Navajo DOT GIS section. Specifically, standards for data collection, and accommodating transfers among multiple GIS implementations will ultimately allow for shared data to be useful to all agencies. It should also reduce redundancy.

Based on the findings of the Navajo DOT pilot study, the following are generally applicable recommendations related to data governance: A data governance committee should be tasked to:

- Understand data needs throughout the DOT and its partners.

- Identify existing data and how it is used and managed.
- Develop metadata for the existing data sources.
- Identify additional data and how it will be used.
- Develop data standards to define:
 - How data is collected.
 - How data is archived.
 - How data is secured.
 - How data is maintained and by which department(s).
 - How data may be accessed, and by whom, using what methods.
- Establish QA/QC requirements such as accuracy and timeliness measures.
- Identify QA/QC workflows that can be automated and run on a routine basis, and additional manual procedures for checking sample sections of the data.
- Develop documentation and training material on the data governance procedures.

The following address the specific barriers listed earlier:

- **System life-cycle management/sustainability:** All four of the case study States made the decision to fund system maintenance completely using State DOT resources rather than require cost sharing by local agencies. The reason is simple: if locals have to pay for the system, they are more likely to choose non-participation or adoption of a different system. If the statewide system is free for local agencies, their only costs are staff time that would be required no matter what system or method they chose. Free software does not guarantee that every local agency will adopt the State DOT's preferred solution, but it does mean that over time agencies have a compelling reason to adopt that preferred solution—it saves them money.
- **Data management and maintenance:** Every one of the case study projects has survived long enough to require updated data collection efforts. In three of the case studies (Michigan, Ohio, and Wisconsin), the local agencies are responsible for all updates to local data, but the State sets data standards and a State agency maintains the GIS basemap for all users. In Tennessee, the State DOT contracted a vendor to collect the majority of local roadway data.
- **Data security, ownership, and sharing:** This issue is best addressed by a formal data governance policy and data governance board. The result is a standard, shared process for identifying sensitive information, ensuring its security, and establishing standards for data access. Local agencies need to understand those restrictions and be both reassured and bound by the same standards as the State-level data custodians and

data managers. The data governance board should include representatives of all target agencies as well as key members of the broader stakeholder community.

- **System development ownership and custodial responsibilities:** In all four case studies, the State DOT paid for the local roadway database, centralized GIS and analytic tools, and ongoing system maintenance. In two States, the DOT set the requirements for the system and, while local agencies have input in directing the future of the system, the State DOT clearly has the final say. In the other two case studies, the State DOT sets data standards for statutorily-required data but leaves all other decisions about the system in the hands of a user group dominated by local agencies. These latter two States view the hands-off approach as central to their success in convincing local agencies to adopt the system, especially in the early days of development.
- **LRS maintenance for all public roads:** In all four of the case studies, the State maintains the LRS structure and its GIS representation. In Wisconsin, some local users can edit the linear representation of their roads in WISLR, but most submit change requests to Wisconsin DOT who makes the changes for them as part of an annual update process. In Michigan, Ohio, and Tennessee, the State makes all changes. Some States have a continuous update process for State-maintained roads, but have only annual updates from local agencies. The continuous process, though difficult to manage, is best from a data governance perspective. Unfortunately, no system is perfect. Even with a continuous update process, some roadway changes will take too long to enter into the system. The annual process is easier to manage and creates a single official version of the LRS and basemap for each year, but it contains inaccurate information for changed roads for part of the year.
- **LRS temporality and how to handle historical changes and queries:** LRS records should include the effective dates for any changes including a view of the network prior to any change to support recalibration. Recalibration is the process of reconciling event data (signs, traffic counts, etc.) based on a change to the LRS. Propagating the changes properly to the event data is key to ensuring that data integrated from other business units or systems is represented properly on the roadway network. This information is crucial to the ability to perform accurate network screening based on roadway attributes. It also impacts the ability to conduct effectiveness evaluation. The Michigan, Ohio, and Wisconsin case studies provide examples of systems that handle temporal changes now.
- **Query Performance of Distributed Layers/Data:** In recent years, web services and web data feeds have expanded so that data owners can share their data via a service layer that can be used by any authorized individual. In the example of a local agency posting data to the State via a web service, the DOT could pull the data into their

master system and set up maintenance routines allowing for better performance of queries and analysis of the merged datasets, but if they choose not to do that and query the various datasets in their native locations, differences between the State's data system and the distributed data sources using web services.

LESSONS LEARNED

Lessons learned on data governance from the pilot studies and case studies are:

- **The data governance group should be inclusive.** Data governance is often seen as an “IT-led” initiative, but examples like Michigan’s Roadsoft and Wisconsin’s WISLR show that the process is most successful when it addresses the needs of the users and is run by a cooperative group of stakeholders representing data collectors, managers, and users. In Michigan and Wisconsin, State and local safety practitioners determined the contents and functionality of the local data and State data systems, as well as the data standards. In both States, the IT teams work to support those users, not simply to complete a process.
- **Data quality performance measures are important.** As a special class of metadata, the data quality performance measures tell users about the timeliness, accuracy, completeness, uniformity, interoperability, and accessibility of the data. Users want this information so that they can decide for themselves if they can use the data to support a specific decision-making process.

SUMMARY

Step 3: Establish Data Governance Process addresses the need for formal data governance over standards, collection, management, and access. The important items in this step are:

- **Leadership: Identify leadership and establish a data governance committee to monitor data integration.** This committee can serve multiple purposes as described in Steps 1 and 2. The data governance group is the formal partnership that should be tasked with all of the integration-related data governance tasks. The committee should have a formal leadership structure so that the jobs of scheduling meetings and tracking progress are clearly assigned.
- **Quality: Establish a data quality assurance program including data standards and measures of data quality.** The need for data standards is well documented as part of the data integration effort. The best way to ensure data quality is to set and enforce data standards. That is a key component of data governance and thus is it something that the data governance group would be best suited for.
- **Prioritization: Establish clear priorities to address data gaps and needs.** The data gap analysis is likely to identify more gaps than can possibly be remedied in the immediate term. Closing data gaps requires resources. The most efficient solutions are often the costliest (e.g., automating field data collection), but have the largest pay off long term for data quality. The partners need to work together to decide which data gaps are critical in the near term, which can wait for a time or for more efficient solutions to be implemented, and which can be deferred indefinitely.

CHECKLIST

- Leadership:** Identify leadership and establish a data governance committee to monitor data integration.
- Quality:** Establish a data quality assurance program including data standards and measures of data quality.
- Prioritization:** Establish clear priorities to address data gaps and needs.
- Cooperation:** Identify opportunities for cross-organizational collaboration and data sharing and integration, establish MOU.
- Flexibility:** Communicate innovative solutions among stakeholders.
- Utilization:** Promote appropriate data usage among stakeholders.

- **Cooperation: Identify opportunities for cross-organizational collaboration and data sharing and integration, establish data sharing and MOUs.** The call for formal agreements among agencies is part of data governance just as it is part of laying the foundation in Step 1. The goal should be to formalize roles and responsibilities, data sharing, and ownership over any newly created resources. These agreements may be mandatory as part of the data sharing process (i.e., some data sets cannot be shared without safeguards in writing and clear lines of authority over the resulting resource).
- **Flexibility: Communicate innovative solutions among stakeholders.** One of the main benefits of deliberative bodies like a State TRCC or the data governance group is the shared knowledge gained by participants. Experts from multiple disciplines are drawn together to share their solutions to common problems. One of the pilot studies included staff from a non-transportation agency but who had recently completed a very similar data integration effort. Their solutions to problems with addressing and location referencing were invaluable to the rest of the agencies.
- **Utilization: Promote appropriate data usage among stakeholders.** Data access restrictions, and proper analytic use of data, are two of the most commonly referenced barriers to data integration. Some data sources cannot be released publicly based on applicable federal and State laws. Sensitive information also exists at the Tribal level and approvals for sharing are subject to Tribal government approval. Integrated data resources containing sensitive data must adhere to the access controls for the original data source, or include ways around detailed disclosure (such as producing only aggregate, non-specific data). Additionally, integrated data resources are complex. Users of the new data sources may need guidelines for how to properly use the data for analysis not just to avoid unwanted disclosure, but simply to arrive at valid results.

STEP 4 – DEVELOP DATA COLLECTION AND INTEGRATION PLAN

OBJECTIVE

A data collection and integration plan is a detailed list of the actions required to go from the current to the desired condition. It shows the sequence of steps, the responsible parties, and is updated with current status as each step is taken. Developing the data collection and integration plan is the fourth step in the safety data integration process shown in Figure 8. A data collection and integration plan is a project work plan that describes every step and activity along the way to achieving data integration. The plan should address each of the nine steps described in this Informational Guide. More importantly, it should describe the actions, and which agencies are responsible, within each of the nine steps. The plan should be coordinated with the efforts of the data governance group so that, as the plan is developed, the solutions and standards designed by the data governance group are incorporated explicitly.

The data collection and implementation plan is very closely akin to the TRCC’s traffic records improvement plan. Both plans address safety data needs and are developed by a coordinating body of stakeholders. Both should be inclusive of all projects that will impact safety data, regardless of who leads or where the funding comes from. The two plans can be linked together and should be coordinated with other major planning efforts (such as any data components of the SHSP).

Every State has a TRCC to promote cooperation on issues relating to data sharing, data quality, and improved safety analysis. FHWA, the Federal Motor Carrier Safety Administration (FMCSA) and NHTSA work with the TRCCs to help States develop a data quality improvement

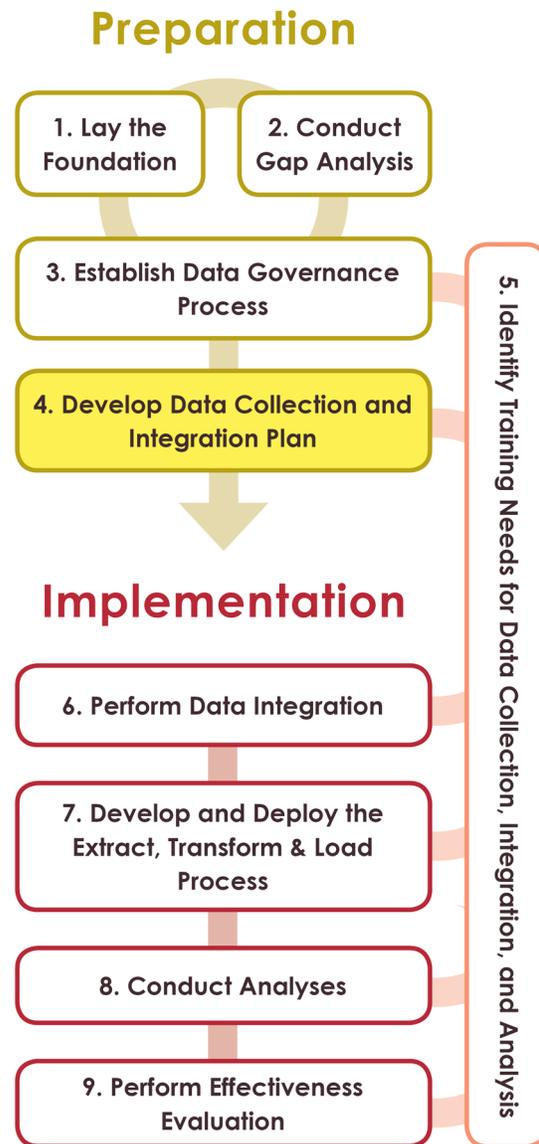


Figure 8. Chart. Step 4: Develop Data Collection and Integration Plan.

management program and a strategic plan for traffic records data improvement. For data integration efforts the partners should develop a data collection and integration plan that addresses the needs identified by the gap analysis and the data governance group. Tribes may require additional help if they do not already have a TRCC or similar cooperative group to direct this effort. To help identify common practices of effective TRCCs, FHWA and NHTSA jointly published the *State Traffic Records Coordinating Committee Noteworthy Practices* (FHWA-SA-15-083).⁽⁸⁾

CHALLENGES

The pilot studies and case studies are rich sources of examples of data integration plans. In particular, the pilot studies each included a project work plan which was vetted by FHWA and the lead agency and their partners. Two example work plans are included in Appendix B. The pilot project work plans are also described in Appendix A. Users of this informational guide can review those work plans to find examples of data collection and integration plans relevant to their own projects. Each of the case studies also included a data integration plan. The following challenges are relevant to creating a data integration implementation plan:

- **Starting the process:** There may be resistance when starting up a data integration project. To kick-start the process, initial communication should be established with the partner agencies in a kickoff meeting as described in Step I (Lay the Foundation). This group of stakeholders may be similar or have the same makeup as the TRCC or even the data governance group. Each participating agency should come prepared to discuss objectives, work plans, and their role in data integration. Each agency also identifies a point-of-contact who will work with the group to develop the integration plan.
- **Establishing roles and commitments for the planning process:** There will be variations among the partners in the kinds of data available and their capabilities in terms of technology and staffing. However, each agency is encouraged to commit to providing data according to their capabilities. In some instances, cooperative relationships based on data integration may already exist. A formal request for the data from each agency will eventually arise; however, at the kickoff meeting it is sufficient to discuss the types of data needed. Early in the plan development, there are three main roles that should be defined and into which each agency should be sorted:
 - **Data Integrators:** These are the agencies that will conduct the steps to integrate data from multiple sources to produce a dataset for analysis.
 - **Data Providers:** These are the agencies that will collect data and share it with the data integrators.

- **Data Maintainers:** These are the agencies that will be responsible for periodic updates to the data.

Note that there is no single role or agency that has responsibility for quality control or data governance. Each of the three roles has a QC component and all are responsible for contributing in a formal data governance process. Additionally, agencies may serve in more than one role—a data provider might also be a data integrator as well as maintain data over time.

- **Data use analysis:** Part of the purpose of a data collection and integration plan is to show how the resulting data resources meet the defined needs of users. A data use analysis will list the desired uses, including all software for business and analytic processes.

RECOMMENDED SOLUTIONS

- **Start the Process:** States already have a TRCC and may, in addition, already have a data governance group. Those two groups provide ready models of the kind of deliberative body that should be tasked with developing the safety data collection and integration plan. It should be a group that is as inclusive as possible and includes the multiple perspectives of IT, collectors, managers and users. As needed, data sharing agreements and MOUs may help to formalize the group and formally task it to deliver a plan that the partner agencies can endorse and adopt.
- **Define roles for integrators, providers, and maintainers:** The Arizona pilot study illustrates how each of the three roles can be filled:
 - **Data Integrators:**
 - Arizona DOT took primary responsibility for housing the target agencies' data files ready for integration. Their goal was to bring all of the data into a centralized database, Arizona Transportation Information System (ATIS), from which extracts could be generated for analysis.
 - BIA serves as a data integrator for Tribes other than the Navajo Nation (which maintains its own data).
 - Navajo Nation acts as a data integrator for multiple agencies within the Tribal area, some of which are inside the State of Arizona.

- **Data Providers:**

- Arizona DOT collects inventory and traffic data on the State-maintained roadways and makes statewide crash data available.
- Pima Association of Governments (PAG) shared a Geodatabase containing their road inventory, including several of the required fields: road category (classification), predominant direction, road name, travel direction, and number of lanes, but no traffic volume data. Locations are coded using a Route/Milepost location referencing method similar to ATIS, although the route identifier does not conform to the standard nomenclature used for ATIS.
- Maricopa Association of Governments (MAG) provides inventory data that is already incorporated into ATIS.
- Navajo DOT provided a Map Package file containing crash data and a separate file of Indian Reservation Roads (IRR) inventory data in geodatabase format. Locations in the Navajo DOT IRR are represented in a Route/Section/Distance location referencing system that is not compatible with the Arizona LRS.
- BIA provided an Excel spreadsheet containing their IRR road inventory in their standard Road Inventory Field Data System Coding Guide format.

- **Data Maintainers:**

- Arizona DOT plans to continue updating data on State-maintained roads.
 - PAG and MAG agreed to update their data periodically in the future by continuing to submit data for ATIS integration.
 - The BIA will continue to compile data from Tribes. A formal data sharing agreement would be required for a long-term update process.
 - Navajo Nation was unsure how well they would be able to meet the need for updated data. Their GIS was not functional at the beginning of the project. If the GIS is not in operation, the data are maintained in a way that is not easily shared or integrated with other data sources. In addition, to make their LRS compatible with Arizona's LRS would mean, at a minimum, developing a repeatable conversion process similar to the one seen in the Wisconsin case study for combining State and local data in WISLR.
- **Data use analysis:** A detailed analysis of the characteristics and uses of data is necessary to mitigate issues integrating a variety of data formats. First, the integration plan should adopt a model that serves the requirements of the business and analysis software applications and other uses of the data. Then the database developer will need to ensure that various applications can use this format or, alternatively, that standard operating procedures are adopted to convert the data to another format. This

information should be readily available from the data governance process and should be based on the gap analysis. Thus, this is not a new product or effort, but a new use of that information as part of the plan development.

LESSONS LEARNED

Lessons learned from the pilot studies relevant to data integration implementation plans include:

- **Start as early as possible.** The integration plan, the gap analysis, and the data governance work are all inter-related. As the various groups are meeting on one topic, it makes sense that they should be aware of the issues for all of the topics. Thus, developing the data integration plan should start simultaneously with the efforts described in the preceding chapters. Plan development should be explicitly tasked to a group, such as the TRCC or the data governance group, and subject to the same types of formal agreements (data sharing agreements and MOUs) as those efforts are.
- **Pick the easiest integrations first.** Arizona's pilot study is instructive in that their end goal was not simply data integration, but to analyze the data using AASHTOWare Safety Analyst™. Their integration plan had to address the desire to test out the analytic software to see if it would produce useful results. Thus, integration was only part of the proof of concept. The project proceeded with the available data in order to get through to the final step of using the analytic tools as a proof of concept. An effective plan will address how the integrated data are to be used.
- **Define and understand task dependencies.** The integration plan is a true implementation plan. Its steps must include the prerequisite tasks from other efforts that address data availability. Its timeline must take those other efforts into account. As an example, if data integration will rely on a centralized GIS and common LRS for all public roads, those two elements must be in place before the data integration step. Before that, however, data standards must be defined and communicated. Task descriptions in the data collection and integration plan should list any precursor and dependent activities so that the impact of schedule changes can be recognized and accounted for.

SUMMARY

Step 4: Develop Data Collection and Integration Plan addresses the need for a business plan that the partnering agencies can follow to achieve successful data integration. The important items in this step are:

- **Develop the data integration plan in conjunction with gap analysis and data governance efforts. Consider tasking an existing, established group with the plan's development.**

This advice is repeated throughout this informational guide. The hope is that data integration partnerships will arise from existing and effective groups, and that the partners will work within existing groups to achieve their data integration goals.

The main reason for this is to avoid a proliferation of related, and potentially competing groups. The responsibilities for data integration steps can be tasks assigned to the TRCC, the data governance group, and others.

A well designed data collection and integration plan will reflect the partners' priorities. Those will likely include meeting federal requirements for an all-public-roads safety data resource and analyses. The plan should also reflect the partners' priorities for the goals they hope to achieve with data integration. Prioritized needs should be reflected in the sequence of tasks in the plan.

- **Identify integrator, provider, and maintainer roles.** The roles are distinct, even a single person or agency may serve more than one role. The reason for clearly identifying these three roles is that they form the basis for task assignments and resource utilization throughout the data integration process. The three have very different needs that should be addressed in the data collection and integration plan.
- **Develop the plan as an implementation plan with full lists of actions within tasks and explicit recognition of task dependencies.** This may also be called a data business plan. The end product is what's important. That should be a plan that lists the detailed tasks and actions within each task that will result in delivery of specified products. Ultimately, the final product is an integrated data resource, but there are many actions to take along the way such as setting data standards, establishing

CHECKLIST

- Start the data integration plan in conjunction with gap analysis and data governance efforts. Consider tasking an existing, established group with the plan's development.
- Identify integrator, provider, and maintainer roles.
- Develop the plan as an implementation plan with full lists of actions within tasks and explicit recognition of task dependencies.

responsibilities for collection and maintenance, collecting and cleaning the data, etc. Each of those actions should have a target date, be assigned to a specific set of partners, and have a known set of predecessor and dependent tasks. Everything in the action plan should be tracked for current status and to address any scheduling or quality problems.

STEP 5 – IDENTIFY TRAINING NEEDS FOR DATA COLLECTION, INTEGRATION, AND ANALYSIS

OBJECTIVES

Identifying training needs is the fifth step in the safety data integration process shown in Figure 9. It draws on information obtained from previous and subsequent steps throughout the process. A training needs assessment is a formal, comprehensive examination of the knowledge, skills, and abilities of existing personnel in comparison to the identified needs for specific tasks such as planning, data collection, data management, and data analysis. Each task has a set of required skills that need to be fulfilled. It is useful for stakeholders to understand the types of safety professionals and organizations involved in data integration, including State safety engineers, data managers, traffic engineers, ITE, etc. This will help stakeholders identify who needs to be contacted if assistance is required. A training needs assessment will involve all relevant stakeholders; anyone who touches the integrated data resource in any way should be assessed for training needs. Those needs identified in the assessment must be prioritized based on what training is already available, which skills are most needed and least available, and the timing of the needs.

Some of the skills needed for data integration include the following:

- Database management
- Data quality measurement and management
- Merging datasets based on common variables
- Location-based, spatial data management, display, and merger
- Knowledge of data standards

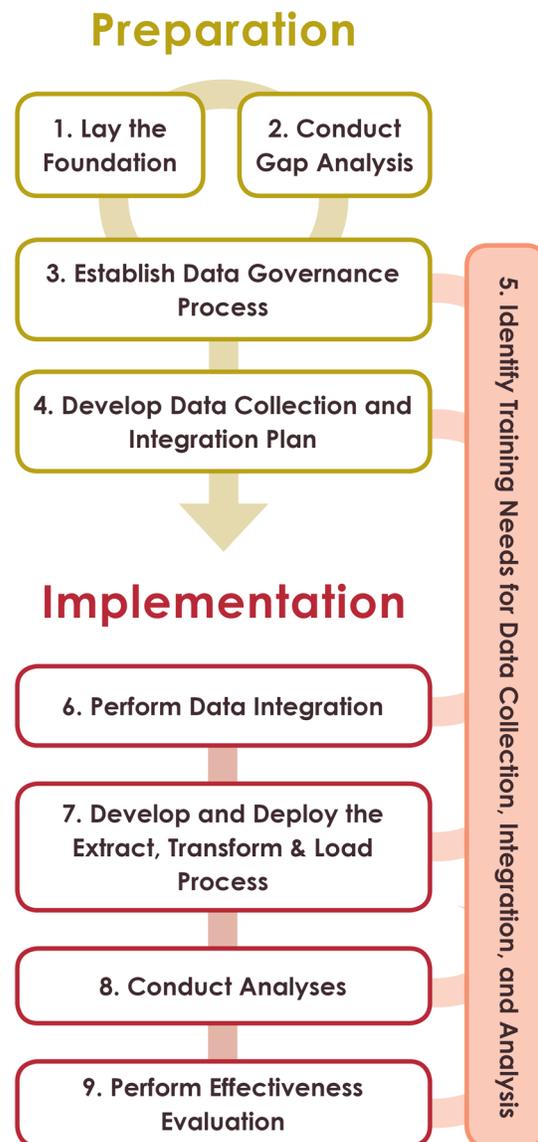


Figure 9. Chart. Step 5: Identify Training Needs.

- System documentation, including metadata, and system inventories
- Safety data analysis tools and methods
- Highway Safety Manual safety management process

The specific training needs for target agencies agency staff depend on their starting capabilities, level of experience with data management and data integration, and the complexity of the required data collection and integration efforts. Training may apply to both data managers and engineers, Training may also include understanding and use of data analyses and data manipulation.

Many of the software tools for safety analysis come with training as part of the purchase price or as optional add-ons. Contractors, LTAPs, and TTAPs can also provide training on those aspects of the data integration efforts that they are supporting. National-level training courses are often available to educate practitioners on how to address common needs.

CHALLENGES

- **Sources of training needs information:** The challenges pertaining to training depends on the needs of the stakeholders. A full training needs analysis can be a costly and time-consuming process. That full formal assessment would consider every legitimate user need, and the capability of every staff member. To make progress in the near term, partnering agencies need some way to identify training needs and find sources of training even while the more complete training needs assessment is being conducted. An experienced training program manager can help target agencies assess skills and knowledge gaps and align those with specific training content as part of the training needs assessment.
- **Focus on practical outcomes:** A formal training needs assessment is designed to identify more training needs than could possibly be met in the short term. Some of the training needs are for things that can be deferred or obtained through new hires and consulting services as an alternative to training existing staff.
- **Prioritize training needs:** The pilot studies identified, but could not possibly meet, all the legitimate training needs that States, Tribes, and local agencies had with respect to data integration. For example, training for data collectors is of critical importance in communicating data standards, and making sure that the data collectors can implement those standards. Michigan and Wisconsin's implementations of pavement condition ratings are examples of the kinds of training that would be required for data collectors of roadway inventory data. Data managers may need training in general data integration processes beyond the tools and techniques passed along by consultants as part of an implementation effort. Users will often need training in what to do with the results of

novel data analyses made possible by the existence of integrated data. The pilot studies did cover some of this by, for example, demonstrating the powerful GIS-based analytic tools. In Arizona the State DOT used its purchased training units for AASHTOWare Safety Analyst™ to make sure that staff knew how to use the tool.

- **Institutional Memory:** In some cases, an agency is reliant on one person who has the most knowledge of data processes. A gap in knowledge develops at an agency if that person leaves. Therefore it is important to ensure others at an agency are trained to prevent knowledge loss.

RECOMMENDED SOLUTIONS

- **Sources of training needs information:** The earlier steps in the data integration process also serve as sources of training needs. The gap analysis described earlier may reveal specific training needs for specific processes and tools. The data governance planning process may also be used to help institute a formal process to address training and technical assistance requests from State and local agencies. This process may identify other entities in the State that can provide training and technical assistance, such as the local LTAP, MPO staff, and universities.
- **Focus on practical outcomes:** In the pilot studies, most of the training centered on the understanding of and implementation of processes and tools specifically implemented as part of the data integration effort. The technical assistance in Arizona, Indiana, Navajo Nation, and Rhode Island centered on data integration tools and GIS. The lead agencies' staff needed training on how to run those tools and use the GIS techniques on their own after the pilot studies ended. This was handled through small group technical training with key staff. Thus there are some training needs that are immediately obvious and are based entirely on mastering the tools and products of the integration process itself.
- **Prioritize training needs:** There are several ways to prioritize training, but one of the most effective is to identify what training is available already that meets at least some of the identified needs. This focus on “low hanging fruit” allows the data integration partners to make progress immediately on at least some aspects of the data integration tasks while allowing time for planning a more complete approach to filling the most important training needs. Where new tools are implemented as part of the data integration effort, the partners are advised to include training as part of the development effort. That way agency staff can learn the processes and tools as they are deployed, and right when the training will do the best.

- **Continued training opportunities:** To ensure there is no loss of institutional memory, opportunities should be identified to keep a group of staff trained on data integration processes.

LESSONS LEARNED

The pilot studies and case studies provided the following lessons learned on training needs:

- **Training needs assessment may be a continuous process.** In the midst of a major data integration effort, it is very easy to lose sight of the training needs that are being created by all the new tools, the expanded user base, and the new analytic results. The pilot studies, for example, barely touched on the end users' needs as those would not become evident until the newly integrated data were used in analyses. In Rhode Island and Arizona, the analytic tools had already been selected (at least for the purposes of testing), so those States had the opportunity to identify training on their selected tools (AASHTOWare Safety Analyst™ and the Esri Roads & Highways suite, respectively). In the other pilot studies, the decisions of which analytic tools to use are for a future time. In each case however, when the partners identified a specific tool or desired analysis, that choice implied a refreshed gap analysis and training needs analysis. Each decision comes with its own set of training needs.
- **Both training and technical user assistance are needed.** As the example of Michigan's Roadsoft shows, formal training must often be supplemented by one-on-one help sessions between users and the system support staff. Help desk functions serve as a necessary component of a training package. Such user assistance is also an ongoing need.

SUMMARY

Step 5: Identify Training Needs for Data Collection, Integration, and Analysis addresses the training needs assessment process and training delivery. The important items in this step are:

- **Conduct an assessment of stakeholder needs related to data integration training and technical assistance opportunities.** The recommendation is for both informal and formal training needs assessments. The informal assessment is needed in order to get started meeting immediate needs for training in technical areas related to the data integration effort. At the same time, a more formal training needs assessment should be undertaken as an ongoing process. This formal process helps partners identify their knowledge and skill gaps and plan in advance how to address those needs.

An example framework to identify training needs can be found in the *HSM Training Guide* (<http://safety.fhwa.dot.gov/hsm/training/hsmguide.pdf>). In this framework, training can be stratified at basic, intermediate, and accomplished levels and also associated with specific focus groups, including management, planning, design, operations, and analysis. For data integration training, an agency will need to determine the audiences and the specific integration tasks to be completed by that audience.

- **Develop a formal process to address data integration training and technical assistance requests from State and local agencies.** Requests for training are the result of the training needs assessment. It is anticipated that there will be more needs expressed than any one existing source can meet. An organized approach, such as an implementation plan, to documenting the requests, and

CHECKLIST

- Conduct an assessment of stakeholder needs related to data integration training and technical assistance opportunities.
- Develop a formal process to address data integration training and technical assistance requests from State and local agencies.
- Discuss local agency needs and capabilities with other entities in the State that can provide training and technical assistance, such as the local LTAP/MPO staff/universities.
- Identify needed resources, including staff time, funding, and partnerships to address technical assistance requests.

identifying sources of potential training will help all partners establish reasonable access to training.

- **Discuss local agency needs and capabilities with other entities in the State that can provide training and technical assistance, such as the local LTAP/MPO staff/universities.** While much training can be obtained from available federal resources and selected software's support teams, much of that training may be unknown or inaccessible to local partners. Other resources such as the LTAP/TTAP, MPO/RPRC staff, and university-based centers can help meet the needs. Ideally these potential sources of training would be partners in the data integration effort from the earliest stages.
- **Identify needed resources, including staff time, funding, and partnerships to address technical assistance requests.** The data integration program must address access to analytic outputs for those who lack the skills or knowledge to use those resources themselves. Especially when advanced analytic techniques or licensed software are used, some potential users will need assistance to access and understand the products of those analyses. Setting aside resources to serve those users is a good way to get the most out of the effort that went into creating the integrated data resources in the first place.

STEP 6 – PERFORM DATA INTEGRATION

OBJECTIVES

Performing data integration is the sixth step in the safety data integration process shown in Figure 10. Integration is the act of merging two or more data sources together into a combined resource that is available for use. For safety data, this most often means that the data from multiple spatial data source files (i.e., held in a GIS at the State, Tribal, or local level) are brought into a centralized GIS or otherwise made accessible through a single GIS portal. GIS is not the only way to accomplish safety data integration, but it is the most common, and was used throughout each of the pilot studies and the case studies. An alternative method would treat the spatial data as a linking variable in a standard database merge operation.

All of the safety data integration efforts take place for a purpose. That purpose is to improve safety through improved analysis. For agencies to complete a safety data integration project, they need to have identified and filled their data needs and merged the source files into a resource ready for analysis. The integration step can take several forms, but the end result must always be a database that supports the intended safety data analyses.

CHALLENGES

- Data insufficiency:** In two of the case studies, the integration plan hit a roadblock for lack of data. In one of those cases, the field data collection process was delayed and plagued with data quality problems. Part of the problem was that the data collectors did not understand how the data were to be used and so needed to learn quickly from feedback how to convert the field data into usable information for the DOT. In another pilot study there were a series of data quality problems some of which did not become

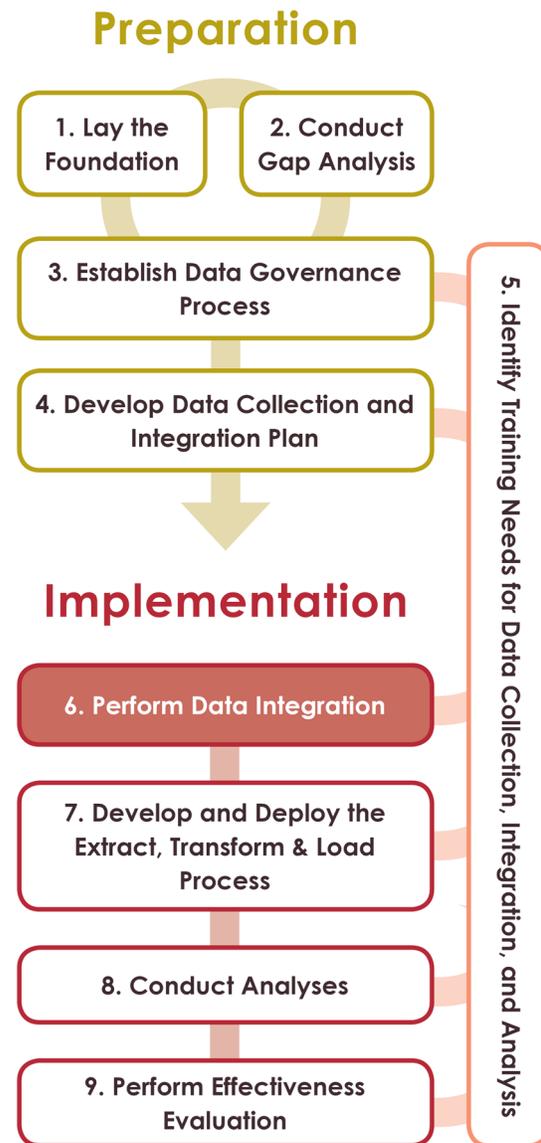


Figure 10. Chart. Step 6: Perform Data Integration.

evident until the integrated data were loaded into the analytic system. The problem was not with integration itself, as that process worked, but with the way the data were coded with respect to segment start and end points. Fixing the segmentation involved some automated (and thus inexpensive and repeatable) processes, and some manual (and thus expensive and unique) processes. One obvious barrier shared by all of the pilot studies was a lack of complete data. For demonstration projects like the pilot studies, this just meant that the final integrated data file was smaller. Long term, however, for integration to be a success, the basic data need to be complete and accurate before they are fully integrated and used in analyses.

In practice, the pilot studies each worked with a sample data set before moving on to full data integration. This is an efficient way to test processes and identify likely data quality problems that may need to be addressed before expending resources on integration.

Some other challenges to the data integration process identified by the FHWA Data Integration Primer include the following:⁽¹⁾

- **Heterogeneous Data:** Data integration often involves synchronizing huge quantities of variable, heterogeneous data resulting from internal legacy systems that vary in data format. Legacy systems may have been created around flat file, network, or hierarchical databases, unlike newer generations of databases that use relational data. Data in different formats from external sources continue to be added to the legacy databases to improve the value of the information. In some data integration cases, the effort becomes a major exercise in data homogenization, which may not enhance the quality of the data offered.
- **Bad Data:** Data quality is a top concern in any data integration strategy. Legacy data must be cleaned up prior to conversion and integration, or an agency will almost certainly face serious data problems later. Legacy data impurities have a compounding effect; if this information is corrupt, so, too, will be the decisions made from it. It is not unusual for undiscovered data quality problems to emerge in the process of cleaning information for use by the integrated system. The issue of bad data leads to procedures for regularly auditing the quality of information used.
- **Lack of Storage Capacity:** The unanticipated need for additional performance and capacity is one of the most common challenges to data integration, particularly in data warehousing. Two storage-related requirements generally come into play: extensibility and scalability. Anticipating the extent of growth in an environment in which the need for storage can increase exponentially once a system is initiated drives fears that the storage cost will exceed the benefit of data integration. Introducing such massive

quantities of data can push the limits of hardware and software. This may force developers to instigate costly fixes if an architecture for processing much larger amounts of data must be retrofitted into the planned system.

- **Unanticipated Costs:** Data integration costs are fueled largely by items that are difficult for the uninitiated to quantify, and thus predict. These include labor costs, software and hardware purchases, unanticipated technology changes/advances, and costs of data storage and maintenance. Unrealistic estimating can be driven by an overly optimistic budget, particularly in these times of budget shortfalls and doing more with less. More users, more analysis needs, and more complex requirements may drive performance and capacity problems. Limited resources may cause project timelines to be extended, without commensurate funding.

RECOMMENDED SOLUTIONS

At its most basic level of description, integration requires source data, a repeatable process for merging two or more data sources, and a way to make the integrated data accessible. In the data integration steps presented in this Informational Guide, data integration is simply the step of implementing the integration plan. If the plan is workable, an integrated dataset will result. The following solutions can help ensure the integration plan is executed smoothly:

- Utilize software technology to minimize obstacles through a series of data access routines that allow structured query languages to access nearly all database management and data file systems—relational or non-relational.
- Have developers and users work together to determine the quality controls that will be put in place in both the development phase and the ongoing use of the system.
- Plan for future data needs and associated data storage.
- Extraordinary care in planning, investing in expertise, obtaining stakeholder buy-in and participation, and managing the process will each help ensure that cost overruns are minimized and, when encountered, can be most effectively resolved.

An example process for establishing a data integration process comes from the Indiana LTAP/Putnam County pilot study as they tested a centralized GIS database. The integration was completed in four phases:

- **Phase I – GIS Assessment –** The project team identified current capabilities, needs, processes and objectives of the LTAP and local agencies (in this case, Putnam County). In addition, the assessment determined system architecture, infrastructure (hardware, software, network), and the technology needs.

- Phase II - GIS data development and integration – The project team collected GIS information and prepared the local information for inclusion in the integrated database. This information included administrative/geographic boundaries, roadway inventory information, and maintenance data, amongst others, in determining a transportation improvement or maintenance. In other projects, this also included traffic and crash data.
- Phase III – GIS website development. The project team developed and implemented a GIS website to provide information to local, State, and federal agencies and the public. This same system included a secured website for internal staff to access information.
- Phase IV – Asset Data Collection – The project team developed an outline-level plan for future asset data collection such as bridges, culverts and signage on priority roads.

LESSONS LEARNED

- **It is possible to successfully integrate poor quality data.** The integration methods rely on the presence of very few data elements—spatial (GIS/LRS) coordinates are linking variables and whatever else the file contains simply follows along as the now integrated data. Integration will fail if the spatial data are incorrect or missing, but otherwise every piece of data (good or bad quality) is brought into the integrated dataset. Cleaning the data thoroughly before using it in an integrated way is more cost effective because it avoids having to redo the integration step multiple times after each new correction of the source file(s).
- **Plan for an iterative process.** The integration plan should include sample-based testing and have the flexibility to react to failed integration. Especially the first time through the process, there will be problems that were not anticipated in the original planning. Allowing for time and resources to address these inevitable but unforeseen problems is helpful.
- **The lead agency still needs help from partners.** When the data integration fails, the fault usually lies in the source files. The data integrators need access to people who know the source files very well and who can identify and implement fixes to the data.

SUMMARY

Step 6: Perform Data Integration addresses the procedures and tools used to create the integrated data resource. The important items in this step are:

- **Follow the data integration plan and understand it will be iterative.** The data integration process will necessarily follow a plan. The recommendation is that this effort should be based on the formal plan developed in Step 4. That plan should be flexible so that as implementation proceeds, it can be adjusted to take advantage of new knowledge and so that it can react to changes in schedules, staff availability, and partners' priorities.
- **Maintain cooperation among stakeholders.** In the earliest steps of the data integration process, partners developed agreements (formal and informal) as they decided how best to approach the project. As the time comes to create the integrated data resource it will test those agreements in the one way that truly matters—will each partner deliver their promised data, staff time, and other resources. The partners should be refreshing their MOUs as they fall out of date or when agency leadership changes. In addition, revisiting commitments as the dates for data integration approach is a good way to make certain that the plan is still viable.
- **Use software tools to enhance data routines.** Data integration is an iterative process involving multiple attempts to find proper matching data from each source file. There are software tools that can help automate some parts of the process and provide standardized measurements of how reliable the matches are. Besides a more efficient process, using data integration tools can help to standardize the process so that data integration success does not rely on the continued availability of just one staff person who knows what to do.
- **Anticipate future data and storage needs.** Some integrated data resources can become very large (in the Terabyte range) once several multi-year data files are merged. If the data integration plan calls for creation of a new, stand-alone database, there will be storage and computer processor needs associated with managing that large file. If data

CHECKLIST

- Follow the data integration plan and understand it will be iterative.
- Maintain cooperation among stakeholders.
- Use software tools to enhance data routines.
- Anticipate future data and storage needs.
- Monitor and manage the data integration plan.

security is a concern, the data governance plan may include installation of a dedicated server. As the system stays in place long term, the needs for storage and computing power will grow. Planning ahead for these needs should be part of the data collection and integration plan, and the data governance plan. Periodic review for new needs is a good way to avoid surprises.

- **Monitor and manage the data integration plan.** This item is brought forward from Step 4 as this is the crucial step (data integration) and the point in the process where most of the partners will be moving into unfamiliar territory. There will be missed deadlines, data files that fail integration because of poor accuracy or missing data, and a host of other events that need to be recognized and dealt with. Likewise, the plan will need to be kept up to date with frequent status reports on each active task, and adjusted due dates.

STEP 7 – DEVELOP AND DEPLOY THE EXTRACT, TRANSFORM, AND LOAD PROCESS

OBJECTIVES

Extract, Transform, and Load (ETL) is the process by which the integrated safety data are made available for analysis. This is a formal, documented, sequence that gets data out of the native system(s), transforms it into a format required by the analytic software, and loads the resulting file(s) into the analytic software. Developing the ETL process is the seventh step in safety data integration shown in Figure 11. The ETL process moves the integrated data into the analytic tool. Extraction is the process of creating an analyzable subset of the total dataset (for example, taking one year’s data from a multi-year dataset, or taking the data for one roadway type out of a file of all data for all roadway types). The transformation process involves multiple transformations to the data and might include recoding variables, converting from the native database to a spreadsheet or flat file, and restructuring the data file to match the requirements of the analysis tool’s data import process. ETL is required because, typically, the analytic software is a separate program from the database used to manage the integrated data resource. It has its own data structure and quality requirements. The ETL processes is designed to meet these requirements. The ETL process requirements may be developed as part of the data governance activities described earlier.

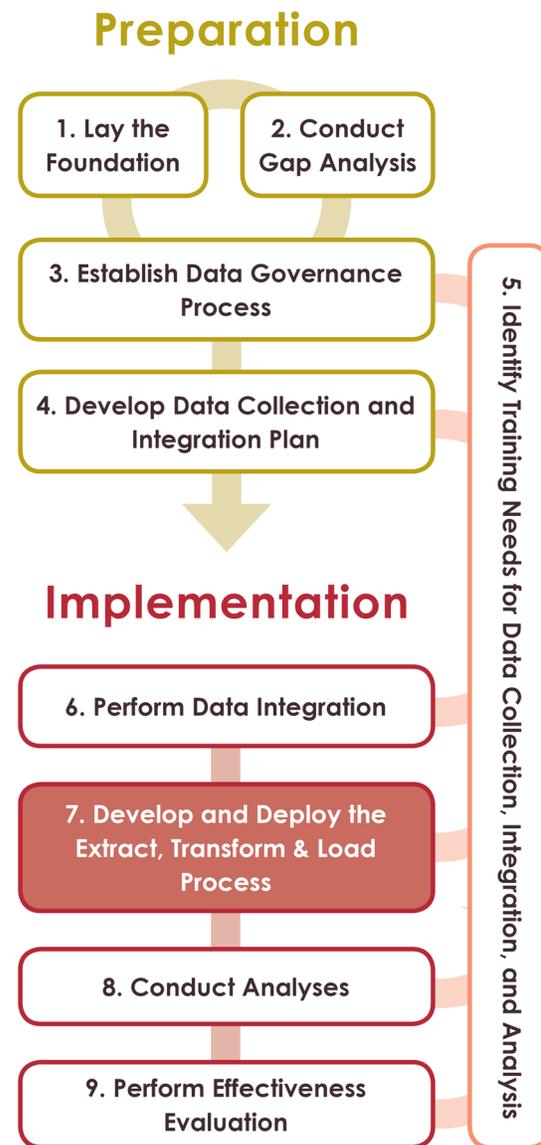


Figure 11. Chart. Step 7: Develop and Deploy the ETL Process.

Data Extraction Process

Safety analysis tools and advanced safety analysis techniques are specific to classes and types of locations as well as types of crashes. For example, in a safety program focusing on safety in curved sections of low-volume rural roads, analysts and engineers are concerned about run-off-

road crashes. They are interested in a specific set of countermeasures (e.g., warning signs, edge lines, rumble strips, and shoulder widening, pavement edge smoothing, etc.). The analyst needs only a subset of locations (rural road curved segments) and their associated crashes and traffic volume data. Data extraction is the ability to pull from a database only those records of interest.

Data extraction tools exist in most data management software products, from spreadsheets to high powered statistical data analysis packages. Some common features are:

- **Filter Cases, aka Select Cases:** Using a data filter, the user specifies a model (a set of target values of one or more data elements), and the software selects only the matching cases. No records are removed from the database, but only the selected cases are used in any subsequent analysis.
- **Data Extract, aka Subsetting:** After filtering, many data management tools give the user the option of creating a subset of the original data file for storage. This is useful when the analytic database must be retained or shared as a final copy is available separate from the original database.
- **Exclude Cases:** This is a filtering operation that is the reverse of selection—it results in dropping any case matching the model. It is most commonly used in data cleansing operations (e.g., dropping cases with missing values or those with codes for “unknown” or values not of interest in the analysis. Most software offers deselection as an option as part of the filtering operation.
- **Join and Filter:** This is a feature of more advanced data management tools. It allows the analyst to link two or more data tables and filter simultaneously to arrive at a dataset with cases merging data and then selecting to match a criterion. For example, if an analyst needs a dataset of 2-lane rural run-off-road crashes on roads with narrow shoulders, the selection criteria apply to the roadway inventory file and the crash records database. The filtered locations show up in a results table including only those 2-lane rural roads with narrow shoulders. That list of locations serves as a filter for extracting crash and traffic volume data only for those locations. Putting the two together results in a merged analysis dataset of crashes on 2-lane rural roads with narrow shoulders.
- **Spatial Filtering:** This is a feature supported by GIS and data management software using spatial data elements. A common example is analysis of intersection safety. The analyst may specify a distance from the intersection in the search for intersection-related events in the crash database. This requires spatial filtering based on the locations of crashes and the geocoded location of the nearest edge of the intersection. Without

spatial filtering, analysts might attempt this analysis relying on distance-from data element in the crash database; however, the spatial analysis tools offer a more precise way to measure that distance. Spatial filtering is also used in conducting safety analyses for specific jurisdictions or areas within a State (e.g., each county).

Data extraction may not necessarily result in a complete desired data set, therefore, assumptions may need to be made to account for additional data. For example, an intersection inventory may include notation of *all* signal-controlled intersections. For data extraction purposes, the target agencies will need to decide if the lack of data on signal presence means that the remaining intersections are controlled by signage only or uncontrolled. Data extraction should result in complete data for the intended analysis. Unknowns in the data element used in the selection criteria are allowable as long as assumptions about the selected data are valid.

Data Transformation

Once a suitable extract is produced, the analyst must change data into the form expected by the analytic software. This is generally software dependent and is based on the specifications for data import files. Transformation can affect individual data elements formats (e.g., traffic volumes should be expressed as total number of vehicles whereas the State data file is expressed as a number times 1000). Transformation can also imply a specific file layout, file type, or database type.

The transformation process is typically performed by a software tool programmed specifically to change the data in the statewide integrated database to meet the data requirements of the analytic software. It is highly repeatable and should only fail if the data contain serious errors or data type mismatches that generate unrecoverable errors.

Data Load

After extraction and transformation, the resulting data file should be ready to load into the analytic software tool. The load process is often manually controlled and involves the software tool user specifying a data file, from a specified location, and clicking the software controls to import the data. The load may succeed even if the data contains errors. Some analytic software tools run a data quality check as part of the load process, while others produce error reports only if the use of the data fails. In Arizona's pilot study, the ETL process resulted in a data file within AASHTOWare Safety Analyst™. Before the data could be analyzed, the software ran an error check and produced an error report. Any file with an error was rejected by the system and could not be analyzed.

GIS-Based Data Integration

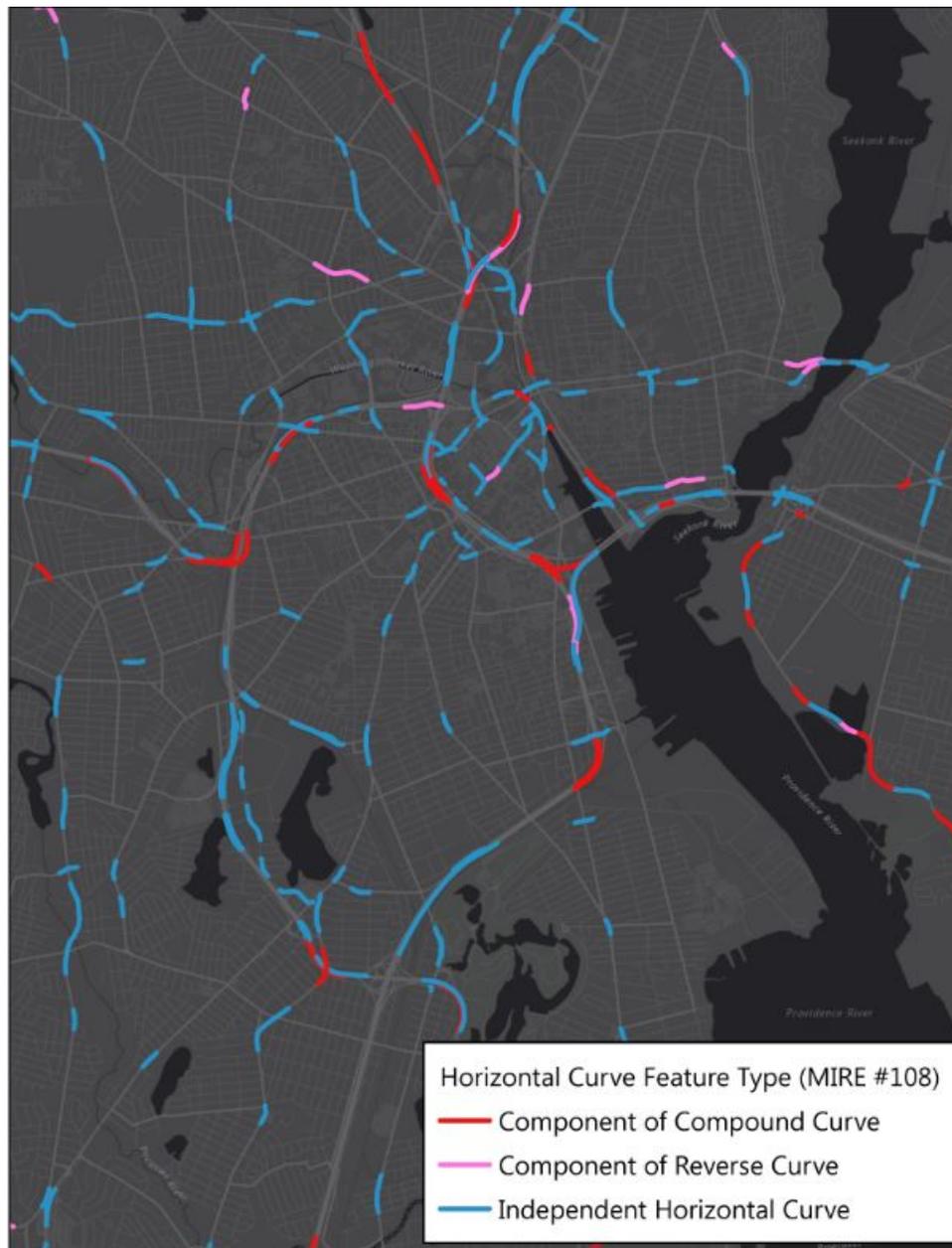


Figure 12. Map. Sample GIS map showing horizontal curve locations.

(Source: VHB)

The GIS map itself is a digital representation of the real world. For highway analysis, it must include the locations of roadways to a sufficient level of accuracy such that to-be-mapped items that exist in the roadway in the physical world will also land on the roadway's representation in the map. Accomplishing this step for the LRS means that any data using the LRS location codes within a linear referencing method automatically maps correctly to the GIS representation of

the roadway network. Figure 12 shows an example of a GIS map displaying locations with horizontal curves.

Spatial data not keyed to the LRS can still enter into the GIS, however it likely will have coordinates that place it somewhat off from the streets as represented in the GIS based on the geocoded LRS. An example is local roadway data collected using GPS coordinates. In many states, this information predates the inclusion of local roadways in the LRS. When the locations of key features (intersections, for example) are mapped in the GIS, they may not align with the LRS that is also displayed in the GIS. Some corrections will be required. Dynamic segmentation also provides methods of defining the start of a new event segment based on a change in one or more key attributes (e.g., number of lanes, lane width, median type, etc.) rather than using a standard segment length (e.g., every 0.10 miles).

In the case of integrating GIS data layers and crash data from various entities and sources that may have different LRSs, conflation and other GIS-centric processes are key to building that single set of data that is easily queried and analyzed. An example process or workflow may be the following:

1. Match the various roadway datasets to a master LRS based dataset or road network.
2. Combine the crash/safety data from the sources based on latitude/longitude or the LRMs in the LRSs and dynamically segment the crashes to the master LRS.
3. Provide the data via services and database objects to query and analysis applications.

Spatial Data Analysis, Filters, and Extraction

In the GIS environment, analysis and data extraction can be identical processes. The GIS tool allows the user to set both spatial and data selection filters. Spatial filters are selections like county boundaries, proximity to a single point; or similar. Other selections would include filtering on any non-spatial data element in the available database(s) such as date, roadway type, minimum or maximum traffic volume, etc. Figure 13 shows an example of querying crash data based on a spatial buffer to create an intersection influence zone. It shows intersection type and crashes. A spatial buffer or query is the ability of the GIS to compare data in one layer versus another, be it via a polygon shape or a buffered line or point. The process of buffering creates a polygon shape that the GIS then uses to clip out features and data in another layer that interacts with that polygon. In this example, an intersection influence zone is defined, the point buffer is generated and the crashes that fall within that influence zone are returned.

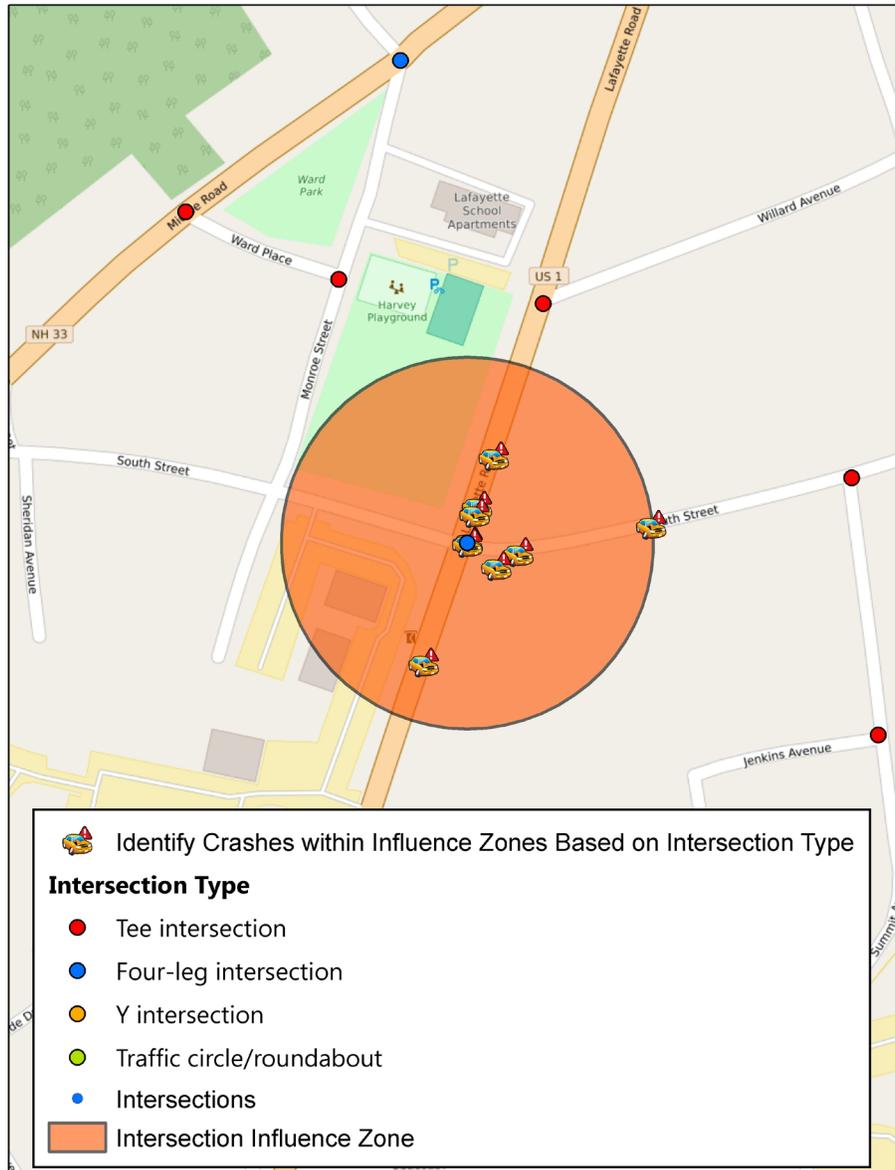


Figure 13. Map. Sample GIS map intersection types and crashes.

(Source: VHB)

Spatial analysis uses the GIS and analytic tools to assess the spatial relationships among various data elements and features in the roadway network. For example, a spatial analysis could examine crash locations in relation to curves. As discussed in the previous section, roadway features and crashes are in two different layers in the GIS. The GIS can map these two layers simultaneously to give a visual representation of their relationship (i.e., does it *look* like curves are a likely place for crashes). A spatial analysis offers more precision answering the same question. With spatial analysis, the precise location of each crash along a roadway is known, as

is the location and degree of curvature for each curve. The analyst might filter for crashes of a type that would be most likely in a control loss situation (e.g., run-off-road, cross median, side-swipe same direction) to see if these crashes fall disproportionately at or near curves using measured distances of the crash location from a defined point on the curve. The analyst could further control for roadway class, speed limit, traffic volume, and other variables that affect crash frequency. The result is an analysis that specifically relates a spatial descriptor of the roadway (curve location) with crashes to determine what percentage of crashes occur within a defined number of feet of a curve.

Other examples of spatial analysis exist. At the local level, analysts might be interested in the association between crashes and driveways along a roadway. Without spatial analysis, this type of analysis typically relied on the density of driveways as an indicator of exposure to risk of crashes. With spatial analysis, the analyst can filter for crashes that occur at or near driveways so long as the driveway locations and crash locations appear in the database. This analysis might also filter for typical crash types expected from driveway incursions into traffic (e.g., sideswipe, rear end, etc.). Figure 14 shows a sample crash application and the potential filters that can be used to query against crash data.

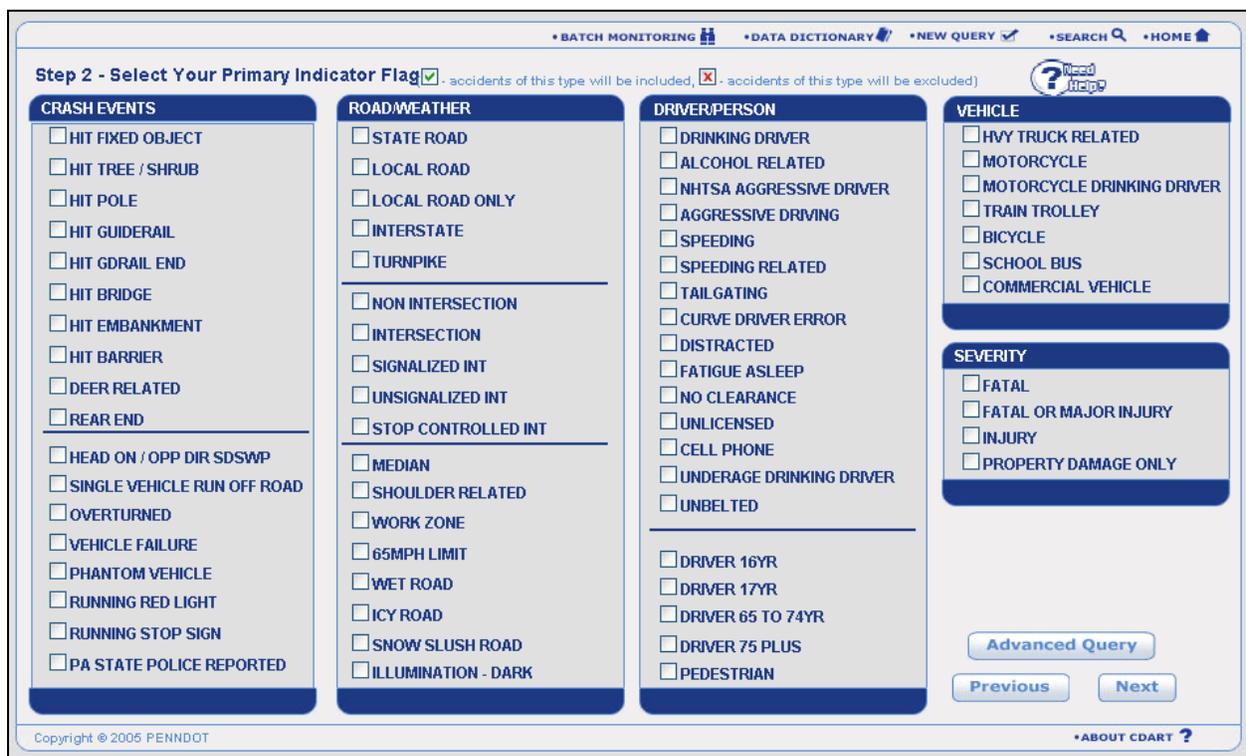


Figure 14. Screenshot. Sample crash application filters for performing analysis.

(Source: GeoDecisions, Inc.)

Spatial safety analyses can also examine area wide associations. In a recent study, Middleton et al. (2014) used spatial analysis to determine if motorcycle crashes are spatially associated with traffic volume. Traffic counter locations and motorcycle crash locations combined to provide a spatial weighting factor that measured the linear distance between crash sites and the nearest count location. Motorcycle counts and total counts spatially correlate with the weighted crash counts, indicating that traffic volume is a reasonable predictor of crash locations. This gives an area-wide indication of the spatial correlation that States could use to determine where to place count stations in order to get a more accurate picture of motorcycle traffic volume based on the historic crash frequency and locations.

CHALLENGES

- **User feedback:** Integrated data systems for State and local asset management and safety analysis support a broad range of users. A challenge from a data management perspective is when more users of the data result in more people “touching” the data and trying different ways to extract, translate and load data. That activity leads to uncovering data deficiencies, however, it also builds a natural constituency for data improvement efforts. The broader the user base, then the more support there is for investing in data quality.
- **Access to metadata:** Users vary in their analytic capabilities, knowledge of data management practices, and level of sophistication when using and interpreting data or analytic results. System designers have to take this user variability into account and the result is often a tiered approach to user access, support, and knowledge sharing. For example, designers may constrain the system’s crash data analysis capabilities so that users avoid common errors such as mistakenly counting people or vehicles instead of crashes (or vice versa). To serve the needs of more advanced users, designers might offer a higher-tier analytic tool, or provide data extracts so that users can use their own analysis software. In the most comprehensive systems, users have the option of learning more about each data element and analytic function supported in the system. For those who opt for a data extract, system documentation explains the data structure and the definition of each data element (i.e., a data dictionary). Users with lower levels of data sophistication might not be able to interpret a data dictionary or to find the relevant information for their purposes. A more useful type of metadata for these users describes in operational terms how each data element is collected—an operational definition.

When integrating data, metadata for each component system and for the data sources in combination takes on added importance. Even when each data source is *internally* consistent, the ability to sync data among multiple sources depends a great deal on when the data were collected, how they were collected, and how reliable each data

element is. For example, analyzing local intersection safety could rely on geometric data from multiple agencies, each used to collecting the data in their own way. Simple definitional items like where does the intersection begin and end, how are skew angles measures, and are horizontal and vertical alignment provided may differ between local agencies. Analyzing integrated data on intersection features and their relationship to safety requires that the analyst know something about the multiple source files and their contents. Metadata will tell users if they can construct a valid analysis or need to make adjustments to the data before drawing conclusions.

RECOMMENDED SOLUTIONS

- **User feedback:** One way system designers and data custodians can serve users' needs during the extract, translate, and load process is by providing a convenient, well-managed feedback utility. Users encounter data gaps and errors that system managers and data custodians need to know about. A feedback system allows users to provide specific, record-level and data-element level information about any problems they encounter. The most capable user feedback systems track these problem reports based on date submitted, person who submitted the issue, the nature of the problem, which staff is assigned to address the problem, and when and how the problem was resolved. Ticklers alert managers to problems that remain unresolved after their projected resolution date. Close-out reports go back to the original user to ensure that the resolution meets the user's needs. All of this information rolls up into a periodic management report that identifies most frequent errors so that they can be brought to the attention of data collectors and become part of the formal data quality management process. That process also works in harmony with the data governance processes established for the system to log data alterations and ensure the changed records propagate appropriately to all other related systems.
- **Access to metadata:** All users should have access to metadata describing any known quality problems with the data. Some systems provide a pop-up warning whenever a user selects specific data elements for analysis. For example, in safety analyses the proportion of vehicle occupants wearing a seatbelt at the time of the crash is considered inaccurate because it is based on self-reports for all but seriously or fatally injured people—a small percentage of all crash-involved vehicle occupants. If a crash data user selects the variable coding for occupant protection device use, it is appropriate to provide metadata about the reliability of that data element.

LESSONS LEARNED

The case studies and pilot studies generated the following lessons learned about data extract, translate, and load processes:

- **ETL processes are a good source of data quality information.** When analysts attempt to pull data from an integrated database, they will often have a rough idea of how many cases they should end up with given their particular query filters. They also may have a specific set of cases in mind that they can look for in the resulting extract. Transformation is a software process that will readily spot certain types of data errors (missing values, values out of range, data type violations, and database key errors). The load process may, depending on the analytic software, also result in a useful error list.
- **Spatial and non-spatial data extraction are similar.** Except for the ability to select cases based on map coordinates and visualization, the spatial and non-spatial data extraction techniques are identical. The user gives the system inclusion and exclusion criteria, and the resulting extract should match those selections. The power of GIS for data extraction is for spatial data visualization. The ability to select cases based on spatial distances is often useful and terribly difficult to achieve any other way.

SUMMARY

Step 7: Develop and Deploy the Extract, Transform, and Load Process addresses the need for processes and tools to move integrated data from the resource file(s) into analytic software. The important items in this step are:

- **Extract.** This is the process of developing an analytic subset from the full, data resource(s). It includes the following activities.
 - **Filter or select cases.** This a process of identifying the use cases for an analysis based on a set of criteria. Filters and selection criteria are specific values of data elements that define which cases are included in the file for analysis. They may be date ranges, locations, values of a variable (e.g., median present), a range (e.g.,

CHECKLIST

- Extract
 - Filter or select cases.
 - Create data extract.
 - Exclude cases.
 - Join and filter.
 - Filter spatially.
- Transform
 - Program software tool to transform data to meet the needs of the analytic software.
- Load
 - Load the translated data into the analytic software tool.
- User Feedback
 - Develop utility to collect feedback from users.
 - Ensure that relevant metadata are provided to users.

horizontal curve = 5 to 15 degrees). Multiple selection criteria may be used. Logical AND and OR operations may be applied to arrive at the final selection criterion.

- **Create data extract.** This process results in the data file for analysis. Analysts should specify which data elements they want in the file and what format and file type to use.
- **Exclude cases.** These are the opposite of use cases. The data extract should be devoid of any cases that meet predefined exclusion criteria.
- **Join and filter.** Database files may be joined and further filtered at this step. It is essentially the application of further selection criteria applied across multiple tables within a database.
- **Filter spatially.** Spatial filtering is listed as a separate step because it may be accomplished using different tools than those used to create the subset of use cases as described in the preceding steps. In spatial filtering, the analyst may select an area on a map to determine which cases are in the final analytic data file.
- **Transform.** This is the process that translates the native data file's format and file type into the format and file type required by the analytic software. The analytic software will support one or more specific import file types and a set of expected data elements.
 - **Program software tool to transform data to meet the needs of the analytic software.** This step assumes that the database management tools used to create the integrated data resource and data extract may not have the capability to create the exact file required for import into the analytic software. If needed, the data integrator will need to program a tool that automates the data transformation process.
- **Load.** This is the process of loading the import file into the data analysis software tool.
 - **Load the translated data into the analytic software tool.** This may require multiple attempts as the data file may not pass software edit checks the first several times. Eventually, as the data integrator reviews data errors with the partners, and the partners address the errors, a clean data file will be accepted into the analysis software.
- **User Feedback.** This is the process of documenting users' experiences with the new data resource and the analytic tool or its output. User feedback can help identify data gaps, training needs, and needs for enhanced analysis tools. User feedback should be shared with the data governance group and incorporated into updated versions of the data integration plan and the training plan.

- **Develop utility to collect feedback from users.** Ideally, users will have an automated way to leave feedback so that their communications are logged, a tickler is set up for any requested actions, and support staff can easily review and respond to questions or requests for assistance. The feedback should be sorted by type of issue and stored for reference in updates to the plans.
- **Ensure that relevant metadata are provided to users.** The data governance group is responsible for defining the metadata for the overall data integration effort. The resulting integrated data resource should have its own set of metadata. A subset of this information is important for users to know as they attempt to use the data or the results. In particular, users should know about the timeliness, accuracy, completeness, and uniformity of the data. They should also know the strength of each match in the final integrated data resource.

STEP 8 – CONDUCT ANALYSES

OBJECTIVES

With data extracted, transformed, and loaded into the relevant data systems, target agencies are able to conduct safety analyses with newly integrated data. The activity of conducting analyses is the eighth step in the safety data integration process shown in Figure 15. Analyses include traditional safety analyses, advanced analyses such as those in the Highway Safety Manual, novel analyses that an agency might create for its own purposes, and analyses aimed at making safety information available for decision-making in other areas beyond the core safety management area. These other areas can include planning, roadway design decisions and exceptions, maintenance, and traffic operations.

Collecting accurate roadway and crash data, integrating all statewide data into a central database, and linking data spatially facilitates the analysis of these data to determine the most effective approach to improving road safety. The ultimate goal of this process is to save lives and reduce the number of severe injuries.

The key to effective and efficient safety data analysis is choosing the appropriate analysis. In other words, there is a need to “right-size” the analysis to match project needs. For example, it is unlikely that an agency would spend a substantial amount of time and money to analyze the safety impacts of proven, low-cost countermeasures such as rumble strips. Conversely, an agency may perform an in-depth analysis as part of an interchange justification report to demonstrate the safety performance of different alternatives.

This chapter provides an overview of the processes and the tools available to target agencies identify safety issues and determine the best approach to address or mitigate them.

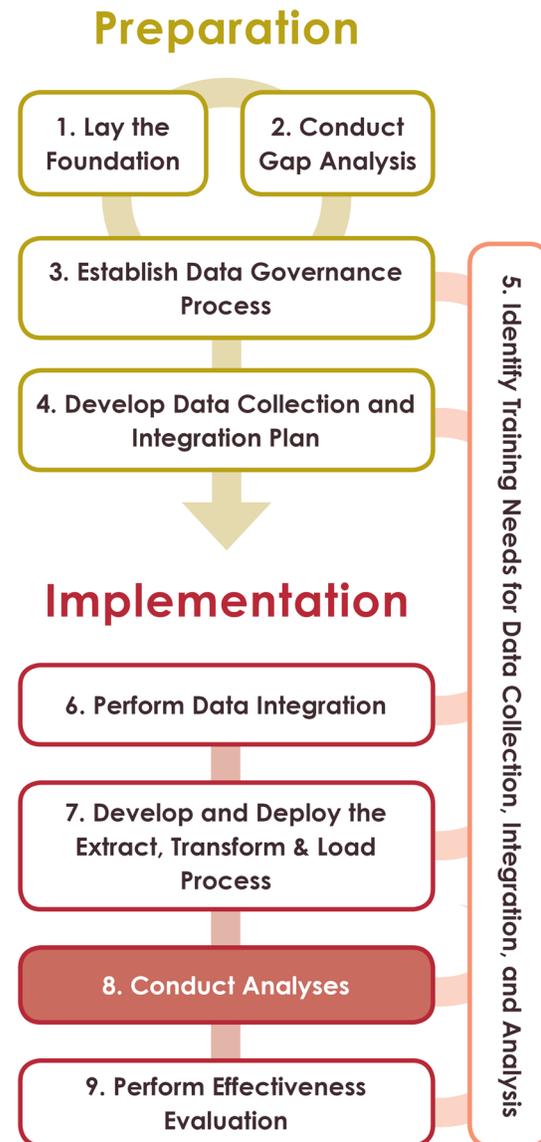


Figure 15. Chart. Step 8: Conduct Analyses.

Types of Safety Analyses

Each of the project stages of planning, design, and operations have distinct safety analyses. Safety analyses can also be identified based on each part of the six-step safety management process described in the *Highway Safety Manual*, which is shown on the left in Figure 16.⁽⁶⁾ The safety management process is not meant to be a standalone process; it can be integrated into the regular planning, design, and operations processes within a target agency as shown on the right of Figure 16. Other resource documents, including the *Highway Safety Improvement Program (HSIP) manual* define the steps differently; however, the depiction in Figure 16 helps to show that the process is cyclical and points to the activities required to manage safety at the State and local level.⁽¹⁰⁾

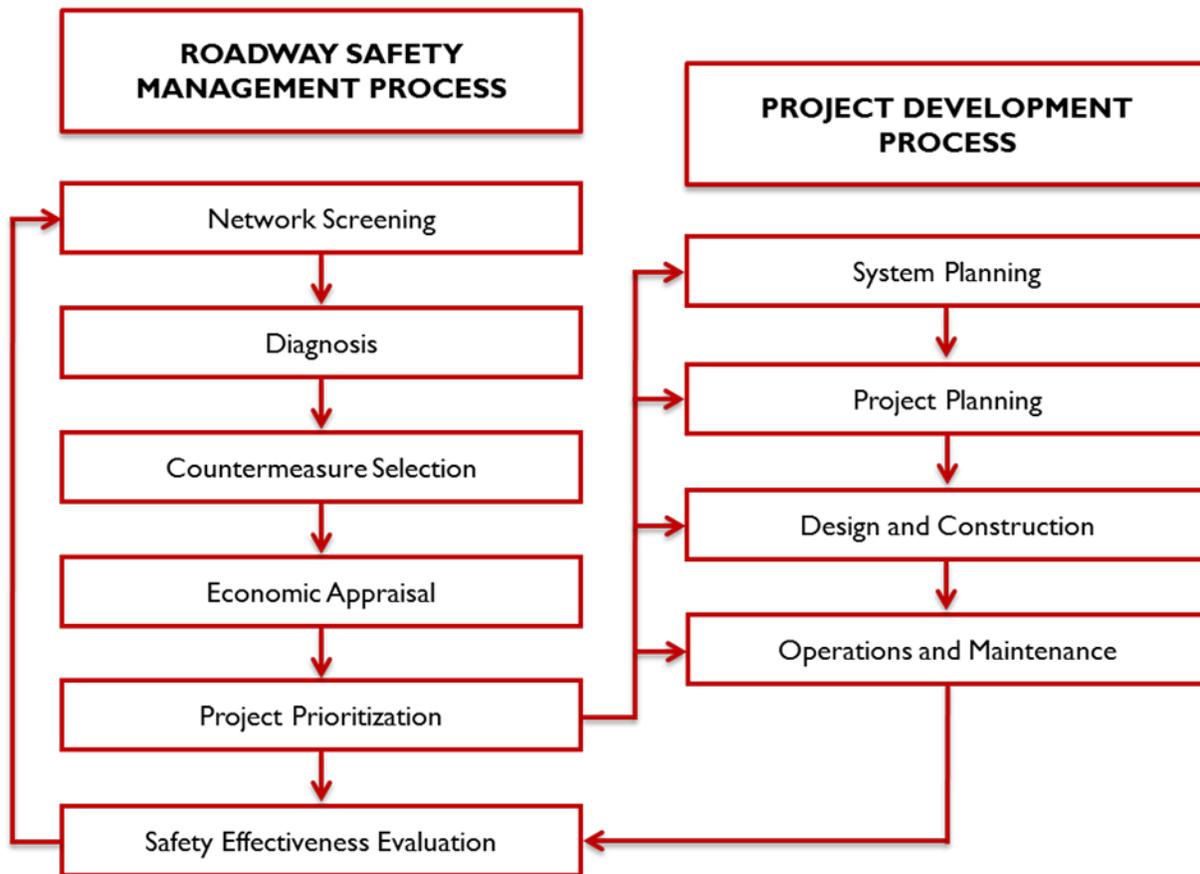


Figure 16. Chart. Integrating the roadway safety management and project development process.

(Source: Reliability of Safety Management Methods: Countermeasure Selection. FHWA: <http://safety.fhwa.dot.gov/rsdp/downloads/fhwasal6039.pdf>)⁽¹¹⁾

Beginning with network screening, which takes place during project planning, safety analysts identify candidate locations for safety improvement. Using advanced methods, such as those covered in the *Highway Safety Manual*, analysts compare among similar locations to develop lists

of sites with higher-than-expected crash frequency or severity. In the diagnosis step, which also takes place during project planning, analysts investigate the types and patterns of crashes and identify the most likely problems at each of the candidate locations. As part of project design, engineers incorporate countermeasures in their designs to address or mitigate the identified problems and crash types. The use of data-driven safety analysis can help justify design decisions and exceptions. Economic appraisal typically takes the form of a benefit/cost ratio, comparing the dollar value of lives saved and injuries avoided to the cost of implementing the countermeasure. Alternatively, analysts may consider the magnitude of project benefits such as total crashes reduced, injuries prevented, or lives saved. In project prioritization, decision-makers compare benefit/cost ratios of the various projects in combination with other factors such as environmental, operational, and cost considerations as well as political influence and agency goals to arrive at a final selection of projects for funding. After project implementation, analysts conduct safety effectiveness evaluations to inform future decisions. The *Highway Safety Manual* describes multiple methods of program and project evaluation.⁽⁶⁾

How to Choose the Right Tool?

FHWA's resource called *Scale and Scope of Safety Assessment Methods in the Project Development Process*⁽¹²⁾ is designed to help analysts identify HSM safety assessment methods for various project applications. Each chapter is focused on a project development phase (planning and scoping, environmental analysis, alternatives identification, design) with relevant safety assessment examples for the phase. Choosing the right-sized analysis will help agencies to allocate resources strategically and avoid unnecessary or excessive analysis costs.

The FHWA *Roadway Safety Data and Analysis Toolbox* describes safety analysis tools to help agencies implement advanced analytic techniques as described in the *Highway Safety Manual*.^(6, 13) The Toolbox provides detailed descriptions of a wide selection of tools including commercial products. The Toolbox will be updated with new tools over time. The Primer and its How-To Video for the Toolbox provides a tutorial on using the Toolbox, and also helps users select an appropriate tool based on their desired activity and data capabilities. Appendix C provides instructions on how to use the Toolbox.

CHALLENGES

Aside from choosing the right-sized analysis, there are additional challenges and limitations associated with the adoption of roadway and safety analysis tools as adapted from FHWA's *Traffic Analysis Tools Data Integration Primer*.⁽¹⁴⁾

- **Limited empirical data.** While data integration efforts may improve the availability of empirical data for certain tools, data may still be lacking for certain analyses. Data

collection can often be a costly component of a study. The best approach is to look at the ultimate goals and objectives of the analysis and focus data collection on critical data.

- **Limited funding.** Limited funding for conducting the study, purchasing tools, running analytical scenarios, and training the users is often a consideration with data analysis. The analysis tools may be costly. Software licensing and training fees can also add to that cost. The analysis of more scenarios also costs money and time. With cost constraints, it is critical to identify the point of diminishing returns for the investment.
- **Training limitations.** Some tools have steep learning curves and, as a result, some practitioners do not receive adequate training or lack the time to learn a new program.
- **Data input and the diversity and inconsistency of data.** Each tool uses unique analytical methodologies, so the data requirements for analysis can vary greatly from tool to tool.
- **Lack of understanding of the limitations of analytical tools.** Often, limitations and “bugs” are not discovered until the project is underway. It is important to learn from experiences with past projects or to communicate with fellow users of a particular tool or tool category in order to assess the tool’s capabilities and limitations.
- **Lack of features.** Some analytical tools are not designed to evaluate the specific strategies and countermeasures the users would like to implement. Tools should be flexible so that advanced users may customize the tools.
- **Tendency to use simpler analytical tools.** Because of lack of resources, past experiences, or lack of familiarity with other available tools, agencies may prefer to use a tool currently in their possession, even if it is not the most appropriate tool for the project.
- **Long run times.** Depending on the computer hardware and the scope of the study (i.e., area size, data requirements, duration, analysis time periods, etc.), an analytical model run may range from a few seconds to several hours. The most effective approaches to addressing this issue involve using the most robust computer equipment available and carefully limiting the study scope to conform to the analytical needs.

RECOMMENDED SOLUTIONS

The Primer on Safety Data and Analysis Toolbox provides a stepwise process to identify an appropriate tool to support a given analysis task based on the user’s needs and capabilities. By following the steps, the user will consider the different types of questions encountered during safety data analysis identification such as how to conduct network screening, how to estimate the safety impacts of design exceptions, and how to evaluate the performance of improvement, among others. The steps also help the user understand the roles, responsibilities, and tasks

supported by the tools in the Toolbox. A number of the challenges associated with adopting analysis tools can be mitigated by following the steps in the Primer:

Step 1: The user identifies his/her role and responsibilities based on those presented in the Primer.

Step 2: The user selects a specific task that best describes his/her task to be accomplished.

Step 3: The user reviews a list of relevant safety data and analysis tools based on the self-identified task at hand. The user explores or refines the list of relevant safety data and analysis tools via the Toolbox, following one of two approaches:

a. **Explore:** The user explores the various available tools from the Toolbox, reading the detailed summaries for each applicable tool. As part of the exploration, the user will learn about the general types and capabilities of tools to complete their task as well as the system and data requirements to employ the tools. The user then considers his/her existing data capabilities and resources with respect to the specific system and data requirements for each tool in order to select the most applicable tool(s) from the Toolbox.

b. **Refine:** The user proceeds to the Toolbox with the intent of refining the potential list of tools. The user may choose this option for one of many reasons (e.g., the list of potential tools is too long to practically explore each one). Table I presents a list of high-level categories (referred to as "tags") that are available in the Toolbox as filters to help refine the list of applicable tools. The Primer provides a list of all tools and the applicable tags. The user can identify the relevant tags for each tool and can then enter those tags to search the Toolbox for related tools. The user can apply additional tags as desired to further refine the list of tools.

By following this three-step process, the analyst is able to select tools that use quality integrated data, fit within the capabilities of its users, and have a feature set that allows for desired analyses to be completed.

Agencies should consider training needs associated with the adoption of tools. The FHWA Resource Center and State LTAP/TTAP offer training for a number of advanced analytical tools. Also, an agency will need to identify budgetary constraints that will narrow the choice of tools to a smaller subset. It is also important to note there are hardware constraints that may limit the performance of some tools, therefore, long-term investments may need to be made to ensure adequate computing performance.

LESSONS LEARNED

The project case studies each included analyses performed using the integrated data. Of the pilot studies, the Arizona DOT and Indiana LTAP projects provide some examples of analyses conducted using the integrated data. Arizona DOT used AASHTOWare Safety Analyst™ on roadway segment data for network screening. Indiana LTAP used Putnam County data in Esri ArcGIS analysis tools to analyze crashes and generate heat maps. These examples are presented in Appendix A. Together the case studies and pilot studies provided lessons learned with respect to safety analysis using integrated data:

- **Analytic users may be found at all agency types.** In Michigan, almost all of the local analysis is conducted by local users accessing their own data. In Wisconsin, analysis takes place at the local and State levels, depending on the type of analysis. Asset management analyses are local; safety analyses are done by the State. Ohio and Tennessee both conduct analyses primarily at the State level. The Indiana LTAP pilot study was focused on a single county and the LTAP and county were the locus of the analyses. In Arizona, the State DOT conducted analyses.
- **Agencies need help deciding which tools to test and adopt.** The pilot studies demonstrated clearly that target agencies safety practitioners do not always know what tools are available to them, or what the capabilities and requirements of the tools are. FHWA and others have developed guides describing the various tools. Making the decision of which software tools to adopt is prone to pitfalls. If data integration is delayed, planned demonstrations of the use of integrated data may be put on hold, indefinitely. Getting support of upper management for significant software purchases can be difficult if the data are found to be of insufficient quality.
- **Changing analytic tools changes what has to be accomplished in many of the other steps in the integration process.** Portions of the extract, translate, load process are software dependent. Some of the training needs relate to using a specific software analysis tool and understanding its output reports.

SUMMARY

Step 8: Conduct Analyses addresses the need for advanced analyses for the project development process and the safety management process elements of network screening, diagnosis, countermeasure selection, economic appraisal, and evaluation. The important items in this step are:

CHECKLIST

- Review the Primer on Safety Data and Analysis Toolbox and the Scale and Scope of the HSM.
- Conduct analyses and produce analytic reports. Support users.

- **Review the Primer on Safety Data and Analysis Toolbox and the Scale and Scope of Safety Analysis in the project development process.** The Toolbox has useful descriptions of a broad range of analytic tools. The Scale and Scope document features analysis approaches applicable for a particular project development phase. Agencies should select those tools that would work best for their needs taking into account the quality and availability of specific data as required by the software.
- **Conduct analyses and produce analytic reports. Support users.** This step is when analysts (and perhaps general users) access the analytic software to obtain output. Analysts and users must understand the outputs from the software in order to make efficient and valid use of it. Support and training are necessary components of this step. User feedback (discussed in Step 7) is also important here.

STEP 9 – PERFORM EFFECTIVENESS EVALUATION

OBJECTIVES

Performing effectiveness evaluation for data integration, not the evaluation of individual projects, is the ninth and final step in the safety data integration process shown in Figure 17. Target agencies have a vested interest in making quantifiable safety improvements. Effectiveness evaluation includes activities designed to tell practitioners which activities paid off by reducing crash frequency and severity. For data integration efforts, practitioners will want to make the connection between the integrated data files and improved safety decisions. Decision makers will want proof to show that the integration efforts were worthwhile and how their agencies could benefit from their expansion.

CHALLENGES

Figure 2 in the earlier overview section displays the logical relationship between data and improved safety as presented by the Data Driven Safety Analysis process.

The barrier to proving the link for integrated safety data is, in part, a lack of multi-year projects where integrated data have been in use for a period of time. Eventually, it should become feasible to quantify the links shown in the figure. That analysis is of interest for all target agencies because it will help to justify spending on improved data and more extensive data integration.

From a target agency perspective, the ideal would be to quantify precisely the impact of having integrated target agencies safety data on overall safety (crashes, injuries, and fatalities). The capability to develop that quantitative analysis does not exist today, but that is not for a lack of an ability to make the link between better data, better decisions, and ultimately improved

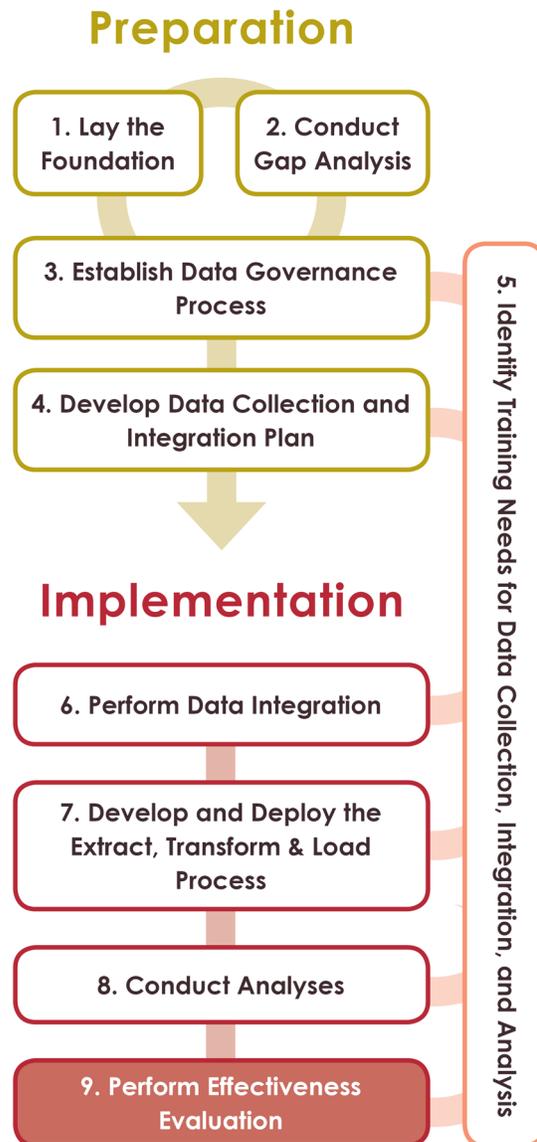


Figure 17. Chart. Step 9: Perform Effectiveness Evaluation.

safety. The hurdle to be overcome is in having sufficient examples where integrated data were used in advanced analyses and where the result of the analysis pointed to opportunities for improved safety beyond what we would have expected using the non-integrated data and older methods of analysis. It will take time to build those use cases.

RECOMMENDED SOLUTIONS

FHWA's *State, Tribal, and Local Safety Data Integration Strategic Plan* is an internal document that expresses the agency's goals for evaluating nationwide adoption of data integration methods, tools, and analyses.⁽¹⁵⁾ For users of this informational guide, the important items are addressed in the Strategic Plan in the following measures of success:

1. Use of predictive methods on all public roads.

Ultimately, each target agency should increase its proportion of road segments and intersections with complete data in the statewide databases, increase its proportion of road segments and intersections included in one or more advanced safety analyses, and make greater use of advanced safety analyses in project-level decision-making.

2. Number of agencies included in data sharing agreements.

This should include all target agencies that manage any portion of the roadway network. Data sharing may include data collection and data maintenance for roadway location spatial data, roadway inventory, traffic volume, crash data, and/or a specified set of asset management data.

3. Implementation of data governance processes for safety and asset management data.

Ideally, target agencies will have fully documented each of the key data sets (roadway inventory, traffic, crash, and asset management), and subject these datasets to a set of data standards (verified by a formal quality control process). Additionally, a formal data governance process will manage changes to those systems.

LESSONS LEARNED

- Target agencies will need more years of data and advanced analyses before they can precisely quantify the benefit of integrated data for improved safety.
- The adoption of noteworthy practice recommendations is potentially a way to measure the success of data integration.

SUMMARY

Step 9: Perform Effectiveness Evaluation addresses the need to quantify improvements in order to demonstrate the impact of data integration programs on safety. The important items in this step are:

- **Increase the use of predictive methods as data integration processes mature.** This is a reasonable goal for safety practitioners working with integrated data resources. Target agency adoption of rigorous analyses using predictive methods promotes data driven decision-making, more efficient resource allocation, and ultimately, reduced fatalities and serious injuries.

Increase the number of agencies included in data sharing agreements and data governance processes. This informational guide focuses on the three core safety data sets of crash, roadway inventory, and traffic volume data. There are several additional data sources (driver, vehicle, citation and adjudication, and injury surveillance) that could be used to enhance the integrated data resources at a target agency. Integrating more data will mean bringing in more partners including those who can supply, manage, or use the additional resources.

CHECKLIST

- Increase the use of predictive methods as data integration processes mature.
- Increase the number of agencies included in data sharing agreements and data governance processes.

CONCLUSION

This Informational Guide presents information on how to complete a safety data integration project. Throughout, the data integration process has been described as a series of nine discrete steps. It is clear, however, that the steps must interact. Figure 18 shows a schematic of the nine steps and their relationships.

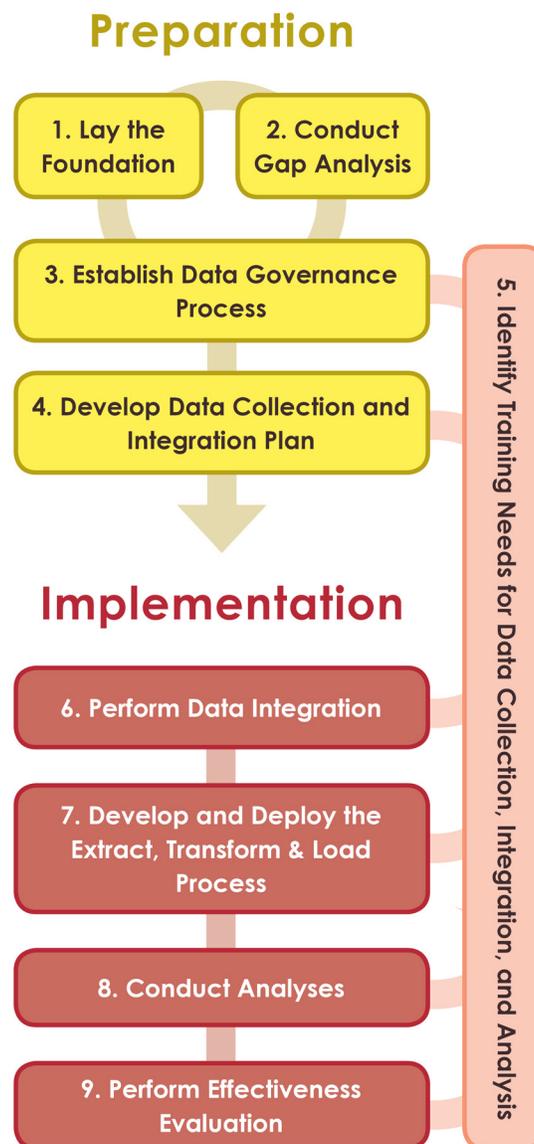


Figure 18. Chart. Interactions among the steps in data integration.

From this view, there are four steps involved in preparing to conduct a data integration. Three of those steps form an inter-related cycle comprised of forming partnerships, identifying gaps,

and managing a formal data governance process, all of which feed into the fourth step of creating data integration plan.

Once the partner agencies have an integration implementation plan, they are ready to perform the integration, conduct analyses, and (eventually) evaluate effectiveness. All along the way, training needs are likely to be identified. These may be technical in nature and relevant to only a small audience (e.g., those who run the ETL process) or they may be quite large, potentially including members of the public who have access to the data and online analytic tools.

Throughout this Informational Guide, States, Tribes, and local agencies are encouraged to work cooperatively using a formal process to meet their needs for safety data and advanced analyses. Only through partnerships will data integration happen. Partnerships are necessary for its maintenance as well.

There are many paths to integrated safety data. This Informational Guide is impartial with respect to which method(s) are best. States, Tribes, and local agencies are encouraged to work together to decide what is best for their situation and capabilities.

BENEFITS OF USING THIS INFORMATIONAL GUIDE

States, Tribes, and local agencies may use this informational guide to help them plan their safety data integration efforts in a logical, step-wise manner. To make the guide as useful as possible, it is organized around the nine steps identified through case studies and pilot studies developed throughout the project. As shown, those steps are not stand-alone—they are part of a larger whole and interact. This guide presents them as separate steps so that partner agencies can identify their own position in the nine-step process and find useful examples as efficiently as possible.

The ultimate goal of the data integration effort is to reduce crashes, fatalities, and serious injuries. By making better data available to decision makers, the data driven decision-making framework implies that agencies will make more efficient spending decisions which in turn results in more lives saved per dollar spent on safety. It is also important to recognize that this view of safety decision-making extends to efforts in Planning, Design, Asset Management, and other business areas of a DOT, not just those specifically aimed at improving safety. Integrating data is an important process because it helps multiple business units incorporate safety into their own decision-making. This guide focuses on integrating safety data; however, the partner agencies are encouraged to develop plans that best fit their needs for all data for decision-making.

REFERENCES

1. FHWA Data Integration Primer. Publication FHWA-IF-10-019, Federal Highway Administration, Washington, D.C. Available online at: <https://www.fhwa.dot.gov/asset/dataintegration/if10019/if10019.pdf>
2. FHWA Guidance on State Safety Data Systems (2016). Federal Highway Administration, Washington, D.C. Available online at: http://safety.fhwa.dot.gov/legislationandpolicy/fast/ssds_guidance.cfm
3. Cambridge Systematics, Inc. (2010). "Target Setting and Data Management to Support Performance-Based Resource Allocation by Transportation Agencies." NCHRP Report 666, Transportation Research Board, Washington, DC.
4. Sawyer, M., Lefler, N., Soika, J., Carter, D., Scopatz, R., Gross, F., Rothenberg, H., Miller, J., Bahar, G., & Eccles, K. (2012). United States Roadway Safety Data Capabilities Assessment (No. FHWA-SA-12-028).
5. Lefler, N., Council, F., Harkey, D., Carter, D., McGee, H., & Daul, M. (2010) Model Inventory of Roadway Elements – MIRE, Version 1.0. (No. FHWA-SA-10-018). Available online at: http://safety.fhwa.dot.gov/tools/data_tools/mirereport/mirereport.pdf
6. National Research Council (US). Transportation Research Board. Task Force on Development of the Highway Safety Manual, Transportation Officials. Joint Task Force on the Highway Safety Manual, & National Cooperative Highway Research Program. (2010). Highway Safety Manual (Vol. 1). AASHTO.
7. Fiedler, R., Eccles, K., Lefler, N., Fill, A., & Chan, E. (2013). MIRE Fundamental Data Elements Cost-Benefit Estimation. (No. FHWA-SA-13-018). Available online at: http://safety.fhwa.dot.gov/rsdp/downloads/mire_fde_cbe_finalrpt_032913_updated.pdf
8. Scopatz, R., Lefler, N., & Peach, K. (2015). State Traffic Records Coordinating Committee Noteworthy Practices (FHWA-SA-15-083). Available online at: <https://www.transportation.gov/government/traffic-records/state-traffic-records-coordinating-committee-noteworthy-practices>
9. The FHWA Data Governance Plan. Federal Highway Administration, Washington, D.C. Available online at: <https://www.fhwa.dot.gov/datagov/dgpvolume%201.pdf>
10. Herbel, S., Laing, L., & McGovern, C. (2010). Highway Safety Improvement Plan (HSIP) Manual. FHWA-SA-09-029, Federal Highway Administration, Washington, D.C.
11. Reliability of Safety Management Methods: Countermeasure Selection.. Publication FHWA-SA-16-039, Federal Highway Administration, Washington, D.C.
12. FHWA Scale and Scope of Safety Assessment Methods in the Project Development Process. Publication FHWA-SA-16-106, Federal Highway Administration, Washington, D.C.
13. Gross, F., & Bahar, G. (2015). Primer on Safety Data and Analysis Toolbox. FHWA-SA-15-051, Federal Highway Administration, Washington, D.C. Available online at: <http://safety.fhwa.dot.gov/rsdp/toolbox-home.aspx>

14. Traffic Analysis Toolbox. (FHWA-HRT-04-038). Federal Highway Administration, Washington, D.C. Available online at:
http://ops.fhwa.dot.gov/trafficanalysistools/tat_vol1/index.htm
15. Scopatz, R. (2016, in preparation) State, Tribal, and Local Safety Data Integration Strategic Plan Federal Highway Administration, Washington, D.C.

APPENDIX A – PILOT STUDY REPORTS

PILOT STUDY: ARIZONA DEPARTMENT OF TRANSPORTATION STATE AND LOCAL SAFETY DATA INTEGRATION

In this initiative, the project team developed automated scripts and database objects to extract, transform, and load data from ADOT and local agency sources into an AASHTOWare Safety Analyst™ minimum data schema. The project team built the extract, transform, and load processes with the intention of providing the end-user an appropriate level of control over the process. A business user can run the entire process in two simple steps: one step to extract and manipulate the data, and a second step to create files in the format required for import into Safety Analyst.

Safety Analyst has very stringent data validation controls. The project team found that only a subset of the data in all the data sources used in this project qualifies for use in Safety Analyst analytical tools.

The data sources within ADOT were the ATIS Roadway Inventory database (ATIS) and Safety Data Mart (SDM) databases. In addition, agencies providing data for this project were Maricopa Association of Governments (MAG), Pima Association of Governments (PAG), Bureau of Indian Affairs (BIA), and Navajo Division of Transportation (DOT). This summary provides details on the sources of data, design and execution of the extract, transform, and load process, project outcomes, and barriers to integration encountered during the project.

Source Data

The project team reviewed the available data from participating agencies and identified the useable data for this effort. The ATIS database—which includes MAG data—records traffic volume for about half of the segments. Ramp segments were excluded because traffic volume data are not recorded in ATIS at the time of this project. The PAG data set had no traffic volume data, and Navajo DOT had traffic volumes for about 15 percent of segments. Since only roadway segments with recorded traffic volume data qualify for import into Safety Analyst, these local data sets were not usable. Navajo DOT data is recorded in a location referencing system that differs significantly from ATIS. BIA data is not geo-referenced in any usable location system. Therefore, neither of these data sets qualified even when they contained all three safety data components.

Table 5. Data source summary.

Data Source	Location System	Traffic Data?	Crash Data?	Data Structure
ATIS	Route/Milepost	Yes – About half of roadway segment mileage	Yes – SDM	MS SQL database
PAG	Route/Milepost	No	No	Geodatabase
MAG	see ATIS	see ATIS	Yes – SDM	see ATIS
Navajo DOT	Route/Section/Distance	Yes	Yes	Geodatabase
BIA	Route/Section	Yes – About 15% of segments	No	Excel

OVERVIEW OF THE EXTRACT, TRANSFORM, AND LOAD PROCESS

The two key business drivers that steered the project team’s design approach are as follows:

1. To the extent possible, the process must use a common and transparent software tool set and must be transportable to different deployment environments.
2. Given the necessary software tool set, the process must be simple for the user to execute.

The objective for this project was to build the extract, transform, and load process model with which the AASHTOWare Safety Analyst™ user, with minimal training, can clearly see and understand the data transformation logic, and can execute the process without assistance of a programmer.

To accomplish this objective, the project team designed the extract, transform, and load process as a series of structured query language (SQL) views and spatial query operations controlled by Python scripts. A view is a pre-defined SQL query that is stored in the database management system. The view defines an SQL SELECT statement, which can select and manipulate data from tables and other views in the database.

Technical Platform

Figure 19 illustrates the general design of the extract, transform, and load process:

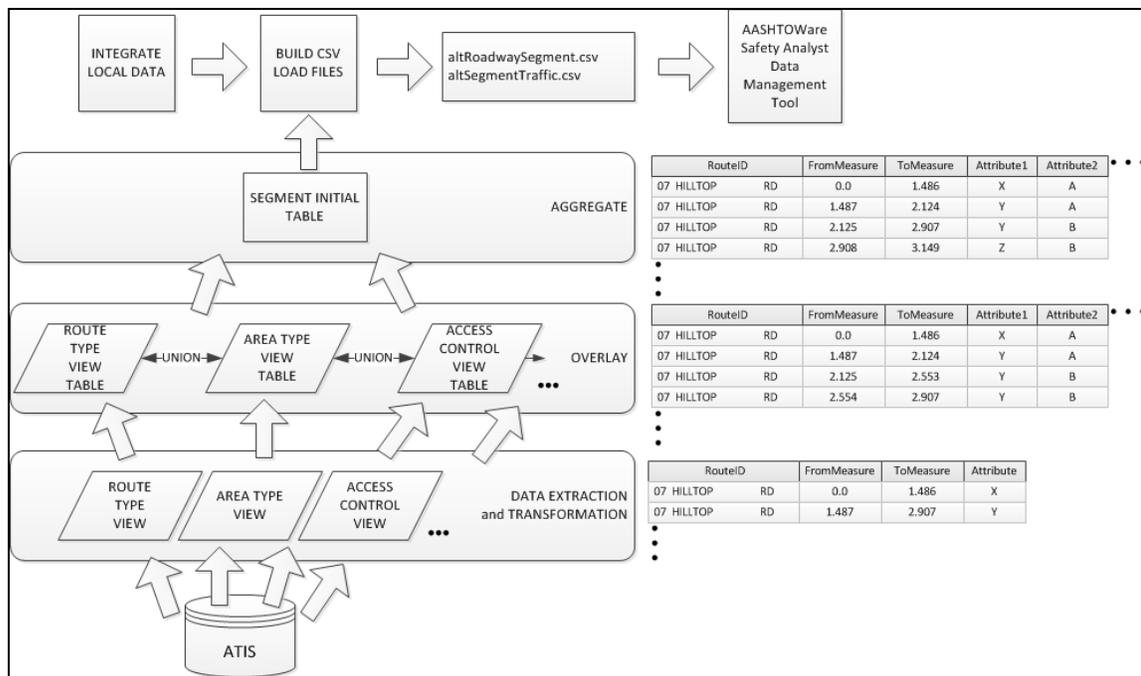


Figure 19. Chart. Process overview.

Data Extraction and Transformation

The extract, transform, and load process design uses a consistent methodology that is simple for the user to manage, and contains the detailed logic to extract and transform ATIS data at the elementary level, and then consolidate and format the data for loading into the AASHTOWare Safety Analyst™ database structure.

Overlay Route Events

Different features can change at different frequencies and at different locations along a given roadway segment. For example, Access Control tends to stay constant over many miles, while Area Type varies several times between urban and rural over the same distance; and median type may change with even greater frequency.

The overlay step merges the event segments for each attribute, producing an output event table that represents the union or intersection of the inputs.

Aggregate Contiguous Route Events

In some instances, the preceding steps can result in consecutive segments with identical attributes. The aggregate step eliminates the extra records by combining the consecutive records into one event segment with FromMeasure and ToMeasure values adjusted to span the length of the combined segments.

Integrate PAG Data

A separate process prepares the PAG data. Using the PAG Geodatabase for input, the process performs a spatial intersect with ATIS to select only the PAG road segments that are **not** already recorded in ATIS. Once this operation is complete, the same “aggregate” operation as described in the preceding paragraphs is executed to combined consecutive identical segments.

Build Load Files

The final step takes the SEGMENTINITIAL and SEGMENTINITIALPAG tables as input, and produces output files in comma-separated values (.csv) format. This step also performs some final data filtering – When there are attribute data fields that contain a value that represents “Other” or “Unknown”, the process assigns a default data value that is meaningful in AASHTOWare Safety Analyst™. This practice cuts down on the number of records flagged in the Safety Analyst Data Management Post-Processing step.

System Requirements

The project team developed the extract, transform, and load process on the ADOT GISSQLPROD server in Microsoft SQL Server Management Studio 2012. Microsoft SQL Server Management Studio is the only software program that ADOT personnel will need to maintain and to run the extract, transform, and load process.

END-USER INSTRUCTIONS

The last step—inputting the final result table and creating the .csv files—is relatively simple and not particularly taxing on system resources. Instructions are given to explain how the end-user can run both sub-processes without any special software tools.

PROCESS DETAILS

Users can perform SQL queries on SEGMENTINITIAL (and SEGMENTINITIALPAG) to examine or audit the data output by the extraction, transformation, overlay, and aggregation operations prior to loading the data into the AASHTOWare Safety Analyst™ database.

ACCIDENT DATA IMPORT

The project team configured AASHTOWare Safety Analyst™ Data Management Tool, shown in Figure 20, to import the required accident data elements directly from the SDM database.

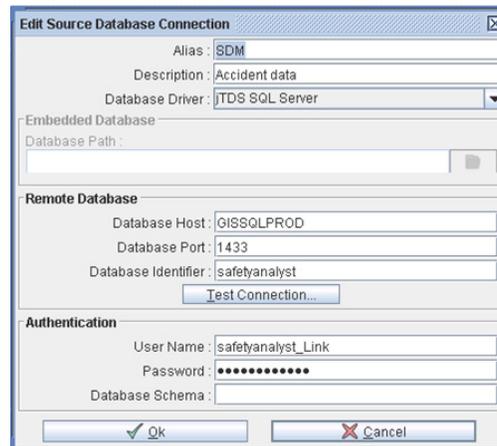


Figure 20. Screenshot. Data management tool.

CONCLUSIONS

Testing and Error Correction

The errors and warnings from the Safety Analyst data importing process resulted in specific groups of errors:

- Overlapping Segments.
- Negative value at Segment Start Location.
- “Other” and “Unknown” Data Attribute Values.
- Invalid Site Subtype.
- Missing Event Attribute Data.
- Superfluous Warnings Generated in Post Process.

Results of the Import Process

Using the extract, transform, and load process developed, the project team was able to successfully import segment and traffic data extracted from ATIS. Subsequently, 2009 to 2014 accident data was imported using the database-to-database import method, and ran the

combined data—segments, traffic, and accidents—through the Post-processing and Calibration steps without errors. To accomplish this objective, data selection was limited to only segments that had matching traffic data and limited the selection of accidents to only those segments.

The result is a Safety Analyst database containing 26,038 roadway segments (totaling 17,472 miles), 156,228 traffic data records, and 213,802 accident records.

Schema Options

This pilot project used the default schema provided by AASHTOWare Safety Analyst™, and determined that no modifications would be needed. The project team recommend that, in general, ADOT should use defaults provided within the AASHTOWare Safety Analyst™ software tools for the foreseeable future, and only customize the configuration after repeated experience demonstrates both a need and a clearly better alternative.

Potential for Integrating Additional Data Sets

Other data sets studied in this project that did not meet the minimum Safety Analyst data standards but have potential for import include:

- Intersections.
- Ramps.
- Navajo DOT and BIA Data.

Conclusions

In this pilot project, the project team built processes that extract roadway and accident data from ADOT's ATIS and SDM databases, respectively, transforms the coded data into the corresponding Safety Analyst enumeration method, and imports the processed data into the Safety Analyst database. The project team employed multiple techniques to accomplish these tasks: used scripted processes to perform complex data manipulations and produce load file in .csv file format and used the database-to-database import facility of the Safety Analyst Data Management Tool to move data directly between source and target data stores.

PILOT STUDY: INDIANA LOCAL TECHNICAL ASSISTANCE PROGRAM

PURPOSE

The purpose of this pilot project was to provide the Indiana Local Technical Assistance Program (LTAP) with expert assistance with GIS based data-integration and analysis, which would facilitate their ability to assist local agencies. This project provided GIS technical support, data management/governance implementation, and analytic training and support. The goal was to integrate local spatial data logically into a centralized database which Indiana LTAP maintained, and to provide methods for sharing and gathering data via the web or through an internal secured website. For the purposes of conducting the Pilot Study, Indiana LTAP selected Putnam County as the pilot county for evaluating their GIS data for input into a centralized GIS database.

The following was the phased approach for a comprehensive GIS database development:

- GIS Assessment.
- GIS Data Development and Integration.
- GIS Website Development.
- Asset Data Collection.

NEEDS ASSESSMENT

The project team created a GIS Needs Assessment questionnaire to gain information about data availability at the county level, data organization and structure, data management, and what types of data support county government needs. The LTAP distributed the questionnaire to all 92 counties, and several other State agencies that maintain GIS information.

INDIANAMAP DATA SHARING AGREEMENT

In 2008, the Indiana Geographic Information Office (IGIO), in collaboration with several agencies, established a data sharing agreement to support IndianaMap and GIS data exchange better within the State.

The goal was to build a statewide map for the benefit of all Indiana citizens. Since the inception of the agreement in 2008, all 92 counties have agreed to participate in the data sharing agreement. Below are featured categories of the data sharing agreement:

- Financial Incentive.
- Web Feature Service Technology.
- Minimum Subset of Data.
- Data Format.
- Existing Data.

REFINED SCOPE OF WORK AND PROJECT GOALS

In addition to the agency suggestions, several counties expressed concern related to the time needed to complete the needs assessment questionnaire or that they did not see the value in completing the survey. Rather than potentially disrupting county governments' successful participation in the IndianaMap data sharing agreement, the project team decided to terminate the LTAP GIS Needs Assessment. For the purposes of the Pilot Project, Indiana LTAP agreed to use existing GIS data from IndianaMap and INDOT rather than building their centralized GIS database. The following are the assumption categories agreed upon by the project team and Indiana LTAP in order to move the project forward:

- Database Creation.
- Data Evaluation.
- Data Analysis.
- Training.

DATABASE CREATION

GIS SOFTWARE

Adoption of GIS at the county level in Indiana is very high; approximately 50 percent of the counties use Esri GIS software, while the remaining 50 percent use Think GIS® or 39°north GIS software.

Indiana LTAP chose Putnam County for the pilot data integration project because of the following factors: the county's highway department is actively using GIS; the county's geography and roadway network are representative of a typical county in Indiana; and the county's GIS data is likely similar to other counties. In addition, Putnam County is one of the counties utilizing Think GIS®, and it is particularly important for the Indiana LTAP to understand the data integration requirements/capabilities of an agency that is not using Esri GIS software.

Esri Local Government Information Model

A number of data models are available to organizations that want to create or manage a GIS database. Esri designed the Local Government Information Model (LGIM) to save time and resources on the part of the local government agency. The LGIM could help Indiana LTAP in establishing a standardized data model for their clients to organize data, improve data sharing, and provide access to existing tools.

LTAP Database Development

The GIS data provided by Putnam County, INDOT, and Indiana LTAP was loaded into a file geodatabase based on the LGIM data model. Before loading the datasets into the file geodatabase, the project team created a data catalog to inventory each dataset, these included the data type, coverage, attribute schema, owner, and coordinate system. The project team eliminated the majority of the LGIM feature datasets and feature classes, resulting in database with the following feature datasets:

- Administrative Areas (administrative boundaries, county boundaries).
- Capital Planning (capital improvement projects).
- Facilities Streets (road inventory, signs, sidewalk, guardrail).
- Storm water (Culverts).

DATA EVALUATION

To evaluate the data integration between INDOT and Putnam County data, additional processes were required to get the WMS data into a useable format. Accessing the WMS by its representational state transfer (REST) endpoint through the ArcGIS service directory allowed the project team to query and convert the data to a local dataset.

When looking at the various data transferred to the project team for use in data integration and analysis, there were inconsistencies between the county and State level data. To create a cohesive geodatabase, it is important to avoid data duplication between data sets and to identify relationship classes between various data sets.

Summary Data Integration Challenges

From a data integration perspective, there was a lack of local (county) data to integrate with INDOT's GIS database. Putnam County's road inventory does not include detailed attributes defining the characteristics of the roadway. In comparison to INDOT's GIS database, additional

attributes are limited to road surface improvements, such as the type and quantify of the material applied to the road and the associated cost. Below is a list of variations with the data:

- Variations in Roadway Geometry.
- Variations in Route Identification.
- Variations in GIS Software.

DATA ANALYSIS

The project team conducted data analyses on the above-mentioned datasets, enhanced the Putnam County road inventory, and generated a preliminary intersection database. The following sections document the data analysis and additional tools created for this pilot.

ARCGIS CONFLATION TOOL SET

As previously noted, the geometry and the route identification system varied between the available road centerlines layers. As a result, spatial and attribute joins would not work in this situation. To accomplish the data transfer, the project team used Esri's ArcGIS Conflation Toolbox. The Conflation Toolbox contains several tools that help reconcile inconsistencies between multiple data sets to create the most comprehensive data for analysis and mapping.

INTERSECTION DATABASE

In addition to enhancing Putnam County's road inventory database, the project team developed a preliminary intersection database. This point feature class identifies potential intersection locations throughout Putnam County.

The intersection inventory included the following:

- Intersection Type.
- Number of Legs.
- Intersection Geometry.
- Traffic Control.
- Major Road Name.
- Minor Road Name.

CRASH ANALYSIS

Indiana LTAP provided crash data for Putnam County. The Indiana State Police provides the Indiana LTAP with the raw data in an MS Excel file format. In order to analyze the data, Indiana LTAP staff developed an internal *County Road Safety Screening Process* for the purposes of identifying county roads or sections of county roads to target safety improvements. After pre-processing the raw crash data, Indiana LTAP compiles the crash data into a number of different selection criteria.

Indiana LTAP distributes the crash types to counties in shapefile format. In addition to distributing the crash data, Indiana LTAP tabulates the highest-ranking sites and provides crash summary data such as crash rates, percent fatal crashes, and several other statistics.

INDOT Safety staff complete their safety analysis using two custom applications developed by Purdue University. INDOT uses the Safety Needs Identification Package version 2 (SNIP2), and the Road Hazard Analysis Tools (Road HAT).

CRASH DATA ANALYSIS TOOL BOX

To help support Indiana LTAP's crash analysis capabilities and automate existing workflows, the project team developed several Esri ArcGIS analysis tools using Model Builder. The project team developed the tools and their internal processes based available crash data provided for Putnam County. The tools rely on attribute values and field names based on the Putnam County crash file. If the data model for the crash data varies between counties, Indiana LTAP will need to modify the tools.

DATA SHARING WITH ARCGIS ONLINE

ArcGIS Online is a cloud-based GIS platform available through Esri that allows users to create web maps and web applications. Users within an organization can share information both internally and externally. Organizations can share data to ArcGIS Online as hosted datasets, or use their own GIS servers to post information.

For this project, the project team used ArcGIS Online to demonstrate how Indiana LTAP could improve data sharing with their customers. Indiana LTAP already had access to ArcGIS Online through their affiliation with Purdue University, but were not experienced with the technology. The project team created multiple web maps and web mapping applications to show how users can display, share, and edit data within ArcGIS Online.

THE COLLECTOR APP

Another capability of ArcGIS is access to Esri's free mobile data collection application, Collector. Users can access this application on an iOS, Android or Windows device. The technology allows users to edit, collect and update data while out in the field and while viewing the data in real-time through web maps/apps back in the office. Disconnected editing is also supporting in Collector. If a cellular or Wi-Fi connection is unavailable, data can be downloaded directly to the mobile device, edited, and then sync' back up to ArcGIS Online or web server when a connection is available.

For this project, the project team developed two web applications for use with the Collector application to gather and develop data; culverts and sign inventory.

FUTURE TASKS AND RECOMMENDATIONS

Using Putnam County as a pilot example, the results of this project highlight the lack of roadway characteristic information available at the county level. If Putnam County represents a typical county in terms of what is available for roadway data in a GIS format, the data that is available will not support more advanced safety analysis. The following are a compilation of the future tasks and recommendations necessary for more advanced safety analyses:

- Leverage the successful work completed by the IGIO, IGIC, INDOT, and other agencies, to expand the IndianaMap data sharing agreement to include the MIRE Fundamental Data Elements (FDE's) for road centerlines.
- Work with INDOT on the configuration of their Esri Roads & Highways Roadway Characteristics Editor (RCE) for the collection of the MIRE FDE's as an alternative to the IndianaMap data sharing agreement protocol.
- Adopt INDOT's GIS data model for standardizing roadway data throughout the State.
- Collaborate with their constituents to adopt the Esri Local Government Information Model (LGIM) or other standardized data model.
- Work with the Indiana State Police to link crash data locations to INDOT's road network.
- Develop procedures for submitting revised crash data locations to the Indiana State Police.
- Establish a data sharing agreement between Indiana LTAP and INDOT.
- Use the capabilities of ArcGIS Online to share crash data analyses with Indiana LTAP clients.
- Work with county governments to collect roadway data using ESRI collector.

PILOT STUDY: GIS DATA DEVELOPMENT AND INTEGRATION AT NAVAJO DOT

PURPOSE

The primary task associated with this pilot project was to assemble a geodatabase that encompassed all the data that was received from the Navajo Division of Transportation (DOT). A geodatabase is a data management tool that allows data managers/users to store GIS data in a centralized location and provides easy access and management of the data. Esri offers three (3) types of geodatabases: 1) Personal Geodatabase; 2) File Geodatabase; 3) Enterprise Geodatabase.

The Navajo DOT has been using GIS for approximately 3-5 years. The primary GIS platform implemented by the Navajo DOT is ESRI ArcGIS software. Each of the five Navajo DOT Departments has ArcGIS desktop software that is administered by the Navajo DOT GIS Section of the Executive Administration.

The desired outcomes of the Pilot Project are to assist with organizing the Navajo DOT's existing GIS data for input into the enterprise server, evaluate the Navajo DOT's existing GIS data for use in implementing safety analyses, provide recommendations on data governance and enterprise database management practices, and assistance with web mapping applications.

EXISTING DATA INTEGRATION AND DATA INCONSISTENCIES

The road inventory and other relevant GIS information were previously stored on the enterprise server. Additionally, other GIS data were stored outside the server, largely in a shapefile format. The Navajo DOT provided the project team with data that had been previously collected or obtained and used by the DOT. The data that was provided to the project team consisted of a variety of file types. All the incoming data from the Navajo DOT was cataloged by the project team to evaluate the available data.

When looking at the existing data there seemed to be inconsistencies in the way that the GIS files were managed. First, there was no metadata for the files. Metadata is a way to give background information about each specific set of data. It describes the content, origin, quality, condition and other information pertaining to the data. Metadata consist of properties and documentation. Properties come directly from the data source such as coordinate system and documentation is entered by the person editing the data. Because of the lack of metadata in the Navajo DOTs data issues arose when the file names were not descriptive enough to fully grasp what information the shapefile was holding. Cataloging all the shapefiles would have been easier if there had been metadata to go along with the files. Other inconsistencies were:

- Multiple coordinate systems being used in the existing Navajo DOT data.
- Differences in the attribute information for shapefiles that contained similar data.
- Cases where shapefiles seem to have no attribute information, besides the generic object number and shape size that is auto generated within the GIS software.

GEODATABASE CREATION

Once all the data from the Navajo DOT was received and cataloged, a File Geodatabase was created. Within a geodatabase, spatial GIS data are referred to as feature classes. The feature classes within the geodatabase were grouped into feature datasets based on the results of the GIS Needs Assessment. All the existing data that was received was placed in the corresponding feature dataset; for example, the road inventory was placed in the roadways feature dataset and the county boundaries were placed in the land status feature dataset. Once data was moved to the appropriate location, all the data was cataloged by the feature dataset that it was located in. A spreadsheet was created for all the final data in the geodatabase.

STANDARDIZE ATTRIBUTE VALUES (BUILDING DOMAINS)

Domains are used to enforce data integrity since they are a set of defined values that are allowed to populate a specific field. Both the crash data and the Fiscal Year 2013 Road Inventory feature classes had alphanumeric coded values in the attribute table, but no descriptive information. In order to build the domains, the Navajo DOT provided documentation for both feature classes that contained the descriptive information for the coded values. Domains were built for Crash Location Data and Road Inventory Data.

FEATURE CLASSES CREATED BY THE PROJECT TEAM

The airport feature classes that are within the Airport feature dataset were created by the project team. The data was created from hardcopy as-built airport plans that were provided by the Navajo DOT.

Other feature classes that were created by the project team include Navajo Nation agency boundaries, composite calibrated routes layer, calibrated route topology, Fy2013Roads topology and the beginning of sign inventory.

All the feature classes that were created by the project team have Federal Geographic Data Committee: Content Standard for Digital Geospatial Data style metadata established. This metadata style is widely used in North America and across the world. It is divided into seven main sections; Identification Information, Data Quality Information, Spatial Data Organization

Information, Entity and Attribute Information, Distribution Information, Metadata Reference Information. Each main section has different subsections that go into detail about the data. In some cases, the details are auto generated from GIS and in others has to be inputted by the person creating or updating the metadata.

ROADWAY DATA DEVELOPMENT

The Navajo DOT's Roadway Inventory database is the core GIS data layer that is maintained by the Navajo DOT.

The road inventory database, which consists of road centerlines, was originally developed to report roadway ownership to the BIA for inclusion in the annual Indian Reservation Road (IRR) Program, now the Tribal Transportation Program (TTP). The road inventory database represents the Navajo DOT's LRS. The LRS represents homogenous sections, where descriptors are constant throughout the entire section length. The LRS is segmented by agency boundaries, reservation codes, and route numbers. The fiscal year 2013 road inventory data is the Navajo Nations current LRS.

The calibrated routes were also highlighted during this project because historically the Navajo DOT had been using the calibrated route files to locate their crash locations, and not the Road Inventory/LRS. The calibrated routes data were split up by agency and within each agency there were three shapefiles that corresponded to each part of the calibrated route.

The Tribal Transportation Improvement Program (TTIP), is a five-year plan for roadways within the Navajo Nation that are in need of assistance. The Navajo DOT provided shapefiles containing the 2014 TTIP projects and a File Geodatabase containing a feature class that reflected the TTIP projects planned for 2015. Along with the shapefiles and the File Geodatabase, there were project information sheets for various TTIP projects in each agency. The feature class that contained all the TTIP projects across Navajo Nation was adjusted to make the attributes more descriptive and link to the project information sheets. A description, estimated cost and project sheet link attribute were added.

FUTURE DATA LAYERS AND DATA INTEGRATION

Although there is base data for the Navajo DOT GIS program, there are still data layers that would be helpful moving forward. The following data layers would be advised for the Navajo DOT to create:

- Expand Road Inventory to include additional roadway characteristics defined by a common data model such as the Model Inventory of Roadway Elements (MIRE).
- Intersection Inventory.

- Traffic Count Locations.
- Culvert Inventory.
- Railroad Crossings.
- Lane Striping.
- Guardrail and other fixed roadside objects such as a sign inventory.
- Transit (Bus stops/routes).
- ROAD ROW Limits.

Routine training sessions or discussions between the Navajo DOT, Traffic Safety and Law Enforcement divisions, would help to ensure the pertinent crash factors are recorded (such as crash type) for each crash. Personnel from the various divisions should understand how the crash record information is used in support of each division's needs or reporting requirements.

SYSTEMS OF DATA ACCESS- ARCGIS ONLINE

The Navajo DOT expressed the need for a better way to communicate and share safety related information to other Divisions within the Navajo Nation, State DOTs, and the public. Specifically, the need for an external interactive Navajo DOT GIS website, where the public can browse crash data records, transportation betterment projects, and roadway geometry information. There is also need for an internal GIS website, where Navajo DOT staff can access (read/write access) roadway information from the enterprise database from a remote location.

For this project, ArcGIS Online was used to show how the Navajo DOT could share and publish data within agency departments as well as the public. Data that is published or stored on ArcGIS Online can be edited by approved users to maintain the most complete and comprehensive data. Web applications can also be published that provide the public with an opportunity to interact with the data, such as the TTIP project locations. Multiple web maps were created, which are described below, to show how data can be displayed, shared and edited within ArcGIS Online.

NAVAJO DOT ARCGIS ONLINE

As the Pilot Project progressed the Navajo DOT obtained their own temporary organizational license for ArcGIS Online. Web maps created in the ArcGIS Online:

- Crash Locations (Swipe).
- Crash Location (Filtered by Collision Type).

- DOT Crash Location (Time enabled).
- Navajo DOT TTIP 2015.
- Crash Location (Filtered by Cause of Crash).
- Navajo Nation Road Inventory (2013).
- Navajo Nation Airports.
- Navajo Nation DOT (Contains Airport data created by the contractor, TTIP project locations, Road inventory and agency boundaries).

DATA GOVERNANCE/DATA MANAGEMENT PRACTICES

The Navajo DOT manages a wide range of data assets including LRS, crash inventory and traffic data used in safety analyses and, route and asset data used in managing the roadway system in general. With approximately 12,000 miles of public roads, spanning 27,000 miles and crossing into three states; having accurate, complete and uniformed data will aid the DOT in making the best decisions possible.

To create successful data governance practices, a data governance committee should be established within the Navajo DOT to:

- Understand data needs throughout the Division.
- Develop data standards to define:
 - a. How data is collected.
 - b. How data is archived.
 - c. How data is secured.
 - d. How is the data maintained and by which department and unit.
 - e. How accessible is the data internally and externally.
- QA/QC requirements such as accuracy and timeliness measures.
- Develop documentation and training material on the data governance procedures.

A short term goal to establishing data governance practices would be to prioritize the data that is currently available. Identify critical data needs to be dealt with first is an important step. Key considerations should include:

- Identifying the departments who are responsible for managing the data.

- The number of consumers of the data, and the number of departments who require the data.
- Policy requirements for data, which impact funding sources.

Future Data Governance Tasks within The Navajo Nation

In the preceding sections, the establishment of a Data Governance Committee at Navajo DOT was recommended. It is further recommended that a Data Governance Committee, consisting of stakeholders from each Division within the Navajo Nation should be created. Below is a list of goals for the Data Governance Committee:

- Schedule monthly\quarterly meetings to discuss data governance needs.
- Identify common data needs that can be shared between Divisions.
- Identify common software and hardware needed to streamline expenditures.
- Work towards the implementation of a Navajo Nation GIS clearinghouse where data can be accessed by all Divisions within Navajo Nation, outside agencies and the public.

FUTURE TASKS AND RECOMMENDATIONS

The road inventory database, with some enhancements, has the potential to support transportation functions such as work order requests, emergency management routing and 511 traveler alerts, advanced safety analysis, and support more advanced GIS modeling such as network routing. Moving forward the Navajo DOT can take the recommendations from this study to help accomplish tasks that are seen as needed to be completed to create a fully function GIS program.

Data Analysis Recommendations

Based on available data and Navajo DOTs needs/interests, the project team suggests that Navajo DOT begin by conducting a dual approach network screening as outlined in the HSM utilizing both hot spot/site specific and systemic approaches.

The project team recommends that the Navajo DOT implement network screening and other steps found in the safety management process. Listed below are recommendations for achieving those goals:

- Evaluate data layers and attributes incorporated into the GIS database and further refine the dual approach of utilizing both site-specific and systemic network screening.

- Conduct network screening by applying the training provided by the project team and following the process outlined by the HSM.
- Navajo DOT should continue to work with Arizona DOT to incorporate RSAs and work to implement other steps of the data management process such as economic appraisals and safety effectiveness evaluations.

ARCSDE (Spatial Database Engine) and GIS Enterprise Database Recommendations

The original purpose of this project was to assist the Navajo DOT in integrating their existing data into an enterprise database that would allow them to share information with Navajo Nation departments, State, Federal and other agencies. Since the enterprise system was not up and running through the duration of the project the focus shifted away from integration and towards data management. Recommendations and best practices that the project team suggests that the Navajo DOT pursue include those in the following areas:

- Database Administration/ User Authentication.
- Database Backups.
- Editing Tracking.
- Time/Date Stamping Edits.
- Archiving/History within Database.
- Data Sharing.

PILOT STUDY: RHODE ISLAND DEPARTMENT OF TRANSPORTATION

PILOT GOALS

The purpose of this project was to provide the Rhode Island Department of Transportation (RIDOT) with expert assistance in GIS-based data integration in support of implementing advanced analytic methods such as those described in the Highway Safety Manual (HSM) and supported by analytic tools like AASHTOWare Safety Analyst™, the Interactive Highway Safety Design Model (IHSDM), interactive websites, and others.

In support of the pilot purpose, vision, and mission, RIDOT established three overarching goals:

- Develop processes and identify staffing and resources needed to guarantee the ongoing maintenance and utility of the roadway location and MIRE inventory data.
- Manage data integration and assist RIDOT in developing processes for integration of the new MIRE data into Esri Roads & Highways (Roads & Highways) and RIGIS for use in safety analysis.
- Support use of advanced analytic tools/methodologies through example analyses and training on data extraction/integration processes.

ROADWAY INVENTORY MAINTENANCE PRACTICES

Existing Internal RIDOT Processes

A review of processes revealed that there is very little formal documentation of existing business systems. The majority of the business systems are disconnected from one another with their data stored as silos, inaccessible to other business systems. The result is multiple Linear Referencing Systems (LRS) are needed to support each business system.

Existing Business Systems

RIDOT Existing Roadway Inventory

RIDOT maintains an existing roadway inventory referred to as Roads_HIIS in their enterprise GIS server, covering all State and NHS roads. Roads_HIIS is a route based system representing the State's authoritative LRS. The LRS route numbering system and mile post calibration mirror the pavement management system, which is stored in Deighton Total Infrastructure Asset Management System (dTIMS). RIDOT also maintains a second roadway inventory data layer

referred to as Roads_Topo, which is used to synchronize with E-911's road centerline database. The result is a statewide road centerline database.

VUEWorks

RIDOT has been utilizing VUEWorks software for approximately two years for maintenance management and processing of work order requests. All customer service, maintenance, and Transportation Management Center (TMC) calls are entered as service requests related to a location or an asset. As part of the Esri Roads & Highways implementation, VUEWorks will be connected to the Roads & Highways enterprise linear referencing system through the use of web service connections, resulting in the location of work against the LRS.

Crash Data

RIDOT stores their crash data in a centralized Crash Data Repository System and has developed an On-line System Crash Analysis and Reporting (OSCAR) application. Currently about 20 to 30 percent of all crashes are referenced with latitude/longitude coordinates. Approximately 30 to 40 percent of the crashes are referenced to a street address. With the implementation of Esri Roads & Highways, RIDOT plans to update the crash database and populate latitude/longitude coordinate values for all crash locations. The crash database will be registered with the enterprise geodatabase developed for Esri Roads & Highways, resulting in a common LRS for all crash records.

Pavement Management

RIDOT's pavement management system is administered using Deighton's Total Infrastructure Asset Management System (dTIMS) software. Pavement condition data is collected annually through contracted services. This data is loaded into dTIMS based on the route and driven mileage. dTIMS provides RIDOT with the ability to forecast condition of roadways and identify possible treatments.

Americans with Disabilities Act (ADA)

To ensure RIDOT complies with ADA requirements, FHWA requires the department have a transition plan that includes a schedule for providing access features, including curb ramps for walkways. The inventory is managed in SQL server with a custom web application. The application's mapping component utilizes SQL spatial. RIDOT is planning a migration of the inventory into the GIS with the management of the program in VUEWorks. Integration with Esri Roads and Highways will occur after the migration to VUEWorks.

RhodeWays Incident Management System

RhodeWays is RIDOT's incident management system. Operators from RIDOT's Transportation Management Center (TMC) locate incidents through a mapping interface that utilizes RIDOT's ArcGIS services. The application stores incident locations with latitude/longitude coordinates. Incidents are made available through RIDOT's website and the 511 system. Future enhancements, include the integration with Esri Roads & Highways to allow for location referencing to the LRS.

Traffic Volume Database

Traffic data is collected through continuous, seasonal, and short-term counting stations. RIDOT Traffic section collects, processes, and analyzes the data using various applications. The final published data is managed in an Access database where the ADT and AADT are stored. This information is shared through the GIS and linked to the location based on the Traffic Station ID. RIDOT plans on replacing the existing system with an off the shelf enterprise traffic data management system.

Bridge (BrM)

RIDOT manages their bridge infrastructure using AASHTOWare Bridge Management software (BrM). BrM provides RIDOT with a means to manage its bridge inventory and inspections following AASHTO's Guide Manual for Bridge Element Inspection. The bridge inventory data is integrated with VUEWorks and is the authoritative source for bridge data for the GIS.

Other Business Systems

In addition to the business systems used to manage roadway data, several other critical administrative business systems exist that are associated with the roadway business systems. The administrative business systems cover project tracking, financial management, contract management, scheduling, and document management.

Current Initiatives

There are a number of major initiatives currently underway at RIDOT aimed at improving internal roadway data system workflows. The initiatives include data collection activities, LRS management improvements, network architecture reviews, and system integration.

MAINTENANCE OF THE ROAD INVENTORY

The data model between Esri Roads & Highways and the data model used to store the MIRE inventory are significantly different. When the MIRE inventory is complete, the inventory will replace RIDOT's existing road inventory (hereinafter referred to as the Road Inventory) data layers, and will be loaded into Esri Roads & Highways.

Maintenance Process

The Road Inventory includes all State and public roads in order to address FHWA's ARNOLD requirements. The task of maintaining the Road Inventory needs to be a joint effort between RIDOT and the 39 municipalities in the state. RIDOT's responsibility is to coordinate with municipalities to track municipal roadway projects and develop a system for municipalities to submit changes to the Road Inventory.

RIDOT Internal Tracking of Roadway Projects

A major component of the maintenance plan is the ability to track roadway projects that alter the geometry or the characteristics of a state road. The information needed to track a project through its lifecycle does not exist in a single database; the information is spread across RIDOT's various internal business systems. Through the Esri Roads & Highways implementation and improved data governance principles, synchronization of these business units with the LRS, will enable a systematic project tracking process.

Maintenance of the Road Inventory for Local Roadway Projects

The development of a web-based LRS editing system is needed to provide municipalities with a means of maintaining the Road Inventory for local roads. Internally, RIDOT needs to develop a basic tracking system that is populated with information gathered through close coordination with local municipalities. The Road Inventory Coordinator should be responsible for coordinating the local roadway project tracking effort.

In support of the coordination effort, RIDOT identified the Esri Roadway Characteristics Editor (RCE), which is included with RIDOT's implementation of Roads & Highways, as a solution for providing municipalities with a means of editing (maintaining) the Road Inventory with local roadway changes. RCE is a web-based application that provides external users (local municipalities) with LRS event editing capabilities. The RCE is not a tracking system, and can only be used in conjunction with Esri Roads & Highways.

Road Inventory Workflow

Without a functioning version of Esri Roads & Highways with the MIRE data loaded into it, it was not possible to test the procedures outlined above. However, based on the information available to date and an understanding of the editing capabilities within Esri Roads & Highways, a recommended workflow was identified (See Figure 21 below).

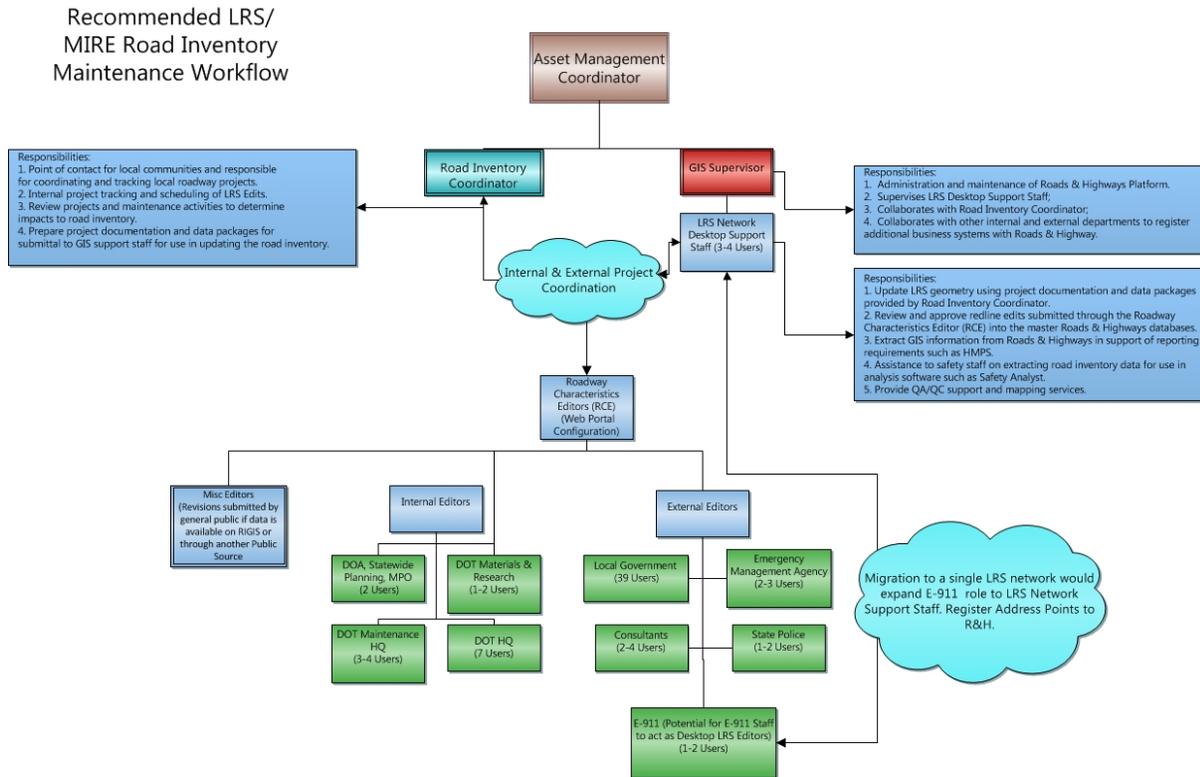


Figure 21. Chart. Recommended LRS/MIRE road inventory work flow.

Strategy for Encouraging Municipalities to Participate in the Road Inventory Maintenance Process

Like State DOT's, municipalities have limited resources (staffing and funding) to complete their required duties and functions. Requiring municipalities to update RIDOT's Road Inventory without an outreach plan is likely to be met with significant pushback. Outreach and education to municipalities pertaining to the federal transportation laws (MAP-21, ARNOLD requirements) requiring states to develop a statewide road inventory is important to obtaining municipal support in maintaining the road inventory. Below are a few recommendations for encouraging municipalities to participate in the maintenance of the Road Inventory:

- Update Rhode Island's TIP application criteria to include local government participation in providing road inventory updates. A municipality that does submit road inventory

changes through the RCE will be penalized when submitting a TIP application for a project in their community.

- Establish a requirement that Highway Safety Improvement Program (HSIP), and Transportation Enhancement (TE) funds are only eligible to municipalities participating in providing road inventory updates.
- Several states have adopted legislation, making it a requirement for local governments to provide changes to the public road updates, which includes acceptance of new roads, closing of public roads, and road transfers.

LRS/Roadway Inventory Maintenance Schedule

RIDOT's Road Inventory was developed during the fall of 2014 and summer of 2015, and in some locations, it will be outdated by the time the inventory is complete. As with any asset, there are various strategies for maintaining a road inventory, which largely depends on the complexity of the inventory. If the inventory is not maintained on a regular basis, the information will be outdated and have a very short lifecycle. The longer an inventory goes unmaintained, the cost and level of effort associated with bringing the inventory back up to date significantly increases.

The recommended maintenance schedule for RIDOT's Road Inventory should consist of continuous maintenance. Continuous maintenance is defined as weekly or sometimes daily to keep the inventory current. It is also recommended that the Road Inventory be published to other state agencies and to the public on a biannual basis, with quarterly snapshots of the data available internally.

RIDOT Transportation Information Systems is responsible for maintaining the Roads & Highways platform, which includes the Road Inventory database. Given the amount of data included in the inventory, additional staffing needs or a reorganization of existing GIS staff, will be required to implement routine maintenance of the MIRE inventory.

DATA QUALITY MANAGEMENT/DATA GOVERNANCE ASSISTANCE

In 2014, RIDOT received assistance from FHWA in the development of the Rhode Island Roadway Data Improvement Program (RDIP), which included recommendations for improving data governance within RIDOT. Released in the fall of 2014, the top priority recommendation within the RDIP called for the establishment of a Data Governance Committee. With the timing not right for establishing a Data Governance Committee and a data governance plan within the timeframe of the pilot program, the project team focused its efforts on a data governance literature review. The goal of the literature review is to provide RIDOT with additional data governance resources that will aid them in implementing the recommendations from the RDIP.

DATA ANALYSIS ASSISTANCE AND TRAINING

The project team delivered a day long training session to RIDOT and their safety partners on March 5, 2015. The following were the primary objectives of the training. The training session helped RIDOT and their safety partners to understand the safety management process and the potential benefits of an enhanced data system. Specifically, the training illustrated how current safety management practices could be enhanced using the additional data. A secondary benefit of the training was the interaction among various safety partners within and outside of RIDOT. Through this training, the analysts gained a better appreciation for the data requirements and the data collectors/managers gained a better appreciation for the use of the data.

APPENDIX B – ADDITIONAL WORK PLANS

FORT BELKNAP: DATA INTEGRATION WORK PLAN

SCOPE

The purpose of this Work Plan is to provide the Fort Belknap Transportation Department with expert GIS-based data integration advice in support of implementing advanced analytic methods such as those described in the Highway Safety Manual (HSM) and supported by analytic tools like AASHTOWare Safety Analyst™, the Interactive Highway Safety Design Model (IHSDM), interactive websites, and others.

VISION

The vision of this project is to implement GIS-based safety data integration and ultimately to support adoption of advanced analytic methods such as those described in the HSM and supported by advanced safety analysis tools. In the long term, the vision extends to adoption and routine use of integrated safety data and advanced analytic methods to promote state-of-the-practice safety decision-making resulting in quantifiable improvements in traffic safety for Fort Belknap.

MISSION

To logically integrate Fort Belknap GIS data with other federal, state, and local data into a database that will support advanced analytic methods and enable the sharing of information via the web or thru an internal secured website.

GOALS

The goals of this Work Plan, once implemented are:

1. Creation of a new and/or modification of an existing GIS database.
2. Integration of safety data from other agencies, with flexibility to add more data sources as they become available in the future.
3. Selection and application of advanced analytic methods for safety analysis, with flexibility to support adoption of new methods in the future.

WORK PLAN TASK DESCRIPTIONS

TASK 1: KICKOFF MEETING

Establishing a project team and scheduling an initial kickoff meeting are an efficient means of updating this plan and developing a shared understanding of the project's goals, tasks, schedule, and milestones. The Fort Belknap Transportation Department currently does not employ a standardized data analysis methodology. During the meetings to develop this initial Work Plan, participants agreed that the Network Screening methodology in the Highway Safety Manual may be the most logical fit and a good first start on implementing advanced safety analysis methods.

TASK 2: FINALIZE PROJECT WORK PLAN

The final Work Plan should include detailed descriptions of tasks, milestones, deliverables, and schedule. The final Work Plan should also provide detailed descriptions of the Fort Belknap Transportation Department's roles and responsibilities, and the work required of other partner agencies.

TASK 3: GIS NEEDS ASSESSMENT & IMPLEMENTATION PLAN

Organizations that are successful in implementing a new initiative typically start with a detailed plan and have a set of specific goals in mind. Before investing any additional resources in implementing their GIS, Fort Belknap should conduct a thorough GIS Needs Assessment and develop an Implementation Plan. The GIS Needs Assessment should achieve the following goals:

- Identify key stakeholders (tribal programs) and their data needs through stakeholder interviews.
- Catalog the most common shared needs among the stakeholders.
- Evaluate the use of GIS to address the shared needs.
- Identify existing GIS resources (internal/external) used by stakeholder groups.
- From a roadway safety and transportation perspective, identify the desired GIS and analytic tools or methods for use by stakeholders and conduct a gap analysis to determine how to achieve those goals.
- Develop a GIS Implementation Plan that addresses the needs identified in the GIS Needs Assessment.

TASK 4: GIS-BASED DATA INTEGRATION AND ANALYSIS

There are a number of data models that are currently available to an organization looking to create a GIS database. Esri has defined a robust data model for all local governments that saves time and resources that would have otherwise been a required action of the local government agency. This data model, the Local Government Information Model (LGIM), fits well for a Tribal government and would assist Fort Belknap in implementing GIS tools.

The LGIM consists of a geodatabase schema, designed for local governments that include the most common data sets required at the local level. In addition to the geodatabase design, the LGIM includes maps, web applications, and services designed specifically for local government use. The project team recommends that Fort Belknap adopt the LGIM as a starting point and implement the components of the model that address the needs identified in the GIS Needs Assessment. If there are local needs or goals that are not attainable using the LGIM, a customized data model should be created to meet those needs.

In order to implement the LGIM, Fort Belknap should move forward with the following tasks:

- Assess data needs identified in the GIS Needs Assessment and align those needs with data sets contained within LGIM.
- Customize LGIM where necessary, such as eliminating feature data sets, features and attribute values not applicable to Fort Belknap.
- Prioritize data need development requirements around common datasets among stakeholders.
- Obtain or plan for recommended data elements for advancing roadway safety capabilities, including:
 - a. Road centerline database.
 - b. Crash Database
 - c. Intersection Inventory
 - d. Traffic Volumes Linkable to GIS

TASK 5: DEVELOP TOOLS TO FACILITATE SAFETY ANALYSIS AND DATA SHARING

There are a variety of methods to communicate and share safety related information to other departments within Fort Belknap, other agencies such as State DOTs or the BIA, and the public. One of those methods is the use of a GIS website where the public can browse crash data records, transportation improvement projects, and roadway geometry information. An

internal website could also allow for Fort Belknap departments and staff to share data and also provide authorized staff with read/write database access from a remote location.

TASK 6: ESTABLISH DATA GOVERNANCE PROCEDURES

Accurate, complete, and uniform data are necessary to make the quantifiably best decisions regarding transportation safety. Data governance is the process of adapting disciplines to manage data and information.

A primary goal when creating the data governance practices is to use an effective, standard approach when managing data quality, updating existing data, and creating new data. Some of the benefits of establishing data governance procedures include:

- Replicability so that staff members have set procedures to follow for creating, updating, and sharing data.
- Quality assurance resulting in increased user confidence.
- Efficiency as only those datasets with a clear purpose are established/maintained and that updates are scheduled on a regular basis.

To create successful data governance practices, the Fort Belknap Transportation Department should incorporate the following elements:

- Understand data needs throughout the Department
- Develop data standards to define:
 - a. How data is collected.
 - b. How data is archived.
 - c. How data is secured.
 - d. How is the data maintained and by which department and unit.
 - e. How accessible is the data internally and externally.
- QA/QC requirements such as timeliness, accuracy, completeness, uniformity, integration, and accessibility measures.
- Develop documentation and training material on the data governance procedures.

A recommended short term goal in establishing data governance practices would be to prioritize the data that is currently available. Next, it would be important to identify critical unmet data needs that are high priority. Key considerations should include:

- Identifying the departments who are responsible for managing the data.

- The number of consumers of the data, and the number of departments who require the data.
- Policy requirements for data, which impact funding sources.

NORTH CAROLINA / STATE DATA INTEGRATION FOR TRAFFIC VOLUME DATA: PROJECT WORK PLAN

INTRODUCTION

The North Carolina Department of Transportation collects traffic data on all primary routes within the State. This includes the Interstates and State highways (NC Routes). The NCDOT Traffic Survey Group provides the data to support planning, design, and safety research initiatives within NCDOT for permanent count locations. For local roadways the State's eighteen largest urban areas are counted on an even/ odd biennial basis.

The State is looking to implement GIS solutions, such as Esri Roads and Highways, which can provide more analysis tools to support research in the safety realm and beyond. This project Work Plan describes how North Carolina can incorporate local traffic data with the State-provided data. The project will also address use of the traffic data outside of the GIS environment. Non-spatial data can be integrated as well, either using the GIS databases or through other means. While safety is a key focus area, the project work plan is designed to address the many uses of traffic volume data throughout NCDOT and its local partner agencies. Traffic data has importance enterprise-wide and NCDOT hopes to make the resource available for all legitimate users.

PURPOSE, VISION, AND MISSION

The purpose of this project is to provide NCDOT with a plan to integrate local traffic data into the statewide traffic data system. The data integrated will include all traffic data such as turning movement counts, railroad counts, annual average daily traffic (AADT), travel time, speed and classification counts. The process will use NCDOT's GIS solutions in order for the State to obtain more complete and timely traffic data. Spatial and non-spatial data will be incorporated.

BRIEF DESCRIPTION AND OVERALL SCOPE

This project will assist NCDOT and selected local agencies to:

- Develop processes and identify staffing and resources needed to integrate local traffic count data into the State system. This will include data governance considerations related to data standards and integration. It will also involve stakeholder (both internal and external to the NCDOT) coordination for all processes and responsibilities.
- Manage data integration and assist the NCDOT Traffic Mobility and Safety (TMS) group in developing processes for integrating local traffic data into Esri Roads and Highways for use in safety analysis.

- Support use of advanced analytic tools and methodologies through example analyses and training on data extraction, transform, and load integration processes.

DETAILED TASK DESCRIPTIONS

Task 1: Kickoff Meeting

The purpose of this meeting is to finalize the project scope resulting a mutual understanding of all tasks, milestones, deliverables, and overall schedule. This meeting will give the agency project team time to ask for clarifications in the project scope.

Task 2: Final Project Work Plan

The project team will make changes to the Draft Data Integration Work Plan based on the results of the kickoff meeting and comments received on the Draft version of the plan. This Work Plan includes detailed descriptions of tasks, milestones, deliverables, and schedule. The Work Plan also provides detailed descriptions of the NCDOT's roles and responsibilities, and the work required of the local agency partners.

Task 3: Traffic Data Collection Practices

This task consists of two main components: (1) internal NCDOT processes and (2) coordinated State and local processes. The NCDOT will develop an updated methodology for the integrated data set description taking into account the need for both the internal processes and coordinated action among the State and local agencies. The internal processes will be developed mainly by the TMS group itself, with advice and assistance from other departments. The coordinated State-local methodology will be designed with assistance from a representative panel consisting of NCDOT staff, local agency and/or MPO staff.

Task 3.1: Internal NCDOT Processes

NCDOT's internal process development will establish the following items through discussions with project staff and other NCDOT stakeholders:

- Data elements to be collected/updated.
- Data standards and data quality measurements.
- Frequency and consistency of updates (periodic or continuous).
- Responsibility for update submissions (balance of State and local action).
- Responsibility for managing the update process.

- Reporting requirements for update process management, including performance measurement (e.g., timeliness, completeness, accuracy, etc.).
- Preliminary testing of the update process.
- Documentation of the NCDOT responsibilities and activities.
- Estimated budget and staffing levels to meet NCDOT responsibilities.
- Data integration procedures and ETL processes for data analysis.

Task 3.2: Coordinated NCDOT and Local Agency Update of Processes

NCDOT will work with stakeholders (to be identified by the NCDOT) developing coordinated update processes to establish the following:

- Content and format for update submissions to the NCDOT.
- Schedule(s) for updates.
- Required management reports for the NCDOT to provide to external partners.
- Documentation of the NCDOT and local partners' responsibilities.
- Technical requirements and options for local agencies to meet update requirements.
- Estimated budget and staffing levels (where appropriate) required to support the local agency data submissions.
- Local data quality assurance processes and reporting.

Task 4: Data Quality Management/Data Governance Assistance

The State will use data quality management and performance measurement experts to get the NCDOT and local partners developing standards for data submissions and updating the volume data. The NCDOT will use the MIRE data dictionary and include operational parameters necessary for successful data integration. Agency staff and local agency involvement is necessary for essential documentation to be created.

Task 5: Data Integration Assistance

The project team will use expert GIS and IT support to establish procedures and accomplish data integration. This technical support will assist agency staff to meet the needs identified in the previous tasks and reflected in the data governance plan. The project team will outline crash and traffic volume data integration into the NCDOT data set for safety analysis.

Task 6: Data Analysis Assistance and Training

The NCDOT should look into training on the use of analytic techniques from the HSM. The training cannot include specifics on usage of some tools (such as AASHTOWare Safety Analyst™); however, training on the extract, transform, and load process for sharing data with such tools, and interpreting the output of safety analyses performed with existing tools, will be required no matter which analytic tools are selected. The data analyses should consist of at least one analysis in each of the following areas: network screening (of a portion of the network), countermeasure selection, systemic improvement, and effectiveness evaluation.

POTENTIAL RISKS:

- **Data standards between state and local agencies are incompatible.** We anticipate that some local agencies' data will not meet the established traffic data collection and reporting requirements. This may include the frequency of counts (update frequency), number of days of data at each site, and accuracy level for location information barring integration via the State's linear reference system or enterprise GIS.

An inclusive data governance process beginning early in the project can minimize this risk. The project team can examine and propose remedies to local agencies' capabilities or incompatibilities with the State systems. Ultimately, however, it is likely that a phased approach will yield the best results statewide. NCDOT will want to phase its access to local traffic volume data by first working with the agencies that have the highest quality data and that are supported with strong information technology capabilities, especially with respect to GIS and spatial data. Later phases will bring in data from those agencies needing more assistance, and perhaps, those needing the State to take over data collection on their behalf.

- **Incompatible GIS basemaps.** While modern GIS software can overcome many barriers to integrating spatial data, there are still technical issues to solve when multiple maps are joined together (e.g., edge matching). These problems are not insurmountable, but they do require staff time and commitment to an iterative process working between the State and local GIS teams to resolve the issues.

This risk is avoided if all users are on the same GIS platform, however, that is not a reasonable expectation in a statewide data integration effort because many local agencies have existing GIS implementations that do not match the State's system. The TMS group will need to allocate time to merging data from multiple local source files. Ideally, the State will develop a repeatable process that can be automated (to a degree) resulting in rapid and correct integration of the local data in the future.

- **Maintenance responsibilities are not clearly defined.** Experience has proven that many States learn late in the project that the need to maintain the data and perform periodic updates was not part of the original data integration plan. This can result in delays, confusion, and cost increases as the local partners and the State negotiate the maintenance process after the fact.

If the State commits to a formal data governance process the responsibility for periodic updates can be added to the data collection and data management protocols. This should be done in cooperation with the local partners so that it is part of the plan from the start.

APPENDIX C – ROADWAY SAFETY DATA AND ANALYSIS TOOLBOX

The FHWA Roadway Safety Data and Analysis Toolbox (<http://safety.fhwa.dot.gov/rsdp/>) is a web-based repository of safety data and analysis tools. Use the Toolbox to identify an appropriate tool for your diagnosis needs. A Primer is available to understand the overall scope and functionality of the Toolbox as well as the roles, responsibilities, and tasks supported by tools in the Toolbox.

USING THE ROADWAY SAFETY DATA AND ANALYSIS TOOLBOX

There are two primary options for searching the Toolbox. The first is a predefined query using the four large icons in the upper right of Figure 22 (Manage, Analyze, Collect, and Research). The second is an advanced search option where users can search keywords and apply filters to customize their search as shown in the lower left of Figure 22.

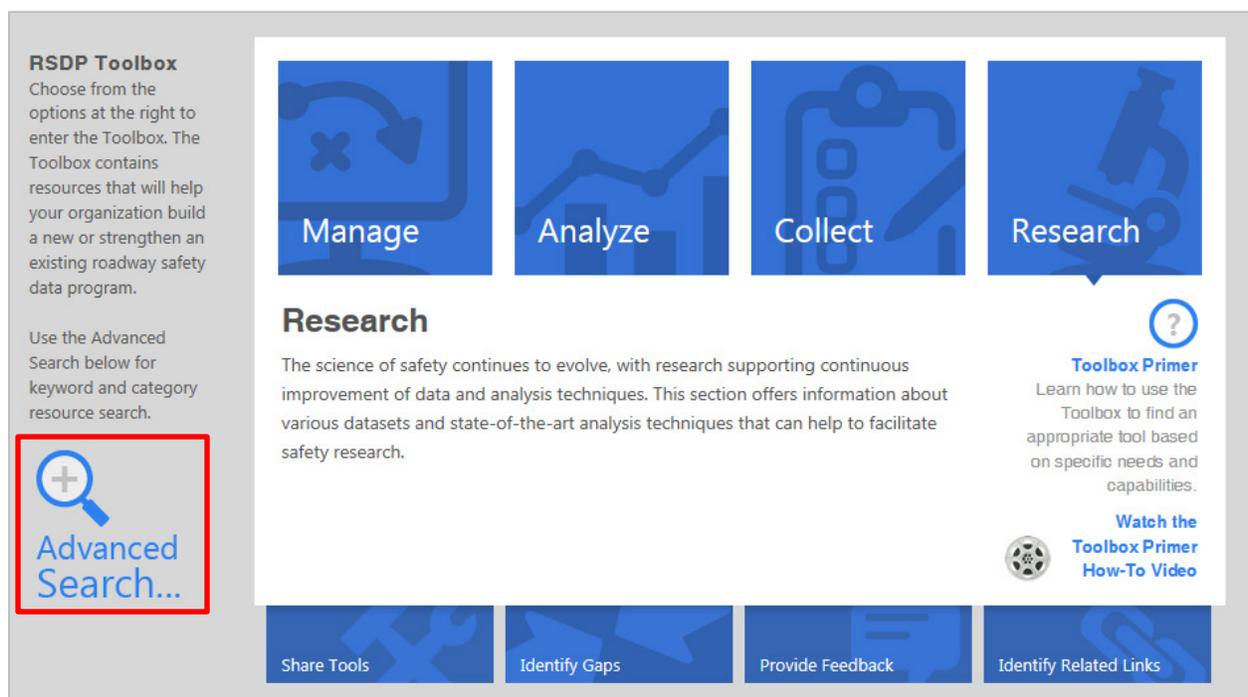


Figure 22. Screenshot. Roadway Safety Data and Analysis Toolbox.

The following is a brief demonstration of the stepwise process to identify an appropriate tool to support diagnosis.

- I. Click the 'Advanced Search' icon, highlighted in the lower left of Figure 22.

2. From the advanced search page (Figure 23), leave the keyword blank and click the search button. This returns a list of all tools in the Toolbox.

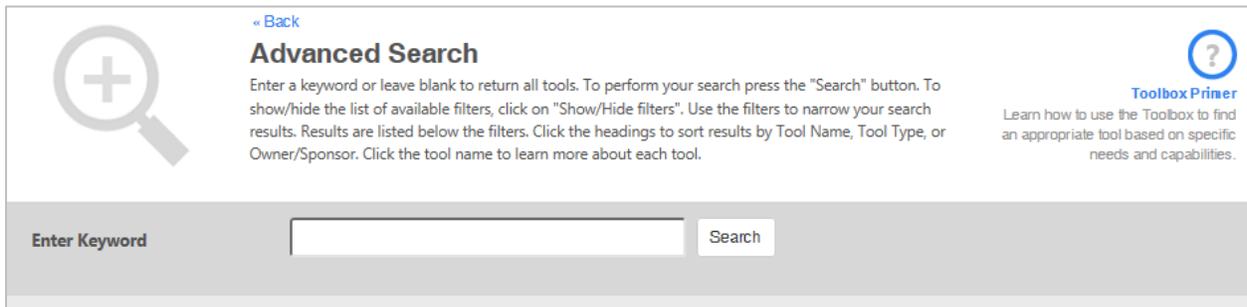


Figure 23. Screenshot. Advanced search feature.

Click the 'Show/Hide Filters' button, highlighted in the upper left of Figure 24. This reveals a list of filters to refine the general search. Use the 'Safety Management Process' filter to select 'Diagnosis' as the primary area of interest as shown in Figure 24. Apply additional filters as needed to refine the results. For example, apply the 'Tool Type' filter to narrow the list of tools to application guides, information guides, software, information sources, or databases.

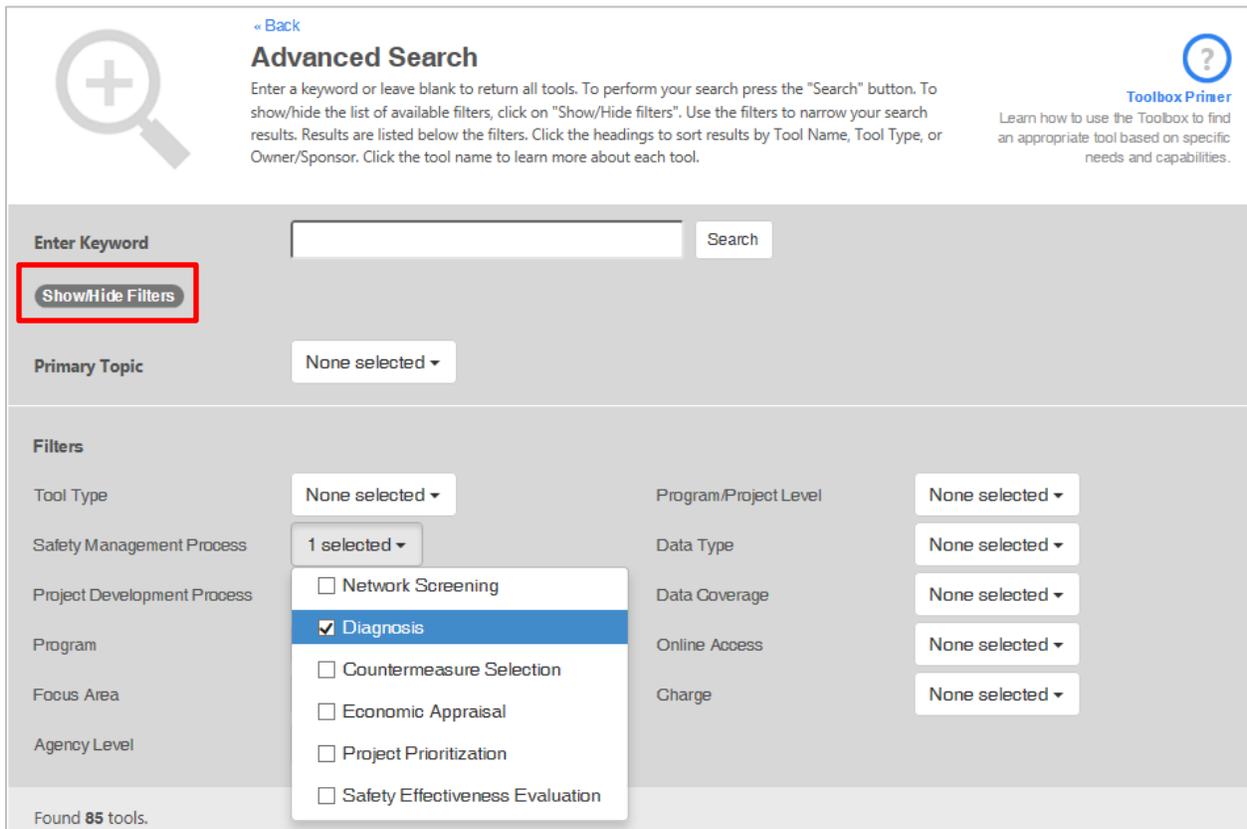


Figure 24. Screenshot. Filter options from advanced search page.

Using the stepwise process described in this section, the Toolbox returns guides such as the Road Safety Audit Guidelines, Integrated Safety Management Process, and Improving Safety on Rural Local and Tribal Roads: Site Safety Analysis User Guide. Related software tools from the Toolbox include AASHTOWare Safety Analyst™, Interactive Highway Safety Design Model, and Pedestrian and Bicycle Crash Analysis Tool.

APPENDIX D – DATA SHARING AGREEMENT AND MEMORANDUM OF UNDERSTANDING

SAMPLE DATA-SHARING AND USAGE AGREEMENT

<State> Department of Transportation and the <City> Department of Public Works

This agreement establishes the terms and conditions under which the <State> DOT and <City> DPW can acquire and use data from the other party. Either party may be a provider of data to the other, or a recipient of data from the other.

1. **Personally Identifying Information (PII)** The confidentiality of data pertaining to individuals will be protected as follows:
 - a. The data recipient will not release the names of individuals, or information that could be linked to an individual, nor will the recipient present the results of data analysis (including maps) in any manner that would reveal the identity of individuals.
 - b. The data recipient will not release individual addresses, nor will the recipient present the results of data analysis (including maps) in any manner that would reveal individual addresses.
 - c. Both parties shall comply with all Federal and State laws and regulations governing the confidentiality of the information that is the subject of this Agreement.
 - d. In analytic results tables any cell with fewer than five (5) counts will be blanked out unless prior approval is obtained from the data provider.
2. The data recipient will not release data, data extracts, or merged data to a third party without prior approval from the data provider.
3. The data recipient will not share, publish, or otherwise release any findings or conclusions derived from analysis of data obtained from the data provider without prior approval from the data provider.
4. Data transferred pursuant to the terms of this Agreement shall be utilized solely for the purposes set forth in the “Data Integration Agreement” section below.
5. All data transferred to <CITY> DPW by <STATE> DOT shall remain the property of <STATE> DOT and shall be returned to <STATE> DOT upon termination of the Agreements.
6. Any third party granted access to data, as permitted under condition #2, above, shall be subject to the terms and conditions of this agreement. Acceptance of these terms must be provided in writing by the third party before data will be released.

IN WITNESS WHEREOF, both the <State> DOT, through its duly authorized representative, and <City> DPW, through its duly authorized representative, have hereunto executed this Data Sharing Agreement as of the last date below written.

 Executive Director
<State> DOT

 Executive Director
<City> DPW

Date: _____

Date: _____

DATA INTEGRATION AGREEMENT

<State> Department of Transportation and the <City> Department of Public Works

As defined by the <State> Safety Data Integration Partnership (SDIP) this agreement establishes the terms and conditions under which the Partnership will manage integrated safety data resources including State, Tribal and local data files of crash, roadway inventory, and traffic volume data. Roles are defined for Data Providers, Integrators, and Users. An agency may play more than one role in the data integration.

Data Providers collect safety data and make it available to integrators and users. They are responsible for any data collection in their jurisdiction, including periodic updates as required under the SDIP data collection and integration plan.

Data Integrators are responsible for creating merged safety data sets from all submitted sources. This is to be accomplished in accordance with the data sharing agreements established between each data provider and the data integrator. Any integrated data files created as part of this process are subject to data custodial requirements of each of the agencies that contributed data to the final product. The Data Integrator is responsible for security of the integrated data, access control, and managing the approval process for users requesting analyses or access to the data.

Data Users are responsible for conducting analyses in accordance with the SDIP and <State> data governance guidelines, and for adhering to all data sharing agreements and data integration guidelines established through inter-agency agreements and managed by the Data Integrator.

The following rules apply to integrated safety data managed under the auspices of the SDIP:

- 1) Data Users shall not attempt to uncover the identity of any person in any of the safety data records. If a user discovers a method by which personal information can be gleaned from the integrated data directly, or indirectly, that user will contact the data integrator and provide specific information on how it might be done. The data integrator is responsible for addressing the potential for release of personally identifying information in accordance with SDIP guidance, <State> data governance guidance, and applicable State and Federal laws.
- 2) Data Providers shall adhere to the data update schedule and the data standards defined by the SDIP and the <State> data governance group for each safety data element they provide for the integration effort.
- 3) Data Integrators shall ensure that all submitted data are kept secure, that only authorized users have access to the integrated data, and that any request for data is subject to a formal approval process as defined by the <State> data governance group.

Note: This document is intended for use in Safety Data Sharing agreements as an added statement of requirements. It is adopted by reference. The agreement is subject to change with the approval of the SDIP and the <State> data governance group.

<State> DOT, and SDIP chair

<State> Data governance group chair

MEMORANDUM OF UNDERSTANDING for the Washington Traffic Records Committee



THIS MEMORANDUM OF UNDERSTANDING (MOU), which shall be effective upon execution by signature of all parties, is entered into among the following agencies, collectively referred to as the parties:

- Washington Traffic Safety Commission (WTSC) pursuant to the authority of chapter 43.59 R.C.W.;
- Administrative Office of the Courts (AOC) pursuant to the authority of chapter 2.56 R.C.W.;
- Washington State Patrol (WSP) pursuant to the authority of chapter 43.43 R.C.W.;
- Washington State Department of Transportation (WSDOT) pursuant to the authority of chapter 47.01 R.C.W.;
- Washington State Department of Licensing (DOL) pursuant to the authority of chapter 43.24 R.C.W.;
- County Road Administration Board (CRAB) pursuant to the authority of chapter 36.78 R.C.W.;
- Washington State Department of Health (DOH) pursuant to the authority of chapter 43.70 R.C.W.;
- Washington Association of Sheriffs & Police Chiefs (WASPC) pursuant to the authority of chapter 36.28A R.C.W.; and
- Washington State Office of Financial Management (OFM) pursuant to the authority of chapter 43.41 R.C.W.;

WHEREAS the Washington Traffic Safety Commission is responsible for the planning, development, administration, and coordination of an integrated framework for traffic safety planning and action among all agencies and organizations in Washington and the successful implementation of traffic safety programs must involve the combined efforts of a number of agencies and organizations to be successful;

WHEREAS traffic records data is integral to the completion of such agencies and organizations' shared mission to reduce the number of fatalities and injuries and the severity of injuries related to trauma; and

WHEREAS the parties wish to improve the timeliness, accuracy, completeness, uniformity, integration and accessibility of traffic records data to identify priorities for national, state and local highway and traffic safety programs; and

WHEREAS the parties seek to make such improvements and to enhance interoperability among Washington's traffic records systems and other state and national systems; and

WHEREAS in support of such purposes the parties named above have established an interagency highway safety data and traffic records coordinating committee, entitled the Washington Traffic Records Committee (TRC); and

NOW, THEREFORE, in furtherance of the foregoing and mutual public benefits derived there from, it is agreed as follows:

Section 1 – OVERVIEW & PURPOSE

I. Traffic Records.

The parties recognize that Washington's traffic records system is a virtual system comprised of the hardware, software and accompanying processes that capture, store, transmit, and analyze the following types of data:

- Collisions
- Citations & Adjudication
- Drivers & Registered Vehicles
- Traffic Fatalities
- Motor Carriers (Commercial Vehicles)
- Injury Surveillance (Emergency Medical Services, Emergency Department, Trauma, Hospital inpatient, Death Records)
- Roadway (Traffic Volume, Features Inventory, Geometrics, etc.) and Location (Geographic Information Systems)

Each component of Washington's traffic records system provides key information to support decisions regarding public and transportation safety. The traffic records system provides critical data for problem identification and for the development of policy and countermeasure programs. Information derived from these systems is equally valuable in evaluating program effectiveness and documenting progress toward key measures of performance to enhance management and accountability in public service. Timely, accurate, integrated, and accessible traffic records data is crucial to Washington's efforts to improve public safety.

II. Mission.

The Washington Traffic Records Committee enhances transportation safety through coordinated projects to provide more timely, accurate, integrated and accessible traffic records data.

III. Goals.

The parties agree to cooperate in good faith to achieve the goals following:

1. To provide an ongoing statewide forum for traffic records and support the coordination of multi-agency initiatives and projects.

2. To leverage technology and appropriate government and industry standards to improve the collection, dissemination, and analysis of traffic records data.
3. To improve the interoperability and exchange of traffic records data among systems and stakeholders for increased efficiency and enhanced integration.
4. To promote the value of traffic records data and encourage training opportunities to maximize the effectiveness of the data for decision and policy making.

Section 2 – OPERATIONAL AUTHORITY

The Washington Traffic Records Committee operates under the authority of the agencies with either a custodial or contributive responsibility for the collection, management, use, or support of one or more components of Washington’s traffic records system.

Section 3 – ORGANIZATIONAL STRUCTURE

The Washington Traffic Records Committee is comprised of two separate bodies, the Oversight Council and the Traffic Records Workgroup, the missions of which are set forth below. The Oversight Council and the Traffic Records Workgroup serve in distinct capacities as outlined in Section 4 – DUTIES AND RESPONSIBILITIES.

I. Oversight Council.

The Oversight Council provides policy oversight and program direction in creating and approving strategies and projects to improve Washington’s traffic records system. The Oversight Council ensures strategic and project alignment with individual agency priorities, standards, and practices and performs an annual evaluation of Washington’s traffic records strategic plan.

II. Traffic Records Workgroup.

The Traffic Records Workgroup functions as a technical and managerial forum for the discussion and examination of statewide traffic records issues. The Traffic Records Workgroup is responsible for developing the state’s traffic records strategic plan and for creating, coordinating, and implementing improvement projects.

III. Administration.

The Washington Traffic Safety Commission shall provide the necessary support to assist and coordinate the Oversight Council and the Traffic Records Workgroup in fulfilling the mission and goals of Washington’s Traffic Records Committee. This support shall include a coordinator to manage federal traffic records grants and to serve as liaison for traffic records activities in Washington.

Section 4 – DUTIES AND RESPONSIBILITIES

I. Oversight Council

The duties and responsibilities of the Oversight Council shall include the following:

1. To provide policy oversight and program direction for statewide traffic records activities.
2. To provide a policy level stakeholder forum for review and discussion of proposed traffic records projects to assess and provide comment on system wide impacts.
3. To review and take action on strategic, project, or legislative recommendations provided by the Traffic Records Workgroup.

4. To promote communication and coordination of traffic records among and within participating agencies.
5. To conduct an annual evaluation for approval of Washington's Traffic Records Strategic Plan.

II. Traffic Records Workgroup.

The duties and responsibilities of the Traffic Records Workgroup shall include the following:

1. To create and maintain the Washington Traffic Records Strategic Plan and Resource Manual.
2. To establish goals, objectives, and strategies to improve the traffic records system.
3. To provide a technical stakeholder forum for review and discussion of proposed traffic records projects to assess and provide comment on system wide impacts.
4. To provide administrative and technical guidance in the planning, coordination, and implementation of traffic records improvement projects.
5. To identify performance measure benchmarks and targets to evaluate the effectiveness of strategies and projects aimed at improving Washington's traffic records system.
6. To recommend procedural, content, and format changes to the Police Traffic Collision Report (PTCR) and related data collection software applications to improve the quality, completeness, and uniformity of statewide collision data.
7. To review current laws and proposed legislation to assess traffic records system impacts.
8. To evaluate new technologies and potential implications for the traffic records system.
9. To conduct periodic audits or assessments of Washington's traffic records system.

Section 5 – MEMBERSHIP

I. Members.

a. The Oversight Council shall include the members following:

1. Washington Traffic Safety Commission, Director
2. Administrative Office of the Courts, Judicial Services Division Director
3. Washington State Patrol, Assistant Chief Technical Services Bureau
4. Washington State Department of Transportation, Transportation Data and GIS Office Manager
5. Washington State Department of Licensing, Chief Information Officer
6. County Road Administration Board, Intergovernmental Policy Manager
7. Washington State Department of Health, Assistant Secretary
8.

9. Washington Association of Sheriffs & Police Chiefs, Chief or Sheriff
10. Washington State, Office of the Chief Information Officer, Sr. Policy Consultant

b. The Traffic Records Workgroup shall include the members following:

The Traffic Records Workgroup shall be made up of representatives from the agency offices or divisions listed below at positions 1 through 15. Each representative shall be appointed and serve at the discretion of the Oversight Council Member representing that member's respective agency. It is anticipated that the National Highway Traffic Safety Administration, the Federal Motor Carrier Safety Administration and the Federal Highway Administration shall also be members of the Traffic Records Workgroup (known collectively as the U.S. Department of Transportation agencies). The parties contemplate that members representing U.S. Department of Transportation agencies (positions 16 – 18) shall be appointed by and serve at the discretion of the Region or Division Administrator of their respective agencies. Members representing U.S. Department of Transportation agencies shall be non-voting members of the Traffic Records Workgroup.

1. Washington State Patrol, Technical Services Bureau
2. Washington State Patrol, Field Operations Bureau
3. Washington State Department of Licensing, Driver Records
4. Washington State Department of Licensing, Title & Registration
5. Administrative Office of the Courts, Information Services Division
6. Administrative Office of the Courts, Judicial Services Division
7. Washington Traffic Safety Commission, Traffic Records Program Manager
8. Washington Traffic Safety Commission, Research & Data Division
9. Washington State Department of Transportation, Statewide Travel & Collision Data Office
10. Washington State Department of Transportation, GIS & Roadway Data Office
11. Washington State Department of Health, Community Health System
12. Washington State Department of Health, Center for Health Statistics
13. Washington Association of Sheriffs & Police Chiefs
14. County Road Administration Board
15. Washington State Office of the Chief Information Officer
16. National Highway Traffic Safety Administration, Pacific Northwest Region Office
17. Federal Motor Carrier Safety Administration, Washington Division

18. Federal Highway Administration, Washington Division

II. Term of Appointed Members.

Appointed Members for both the Oversight Council and the Traffic Records Workgroup shall serve at the pleasure of their respective appointing agencies.

III. Resignation.

An appointed Member may resign at any time by delivering written notice to the Chairperson, or by giving oral notice of resignation at any meeting. Upon resignation, the resigning Member shall recommend a replacement to fill the resulting vacancy.

IV. Vacancies.

The Chairperson or Co-Chairs shall bring a vacancy in either the Oversight Council or the Traffic Records Workgroup to the attention of the agency whose appointed member has vacated his or her position. A replacement shall be named at the discretion of such appointing agency within three months of the vacancy announcement.

V. Replacement of Appointed Members - Absenteeism.

a. Oversight Council

Any appointed Member of the Oversight Council who misses three (3) consecutive meetings will have such absences called to the Member's attention by the Chairperson. The Chairperson may advise the appropriate agency of continuing absenteeism and request that the appropriate agency replace the appointed Member.

b. Traffic Records Workgroup

Any appointed Member of the Traffic Records Workgroup who misses five (5) consecutive meetings will have such absences called to the Member's attention by the Co-Chairs. The Co-Chairs may advise the appropriate agency of continuing absenteeism and request that the appropriate agency replace the appointed Member.

VI. Stakeholders.

The Oversight Council may appoint stakeholder representatives to either the Oversight Council or the Traffic Records Workgroup as additional, voting or nonvoting members.

Section 6 – CHAIR PERSONS and CO-CHAIRS

I. Chairpersons.

a. Oversight Council.

The Director of the Washington Traffic Safety Commission shall act as the permanent Chairperson and coordinator for the activities of the Oversight Council.

b. Traffic Records Workgroup.

There shall be two Co-Chairs of the Traffic Records Workgroup. One Co-Chair shall be the Traffic Records Coordinator from the Washington Traffic Safety Commission. The second Co-Chair shall be elected from among the membership of the Traffic Records Workgroup and shall serve for a period of one year. The second Co-Chair may be re-elected upon the expiration of his or her term. The presiding facilitator of a particular Traffic Records Workgroup meeting shall be determined by the Co-Chairs on a meeting by meeting basis.

II. Acting Chairpersons.

a. Oversight Council

In the case of the absence of the Chairperson of the Oversight Council, the Chairperson may designate in advance of a particular meeting an Acting Chair to preside at the meeting. In the case of the absence of the Oversight Council Chairperson, and when an Acting Chair has not been designated, the Oversight Council may delegate the powers or duties of such officer to any Member for a particular meeting. In the case of a vacancy of the Chairperson of the Oversight Council, the Deputy Director of the Washington Traffic Safety Commission shall be the Acting Chair until the vacancy is filled.

b. Traffic Records Workgroup

In the case of an anticipated absence or vacancy of one or both of the Co-Chairs of the Traffic Records Workgroup, the Co-Chairs may designate in advance of a particular meeting an Acting Chair or Chairs to preside at the meeting. In the case of the absence of one or both of the Traffic Records Workgroup Co-Chairs, and when an Acting Chair or Chairs has not been designated, the Traffic Records Workgroup may delegate the powers or duties of such officer to any Member or Members for a particular meeting. In the case of vacancy of one of the co-chairs, the other Co-Chair shall preside at the meetings until such time as the vacant Co-Chair position is duly appointed or elected, depending on which Co-Chair position is vacant. In the case of the vacancy of both Co-Chairs, the Traffic Records Workgroup may delegate the powers and duties of the vacant Co-Chairs to any members of the Traffic Records Workgroup for a particular meeting, until such time as one or both of the vacant Co-Chair positions is duly appointed or elected, depending on which Co-Chair position is vacant.

Section 7 – MEETINGS

I. Regular Meetings.

a. Oversight Council.

Regular meetings of the Oversight Council shall be held each quarter at a time and place designated by the Chairperson.

b. Traffic Records Workgroup.

Regular meetings of the Traffic Records Workgroup shall be held monthly at a time and place designated by the Co-Chairs.

II. Special Meetings.

a. Oversight Council

Special meetings of the Oversight Council shall be held at the call of the Chairperson or upon request of any three (3) voting Members.

b. Traffic Records Workgroup

Special meetings of the Traffic Records Workgroup shall be held at the call of either Co-Chairs or upon request of any five (5) voting Members.

III. Quorum.

a. Oversight Council.

A quorum for the transaction of business shall constitute not less than five (5) members of the Oversight Council, and shall include within those five members the presiding Chairperson or designated Acting Chair. The Members present at an Oversight Council meeting at which a quorum is not constituted may elect to proceed only with informational and procedural portions of the meeting.

b. Traffic Records Workgroup

A quorum for the transaction of business shall constitute not less than nine (9) members of the Traffic Records Workgroup, including the presiding Co-Chair or designated Acting Chair. The Members present at a Traffic Records Workgroup meeting at which a quorum is not constituted may elect to proceed only with informational and procedural portions of the meeting.

IV. Meeting Conduct.

a. Discussion.

Only Members of the Oversight Council or Traffic Records Workgroup, appointed Stakeholders, and such other persons as are recognized by the presiding Chairperson shall be permitted to participate in discussion of matters of business, unless otherwise authorized by a majority vote of the Members.

b. Chairperson Votes.

The Chairperson of the Oversight Council and the Co-Chairs of the Traffic Records Workgroup shall have the same voting rights as any other Member of the Oversight Council and the Traffic Records Workgroup.

c. Member Voting.

1. Oversight Council.

Only Members of the Oversight Council may vote. Members may not vote by proxy or through a designee.

2. Traffic Records Workgroup.

Only individuals representing the agency offices or divisions listed in Section 5.1.b may vote. The vote may occur by proxy.

d. Majority Vote.

The action by vote of the majority of the Members present at a meeting at which there is a quorum shall be the act of Oversight Council or the Traffic Records Workgroup.

e. Robert's Rules of Order.

The parties agree to use their best efforts to apply Robert's Rules of Order to meetings of the Oversight Council and the Traffic Records Workgroup and any of its ad hoc or standing subcommittees except as otherwise provided in this MOU.

V. Attendance by Communication Equipment.

Meeting attendance may be by means of conference telephone call or any other communications equipment that allows all persons participating in the meeting to speak to and hear all participants. Participation by such means shall constitute presence in person at a meeting.

VI. Meeting Notices.

Advance notice of all regular and special meetings of the Oversight Council and the Traffic Records Workgroup shall be provided by mail, facsimile transmission or email.

VII. Meeting Minutes.

Minutes shall be made of all Oversight Council and Traffic Records Workgroup meetings. Minutes of Oversight Council meetings will be promptly distributed to members for review and approval at the following meeting. Minutes from Oversight Council and Traffic Records

Workgroup meetings will be regularly available on the Washington Traffic Records Committee website (<http://www.trafficrecords.wa.gov>).

Section 8 – SUBCOMMITTEES

I. Ad Hoc or Standing Subcommittees.

The Traffic Records Workgroup may, by the authorization of the Oversight Council, establish such ad hoc or standing subcommittees as deemed appropriate. The ad hoc or standing subcommittee membership and chairperson shall be designated by the Traffic Records Workgroup.

II. Subcommittee Authority.

The Traffic Records Workgroup may delegate project planning, coordination, and implementation authority to ad hoc or standing subcommittees as deemed appropriate.

III. Procedures.

Ad hoc or standing subcommittees shall follow all Traffic Records Workgroup procedures as defined in this MOU.

Section 9 – AMENDMENTS

I. By Oversight Council.

The Oversight Council shall have power to make, alter, and amend this MOU and shall not be effective unless in writing and signed by all members of the Oversight Council.

II. By Undersigned Parties.

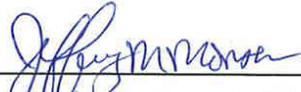
The undersigned Parties shall have power to make, alter, amend, and repeal this MOU upon written agreement, signed by all parties to this MOU.

Section 10 – GOOD FAITH

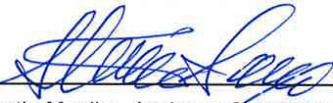
I. The parties agree to conduct all activities and perform all obligations in good faith and to work cooperatively with one another to accomplish the goal of providing timely, accurate, integrated and accessible traffic records data.

SIGNED AND ACCEPTED:

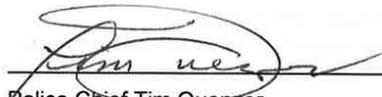

Darrin Grondel, Director
Washington Traffic Safety Commission
Date: 6/24/14

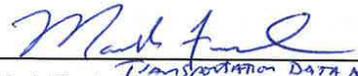

Jeff Monson, Intergovernmental Policy
Manager
County Road Administration Board
Date: 5/5/14


~~JUDICIAL SERVICES DIR.~~
Dirk Marler, Chief Information Officer
Administrative Office of the Courts
Date: 7/25/14


~~Steven Saxe, Director, OCHS~~
Martin Mueller, Assistant Secretary
Department of Health
Date: 8/12/14


Assistant Chief Shawn Berry
Washington State Patrol, Technical
Services Bureau
Date: 5/5/14


Police Chief Tim Quenzer
Washington Association of Sheriffs &
Police Chiefs
Date: May 5, 2014


~~Transporation Data and~~
Mark Finch, Statewide Traffic & GIS
Collision/GIS & Roadway Data Office
Manager
Department of Transportation
Date: 5/5/14


Scott Bream, Sr. Policy Advisor
Office of the Chief Information Officer
Date: 6/4/2014


Melissa Spencer, Chief Information Officer
~~Melissa Spencer, Asst. Dir.~~
Department of Licensing
Date: 8/14/14

APPENDIX E – SAMPLE SURVEY

GIS Needs Assessment

Enterprise GIS Integration & Analysis

Section 1: Background GIS Information -To be completed by each department within the DOT

First Name

Last Name

Title

Department

Phone

Email

[Redacted]

- Less than one year
- 1-2 years
- 3-5 years
- More than 5 years
- Not currently, but plan to implement GIS in one year
- Not using GIS
- Don't know

[Redacted]

- ESRI Desktop Software
- ESRI Server Software
- MapInfo
- Caliper Maptitude or TransCAD

- Cube GIS
- Quantum
- Other

[Redacted]

[Redacted]

- Planning stage
- Limited Use
- Extensive use
- Other _____

[Redacted]

- Government agencies
- Commercial sources
- Developed in-house
- Other _____

[Redacted]

- Yes
- No

[Redacted]

Q8. What GIS data do you wish you had access to, but don't have access to currently? (Sources can be within your department or from outside agencies).

[Redacted]

- Project
- Department
- Enterprise-wide
- Other

[Redacted]

- Yes _____

- No

11. Do you need GIS data beyond your jurisdictional boundaries?

- Yes
- No

Q12. Do you have access to GIS data beyond your jurisdictional boundaries? If No, please specify.

- Yes
- No _____

Q13. Do you have access to data outside your department?

- Yes. But only publish data for internal use
- Yes. Publish data to the public _____
- No. We don't use web-GIS service

Q14. Does your department provide GIS data to other organizations?

- Government Agencies
- Non-profit organizations
- Private companies
- Other _____
- No

Q15. Number of GIS Employees and % effort dedicated to GIS Activities

GIS Staff (does not include users)			
GIS Coordinator or Manager			
GIS Analyst / Programmer / Developer			
GIS Users			
Other			

Q16. Existing GIS data maintained by your department

Layer Name	Format (Shapefile, File/Personal Geodatabase, Other)	Geometry Type (Polygon, Line, Point, Table)	Metadata or Data Dictionary	Status (Complete, Partial, Limited Coverage, etc.)	Scale	Maintenance Schedule

Section 2: DOT GIS System, Technical Details – To be filled out by DOT GIS Coordinator

Q1. Please list current GIS Software (including extensions) and number of licenses?

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.

Q2. Please list available GIS Hardware (both enterprise and desktop level)?

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.

Q3. Please indicate your existing GIS Server network connection speed?

Q4. How are web-mapping services currently used?

Q6. Do you have an ESRI ArcGIS Online account?

Q7. Have you developed any ArcGIS Online web-mapping applications?

Q8. Do any GIS sharing agreements exist between internal departments?

Q9. Do any GIS sharing agreements exist between the DOT and outside agencies?

Q10. Is there a data catalog with data definitions, standards, policies, and procedures for the collection and use of data available electronically in the DOT and is it accessible to users?

Q11. Which databases have standard business rules? (Check all that apply.)

- Roadway inventory data
 - Traffic Data
 - Crash data
 - Citation/adjudication data
 - Injury data
 - Driver data
 - Vehicle data
 - Other (please specify).
-

Section 3: Roadway Data/Standards

Q1. Does the DOT have a basemap of the roadway network?

<input type="checkbox"/> Yes, linear referencing system where each inventory record is a “homogeneous section” (i.e., all inventory elements/descriptors are constant through the entire section length) and each record is defined by an “address” such as the route and milepost of the beginning of each section
<input type="checkbox"/> Yes, GIS system where the inventory elements/attributes for each ft. (or each x ft.) of road are stored spatially and can be linked to the GIS base roadway network
<input type="checkbox"/> Yes, GIS system where the inventory elements are generally consistent along a segment, however, some may vary slightly and the average, min, or max is used for that segment
<input type="checkbox"/> Yes, other (please describe)
<input type="checkbox"/> No (skip #2)
<p>Q2. What roadways are covered in the basemap?</p> <input type="checkbox"/> DOT-owned roadways: <input type="checkbox"/> Tribal <input type="checkbox"/> Other

Q3. What types of primary roadway data are included in your roadway data inventory system? If collected, please briefly describe how the data were originally collected.

Data collection techniques/technologies may include: As-built plans (AB), Field survey (FS), Instrumented vehicle (IV), Aerial Photos (AP), and Other, please describe (O, description).

MIRE Elements	Collected	How it was collected
I. Roadway Segment Descriptors		
Segment Cross Section		
Roadside Descriptors		
Segment Traffic Flow Data		
Segment Traffic Operations/Control Data		
II. Roadway Alignment (Curve and Grade) Descriptors		
III. Roadway Junction Descriptors		
At-Grade Intersection/Junctions		
Interchange and Ramp Descriptors		

Q4. What percentage of your system has roadway inventory data that are

a) DOT roadways:

b) Tribal:

c) Other:

[Redacted]

Annually

Quarterly

Monthly

Daily

Q6. What percentage of the MIRE elements are in the inventory file, what percent of roadways are they collected on, and what elements are planned for future collection?

Additional information on MIRE can be found at <http://www.mireinfo.org>.

MIRE Elements (# of elements)	Percent of Elements		Percent of Roadways Collected On		Planned Future Collection	
	DOT %	Other %	DOT (All, Most, Some)	Other (All, Most, Some)	DOT (Y/N)	Other (Y/N)
I. Roadway Segment Descriptors						
I.a. Segment Location/Linkage Elements (18)						
I.b. Segment Roadway Classification (4)						
I.c. Segment Cross Section						
I.c.1. Surface Descriptors (8)						
I.c.2. Lane Descriptors (12)						
I.c.3. Shoulder Descriptors (11)						
I.c.4. Median Descriptors (8)						
I.d. Roadside Descriptors (13)						
I.e. Other Segment Descriptors (4)						
I.f. Segment Traffic Flow Data (12)						
I.g. Segment Traffic Operations/Control Data (15)						
I.h. Other Supplemental Segment Descriptors (1)						
II. Roadway Alignment Descriptors						
II.a. Horizontal Curve Data (8)						
II.b. Vertical Grade Data (5)						

MIRE Elements (# of elements)	Percent of Elements		Percent of Roadways Collected On		Planned Future Collection	
	DOT %	Other %	DOT (All, Most, Some)	Other (All, Most, Some)	DOT (Y/N)	Other (Y/N)
III. Roadway Junction Descriptors						
III.a. At-Grade Intersection/Junctions						
III.a.1. At-Grade Intersection/Junction - General Descriptors (18)						
III.a.2. At-Grade Intersection/Junction Descriptors - Each Approach (40)						
III.b. Interchange and Ramp Descriptors						
III.b.1. General Interchange Descriptors (8)						
III.b.2. Interchange Ramp Descriptors (17)						

Q7. What supplemental datasets are included in your roadway data inventory system? If collected, please briefly describe how the data were originally collected.

Data collection techniques/technologies may include: As-built plans (AB), Field survey (FS), Instrumented vehicle (IV), Aerial Photos (AP), and Other, please describe (O, description).

Supplemental Data	Collected DOT (All, Most, Some, None)	Collected Other (All, Most, Some, None)	How is it Collected	How is it Linked: Tribal/Local
Access management				
Automated enforcement devices				
Bridge descriptors				
Land use elements related to safety				
Lighting				
Pedestrian				
Bicycle				
Pavement condition				
Railroad grade-crossing descriptors				
Roadside fixed objects				
Safety improvements				
Signs				
Speed				
Other (please describe).				

Q8. What type of procedure do you have to update each of your inventory data types? What is the time lapse between the “open to traffic” date of a new roadway or roadway modification and when the revised data are included in each inventory file?

Inventory Type	Update Procedure	Time Lapse
Roadway Segment		
Traffic		
Intersection		
Interchange		
Ramp		
Curve		
Grade		
Other (please describe)		

Example procedures might include the following:

- Ad hoc procedure – no standardized procedure, but changes to the file are made when they come to the attention of the file maintainer.
- Annual (or less often) survey of entire or part of the roadway system (e.g., the roadway system is re-inventoried over a five-year period).
- On-going “as roadway is modified” process where descriptions or “as-built” plans are submitted to the file maintainer each time a change is made to the roadway or a new roadway segment is opened to traffic. The data for the affected section or location are then updated.
- Updates vary by data type.
- Other (please describe).

Example time lapses might include the following:

- There is no systematic updating process, thus the time varies greatly.

- *Need time to develop.*
- *More than one year.*
- *6 – 12 months.*
- *1-5 months.*
- *1 month or less.*

[Redacted]

[Redacted]

Whether an inventory element/item/file was updated.

How: _____

Which Element/Characteristic of that item was updated?

How: _____

The date when the updating took place.

How: _____

[Redacted]

No one

Internal users (DOT/Other State agencies)

External users

[Redacted]

Q13. If your inventory system contains data on more than one LRS (perhaps from different jurisdictions), are element definitions and coding consistent across all jurisdictions? If not, describe the differences for each data type roadways (e.g., roadway segments, intersections, curves, etc.) included.

[Redacted]

Q15. Are there procedures in place to ensure that the same "site address" (e.g.,

[Redacted]

[Redacted]

- All DOT maintained roadways
- A portion of DOT roadways

[Redacted]

[Redacted]

- Highway Safety Manual
- SafetyAnalyst
- Interactive Highway Safety Design Model
- CMF Clearing House
- Other: Please Specify:

- None

[Redacted]

- Network Screening
- Countermeasure Selection
- Evaluation
- Other. Please Specify:

Q20. Does the DOT maintain records for roadway safety improvement projects?

- Yes
- No

[Redacted]

- Installation date
- Location
- Project type

- Project cost

Q22. Does DOT have the ability to link crash data to the safety improvement project site(s) or interest to? If so, how many years of historical crash data are available?

- No
- Yes. Number of years available: _____

ACKNOWLEDGEMENTS

FHWA Office of Safety Technologies Project Staff:

Ray Krammes, Data and analysis Tools Team Leader

Yanira Rivera, Intern

Stuart Thompson, Task Manager

Advisory Panel Members:

Victor Lund, St. Louis County Public Works Department, Minnesota

Dennis Trusty, North Plains Tribal Technical Assistance Program; United Tribes Technical College

Matthew Enders, Washington State Department of Transportation

Karen Carroll, Iowa Department of Transportation

Bradley Estochen, Minnesota Department of Transportation

Craig Thor, Federal Highway Administration

Rosemarie Anderson, Federal Highway Administration

Pilot Study Representatives:

Arizona DOT pilot study:

Pradeep Tiwari, Arizona Department of Transportation

Maysa Hanna, Arizona Department of Transportation

Kelly LaRosa, Federal Highway Administration Arizona Division Office

Josh Pope, Pima Association of Governments

Paul Casertano, Pima Association of Governments

Vladimir Livshits, Maricopa Association of Governments

Robert Maxwell, Bureau of Indian Affairs

Paul Bonar, Bureau of Indian Affairs

Fort Belknap work plan:

John Healy, Fort Belknap Transportation Department

Dawn Chandler, Fort Belknap Transportation Department

Dennis Trusty, Northern Plains Tribal Transportation Assistance Program

Mike Toland, Bureau of Indian Affairs

James Combs, Montana Department of Transportation

Indiana LTAP pilot study:

Bob McCullough, Indiana Local Technical Assistance Program

Laura Slusher, Indiana Local Technical Assistance Program

Michael Ricketts, Putnam County

Michael Holowaty, Indiana Department of Transportation

Navajo DOT pilot study:

Garren Burbank, Navajo Department of Highway Safety

Luis Melgoza, Federal Highway Administration New Mexico Division Office

Kelly LaRosa, Federal Highway Administration Arizona Division Office

Adam Larsen, Federal Highway Administration Tribal Transportation Program

Rhode Island DOT pilot study:

Robert Rocchio, Rhode Island Department of Transportation

Sean Raymond, Rhode Island Department of Transportation

Jacinda Russell, Federal Highway Administration Rhode Island Division Office

North Carolina DOT work plan:

Brian Mayhew, North Carolina Department of Transportation

For More Information:

Visit www.safety.fhwa.dot.gov/rsdp

FHWA, Office of Safety

Stuart Thompson
Stuart.Thompson@dot.gov
202-366-8090