Final Report

Contract No. BDV29-977-07

Extraction of Basic Roadway Information for Non-State Roads in Florida

Prepared for:

Research Center
Florida Department of Transportation
605 Suwannee Street, M.S. 30
Tallahassee, FL 32399-0450

Prepared by:

Wanyang Wu, Ph.D., Research Associate
Albert Gan, Ph.D., Professor
Priyanka Alluri, Ph.D., P.E., Assistant Professor

Lehman Center for Transportation Research
Florida International University
10555 West Flagler Street, EC 3680
Miami, FL 33174
Phone: (305) 348-3116
Fax: (305) 348-2802
E-mail: gana@fiu.edu

FIU | FLORIDA INTERNATIONAL UNIVERSITY

June 2015

DISCLAIMER

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the State of Florida Department of Transportation.

METRIC CONVERSION CHART

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
LENGTH				
in	inches	25.400	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.610	kilometers	km
mm	millimeters	0.039	inches	in
m	meters	3.280	feet	ft
m	meters	1.090	yards	yd
km	kilometers	0.621	miles	mi
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
AREA				
in^2	square inches	645.200	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	2.590	square kilometers	km^2
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	2.470	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
VOLUME				
fl oz	fluid ounces	29.570	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
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
NOTE: volumes greater than 1,000 L shall be shown in m^3.				

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle	5. Report Date
Extraction of Basic Roadway Information for Non-State Roads in Florida	June 2015
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Wanyang Wu, Albert Gan, and Priyanka Alluri	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Lehman Center for Transportation Research Florida International University 10555 West Flagler Street, EC 3680, Miami, FL 33174	11. Contract or Grant No. BDV29-977-07

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Research Center State of Florida Department of Transportation 605 Suwannee Street, M.S. 30, Tallahassee, Florida 32399-0450	Final Report August 2013 – June 2015
	14. Sponsoring Agency Code 99700-3596-119

15. Supplementary Notes

Mr. Joseph Santos, P.E., of the State Safety Office at the Florida Department of Transportation, served as the Project Manager for this project.

16. Abstract

The Florida Department of Transportation (FDOT) has continued to maintain a linear-referenced "All-Roads" map that includes both state and non-state local roads. The state portion of the map could be populated with select data from FDOT's Roadway Characteristics Inventory (RCI). However, the RCI data are available for only a small portion of the local roads in the All-Roads map, leaving a majority of the local roads in the map without the same data. Given the large number of local roads in the map, it is clearly not cost feasible to collect the data in the field. Methods that make use of existing data as alternatives to field data collection are thus needed.

One potential source of existing data is police crash reports. For every reported crash in Florida, the law enforcement officers record information on more than 300 variables to describe the site and time of the crash, the geometric conditions, the traffic control, and drivers' and pedestrian's characteristics. Accordingly, this project aims to develop methods to extract roadway data recorded in crash reports as a means to both acquiring and continually updating the All-Roads map for local roads in Florida. To the extent possible, the project attempted to extract data for the following four variables that are included in Florida's crash reports: number of through lanes, posted speed limit, shoulder type, and median type.

The data extraction process to acquire data for the four variables in this project includes three steps. The first step involves the extraction of data from crash records for as many road segments as possible. The second step covers the case in which a road segment does not have any crashes. In this step, the values are derived from their adjacent segments based on the assumption that roadway features are likely to be continuous. Finally, the third step focuses on the remaining segments for which data could not be extracted or derived in the first two steps. In this step, the missing data are manually collected using a web-based data collection application that is designed specially to facilitate the process of observing and recording information from satellite images in Google Maps.

17. Key Word	18. Distribution Statement
Non-state Roads, Data Extraction, Police Reports, GIS	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	46	

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

ACKNOWLEDGEMENTS

This research was funded by the Research Center of the Florida Department of Transportation (FDOT) under the direction of Mr. Darryll Dockstader. We are particularly grateful to our Project Manager, Mr. Joseph Santos, P.E., of the FDOT State Safety Office for his guidance and support throughout the project. We are also grateful to Dr. Kaiyu Liu for his assistance in data pre-processing and in the set up of the VRICS web application. The assistance of many student assistants who performed the manual data collection is also gratefully acknowledged.

EXECUTIVE SUMMARY

The Florida Department of Transportation (FDOT) has continued to maintain a linear-referenced All-Roads map that includes both state and non-state local roads. The state portion of the map could be populated with select data from FDOT's Roadway Characteristics Inventory (RCI). However, these RCI data are available for only a small portion of the local roads (i.e., the "off-system" roads) in the All-Roads map, leaving a majority of the local roads in the map without the same data. Given the large number of local roads in the map, it is clearly not cost feasible to try to collect the data in the field. Accordingly, this project aims to develop methods to extract roadway data recorded in crash reports as a means to both acquiring and continually updating the All-Roads map for local roads in Florida. To the extent possible, the project attempts to extract data for the following four variables that are included in Florida's crash reports:

- number of through roadway lanes,
- posted speed limit,
- shoulder type (paved or unpaved), and
- median type (divided or undivided).

The following databases were used to extract data on these four data variables:

- Crash databases:
  - Crash data in the shapefiles format for the years 2003-2012 for both on-system and off-system roads.
  - Roadway Table (RDWTBL) 50 for the years 2003-2012.

- Roadway characteristics inventory databases:
  - FDOT RCI database for the year 2015 for Type of Road variable.
  - FDOT Road Data shapefiles for:
    - Intersections
    - Number of Lanes
    - Shoulder Type – Outside
    - Maximum Speed Limits
  - All-Roads shapefiles for the year 2012 based on the data from NAVTEQ™.

The process to acquire information for the four data variables includes three steps. The first step involves extraction of data from crash records. Crash records for the years 2003 through 2012 are used to extract this information. In the case where there are multiple crashes with different data entries (e.g., different speed limits were recorded in the police reports for crashes that occurred on a segment), the process considered the number of crashes and the timeline of crash occurrences to select the most probable value. The second step covers the case in which a road segment does not have any crashes. In this step, the values are derived from their adjacent segment, and is based on the assumption that roadway features are continuous. Finally, the missing information is manually collected from satellite images in Google Maps. The process was facilitated using an in-house web-based system. The following sections summarize the results for each of the four data variables.

Number of Through Lanes

Although number of through lanes is an entry in the police reports, it was found that the data for the corresponding field, TRWAYLN, in RDWTBL 50 of the Crash Analysis Reporting (CAR) database were completely missing. However, because the All-Roads map already included the LANES variable for the number of through lanes, this part of the project became one of verifying the existing LANES data. The variable NOLANES in the RCI was used to verify the LANES data. The results show that the data was 96.3% accurate. This high percentage was expected as the verified LANES data came originally from the RCI. The 3.7% inconsistency was likely a result of mainly changes in the RCI data between 2012 (when the All-Roads map was updated) and 2015 (RCI data year used in verification) due to roadway construction. As 3.7% of the road links in the All-Roads map did not have the same number of through lanes as identified in the RCI database, these links (11,043 links in total) were manually verified in VRICS.

As the RCI data includes links on state roads and major off-system roads with functional classification codes 1-19, the inconsistencies in the RCI data and the All-Roads map could be because the major roads (i.e., freeway, arterials, and major collectors) were more likely to undergo frequent improvements as a result of upgrade or reconstruction. The remaining links in the All-Roads maps with functional classification codes 20-33 are mainly local roads. As these roads are typically two-lane and do not usually change, it was found that only about 0.5% of these roads needed to be updated or corrected.

Posted Speed Limit

Data for posted speed limit are recorded as SPDLIMIT in RDWTBL 50 and as SPEED in the RCI. Although SPEED_CAT variable in the All-Roads map has information on speed limit, it is not considered a reliable source since it does not provide the specific speed limit. After the data extraction and data derivation steps, 39.4% of the total links (i.e., 616,078 of 1,565,026 links) have information on speed limits. As part of the data verification step, the data obtained from the SPDLIMIT variable in RDWTBL 50 was compared to the data from the SPEED variable in the RCI. Only the links with functional classification codes 1-19 were verified as the data in the RCI includes only these links. The accuracy of this variable was found to be 62.7%. In other words, 62.7% of the total links verified were found to match those from the RCI.

Shoulder Type

Data for shoulder type are recorded as SHLDTYPE in both RDWTBL 50 and the RCI databases. However, the codes in the RCI are more detailed compared to the codes in RDWTBL 50. The police records shoulder type as paved, unpaved, or curbed, while the RCI codes shoulder type using the following codes: raised curb, paved, paved with warning device, lawn, gavel/marl, valley gutter, curb and gutter, other, curb with resurfaced gutter, and managed lane. Therefore, the prospect of using shoulder type extracted from RDWTBL 50 is limited. In addition, manual data collection and data verification of shoulder type cannot be reliably done as the difference between “paved” and “curb” shoulder types is difficult to determine from the existing satellite images. Consequently, RDWTBL 50 was not found to be a sufficiently accurate data source for acquiring the shoulder type data.

Median Type

Data for median type are recorded as TYPTRWAY in RDWTBL 50, DIVIDER in the All-Roads map, and TYPEROAD in the RCI. For the links available in both the RCI and the All-Roads map, the accuracy of the variable DIVIDER in the All-Roads map was found to be only 49.0%. Therefore, this variable is not considered as a reliable source for median type information. On the other hand, the accuracy of this variable in RDWTBL 50 was found to be 77.8%. However, there is a slight inconsistency in how the police define "divided" segments and the RCI's definition. The police code locations with only a physical barrier as "divided", while the RCI codes links separated by turn bays as "divided". Due to this inconsistency, the prospect of using median type information obtained from crash data (i.e., RDWTBL 50) is somewhat limited.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AADT	Average Annual Daily Traffic
API	Application Program Interface
CAR	Crash Analysis Reporting
DHSMV	Department of Highway Safety and Motor Vehicles
FDOT	Florida Department of Transportation
GIS	Geographic Information System
LRS	Linear Referencing System
RCI	Roadway Characteristics Inventory
RDWTBL	Roadway Table
SHS	State Highway System
TRCC	Traffic Records Coordinating Committee
UBR	Unified Basemap Repository
VRICS	Visual Roadway Inventory Collection System

CHAPTER 1
INTRODUCTION

1.1 Project Background and Objectives

The Florida Department of Transportation (FDOT) has continued to maintain a linear-referenced All-Roads map based on the data from NAVTEQ™. This All-Roads map includes data on both state and non-state local roads. The state portion of the map could be populated with select data from FDOT's Roadway Characteristics Inventory (RCI). However, these RCI data are available for only a small portion of the local roads (i.e., the "off-system" roads) in the All-Roads map, leaving a majority of the local roads in the map without the same data. Given the large number of local roads in the map, it is clearly not cost feasible to collect the data in the field. As such, methods that make use of existing data as alternatives to field data collection are needed.

One potential source of existing data is police crash reports. For every reported crash in Florida, the law enforcement officers record information on more than 300 variables to describe the site and time of the crash, the geometric conditions, the traffic control, and drivers' and pedestrian's characteristics. This project aims to develop methods to extract roadway data recorded in crash reports as a means to both acquiring and continually updating the All-Roads map for local roads in Florida. To the extent possible, the project attempts to extract data for the following four variables that are included in Florida's crash reports: number of through lanes, posted speed limit, shoulder type, and median type.

1.2 Data Extraction Process

The process to acquire data for the four variables (i.e., number of through lanes, posted speed limit, shoulder type, and median type) includes the following three major steps:

1. *Extract Data from Crash Records*: This step is designed to extract the available data from the crash records for as many roadway segments as possible. In the case where there are multiple crashes with different data entries (e.g., different speed limits were recorded in the police reports for crashes that occurred on a segment), the process considers the number of crashes and the timeline of crash occurrences to select the most probable value. This method is fine-tuned and its accuracy verified using the RCI data available for state roads.

2. *Derive Data Using Spatial Analysis*: This step addresses the scenarios in which a roadway segment does not have any police reports. The method relies on a Geographic Information System (GIS)-assisted approach with the assumption that roadway features tend to be continuous. For example, a roadway segment is likely to have a posted speed limit of 40 mph if it is determined from the crash records that its immediate upstream and downstream roadway segments both have 40 mph lanes. Again, using the RCI data available for the state roads, this method is fine-tuned to determine the extent of which feature data could reliably be inferred.

3. *Collect Data Manually*: In this third and final step of the process, data for the roadway

segments that cannot be extracted or derived in the first two steps are manually collected. An in-house web-based data collection application, Visual Roadway Inventory Collection System (VRICS), is applied to facilitate the process. VRICS integrates linear-referenced networks with Google Maps to allow the user to quickly navigate to different road segments to visually identify feature information through Google Maps' satellite images and to quickly record the observed information.

1.3 Report Organization

The rest of the report is organized as follows:

- Chapter 2 describes the data sources and data preparation efforts undertaken for this project.

- Chapter 3 focuses on the first step of the data extraction process, i.e., the process of extracting data from crash records. It discusses the GIS application process and the rules used to retrieve data from crash records. The chapter also describes the verification procedure used to compare the data extracted from the crash records and the data available in the RCI database.

- Chapter 4 focuses on the spatial analysis procedures adopted to acquire data for segments that do not have any police reports. It discusses the methodology and results, and also provides the procedure used to verify the accuracy of the derived data.

- Chapters 5 focuses on the manual data collection procedures adopted to collect data for the segments for which data could not be extracted using the first two steps. It briefly introduces the basic functions of the VRICS web-based application. It further provides the results of the manual data collection efforts.

- Chapter 6 provides a summary of this project effort and the relevant findings, conclusions, and recommendations.

CHAPTER 2
DATA SOURCES AND PREPARATION

This chapter discusses the crash and roadway databases used in this project. It also describes the data preparation efforts undertaken to extract the data on number of through lanes, posted speed limit, shoulder type, and median type for the road network in Florida. The following databases were used to extract data for these four variables:

- Crash databases:
  - Crash data in the shapefiles format for the years 2003-2012 for both on-system and off-system roads.[1]
  - Roadway Table (RDWTBL) 50 for the years 2003-2012.

- Roadway characteristics inventory databases:
  - FDOT RCI database for the year 2015 for Type of Road variable.
  - FDOT Road Data shapefiles[2] for:
    - Intersections
    - Number of Lanes
    - Shoulder Type – Outside
    - Maximum Speed Limits
  - All-Roads shapefiles for the year 2012 based on the data from NAVTEQ™.

2.1 Crash Data

The crash data used in this project came from the Crash Analysis Reporting (CAR) system maintained by the FDOT State Safety Office. The database was originally generated by merging crash data from the Department of Highway Safety and Motor Vehicles (DHSMV) with roadway data from FDOT. All reported crashes with a fatality, an injury, or a property damage that occurred on state roads are included in this database. Each crash in the database is given a unique crash identification number in the database.

The database includes data recorded in each police crash report, which covers data for more than 300 variables to describe the site and time of the crash, the geometric conditions, the traffic control, and drivers' and pedestrian's characteristics. The data collected in the police reports are classified into three major categories: crash-level, vehicle-level, and person-level information. RDWTBL 50 is the crash-level data table, while RDWTBL 51 and 52 are the vehicle-level and person-level data tables, respectively. As the variables of interest for this project are specific to each crash (i.e., crash-level information), RDWTBL 50 files for the years 2003-2012 were obtained and used to extract the target data.

[1] https://www3.dot.state.fl.us/unifiedbasemaprepository/
[2] http://www.dot.state.fl.us/planning/statistics/gis/roaddata.shtm

Although the CAR system contains location-specific markers, it does not include the spatial coordinates of the crashes (i.e., latitudes and longitudes of the crash locations). Since number of through lanes, posted speed limit, shoulder type, and median type are all location-specific, crash data in the shapefiles format for the years 2003-2012 from the Unified Basemap Repository (UBR) System maintained by the Florida Traffic Records Coordinating Committee (TRCC) were used. The UBR system provides separate shapefiles for crashes on on-system and off-system roads. These shapefiles were linked to the RDWTBL 50 file using the unique crash identification number.

The on-system crash database includes crashes recorded in the long-form crash reports within the state of Florida that occurred on the Florida State Highway System (SHS). On the other hand, the off-system crash database includes crashes recorded in the long-form crash reports within the state of Florida that did not occur on the SHS. This off-system database includes crashes on the public road network, and excludes crashes in parking lots, on private property, and on forest roads or private roads.

RDWTBL 50 includes relevant crash-level information for every crash in the system. The following four data variables were extracted from the RDWTBL 50:

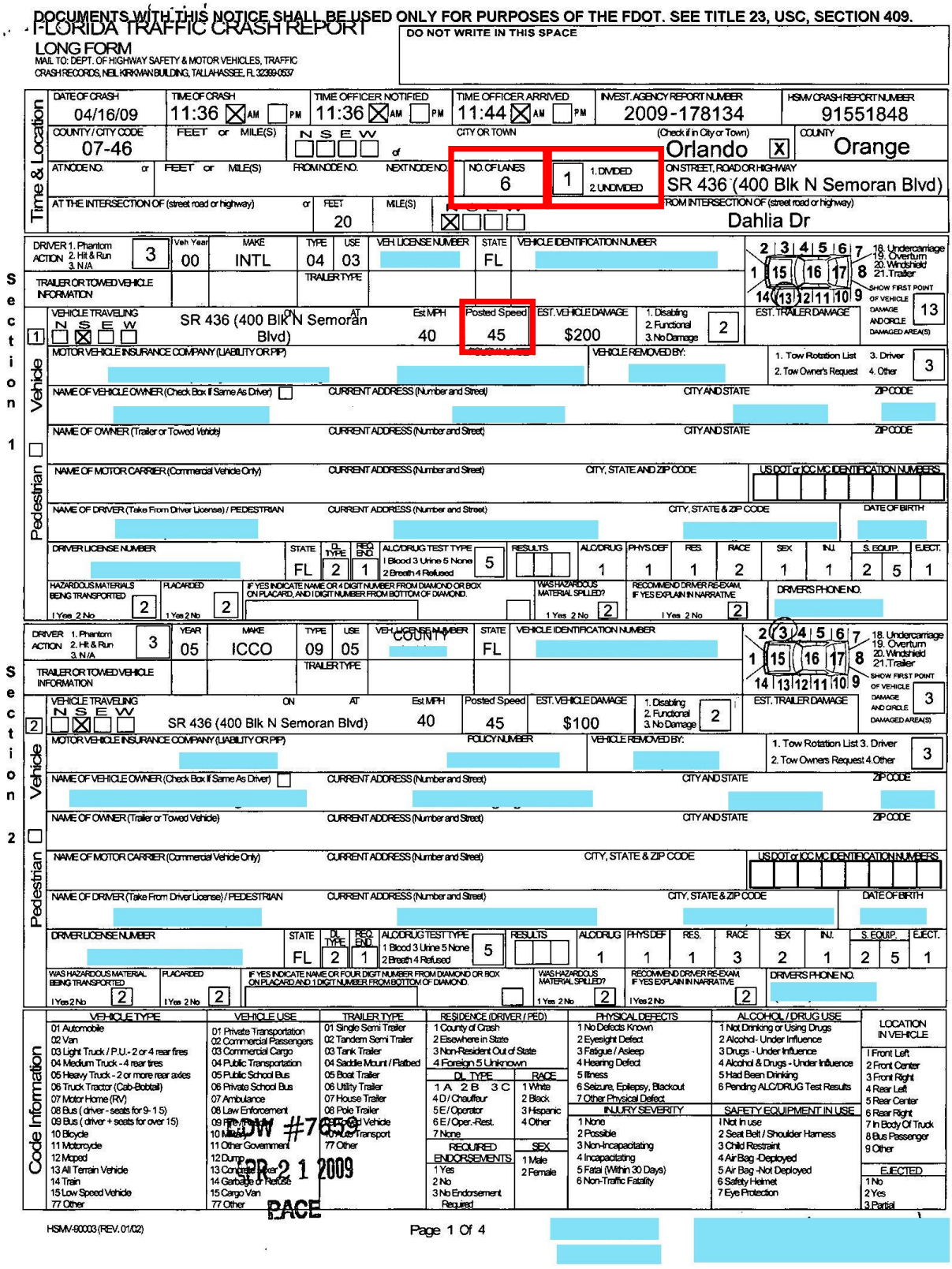
- NOLANES (number of through roadway lanes)
- SPDLIMIT (posted speed limit)
- SHLDTYPE (shoulder type)
- TYPTRWAY (median type)

Figures 2-1 and 2-2 show the first two pages of the crash report form. Note that the four variables of interest are marked in red boxes. As can be seen in these figures, number of through lanes and posted speed limit are given in actual numbers, while shoulder type and median type are categorical variables. Specifically, shoulder type is recorded using the following three codes:

- 1: Paved
- 2: Unpaved
- 3: Curb

Similarly, median type is recorded using the following two codes:

- 1: Divided
- 2: Undivided



DOCUMENTS WITH THIS NOTICE SHALL BE USED ONLY FOR PURPOSES OF THE FDOT. SEE TITLE 23, USC, SECTION 409.

FLORIDA TRAFFIC CRASH REPORT

LONG FORM

MAIL TO: DEPT. OF HIGHWAY SAFETY & MOTOR VEHICLES, TRAFFIC CRASH RECORDS, NEIL KIRKMAN BUILDING, TALLAHASSEE, FL 32399-0537

DO NOT WRITE IN THIS SPACE

Time & Location

DATE OF CRASH	TIME OF CRASH	TIME OFFICER NOTIFIED	TIME OFFICER ARRIVED	INVEST. AGENCY REPORT NUMBER	HSMV CRASH REPORT NUMBER
04/16/09	11:36 AM	11:36 AM	11:44 AM	2009-178134	91551848

COUNTY / CITY CODE	FEET or MILE(S)	N S E W of	CITY OR TOWN	(Check if in City or Town)	COUNTY
07-46			Orlando	X	Orange

AT NODE NO.	or FEET or MILE(S)	FROM NODE NO.	NEXT NODE NO.	NO. OF LANES	1. DIVIDED 2. UNDIVIDED	ON STREET, ROAD OR HIGHWAY
				6	1	SR 436 (400 Blk N Semoran Blvd)

AT THE INTERSECTION OF (street road or highway)	or FEET	MILE(S)	N S E W	FROM INTERSECTION OF (street road or highway)
	20		X	Dahlia Dr

Section 1 — Vehicle / Pedestrian

DRIVER 1. Phantom 2. Hit & Run 3. N/A	Veh Year	MAKE	TYPE	USE	VEH. LICENSE NUMBER	STATE	VEHICLE IDENTIFICATION NUMBER
3	00	INTL	04	03		FL	

TRAILER OR TOWED VEHICLE INFORMATION — TRAILER TYPE

18. Undercarriage 19. Overturn 20. Windshield 21. Trailer — SHOW FIRST POINT OF VEHICLE DAMAGE AND CIRCLE DAMAGED AREA(S): 13

VEHICLE TRAVELING N S E W	ON	AT	Est MPH	Posted Speed	EST. VEHICLE DAMAGE	1. Disabling 2. Functional 3. No Damage	EST. TRAILER DAMAGE
1	SR 436 (400 Blk N Semoran Blvd)		40	45	$200	2	

MOTOR VEHICLE INSURANCE COMPANY (LIABILITY OR PIP)	VEHICLE REMOVED BY:	1. Tow Rotation List 2. Tow Owner's Request 3. Driver 4. Other
		3

NAME OF VEHICLE OWNER (Check Box If Same As Driver) | CURRENT ADDRESS (Number and Street) | CITY AND STATE | ZIP CODE

NAME OF OWNER (Trailer or Towed Vehicle) | CURRENT ADDRESS (Number and Street) | CITY AND STATE | ZIP CODE

NAME OF MOTOR CARRIER (Commercial Vehicle Only) | CURRENT ADDRESS (Number and Street) | CITY, STATE AND ZIP CODE | US DOT or ICC MC IDENTIFICATION NUMBERS

NAME OF DRIVER (Take From Driver License) / PEDESTRIAN | CURRENT ADDRESS (Number and Street) | CITY, STATE & ZIP CODE | DATE OF BIRTH

DRIVER LICENSE NUMBER	STATE	DL TYPE	REQ END	ALC/DRUG TEST TYPE 1 Blood 3 Urine 5 None 2 Breath 4 Refused	RESULTS	ALC/DRUG	PHYS DEF	RES.	RACE	SEX	INJ.	S. EQUIP.	EJECT.
	FL	2	1	5		1	1	1	2	1	1	2 5	1

HAZARDOUS MATERIALS BEING TRANSPORTED	PLACARDED	IF YES INDICATE NAME OR 4 DIGIT NUMBER FROM DIAMOND OR BOX ON PLACARD, AND 1 DIGIT NUMBER FROM BOTTOM OF DIAMOND.	WAS HAZARDOUS MATERIAL SPILLED?	RECOMMEND DRIVER RE-EXAM, IF YES EXPLAIN IN NARRATIVE	DRIVER'S PHONE NO.
1 Yes 2 No: 2	1 Yes 2 No: 2		1 Yes 2 No: 2	1 Yes 2 No: 2	

Section 2 — Vehicle / Pedestrian

DRIVER 1. Phantom 2. Hit & Run 3. N/A	YEAR	MAKE	TYPE	USE	VEH. LICENSE NUMBER	STATE	VEHICLE IDENTIFICATION NUMBER
3	05	ICCO	09	05		FL	

TRAILER OR TOWED VEHICLE INFORMATION — TRAILER TYPE

18. Undercarriage 19. Overturn 20. Windshield 21. Trailer — SHOW FIRST POINT OF VEHICLE DAMAGE AND CIRCLE DAMAGED AREA(S): 3

VEHICLE TRAVELING N S E W	ON	AT	Est MPH	Posted Speed	EST. VEHICLE DAMAGE	1. Disabling 2. Functional 3. No Damage	EST. TRAILER DAMAGE
2	SR 436 (400 Blk N Semoran Blvd)		40	45	$100	2	

MOTOR VEHICLE INSURANCE COMPANY (LIABILITY OR PIP)	POLICY NUMBER	VEHICLE REMOVED BY:	1. Tow Rotation List 3. Driver 2. Tow Owners Request 4. Other
			3

NAME OF VEHICLE OWNER (Check Box If Same As Driver) | CURRENT ADDRESS (Number and Street) | CITY AND STATE | ZIP CODE

NAME OF OWNER (Trailer or Towed Vehicle) | CURRENT ADDRESS (Number and Street) | CITY AND STATE | ZIP CODE

NAME OF MOTOR CARRIER (Commercial Vehicle Only) | CURRENT ADDRESS (Number and Street) | CITY, STATE & ZIP CODE | US DOT or ICC MC IDENTIFICATION NUMBERS

NAME OF DRIVER (Take From Driver License) / PEDESTRIAN | CURRENT ADDRESS (Number and Street) | CITY, STATE & ZIP CODE | DATE OF BIRTH

DRIVER LICENSE NUMBER	STATE	DL TYPE	REQ END	ALC/DRUG TEST TYPE 1 Blood 3 Urine 5 None 2 Breath 4 Refused	RESULTS	ALC/DRUG	PHYS DEF	RES.	RACE	SEX	INJ.	S. EQUIP.	EJECT.
	FL	2	1	5		1	1	1	3	2	1	2 5	1

WAS HAZARDOUS MATERIAL BEING TRANSPORTED	PLACARDED	IF YES INDICATE NAME OR FOUR DIGIT NUMBER FROM DIAMOND OR BOX ON PLACARD AND 1 DIGIT NUMBER FROM BOTTOM OF DIAMOND.	WAS HAZARDOUS MATERIAL SPILLED?	RECOMMEND DRIVER RE-EXAM, IF YES EXPLAIN IN NARRATIVE	DRIVER'S PHONE NO.
1 Yes 2 No: 2	1 Yes 2 No: 2		1 Yes 2 No: 2	1 Yes 2 No: 2	

Code Information

VEHICLE TYPE	VEHICLE USE	TRAILER TYPE	RESIDENCE (DRIVER / PED)	PHYSICAL DEFECTS	ALCOHOL / DRUG USE	LOCATION IN VEHICLE
01 Automobile	01 Private Transportation	01 Single Semi Trailer	1 County of Crash	1 No Defects Known	1 Not Drinking or Using Drugs	1 Front Left
02 Van	02 Commercial Passengers	02 Tandem Semi Trailer	2 Elsewhere in State	2 Eyesight Defect	2 Alcohol- Under Influence	2 Front Center
03 Light Truck / P.U. - 2 or 4 rear tires	03 Commercial Cargo	03 Tank Trailer	3 Non-Resident Out of State	3 Fatigue / Asleep	3 Drugs - Under Influence	3 Front Right
04 Medium Truck - 4 rear tires	04 Public Transportation	04 Saddle Mount / Flatbed	4 Foreign 5 Unknown	4 Hearing Defect	4 Alcohol & Drugs - Under Influence	4 Rear Left
05 Heavy Truck - 2 or more rear axles	05 Public School Bus	05 Boat Trailer	DL TYPE: 1 A 2 B 3 C 4 D / Chauffeur 5 E / Operator 6 E / Oper.-Rest. 7 None	RACE: 1 White 2 Black 3 Hispanic 4 Other	5 Had Been Drinking	5 Rear Center
06 Truck Tractor (Cab-Bobtail)	06 Private School Bus	06 Utility Trailer			6 Pending ALC/DRUG Test Results	6 Rear Right
07 Motor Home (RV)	07 Ambulance	07 House Trailer		INJURY SEVERITY	SAFETY EQUIPMENT IN USE	7 In Body Of Truck
08 Bus (driver - seats for 9- 15)	08 Law Enforcement	08 Pole Trailer	REQUIRED ENDORSEMENTS: 1 Yes 2 No 3 No Endorsement Required	SEX: 1 Male 2 Female	1 None 2 Possible 3 Non-Incapacitating 4 Incapacitating 5 Fatal (Within 30 Days) 6 Non-Traffic Fatality	1 Not In use 2 Seat Belt / Shoulder Harness 3 Child Restraint 4 Air Bag -Deployed 5 Air Bag -Not Deployed 6 Safety Helmet 7 Eye Protection	8 Bus Passenger 9 Other
09 Bus (driver + seats for over 15)	09 Fire Apparatus	09 Mobile Home Trailer					EJECTED: 1 No 2 Yes 3 Partial
10 Bicycle	10 Military	10 Motor Vehicle Transport					
11 Motorcycle	11 Other Government	77 Other					
12 Moped	12 Dump						
13 All Terrain Vehicle	13 Concrete Mixer						
14 Train	14 Garbage or Refuse						
15 Low Speed Vehicle	15 Cargo Van						
77 Other	77 Other						

DW #7

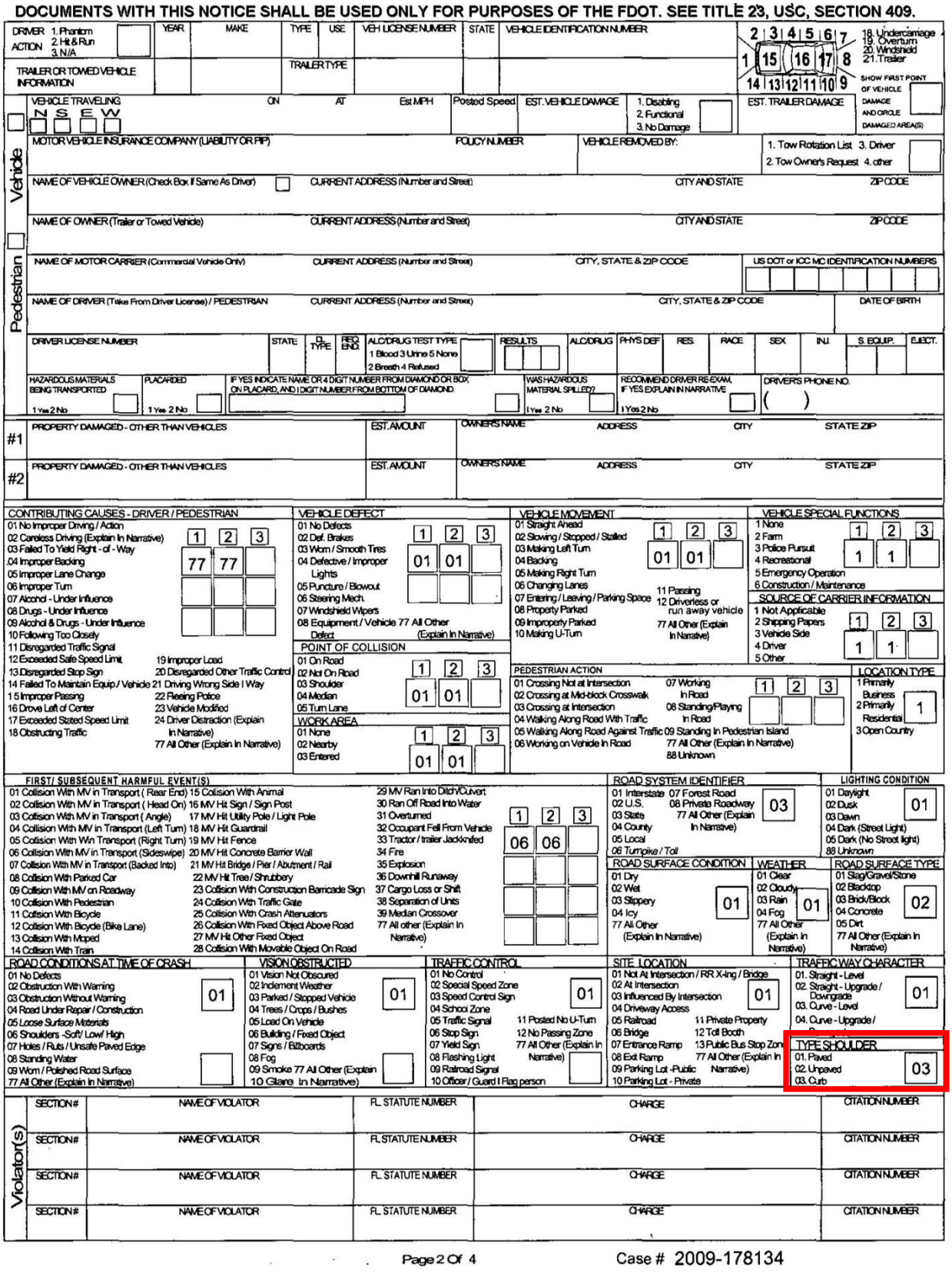
APR 21 2009

PACE

HSMV-90003 (REV. 01/02)

Page 1 Of 4

Figure 2-1: Florida Crash Report – Page 1


DOCUMENTS WITH THIS NOTICE SHALL BE USED ONLY FOR PURPOSES OF THE FDOT. SEE TITLE 23, USC, SECTION 409.

Section 3 Pedestrian / Vehicle

DRIVER ACTION 1. Phantom 2. Hit & Run 3. N/A | YEAR | MAKE | TYPE | USE | VEH LICENSE NUMBER | STATE | VEHICLE IDENTIFICATION NUMBER

TRAILER OR TOWED VEHICLE INFORMATION | TRAILER TYPE

18. Undercarriage 19. Overturn 20. Windshield 21. Trailer

SHOW FIRST POINT OF VEHICLE DAMAGE AND CIRCLE DAMAGED AREA(S)

VEHICLE TRAVELING N S E W | ON | AT | Est MPH | Posted Speed | EST. VEHICLE DAMAGE | 1. Disabling 2. Functional 3. No Damage | EST. TRAILER DAMAGE

MOTOR VEHICLE INSURANCE COMPANY (LIABILITY OR PIP) | POLICY NUMBER | VEHICLE REMOVED BY: | 1. Tow Rotation List 3. Driver 2. Tow Owner's Request 4. other

NAME OF VEHICLE OWNER (Check Box If Same As Driver) | CURRENT ADDRESS (Number and Street) | CITY AND STATE | ZIP CODE

NAME OF OWNER (Trailer or Towed Vehicle) | CURRENT ADDRESS (Number and Street) | CITY AND STATE | ZIP CODE

NAME OF MOTOR CARRIER (Commercial Vehicle Only) | CURRENT ADDRESS (Number and Street) | CITY, STATE & ZIP CODE | US DOT or ICC MC IDENTIFICATION NUMBERS

NAME OF DRIVER (Take From Driver License) / PEDESTRIAN | CURRENT ADDRESS (Number and Street) | CITY, STATE & ZIP CODE | DATE OF BIRTH

DRIVER LICENSE NUMBER | STATE | DL TYPE | REQ END | ALC/DRUG TEST TYPE 1 Blood 3 Urine 5 None 2 Breath 4 Refused | RESULTS | ALC/DRUG | PHYS DEF | RES. | RACE | SEX | INJ. | S. EQUIP. | EJECT.

HAZARDOUS MATERIALS BEING TRANSPORTED 1 Yes 2 No | PLACARDED 1 Yes 2 No | IF YES INDICATE NAME OR 4 DIGIT NUMBER FROM DIAMOND OR BOX ON PLACARD, AND 1 DIGIT NUMBER FROM BOTTOM OF DIAMOND | WAS HAZARDOUS MATERIAL SPILLED? 1 Yes 2 No | RECOMMEND DRIVER RE-EXAM, IF YES EXPLAIN IN NARRATIVE 1 Yes 2 No | DRIVER'S PHONE NO.

#1 PROPERTY DAMAGED - OTHER THAN VEHICLES | EST. AMOUNT | OWNER'S NAME | ADDRESS | CITY | STATE ZIP

#2 PROPERTY DAMAGED - OTHER THAN VEHICLES | EST. AMOUNT | OWNER'S NAME | ADDRESS | CITY | STATE ZIP

CONTRIBUTING CAUSES - DRIVER / PEDESTRIAN — 1: 77; 2: 77; 3:
01 No Improper Driving / Action
02 Careless Driving (Explain In Narrative)
03 Failed To Yield Right - of - Way
04 Improper Backing
05 Improper Lane Change
06 Improper Turn
07 Alcohol - Under Influence
08 Drugs - Under Influence
09 Alcohol & Drugs - Under Influence
10 Following Too Closely
11 Disregarded Traffic Signal
12 Exceeded Safe Speed Limit
13 Disregarded Stop Sign
14 Failed To Maintain Equip / Vehicle
15 Improper Passing
16 Drove Left of Center
17 Exceeded Stated Speed Limit
18 Obstructing Traffic
19 Improper Load
20 Disregarded Other Traffic Control
21 Driving Wrong Side I Way
22 Fleeing Police
23 Vehicle Modified
24 Driver Distraction (Explain In Narrative)
77 All Other (Explain In Narrative)

VEHICLE DEFECT — 1: 01; 2: 01; 3:
01 No Defects
02 Def. Brakes
03 Worn / Smooth Tires
04 Defective / Improper Lights
05 Puncture / Blowout
06 Steering Mech.
07 Windshield Wipers
08 Equipment / Vehicle Defect
77 All Other (Explain In Narrative)

POINT OF COLLISION — 1: 01; 2: 01; 3:
01 On Road
02 Not On Road
03 Shoulder
04 Median
05 Turn Lane

WORK AREA — 1: 01; 2: 01; 3:
01 None
02 Nearby
03 Entered

VEHICLE MOVEMENT — 1: 01; 2: 01; 3:
01 Straight Ahead
02 Slowing / Stopped / Stalled
03 Making Left Turn
04 Backing
05 Making Right Turn
06 Changing Lanes
07 Entering / Leaving / Parking Space
08 Property Parked
09 Improperly Parked
10 Making U-Turn
11 Passing
12 Driverless or run away vehicle
77 All Other (Explain In Narrative)

VEHICLE SPECIAL FUNCTIONS — 1: 1; 2: 1; 3:
1 None
2 Farm
3 Police Pursuit
4 Recreational
5 Emergency Operation
6 Construction / Maintenance

SOURCE OF CARRIER INFORMATION — 1: 1; 2: 1; 3:
1 Not Applicable
2 Shipping Papers
3 Vehicle Side
4 Driver
5 Other

PEDESTRIAN ACTION — 1: ; 2: ; 3:
01 Crossing Not at Intersection
02 Crossing at Mid-block Crosswalk
03 Crossing at Intersection
04 Walking Along Road With Traffic
05 Walking Along Road Against Traffic
06 Working on Vehicle In Road
07 Working In Road
08 Standing/Playing In Road
09 Standing In Pedestrian Island
77 All Other (Explain In Narrative)
88 Unknown

LOCATION TYPE — 1
1 Primarily Business
2 Primarily Residential
3 Open Country

FIRST/ SUBSEQUENT HARMFUL EVENT(S) — 1: 06; 2: 06; 3:
01 Collision With MV in Transport (Rear End)
02 Collision With MV in Transport (Head On)
03 Collision With MV in Transport (Angle)
04 Collision With MV in Transport (Left Turn)
05 Collision With MV in Transport (Right Turn)
06 Collision With MV in Transport (Sideswipe)
07 Collision With MV in Transport (Backed Into)
08 Collision With Parked Car
09 Collision With MV on Roadway
10 Collision With Pedestrian
11 Collision With Bicycle
12 Collision With Bicycle (Bike Lane)
13 Collision With Moped
14 Collision With Train
15 Collision With Animal
16 MV Hit Sign / Sign Post
17 MV Hit Utility Pole / Light Pole
18 MV Hit Guardrail
19 MV Hit Fence
20 MV Hit Concrete Barrier Wall
21 MV Hit Bridge / Pier / Abutment / Rail
22 MV Hit Tree / Shrubbery
23 Collision With Construction Barricade Sign
24 Collision With Traffic Gate
25 Collision With Crash Attenuators
26 Collision With Fixed Object Above Road
27 MV Hit Other Fixed Object
28 Collision With Movable Object On Road
29 MV Ran Into Ditch/Culvert
30 Ran Off Road Into Water
31 Overturned
32 Occupant Fell From Vehicle
33 Tractor / trailer Jackknifed
34 Fire
35 Explosion
36 Downhill Runaway
37 Cargo Loss or Shift
38 Separation of Units
39 Median Crossover
77 All other (Explain In Narrative)

ROAD SYSTEM IDENTIFIER: 03
01 Interstate
02 U.S.
03 State
04 County
05 Local
06 Turnpike / Toll
07 Forest Road
08 Private Roadway
77 All Other (Explain In Narrative)

LIGHTING CONDITION: 01
01 Daylight
02 Dusk
03 Dawn
04 Dark (Street Light)
05 Dark (No Street light)
88 Unknown

ROAD SURFACE CONDITION: 01
01 Dry
02 Wet
03 Slippery
04 Icy
77 All Other (Explain In Narrative)

WEATHER: 01
01 Clear
02 Cloudy
03 Rain
04 Fog
77 All Other (Explain In Narrative)

ROAD SURFACE TYPE: 02
01 Slag/Gravel/Stone
02 Blacktop
03 Brick/Block
04 Concrete
05 Dirt
77 All Other (Explain In Narrative)

ROAD CONDITIONS AT TIME OF CRASH: 01
01 No Defects
02 Obstruction With Warning
03 Obstruction Without Warning
04 Road Under Repair / Construction
05 Loose Surface Materials
06 Shoulders -Soft/ Low/ High
07 Holes / Ruts / Unsafe Paved Edge
08 Standing Water
09 Worn / Polished Road Surface
77 All Other (Explain In Narrative)

VISION OBSTRUCTED: 01
01 Vision Not Obscured
02 Inclement Weather
03 Parked / Stopped Vehicle
04 Trees / Crops / Bushes
05 Load On Vehicle
06 Building / Fixed Object
07 Signs / Billboards
08 Fog
09 Smoke
10 Glare
77 All Other (Explain In Narrative)

TRAFFIC CONTROL: 01
01 No Control
02 Special Speed Zone
03 Speed Control Sign
04 School Zone
05 Traffic Signal
06 Stop Sign
07 Yield Sign
08 Flashing Light
09 Railroad Signal
10 Officer / Guard I Flag person
11 Posted No U-Turn
12 No Passing Zone
77 All Other (Explain In Narrative)

SITE LOCATION: 01
01 Not At Intersection / RR X-ing / Bridge
02 At Intersection
03 Influenced By Intersection
04 Driveway Access
05 Railroad
06 Bridge
07 Entrance Ramp
08 Exit Ramp
09 Parking Lot -Public
10 Parking Lot - Private
11 Private Property
12 Toll Booth
13 Public Bus Stop Zone
77 All Other (Explain In Narrative)

TRAFFIC WAY CHARACTER: 01
01. Straight - Level
02. Straight - Upgrade / Downgrade
03. Curve - Level
04. Curve - Upgrade /

TYPE SHOULDER: 03
01. Paved
02. Unpaved
03. Curb

Violator(s)

SECTION#	NAME OF VIOLATOR	FL. STATUTE NUMBER	CHARGE	CITATION NUMBER
SECTION#	NAME OF VIOLATOR	FL. STATUTE NUMBER	CHARGE	CITATION NUMBER
SECTION#	NAME OF VIOLATOR	FL. STATUTE NUMBER	CHARGE	CITATION NUMBER
SECTION#	NAME OF VIOLATOR	FL. STATUTE NUMBER	CHARGE	CITATION NUMBER

Page 2 Of 4 Case # 2009-178134

Figure 2-2: Florida Crash Report – Page 2

2.2 RCI Database

The RCI database includes various physical and administrative data related to the roadway networks that are either maintained by or are of special interest to FDOT. For this project, the following FDOT Road Data shapefiles were downloaded and used to extract the required data:

- Intersections
- Number of Lanes
- Shoulder Type – Outside
- Maximum Speed Limits

Note that the RCI database is the original data source for these shapefiles. Table 2-1 lists the five data variables in the RCI database (also available in the shapefiles) that are of interest to this project. These variables were used to verify the accuracy of the data retrieved from crash records.

Table 2-1: Variables of Interest in RCI Database

Variable	Definition	Feature Type	Year
NOLANES	Number of Through Lanes	polyline	2015
SPEED	Speed Limit	polyline	2015
SHLDTYPE	Shoulder Type	polyline	2015
TYPEROAD	Median Type	dBase	2014
INTERSECT	Intersection	Point	2014

The variable SHLDTYPE, with the following ten codes, is used to determine shoulder type. Figure 2-3 shows the shoulder type codes in the RCI database.[3]

- 0 – Raised Curb (no shoulder or width exists)
- 1 – Paved (including paved parking and bike slots)
- 2 – Paved with Warning Device (any device that serves to warn, guide, or regulate the motorist)
- 3 – Lawn (number of feet to support roadbed)
- 4 – Gravel/Marl
- 5 – Valley Gutter (not a barrier)
- 6 – Curb & Gutter
- 7 – Other (to warn, guide, or regulate the motorist)
- 8 – Curb with Resurfaced Gutter
- 9 – None (Managed Lane)

[3] http://www.dot.state.fl.us/planning/statistics/rci/rcihandbbok.pdf

0 – Raised Curb | 1 – Paved | 1 – Paved

2 – Paved with Warning Device | 2 – Paved with Warning Device | 2 – Paved with Warning Device

2 – Paved with Warning Device | 2 – Paved with Warning Device | 2 – Paved with Warning Device

3 – Lawn | 4 – Gravel/Marl | 5 – Valley Gutter

6 – Curb & Gutter | 8 – Curb with Resurfaced Gutter | 9 – None

Figure 2-3: Shoulder Type Codes in RCI Database

The variable TYPEROAD, with the following four codes, was used to determine the median type. Figure 2-4 shows the median type codes in the RCI database.[4]

- 0 – Not divided
- 2 – Divided (painted or physical)
- 4 – One-way
- 6 – Reversible

0 - Not divided

2 - Divided (painted or physical)

4 - One-way

6 – Reversible (Selmon Expy not elevated)

6 – Reversible (Selmon Expy elevated)

Figure 2-4: Median Type Codes in RCI Database

2.3 All-Roads Map

The shapefiles for the All-Roads map for the year 2012 is based on the data provided by NAVTEQ™. NAVTEQ™ provides a dataset called NAVSTREETS which contains the most navigable attributes available in a database. This dataset includes access features including expressway ramps; complete and correct connectivity of all roadways; one-way streets; physical, logical, and legal turn restrictions; construction projects; and physical and painted lane dividers. Figure 2-5 shows the basic data table of the NAVSTREETS map.

[4] http://www.dot.state.fl.us/planning/statistics/rci/rcihandbbok.pdf

Table

Delivery_Jan_2015

FID	Shape *	LINK_ID	ST_NAME	FEAT_ID	ST_LANGCD	NUM_STNMES	ST_NM_PREF	ST_TYP_BEF	ST_NM_BASE	ST_NM_SUFF
0	Polyline	16973829	LOXAHATCHEE RIVER RD	717723303	ENG	1			LOXAHATCHEE RIVER	
1	Polyline	16973830	FOX RUN DR	732938875	ENG	1			FOX RUN	
2	Polyline	16973831	LOXAHATCHEE RIVER RD	717723303	ENG	1			LOXAHATCHEE RIVER	
3	Polyline	16973832	LOXAHATCHEE RIVER RD	717723303	ENG	1			LOXAHATCHEE RIVER	
4	Polyline	16973833	LOXAHATCHEE RIVER RD	717723303	ENG	1			LOXAHATCHEE RIVER	
5	Polyline	16973838	OLD INDIANTOWN RD	733353021	ENG	1			OLD INDIANTOWN	
6	Polyline	16973839	189TH CT N	719726339	ENG	1			189TH	N
7	Polyline	16973840	120TH TER N	717782493	ENG	1			120TH	N
8	Polyline	16973841	188TH ST N	721103854	ENG	1			188TH	N
9	Polyline	16973842	125TH AVE N	717367376	ENG	2			125TH	N
10	Polyline	16973843	134TH WAY N	717782600	ENG	1			134TH	N
11	Polyline	16973844	187TH PL N	719726347	ENG	1			187TH	N
12	Polyline	16973845	COCHISE TRL	735343864	ENG	2			COCHISE	
13	Polyline	16973846	127TH DR N	719726346	ENG	1			127TH	N
14	Polyline	16973847	187TH PL N	719726347	ENG	1			187TH	N
15	Polyline	16973848	125TH AVE N	717367376	ENG	2			125TH	N
16	Polyline	16973849	OLD INDIANTOWN RD	733353021	ENG	1			OLD INDIANTOWN	
17	Polyline	16973850	120TH TER N	717782493	ENG	1			120TH	N
18	Polyline	16973851	OLD INDIANTOWN RD	733353021	ENG	1			OLD INDIANTOWN	
19	Polyline	16973852	OLD INDIANTOWN RD	733353021	ENG	1			OLD INDIANTOWN	
20	Polyline	16973853	120TH TER N	717782493	ENG	1			120TH	N
21	Polyline	16973854	127TH DR N	719726346	ENG	1			127TH	N
22	Polyline	16973855	186TH CT N	719726349	ENG	2			186TH	N
23	Polyline	16973856	125TH AVE N	717367376	ENG	2			125TH	N
24	Polyline	16973857	134TH WAY N	717782600	ENG	1			134TH	N
25	Polyline	16973858	185TH PL N	719726590	ENG	1			185TH	N
26	Polyline	16973859	127TH DR N	719726346	ENG	1			127TH	N

Figure 2-5: NAVSTREETS Data Table Sample

For this project, the following seven data variables were extracted from the NAVSTREETS dataset:

1. LANE_CAT (lane category)
2. FROM_LANES (number of lanes from the reference node)
3. TO_LANES (number of lanes toward the reference node)
4. SPEED_CAT (speed category)
5. FR_SPD_LIM (from reference speed limit)
6. TO_SPD_LIM (toward reference speed limit)
7. DIVIDER (presence of physical blocking divider)

Lane Category classifies a road based on the number of through lanes in each direction. It has the following four codes:

- (space) - Not Applicable
- 1 - One Lane
- 2 - Two or Three Lanes
- 3 - Four or More Lanes

Speed Category classifies the general speed trend of a road based on posted or legal speed. Speed Category values represent the combination of several factors besides legal speed limit (e.g., physical restrictions, access characteristics, etc.). Therefore, Speed Category values can differ from posted speed limit values, which represent the legal speed limit. It has the following eight codes:

- 1 - > 80 mph
- 2 - 65-80 mph
- 3 - 55-64 mph
- 4 - 41-54 mph
- 5 - 31-40 mph
- 6 - 21-30 mph
- 7 - 6-20 mph
- 8 - < 6 mph

The variable DIVIDER identifies the presence of a physical traffic blocking divider. It has the following five codes:

- A is applied when the link and both nodes are divided. Additionally, right turns and U-turns are not allowed to/from the divided link to/from any link at either node or to driveways along the link.
- 1 is applied when the link and reference node are divided. Right turns and U-turns are not allowed from the divided link to/from any link at the reference node or to driveways along the link.
- 2 is applied when the link and non-reference node are divided. Right turns and U-turns are not allowed from the divided link to/from any link at the non-reference node or to driveways along the link.
- N is applied when the link is not divided and the link is navigable.
- (space) is Not Applicable, and is applied to non-navigable links.

The All-Roads map for the year 2012 was extended by FDOT by adding the following RCI variables: roadway ID, milepost, AADT, roadside information, and number of through lanes.

CHAPTER 3
DATA EXTRACTION FROM CRASH RECORDS

This chapter describes the first of the three-step process which is to extract data from crash records for as many roadway segments as possible. The method first applied GIS techniques to extract data for roadway segments that experienced at least one crash during the analysis period. Where multiple crashes occurred on a specific roadway segment, a rule based on the crash year and crash frequency was applied to select the most probable value. The accuracy of the variables in the final merged dataset was then verified using the RCI data available for the state roads.

3.1 Data Processing

The crash data includes three different datasets: two shapefiles for on-system and off-system roads, and a .txt dataset for crash-level information (i.e., RDWTBL 50). The following steps were used to extract the required roadway characteristics data from the crash records:

1. Open ArcGIS 10.2.2 and add 2003-2012 on-system crash shapefiles, 2003-2012 off-system crash shapefiles, FDOT intersection layer, and FDOT All-Roads map.

2. Combine 2003-2012 on-system crash shapefiles and 2003-2012 off-system crash shapefiles into one crash shapefile.

3. Create a 50-ft buffer around intersections. Figure 3-1 shows the screen capture of the buffer tool in ArcGIS 10.2.2. Figure 3-2 shows the screenshot of the All-Roads map with buffers created around the intersections.

4. Exclude the crashes within the buffer. Since the crashes at intersections do not give information on the data variables of interest, these crashes are excluded from the analysis. Note that although some of the crashes that occurred at intersections have information on these data variables, it is difficult to determine the roadway where this crash had occurred.

5. Create a 10-ft buffer around undivided streets and a 40-ft buffer around divided streets. Figure 3-3 shows the screenshot of the All-Roads map with 10-ft buffers around undivided sections and 40-ft buffers around divided streets.

6. Once the buffers are created, intersect this layer with the crash layer. Figure 3-4 gives a screenshot of the All-Roads layer overlaid with the buffer and crash layers.

7. Identify and exclude the crash records with inconsistent side of road by comparing ROADSIDE variable in the All-Roads map and SIDEOFROAD variable in the crash layer.

8. Identify and exclude the crash records that are in close proximity to more than one roadway section.

9. Export the attributes table of the crash shapefiles to Microsoft Excel.

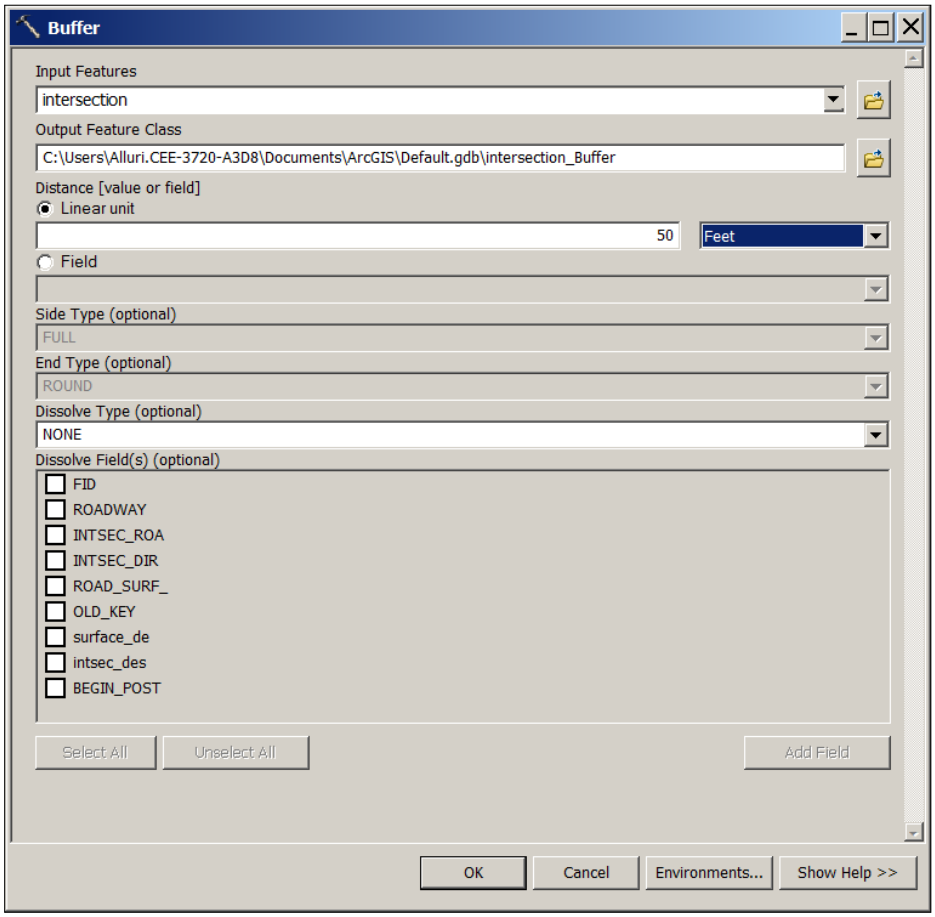


Figure 3-1: Screen Capture of the Buffer Tool in ArcGIS 10.2.2

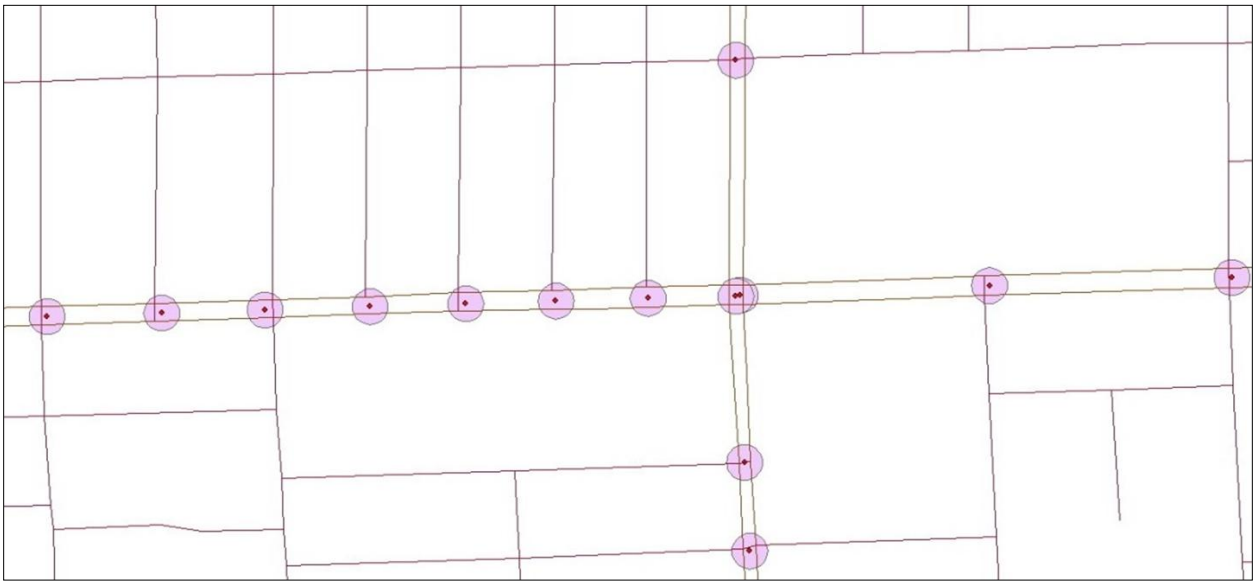


Figure 3-2: Screenshot of All-Roads Map with 50-ft Buffer Around Intersections

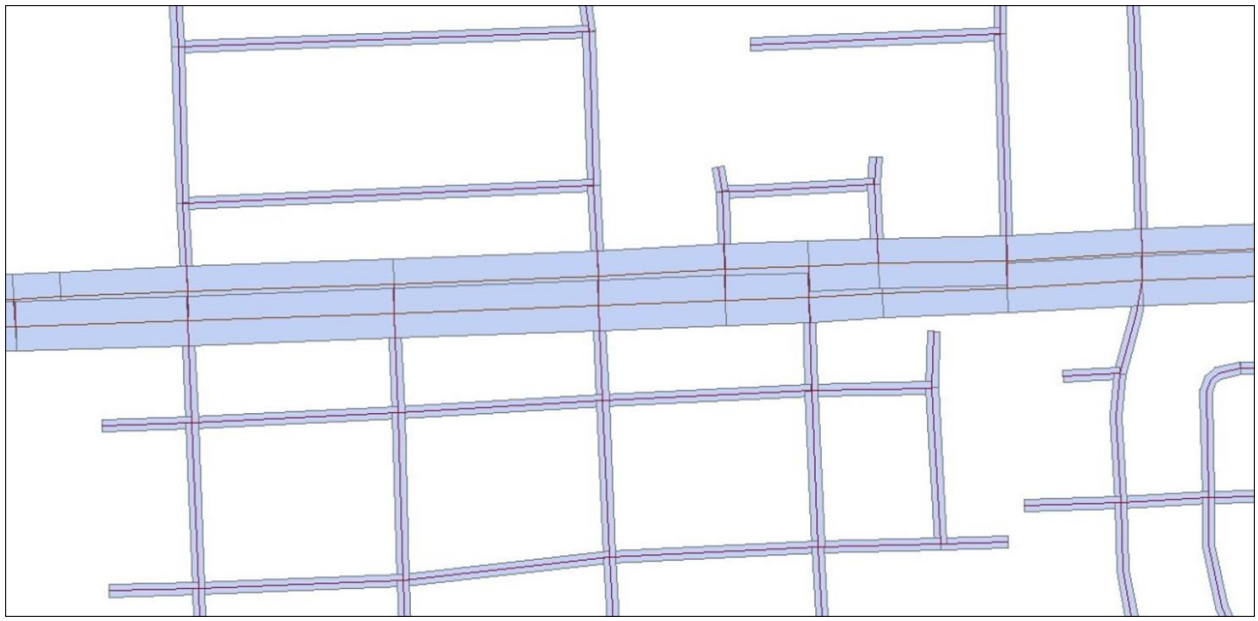


Figure 3-3: Screenshot of All-Roads Map with 10-ft Buffer Around Undivided Roads and 40-ft Buffer Around Divided Roads

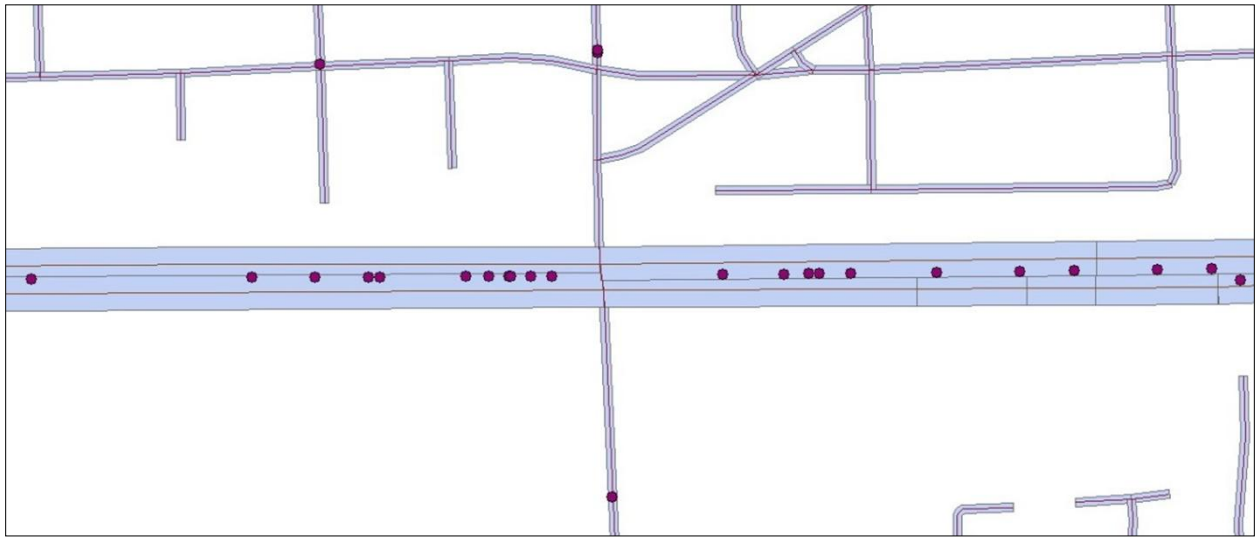


Figure 3-4: Screenshot of All-Roads Map with Buffer and Crash Layers

Overall, the All-Roads map has data for 1,565,026 road links. Of these, 232,013 links have at least one crash during 2003-2012. These represent about 15% of the total road network. In total, during 2003-2012, 679,973 crashes occurred on 232,013 links. The number of crashes that have data for each variable is given below:

- 668,777 crashes (i.e., 98.3% of total crashes) have data on median type (TYPTRWAY).
- 668,811 crashes (i.e., 98.3% of total crashes) have data on shoulder type (TYPESHLD).

- 595,643 crashes (i.e., 87.6% of total crashes) have data on posted speed limit (SPDLIMIT).
- None of the crashes have data on number of through lanes (TRWAYLN).
-

Although data for number of through lanes are recorded in police reports, they are not recorded in the corresponding field TRWAYLN (number of through lanes) in RDWTBL 50. Since these data are missing in the crash database, it could not be obtained using the data extraction or data derivation steps. However, since the All-Roads map includes number of through lanes under its LANES variable, this project does not include Step 1 and 2 of the process. Instead, it focuses on Step 3 of the process which is to manually verify and update the existing data. Table 3-1 provides the sample crash data with the extracted roadway characteristics information. The table includes the following columns:

- LINK_ID: Link identification number in the All-Roads map
- CALYEAR: Year when the crash occurred
- CRASHNUM: Crash number
- TYPTRWAY: Type of travel way (1 is divided and 2 is undivided)
- TRWAYLN: Number of through roadway lanes
- TYPESHLD: Type of shoulder (1 is paved, 2 is unpaved, and 3 is curb)
- SPDLIMIT: Posted speed limit

Table 3-1: Sample Crash Data with Extracted Roadway Characteristics Information

LINK_ID	CALYEAR	CRASHNUM	TYPTRWAY	TRWAYLN	TYPESHLD	SPDLIMIT
100000030	2012	22561662	2	-	1	30
100001460	2006	23048332	2	-	2	35
100001470	2006	23057330	2	-	2	45
100001500	2006	23047308	1	-	3	45
100004190	2007	135857653	1	-	3	45
100006610	2007	22905136	1	-	3	45
100008670	2008	134888595	2	-	2	25
100011870	2008	761578183	2	-	2	45
100022000	2010	888102752	2	-	3	45
100027180	2010	23238906	2	-	2	45
100028040	2008	782992718	2	-	1	45
100028080	2008	23179614	2	-	2	25
100038570	2011	755413950	1	-	3	45
100040390	2008	771546783	1	-	3	45
100040980	2008	838794886	1	-	1	45
100041010	2008	23163354	1	-	2	
100074640	2012	22405311	2	-	2	15
100074680	2012	762079448	2	-	2	
100074750	2012	22405540	2	-	2	15
100074800	2012	875175606	1	-	2	30

3.2 Data Extraction

As discussed in Section 3.1, the crash data from 2003 to 2012 were first merged with the roadway characteristics data to extract information on the following three data variables: posted speed limit, shoulder type, and median type. The next step was to extract this information for all the links. Some of the links experienced multiple crashes, and in several instances, the data in these crash records was inconsistent. Table 3-2 gives an example for the speed limit variable. As can be seen from the table, link 16982194 experienced five crashes between 2003 and 2012. The two crashes that occurred in 2003 and 2004 recorded a speed limit of 40 mph, the crash report from 2006 recorded a speed limit of 45 mph, while the crash reports from 2009 and 2010 each recorded a speed limit of 50 mph. This is one of the several scenarios where inconsistent data were found.

Table 3-2: Sample Road Data with Multiple Crashes and Different Speed Limits

LINK_ID	CALYEAR	CRASHNUM	SPDLIMIT
16982194	2003	732196010	40
16982194	2004	701661050	40
16982194	2006	768858890	45
16982194	2009	804573380	50
16982194	2010	819627230	50

An algorithm was developed to select the most accurate value of each variable for each link for which multiple values were extracted from the crash records. The algorithm is based on the following rules:

- Select the most recent year's data.
- If the most recent year has multiple crash records, select the value with the highest frequency in the most recent year.
- If the multiple values have the same highest frequency in the same year (for example, year 2012 has one record with a speed limit of 30 mph and another record with a speed limit of 35 mph), ignore the most recent year and select the data from the preceding year.

The detailed steps of the algorithm are described below:

1. Open crash data with the extracted roadway characteristics data table, sort the table by road link and CALYEAR separately for the following three data variables: median type, shoulder type, and posted speed limit.
2. For each link, retrieve only the crash data from the most recent year.
3. For each link, calculate the values of the fields and their frequencies for median type, shoulder type, and posted speed limit.

4. For each of the three data variables, select and retain the value of the fields with the highest frequency in the most recent year.

5. If multiple values have the same highest frequency in the most recent year (for example, year 2012 has one record with a speed limit of 30 mph and another record with a speed limit of 35 mph), ignore this year and select the information from the preceding year.

6. Repeat Steps 3-5 for each link.

7. Generate the final table with the link, median type, shoulder type, and posted speed limit values extracted from crash data.

Overall, of the 1,565,026 links in the All-Roads map, 232,013 links (i.e., 14.8%) experienced at least one crash during 2003-2012. The number of links that have data for each variable is given below:

- 226,300 links (i.e., 14.5% of total links) have data for median type.
- 229,778 links (i.e., 14.7% of total links) have data for shoulder type.
- 210,203 links (i.e., 13.4% of total links) have data for posted speed limit.

3.3 Data Verification

During the data extraction process, the data from the crash records were extracted for the links in the All-Roads map for which the crash data are available. In the data verification process, the data from the RCI database were compared with this extracted data. Note that the RCI database includes data on all state roads and a few off-system roads, while the All-Roads map includes data on the entire road network (i.e., both state roads and non-state roads) in Florida. More specifically, the RCI database includes data on links with the following functional classification codes:

- 01 – RURAL – Principal Arterial–Interstate
- 02 – RURAL – Principal Arterial–Other
- 06 – RURAL – Minor Arterial
- 07 – RURAL – Major Collector
- 08 – RURAL – Minor Collector
- 09 – RURAL – Local
- 11 – URBAN – Principal Arterial–Interstate
- 12 – URBAN – Principal Arterial–Other Freeways and Expressways
- 14 – URBAN – Principal Arterial–Other
- 16 – URBAN – Minor Arterial
- 17 – URBAN – Collector
- 19 – URBAN – Local

On the other hand, the links in the All-Roads map have functional classification with codes from 01 to 33, where the codes from 01 to 19 are similar to the codes in the RCI database and the remaining codes (i.e., from 20 to 33) constitute the remaining local roads that are not identified

in the RCI database. For this reason, the data verification process only compares the data for the links that are available in both the All-Roads map and the RCI database.

Table 3-3 lists the four variable names in the three data sources (i.e., RDWTBL 50, All-Roads map, and RCI). Note that there are two variables for number of through lanes in the All-Roads map. The variable LANE_CAT was developed by NAVTEQ™ and the variable LANES was added by FDOT in 2012. Four variables, FROM_LANES, TO_LANES, FR_SPD_LIM and TO_SPD_LIM, discussed in Section 2.3, were not used in data verification process because these four variables have only limited number of valid links. The FROM_LANES and TO_LANES variables only have approximately 200,000 valid links out of 1,565,026 total links (i.e., 12.8% of total links). Similarly, the FR_SPD_LIM and TO_SPD_LIM variables only have approximately 20,000 valid links out of the 1,565,026 total links (i.e., 1.3% of the total links).

Table 3-3: Variable Names in Different Data Sources

Variable	RDWTBL 50	All-Roads Map	RCI
Number of Through Lanes	Not Available	LANES LANE_CAT	NOLANES
Posted Speed Limit	SPDLIMIT	SPEED_CAT	SPEED
Shoulder Type	SHLDTYPE	Not Available	SHLDTYPE
Median Type	TYPTRWAY	DIVIDER	TYPEROAD

To verify the data extracted from RDWTBL 50 and the All-Roads map, the RCI GIS layers were first merged into the layer generated in Section 3.2. The merging process is based on the following rules:

- The link in the extracted layer and the link in the RCI layer should have the same Roadway ID.
- The link in the extracted layer and the link in the RCI layer should have the same Roadside (i.e., C - Composite, R - Right, or L - Left).
- The begin milepost of the link in the extracted layer should be greater than or equal to the begin milepost of the link in the RCI layer.
- The end milepost of the link in the extracted layer should be less than or equal to the end milepost of the link in the RCI layer.

Table 3-4 compares the variable codes in the three data sources (i.e., RDWTBL 50, All-Roads Map, and RCI) for the two categorical variables, shoulder type, and median type.

For the LANE_CAT and SPEED_CAT variables, the rule is to test whether or not the corresponding data in the columns LANES and SPDLIMIT fall within the same category (i.e., identified by the LANE_CAT and SPEED_CAT variables).

Table 3-4: Comparison of Variable Codes in Different Data Sources

Variable	RDWTBL 50	All-Roads Map	RCI
Shoulder Type	1 - Paved	N/A	1 - Paved with or without Striping (including paved parking and bike slots) 2 - Paved with Warning Device (raised or indented strips) 5 - Valley Gutter (not a barrier)
	2 - Unpaved		3 - Lawn (number of feet to support road bed) 4 - Gravel/Marl 9 - None (Managed Lane)
	3 – Curb		0 - Raised Curb (no shoulder or width exists) 6 - Curb and Gutter 8 - Curb with resurfaced gutter
Median Type	1 - Divided	A - Both Nodes and Link 1 - Ref Node and Link 2 - Nref Node and Link	2 - Divided (painted or physical) 4 - One-way 6 - Reversible
	2 - Undivided	N - No Divider	0 - Not divided

Table 3-5 shows the number of links with valid records (i.e., links with acceptable codes for the data variables) in the All-Roads map and the corresponding links in the RCI. As information on number of through lanes is complete for all links in the All-Roads map, all the links in the RCI database with valid data on number of through lanes were identified for verification. These constitute approximately 300,000 links (i.e., 20% of total links). On the other hand, only around 200,000 links in the All-Roads map have data extracted for all the remaining three variables, and only around 100,000 links in the RCI were identified for verification (i.e., around 45% of total links with extracted data).

Table 3-5: Number of Links in All-Roads Map and RCI After Data Extraction Step

Variable	Links with Extracted Data (Source: All-Roads map)	Total Links Identified for Verification (Source: RCI)	Percent of Links Identified for Verification
Number of Through Lanes	1,565,026	302,245	19.3%
Posted Speed Limit	210,203	83,612	39.8%
Shoulder Type	229,778	106,471	46.3%
Median Type	226,300	114,147	50.4%

Table 3-6 shows the results from the data verification process after the data extraction step for all four variables. For number of through lanes, the variable LANES was found to have the highest accuracy of 96.3%. This is expected as the verified LANES data came originally from the RCI. The 3.7% inconsistency was likely a result of mainly changes in the RCI data between 2012 (when the All-Roads map was updated) and 2015 (RCI data year used in verification) due to roadway construction. It is noted that even though the number of through lanes and speed limits in NAVSTREETS are given in ranges, the accuracies of these variables were still found to be lower than those from RDWTBL 50 and the LANES column updated by FDOT. The variable LANE_CAT was found to have the lowest accuracy of 44.7%.

Table 3-6: Data Verification Results after Data Extraction Step

Variable		Total Links Verified	Links Matching RCI	Percent Matched
Number of Through Lanes	NOLANES in RCI vs LANES in All-Roads Map	302,245	291,202	96.3%
	NOLANES in RCI vs LANE_CAT in All-Roads Map	302,245	135,164	44.7%
Posted Speed Limit	SPEED in RCI vs SPDLIMIT in RDWTBL 50	83,612	55,554	66.4%
	SPEED in RCI vs SPEED_CAT in All-Roads Map	83,612	72,197	86.3%
Shoulder Type	SHLDTYPE in RCI vs SHLDTYPE in RDWTBL 50	106,471	87,386	82.1%
Median Type	TYPEROAD in RCI vs TYPTRWAY in RDWTBL 50	114,147	88,280	77.3%
	TYPEROAD in RCI vs DIVIDER in All-Roads Map	114,147	55,917	49.0%

For posted speed limit, the variable SPDLIMIT in RDWTBL 50 was found to have an accuracy of 66.4%. Although this is lower than the 86.3% achieved by the SPEED_CAT variable in the All-Roads map (provided by NAVTEQ™), the SPDLIMIT variable is still considered more reliable, as the SPEED_CAT variable does not provide the specific speed limit of the roadway. For example, a link with SPEED_CAT = 3 could either be a 55 mph section or a 60 mph section.

As can be seen from Table 3-6, data for shoulder type are available only in RDWTBL 50, and their accuracy was found to be 82.1%. The accuracy of the variable median type (TYPTRWAY) in RDWTBL 50 was found to be 77.3%, while the accuracy of its counterpart (DIVIDER) in the All-Roads map was found to be a low 49.0%.

CHAPTER 4
DATA DERIVATION USING SPATIAL ANALYSIS

This chapter describes the second step of the three-step process, which is to derive the missing data for roadway segments from adjacent segments that were already populated with data from the first step. It relies on the spatial analysis capabilities of GIS with the assumption that roadway features tend to be continuous. Similar to the verification procedure described in Section 3.3, the accuracy of the four variables in the final dataset was verified using the RCI data available for the state roads.

4.1 Data Derivation

The data extraction step described in Chapter 3 yielded information for only 14.9% of total links. Figure 4-1 shows an example of the data extraction results in ArcGIS. In this figure, the links in thick red color represent those with the information extracted from RDWTBL 50 and the All-Roads map. The links in gray color did not experience any crashes from 2003-2012, and hence, information is missing for these links. The links within the blue rectangle are in close proximity and they belong to the same roadway and share the same roadway characteristics. Accordingly, it can be assumed that the links in gray color are likely to have the same roadway geometry and traffic properties as the links in red color within the blue rectangle area.

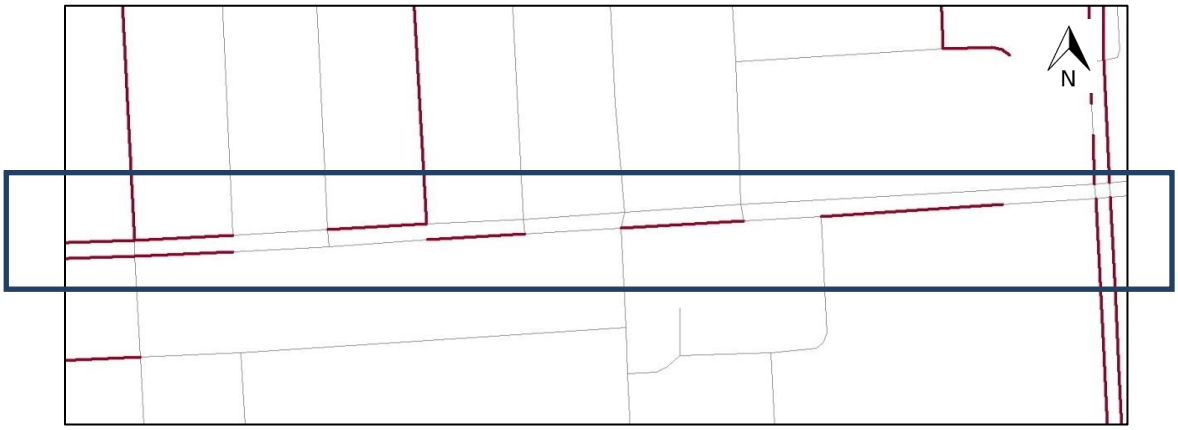


Figure 4-1: Screenshot of Links After Data Extraction Step

Based on the above continuity assumption, an automatic filling algorithm implementing the following rules was developed to derive the missing data:

- The link with missing data has the same Roadway ID as the link with extracted data.
- The link with missing data is physically connected to the link with extracted data.
- The link with missing data has the same roadway direction as the link with extracted data. This is determined by comparing the variable ROADSIDE in the roadway layer.
- The link with missing data has the same number of through lanes as the link with extracted data. This is determined by comparing the variable LANES in the roadway layer.

The detailed steps of the algorithm are described below:

1. Open the data table after the data extraction process, sort the table by ROADWAY ID, ROADSIDE, and LANES for the following three variables: median type, shoulder type, and posted speed limit.

2. Group all the links with the same ROADWAY ID, ROADSIDE, and LANES.

3. Within each group, select the link with extracted data and the smallest begin/end mile post (BMPADJ/EMPADJ) separately for all the three variables.

4. Based on the smallest begin/end mile post, verify whether or not the connected links have the data for the variables median type, shoulder type, and posted speed limit.

5. If the connected links do not have the data for these three variables, assign the values of the selected link (i.e., those with extracted information) to these links.

6. Repeat Steps 3-5 until all links within each group have a value for the variables median type, shoulder type, and posted speed limit.

7. Repeat Steps 3-6 for each group.

8. Generate the final table with the link, median type, shoulder type, and posted speed limit values.

Figure 4-2 shows the results of the automatic filling algorithm. Within the blue rectangle, the values for all westbound links in green color were derived from the adjacent (i.e., upstream and downstream) red links (i.e., links with extracted information) and the values for all eastbound links in purple color were derived from the adjacent red links.

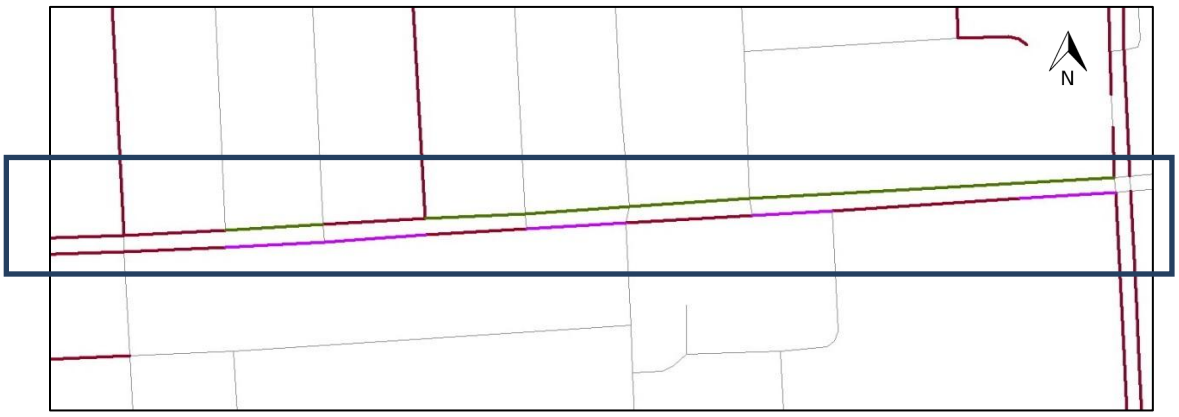


Figure 4-2: Screenshot of Links with Data Derived from Upstream and Downstream Links

After this step, 648,120 links (i.e., 41.4% of total road network) were found to have information for at least one variable. The number of links that have data for each variable is provided below:

- 616,078 links (i.e., 39.4% of total links) have data for posted speed limit.
- 646,772 links (i.e., 41.3% of total links) have data for shoulder type.
- 642,140 links (i.e., 41.0% of total links) have data for median type.

The above discussed statistics in terms of roadway miles are provided here. The links in the All-Roads map constitute 187,882.6 miles. After both the data extraction and data derivation steps, 114,342 miles (i.e., 60.8% of total miles) have data for at least one of the three variables. The total roadway miles that have data for each variable is given below:

- 104,314.7 miles (i.e., 55.5% of total miles) have data for posted speed limit.
- 114,301.7 miles (i.e., 60.8% of total miles) have data for shoulder type.
- 112,783.6 miles (i.e., 60.0% of total miles) have data for median type.

Tables 4-1 and 4-2 provide the results of the data extraction and data derivation steps based on number of links and miles of road network, respectively.

Table 4-1: Data Availability from Data Extraction and Data Derivation Steps Based on Number of Links

Variable	Crashes with Information	Total Links with Extracted Data (Step 1)	Percent Links with Extracted Data (Step 1)	Total Links with Derived Data (Step 2)	Final Percent of Links with Data (Steps 1 and 2)
Number of Through Lanes	0	N/A	N/A	N/A	N/A
Posted Speed Limit	595,643	210,203	13.4%	616,078	39.4%
Shoulder Type	668,811	229,778	14.7%	646,772	41.3%
Median Type	668,777	226,300	14.5%	642,140	41.0%

Note: There are a total of 1,565,026 links in the All-Roads map.

Table 4-2: Data Availability from Data Extraction and Data Derivation Steps Based on Miles of Road Network

Variable	Crashes with Information	Total Miles with Extracted Data (Step 1)	Percent of Miles with Extracted Data (Step 1)	Total Miles with Derived Data (Step 2)	Final Percent of Miles with Data (Steps 1 and 2)
Number of Through Lanes	0	N/A	N/A	N/A	N/A
Posted Speed Limit	595,643	32,264.6	17.2%	104,314.7	55.5%
Shoulder Type	668,811	34,625.5	18.4%	114,301.7	60.8%
Median Type	668,777	34,219.8	18.2%	112,783.6	60.0%

Note: There are a total of 187,882.6 miles in the All-Roads map.

4.2 Data Verification

The data verification procedure adopted to verify the accuracy of the data derived from spatial analysis is similar to the procedure discussed in Section 3.3. Table 4-3 gives the number of links with valid records (i.e., links with acceptable codes for the data variables) in the All-Roads map and the corresponding links in the RCI database after the data derivation step. Note that since the data for number of through lanes are complete for all links, the data were not derived for this variable. The results for number of through lanes are similar to the results presented in Section 3.3. About 600,000 of links in the All-Roads map have data derived for all the remaining three variables, and about 250,000 links in the RCI were identified for verification (i.e., 40% of total links with derived data).

Table 4-3: Number of Links in All-Roads Map and RCI After Data Derivation Step

Variable	Total Links with Data Derived from Spatial Analysis (Source: All-Roads map)	Total Links Identified for Verification (Source: RCI)	Percent of Links Identified for Verification
Number of Through Lanes	1,565,026	302,245	19.3%
Posted Speed Limit	616,078	180,223	29.2%
Shoulder Type	646,772	247,855	38.3%
Median Type	642,140	266,852	41.5%

Table 4-4 provides the final data verification results for all the four variables for the data extraction step, the data derivation step, and the overall combining both steps. As expected, the accuracy of the data derivation step was found to be generally lower compared to the accuracy of the data extraction step. This is observed for both the speed limit and shoulder type variables. However, for the median type variable, the accuracy of the data derivation step was found to be slightly higher than that of the data extraction step. Overall, the accuracies of the data derivation step are somewhat similar to the accuracies of the data extraction step, indicating a good viability of using the automatic filling algorithm (i.e., data derivation step) to increase the number of links with data.

Table 4-4: Data Verification Results After Data Derivation Step

Variable	Step 1: Data Extraction			Step 2: Data Derivation			Overall Steps 1 and 2		
	Total Links Verified	Links Matching RCI	Percent Matched	Total Links Verified	Links Matching RCI	Percent Matched	Total Links Verified	Links Matching RCI	Percent Matched
No. of Through Lanes	302,245	291,202	96.3%	N/A	N/A	N/A	302,245	291,202	96.3%
Speed Limit	83,612	55,554	66.4%	96,611	57,506	59.5%	180,223	113,060	62.7%
Shoulder Type	106,471	87,386	82.1%	141,384	114,616	81.1%	247,855	202,002	81.5%
Median Type	114,147	88,280	77.3%	152,705	119,280	78.1%	266,852	207,560	77.8%

4.2.1 Number of Through Lanes

As discussed in Chapters 2 and 3, the LANES column in the All-Roads map has data for the number of through lanes for all the links. As discussed in Section 3.3, the NOLANES variable in the RCI was compared to the LANES variable in the All-Roads map. The data in the All-Roads map were found to have an accuracy of 96.3% for links with functional classification codes 1-19. Note that these links are either state roads or major off-system roads, and this 3.7% inaccuracy could be because of either road construction projects or incorrect information. The data accuracy of the remaining local roads with functional classification codes 20-33 was manually verified, and the results are discussed in Chapter 5.

4.2.2 Posted Speed Limit

RDWTBL 50 was found to give an accuracy of 62.7% based on a total of 180,223 links verified, or about 11.5% of the total 1,565,026 links in the All-Roads map. It is reasonable to assume that speed limits extracted from RDWTBL 50 for the remaining 88.5% of links could achieve a similar level of accuracy. Figure 4-3 shows the number and distribution of speed limits extracted from RDWTBL 50 for the links on local streets with functional classification codes greater than 19. It can be seen that most of the links have a speed limit of 25 mph or 30 mph. This is expected because a majority of these links are minor local roads. Note that there are still a significant number of links with higher speed limits (i.e., ≥ 45 mph).

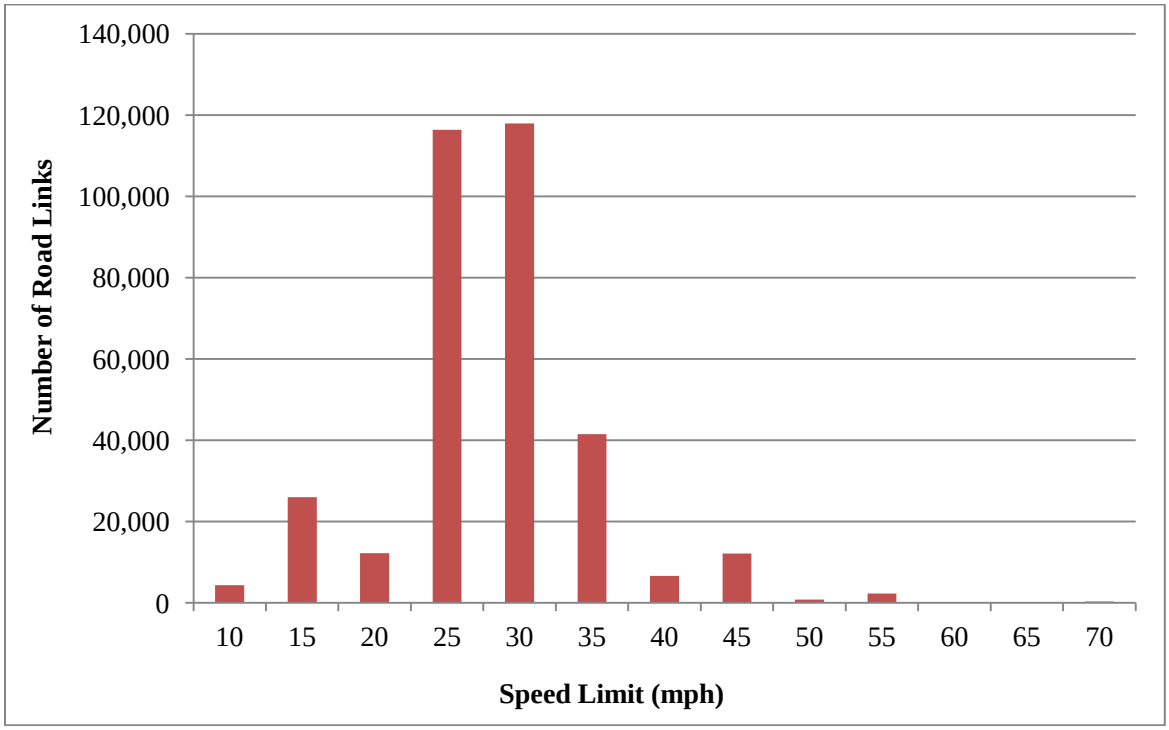


Figure 4-3: Distribution of Speed Limit Data Extracted from RDWTBL 50 for Links with Functional Class Greater Than 19

Figure 4-4 shows an example of a link with a 70 mph speed limit. The link identified in Figure 4-4 is a ramp section and a parking plaza within a freeway service area. This link should not have 70 mph speed limit. It appears that the police officer had incorrectly considered this link to be part of the adjacent freeway, and recorded the freeway speed limit of 70 mph as the default speed limit for this link. Overall, with an accuracy of 62.7%, the data from RDWTBL 50 are not sufficiently accurate to be used for acquiring posted speed limits. However, they may serve as a means to detecting future changes in posted speed limits for the purpose of data updates.

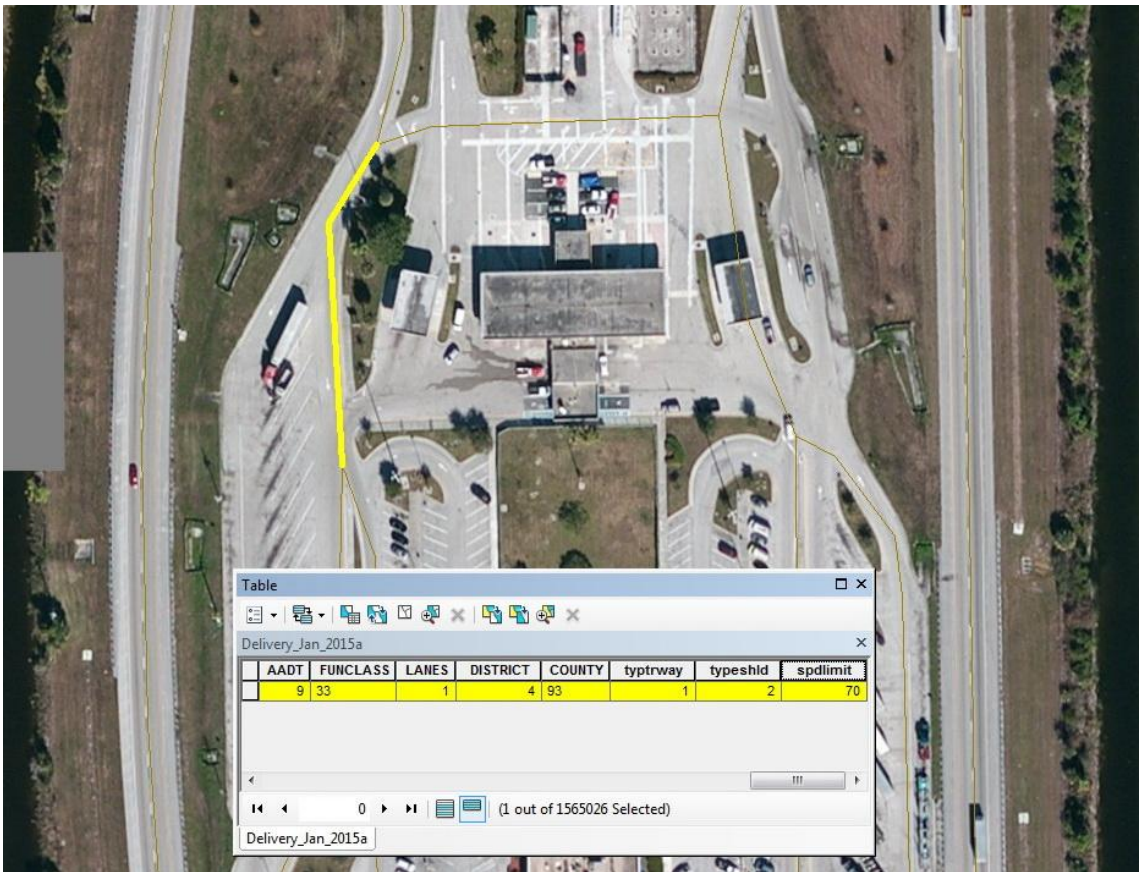


Figure 4-4: Example of a Road Link on Local Road with 70 mph Speed Limit

4.2.3 Shoulder Type

The data from RDWTBL 50 were found to achieve an overall accuracy of 81.5% based on a total of 247,855 links verified, or 15.8% of the total 1,565,026 links. It may be assumed that the data extracted from RDWTBL 50 for the remaining 84.2% of links have a similar level of accuracy. Although the accuracy of the derived data for shoulder type is over 80%, the data were not collected to the required detail in the crash records. The shoulder type variable in the police crash reports has three codes, while the RCI has ten codes (see Table 3-4). For this reason, the prospect of using shoulder type extracted from RDWTBL 50 is considered limited. In addition, manual data collection and data verification of shoulder type cannot be reliably done as the difference between "paved" and "curb" shoulder types is difficult to determine from the existing satellite

images. Consequently, RDWTBL 50 is not considered a sufficiently accurate data source for acquiring shoulder type data.

4.2.4 Median Type

The median type data from RDWTBL 50 were found to have a relatively better accuracy of 77.8% based on a total of 266,852 links verified, or about 17% of the total 1,565,026 links. It may be assumed that median type data extracted from RDWTBL 50 for the remaining 83% of links could achieve a similar level of accuracy. However, there is a slight inconsistency in how the police define "divided" segments and the RCI's definition.

Figure 4-5 gives an example of a link identified as "divided" in the RCI database and "undivided" in RDWTBL 50. It can be seen from the figure that the links in the two directions are separated by a left-turn bay. The police code this link as undivided because there is no raised median. On the other hand, the link was coded as divided in the RCI database because the left-turn bay separates the links in the two directions. Due to such inconsistency, the prospect of using median type information obtained from crash data (i.e., RDWTBL 50) is somewhat limited.

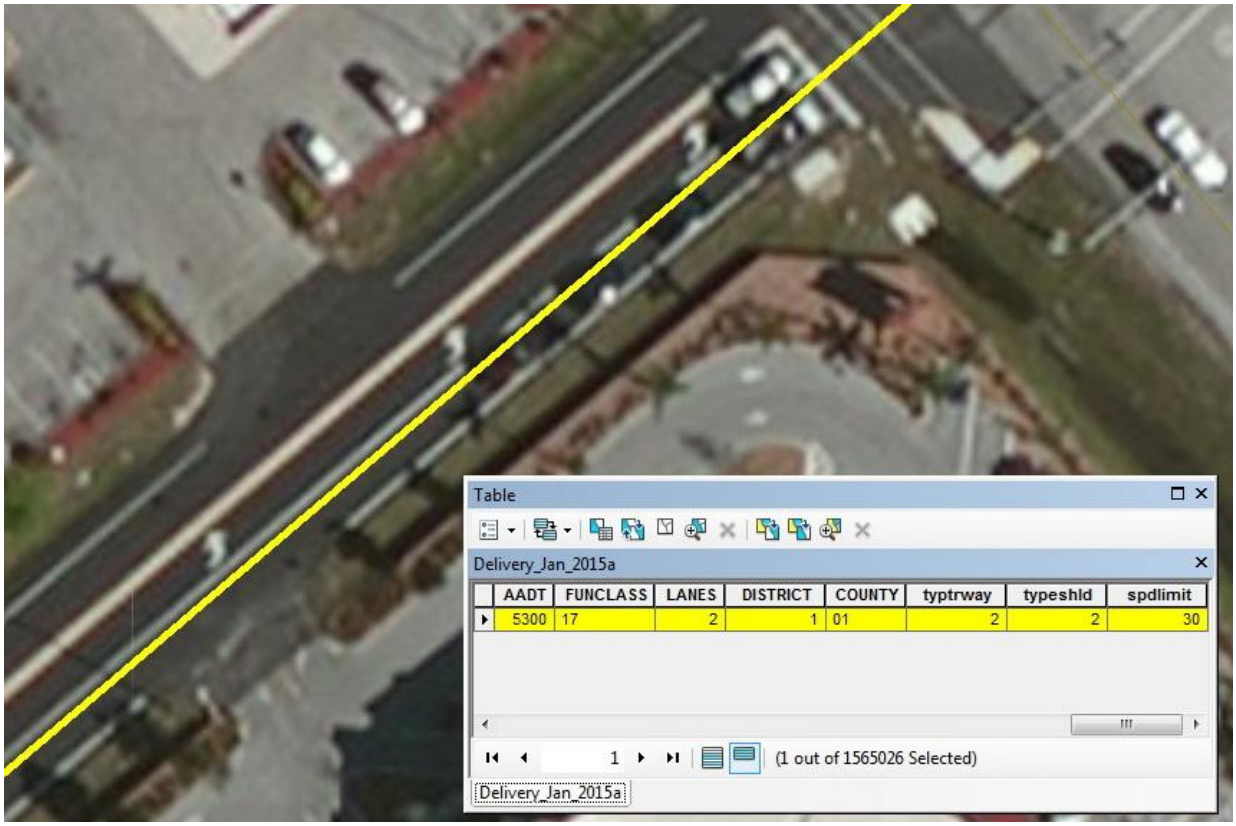


Figure 4-5: Example of a Link Coded as "Divided" in RCI and "Undivided" in RDWTBL 50

CHAPTER 5
MANUAL DATA COLLECTION

This chapter describes the manual data collection and verification process adopted to verify and update data on number of through lanes. As aforementioned, although number of through lanes is recorded in the police reports, data for this variable are not included in RDWTBL 50. However, because the All-Roads map already included the LANES variable for the number of through lanes, this project focused on verifying this existing LANES data. An in-house web-based application, Visual Roadway Inventory Collection System (VRICS), was adapted to facilitate the collection and verification of the LANES data using Google Maps. This chapter first introduces the VRICS application, and then discusses the data collection effort to collect information on number of through lanes. It also provides the descriptive statistics of the data collected.

5.1 VRICS Application

Figure 5-1 shows a screen capture of the main interface of the system. The system reads a linear-referenced roadway segment, converts its coordinates to the Google Maps projection on the fly, and then displays the satellite image of the segment. The system also shows the existing number of through lanes retrieved from the All-Roads map. The user can choose the observed number of through lanes (from 1 lane to 9 lanes) based on the displayed satellite image. If the user cannot determine the number of through lanes, a note can be made to explain the reason (for example, the segment is covered with trees and lanes are not clearly visible, etc.). After completing a segment, the user can quickly have the system jump to and display the next segment to continue with data verification and collection.

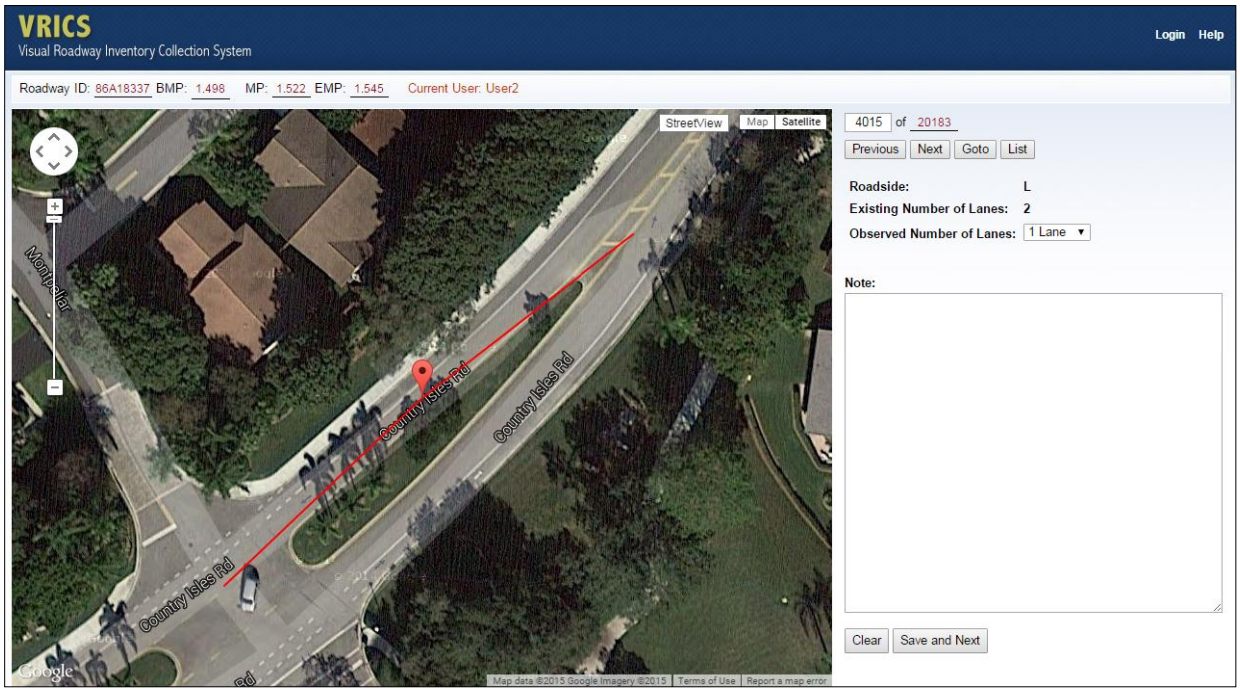


Figure 5-1: VRICS Application Customized to Collect Data for Number of Through Lanes

5.2 Data Verification Results

As shown in Table 4-3, a total of 302,245 links from the RCI were used to automatically verify the accuracy of number of through lanes information in the All-Roads map. Of these 302,245 links, 291,202 were found to match the RCI, and the remaining 11,043 links (i.e., 3.7%) needed to be updated. Information on non-state roads that is not available in the RCI could not be automatically verified. VRICS was customized to verify number of through lanes data for the remaining 1,262,781 links. Of these 1,262,781 links, only 6,061 links (i.e., 0.5%) were updated. Table 5-1 shows the detailed results of the data verification and collection process. In total, 17,104 links were updated for the number of through lanes variable, which accounts for 1.1% of the total 1,565,026 links.

Table 5-1: Data Verification Results for Number of Through Lanes

Process	Total Verified Links	Total Corrected Links	Percent of Corrected Links
Automatic Data Verification after Data Extraction and Data Derivation Steps	302,245	11,043	3.7%
Manual Data Verification using VRICS	1,262,781	6,061	0.5%
Total	**1,565,026**	**17,104**	**1.1%**

As can be seen from Table 5-1, 0.5% of links were updated using manual data verification process, while 3.7% of links were corrected using automatic data verification process. The difference in the two percentages is expected. The links verified manually using VRICS were mostly minor local roads in areas such as local communities and subdivisions. These roads are typically two-lane and do not usually change. On the other hand, the links in the RCI were mostly major roads, which have a varying number of lanes and were more likely to undergo frequent improvements as a result of upgrade or reconstruction. However, some of the corrected links were also due to miscoding in the All-Roads map. Figure 5-2 shows an example of a link with incorrect number of through lanes. The highlighted two links should each have only one through lane based on the satellite image; however, the LANES variable in the All-Roads map for the two links shows two lanes for each link.

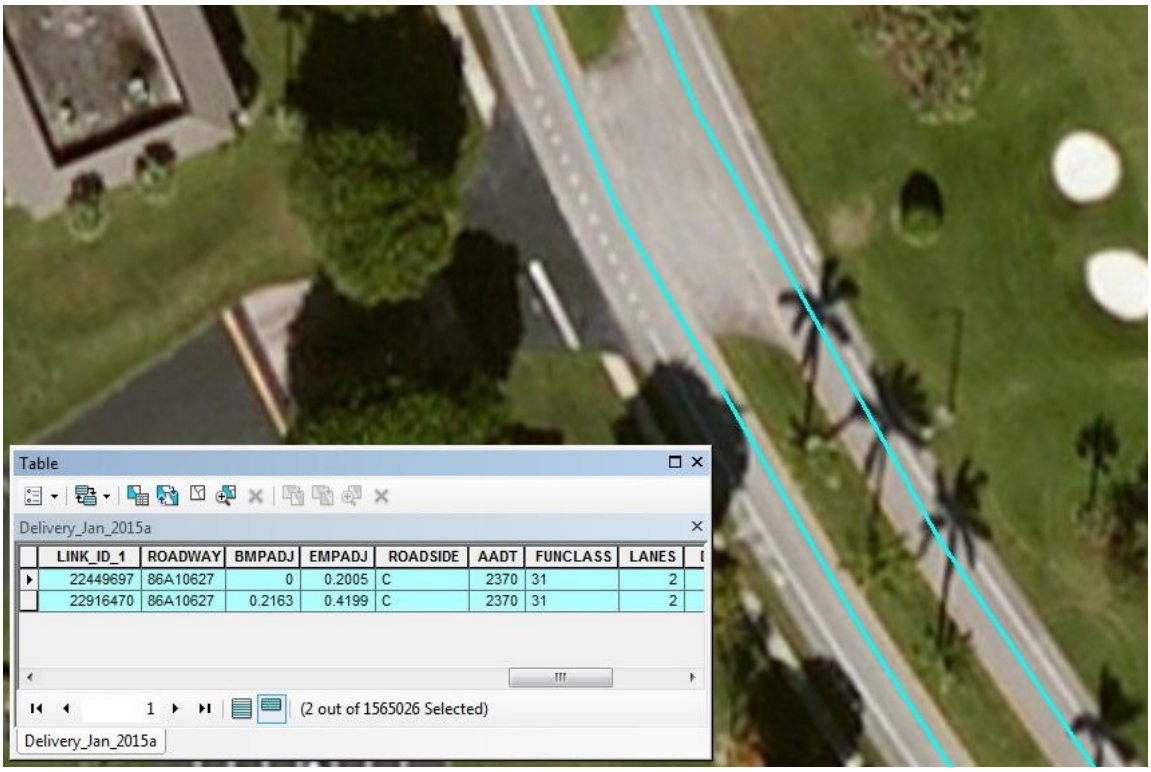


Figure 5-2: Example of a Road Link with a Different Number of Through Lanes in All-Roads Map

CHAPTER 6
CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This project aimed to develop methods to extract roadway data recorded in crash reports as a means to both acquire and continually update the All-Roads map for local roads in Florida. To the extent possible, the project attempted to extract data for the following four variables that are included in Florida's crash reports:

- number of through roadway lanes,
- posted speed limit,
- shoulder type (paved or unpaved), and
- median type (divided or undivided).

The process to acquire information for the four data variables involved three steps. The first was to extract data from crash records. Crash records for the years 2003 through 2012 were used. In the case where there were multiple crashes with different data entries (e.g., different speed limits were recorded in the police reports for crashes that occurred on a segment), the process considered the number of crashes and the timeline of crash occurrences to select the most probable value. The second step was designed to address the case in which road segments did not have any crashes. In this step, the values were derived from their adjacent segment based on the assumption that roadway features are continuous. Finally, as the third and final step of the process, missing data for the remaining segments were manually collected using an in-house web-based application designed to facilitate the review and recording of information from Google Maps' satellite images.

Table 6-1 lists the new variables added to the All-Roads map using the three-step process. The table also provides the field names of the variables obtained during each step of the process along with the variable codes.

6.1.1 Number of Through Lanes

Although number of through lanes is an entry in the police reports, it was found that the data for the corresponding field, TRWAYLN, in RDWTBL 50 of the CAR database were completely missing. However, because the All-Roads map already included the LANES variable for the number of through lanes, this part of the project became one of verifying the existing LANES data. The variable NOLANES in the RCI was used to verify the LANES data. The results show that the data was 96.3% accurate. This high percentage was expected as the verified LANES data came originally from the RCI. The 3.7% inconsistency was likely a result of mainly changes in the RCI data between 2012 (when the All-Roads map was updated) and 2015 (RCI data year used in verification) due to roadway construction. As 3.7% of the road links in the All-Roads map did not have the same number of through lanes as identified in the RCI database, these links (11,043 links in total) were manually verified in VRICS.

Table 6-1: New Variables Included in All-Roads Map

Variable	Field Name	Description	Code
Number of Through Lanes	LANES_S3	Number of Lanes from Step 3 of the process for updated links only	N/A
	LANES_UPTD	Final Number of Lanes for all links	
Posted Speed Limit	SPDLIM_S1	Extracted Maximum Speed Limit from Step 1 of the process	N/A
	SPDLIM_S2	Derived Maximum Speed Limit from Step 2 of the process	
	SPDLIM	Maximum Speed Limit from Steps 1 and 2 of the process	
Shoulder Type	SHLDTYP_S1	Extracted Shoulder Type from Step 1 of the process	1 - Paved 2 - Unpaved 3 - Curb
	SHLDTYP_S2	Derived Shoulder Type from Step 2 of the process	
	SHLDTYP	Shoulder Type from Steps 1 and 2 of the process	
Median Type	TYP_WAY_S1	Extracted Median Type from Step 1 of the process	1 - Divided 2 - Undivided
	TYP_WAY_S2	Derived Median Type from Step 2 of the process	
	TYP_WAY	Median Type from Steps 1 and 2 of the process	

As the RCI data includes links on state roads and major off-system roads with functional classification codes 1-19, the inconsistencies in the RCI data and the All-Roads map could be because the major roads (i.e., freeways, arterials, and major collectors) were more likely to undergo frequent improvements as a result of upgrade or reconstruction. The remaining links in the All-Roads maps with functional classification codes 20-33 are mainly local roads. As these roads are typically two-lane and do not usually change, it was found that only about 0.5% of these roads needed to be updated or corrected.

6.1.2 Posted Speed Limit

Data for posted speed limit are recorded as SPDLIMIT in RDWTBL 50 and as SPEED in the RCI. Although SPEED_CAT variable in the All-Roads map has information on speed limit, it is not considered a reliable source since it does not provide the specific speed limit. After the data extraction and data derivation steps, 39.4% of the total links (i.e., 616,078 of 1,565,026 links) have information on speed limits. As part of the data verification step, the data obtained from the SPDLIMIT variable in RDWTBL 50 was compared to the data from the SPEED variable in the RCI. Only the links with functional classification codes 1-19 were verified as the data in the RCI includes only these links. The accuracy of this variable was found to be 62.7%. In other words, 62.7% of the total links verified were found to match those from the RCI.

6.1.3 Shoulder Type

Data for shoulder type are recorded as SHLDTYPE in both RDWTBL 50 and the RCI databases. However, the codes in the RCI are more detailed compared to the codes in RDWTBL 50. The

police records shoulder type as paved, unpaved, or curbed, while the RCI codes shoulder type using the following codes: raised curb, paved, paved with warning device, lawn, gavel/marl, valley gutter, curb and gutter, other, curb with resurfaced gutter, and managed lane. Therefore, the prospect of using shoulder type extracted from RDWTBL 50 is limited. In addition, manual data collection and data verification of shoulder type cannot be reliably done as the difference between "paved" and "curb" shoulder types is difficult to determine from the existing satellite images. Consequently, RDWTBL 50 was not found to be a sufficiently accurate data source for acquiring the shoulder type data.

6.1.4 Median Type

Data for median type are recorded as TYPTRWAY in RDWTBL 50, DIVIDER in the All-Roads map, and TYPEROAD in the RCI. For the links available in both the RCI and the All-Roads map, the accuracy of the variable DIVIDER in the All-Roads map was found to be only 49.0%. Therefore, this variable is not considered as a reliable source for median type information. On the other hand, the accuracy of this variable in RDWTBL 50 was found to be 77.8%. However, there is a slight inconsistency in how the police define "divided" segments and the RCI's definition. The police code locations with only a physical barrier as "divided", while the RCI codes links separated by turn bays as "divided". Due to this inconsistency, the prospect of using median type information obtained from crash data (i.e., RDWTBL 50) is somewhat limited.

6.2 Recommendations

Among the four variables explored in this project, the number of through lanes and posted speed limits are considered the most important for safety analyses as well as for many other applications. While the All-Roads map includes the number of through lanes, there is not an automated mechanism to either acquire these data for new links or update existing links (other than those links that are in the RCI database). It is not known why such data from the crash reports are missing from the CAR system. However, the authors have been told that the data would be included in the CAR database beginning with the 2011 crash data under its new crash data format. Once the data are included, Steps 1 and 2 of the process presented in this project may be applied to serve as a means to detecting and updating number of through lanes for the All-Roads map.

In this project, data for posted speed limits were obtained from crash report data for only 39.4% of the total links in the All-Roads map. Unfortunately, it was not feasible to collect the data for the remaining links. An extension of the current study would be to explore the use of speed limits data from Google Maps' Roads API[5]. The application uses large samples of real-time speed data from local vehicles to estimate speed limits.

[5] https://developers.google.com/maps/documentation/roads/

Future studies could also explore methods to estimate the posted speed limit from existing speed data. One potential source of speed data is the Waze application[6] which collects speed data reported by local drivers. The application helps drivers reduce travel times by suggesting less congested routes, identifying locations with traffic incidents, etc. The travel speeds of drivers could be calibrated to correlate with the known posted speeds from the RCI for the state roads and from crash reports for the local roads. It is then possible to apply these calibrated relationships to estimate the posted speed limits for similar roadways given their drivers' travel speeds. The calibration can be done for different types of roads and for different areas and regions.

[6] https://www.waze.com/