Collaborative Sciences Center for ROAD SAFETY

FINAL REPORT

National Pedestrian and Bicycle Safety Data Clearinghouse Phase I: Inventory & Framework

May 31, 2019

Krista Nordback
Wesley Kumfer
Seth LaJeunesse
Libby Thomas
Katie Heuser
University of North Carolina Highway Safety Research Center

Julia Griswold
University of California Berkeley

National Pedestrian and Bicycle Safety Data Clearinghouse Phase I: Inventory & Framework

THE UNIVERSITY OF TENNESSEE KNOXVILLE | Berkeley UNIVERSITY OF CALIFORNIA | FAU FLORIDA ATLANTIC UNIVERSITY | Duke UNIVERSITY | THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

U.S. DOT Disclaimer

Acknowledgement of Sponsorship

This project was supported by the Collaborative Sciences Center for Road Safety, www.roadsafety.unc.edu, a U.S. Department of Transportation National University Transportation Center promoting safety.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
CSCRS-R14		

4. Title and Subtitle: National Pedestrian and Bicycle Safety Data Clearinghouse Phase I: Inventory & Framework	**5. Report Date** May 2019 **6. Performing Organization Code**
7. Author(s) Krista Nordback, Wesley Kumfer, Seth LaJeunesse, Libby Thomas, Katie Heuser, Julia Griswold	**8. Performing Organization Report No.**
9. Performing Organization Name and Address University of North Carolina Highway Safety Research Center 730 Martin Luther King Jr. Blvd., Suite 300 Chapel Hill, NC 27599-3430 University of California, Berkeley Safe Transportation Research and Education Center Berkeley, California	**10. Work Unit No.** **11. Contract or Grant No.** CSCRS-R14
12. Sponsoring Agency Name and Address CSCRS	**13. Type of Report and Period Covered** Final Report (April 2018-May 2019) **14. Sponsoring Agency Code** CSCRS

15. Supplementary Notes

Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.
This project produced an online resource: www.pedbikedata.org

16. Abstract

Safety researchers, especially those studying bicycle and pedestrian safety, often struggle to obtain sufficiently high-quality data to conduct robust safety studies. The goal of the project is to create an online centralized data clearinghouse for bicycle and pedestrian safety-related data as a national resource for safety researchers to expedite research on this topic. CSCRS Project R14, the first phase of this work, produced a pilot online pedestrian and bicycle safety data resource clearinghouse available at pedbikedata.org. It allows researchers to search from over 4,000 pedestrian and bicycle safety-related data resources by geographic location (national, state, region, county or city), data type (collision, counts, infrastructure, or other), data format, and availability. The search results provide information on the time period for which the data are available as well as the owner of the data and a link to the online dataset itself. The final report details a system for rating the datasets and shows that there is a great need for higher quality data of all types. This project is continuing with a second phase to improve the clearinghouse and user interface.

17. Key Words	18. Distribution Statement
Pedestrian, Bicycle, Safety, Data, Clearinghouse	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	50	

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

CONTENTS

List of Figures

List of Tables

Introduction

In 2017, 5,977 pedestrians and 783 bicyclists died on roads in the United States [1]. While total traffic fatalities in the United States other than pedestrians and bicyclists fatalities have declined by 6% over the decade (2008-2017), pedestrian and bicyclists fatalities have increased by 32% [1]. In 2017, when pedestrian and bicyclist deaths accounted for 18% of the total traffic fatalities in the United States, they only accounted for approximately 3% of commuters [2].

Despite these concerning trends, pedestrian and bicycle safety treatments are consistently underfunded [3], especially in low-income neighborhoods [4]. Although pedestrian and bicycle fatalities comprise a significant portion of the annual total, any severity of pedestrian and bicycle collisions can be widely dispersed events throughout a network [5]. Confounding this problem is a lack of consistent, centralized data (e.g., exposure and facility infrastructure data) to enable collision prediction. For example, although many cities and local jurisdictions throughout the country have gathered pedestrian and bicycle counts for years, these data are often problematic and not easily accessible [6]. Even when available, collision data for non-fatal crashes is not available in a nationwide database, and even for a single agency, sources can differ, making comparison between jurisdictions more widely difficult [7]. Therefore, predicting vulnerable road user collisions and determining appropriate countermeasures can be unduly difficult for traffic safety professionals.

To address the data shortcomings, a National Pedestrian and Bicycle Data Clearinghouse has been developed to improve data for pedestrian and bicycle analyses in the United States. The goal for this Clearinghouse is to serve as a central, consistent, and open data source for traffic safety professionals to access for planning and safety analysis purposes. The current phase of this project, as described in this report, entails defining the scope of the Clearinghouse, cataloguing available datasets, and building a rudimentary framework for data hosting and acquisition.

The purpose of this report is to document the process used to create the Clearinghouse. The process started by identifying the major hurdles facing pedestrian and bicycle safety researchers, such as lack of reliable exposure data or difficulty accessing crash report narratives. To identify these hurdles, a three-part analysis was conducted: literature search of representative, cutting edge studies detailing common analysis methods and existing data sources used; interviews with leading researchers on vulnerable road user safety; and an inventory of existing publicly available online datasets. A dataset rating system was also developed and applied to the state-wide datasets. Ultimately, this research was conducted to determine what is currently missing in vulnerable road user safety analysis so that our Clearinghouse can fill that gap.

This report is organized into the following chapters:

- The report starts with an introduction to the problem and illustrates the purpose of this study.
- Chapter 2 contains a literature review with details on known data issues, analysis methods, common variables, and available datasets and includes a curated list of available data sources identified in the literature.
- Chapter 3 then presents a summary of interviews conducted with leading researchers to verify conclusions drawn from the literature review and data scan.
- Chapter 4 discusses our inventory of online pedestrian and bicycle safety related datasets.
- Chapter 5 discusses a rating system and how it was applied to the state-wide datasets in the Clearinghouse.
- Chapter 6 synthesizes the previous chapters, identifies data gaps, and discusses the broader implications for our Clearinghouse.
- Chapter 7 discusses the structure of the clearinghouse and documents its online features.
- Chapter 8 provides an overview of the findings of the study.

Definitions

To assist the reader, some terms are defined below.

Clearinghouse – For the purposes of this report, the “Clearinghouse” refers to the online National Pedestrian and Bicycle Safety Data Clearinghouse available at www.pedbikedata.org. The term clearinghouse indicates that the site’s purpose is to connect data users to data sources.

Collisions – In this report and in the Clearinghouse, the term “collisions” is used to indicate crashes between motorists and pedestrians or between motorists and bicyclists. This term is used instead of “crashes” because the term crash seems more indicative of a metal on metal collision and less indicative of the metal on flesh collisions that usually occur between motorists and vulnerable road users. The term also applies to single bicycle, single pedestrian crashes and bicyclist-pedestrian crashes.

Counts – For the purposes of the Clearinghouse, counts refer to a count of pedestrian, bicycle or motor vehicles, or “traffic monitoring”, data that can be used as a basic metric of exposure. While other exposure metrics are commonly used, such as demographics (e.g., population and employment density) and travel survey data [8], count data was specifically identified as a needed data type for safety analysis by interviewees and in the literature.

Geographic scale – The Clearinghouse describes each dataset in terms of four geographic levels city, county, region, state and national geographic areas. International data were not included.

Infrastructure – The Clearinghouse defines data as infrastructure-related when the data are describing physical structures. These include permanent count equipment as well as street networks, sidewalks and bicycle lanes.

Metadata – Metadata are data about data. Information that describes, defines or details aspects of the data.

Literature Review

This literature review explores the state of pedestrian and bicycle data available for traffic safety modeling. Pedestrian and bicycle data are needed for many purposes from identifying high crash locations to planning and design for future pedestrian and bicycle facilities. However, a main focus of this project is specifically to improve pedestrian and bicycle safety research in order to identify risk factors and quantify the effectiveness of treatments. Thus, the needs of traffic safety modeling are key. The purpose of the review is to identify the major gaps in pedestrian and bicycle data for safety needs so that these shortcomings can be accommodated by the Clearinghouse, as possible. Specific topics discussed in this literature review include:

- Data demands for modeling pedestrian and bicyclist safety
- Available datasets and common modeling processes and applications
- Common issues with the available data

Within each of these topics are a myriad of concepts and issues. The literature here is representative of the extant of the current state of the practice, although other, more specific studies may also exist.

Data Demands

As mentioned in the introduction, current traffic safety trends necessitate more robust analyses of pedestrian and bicycle safety. However, the demanding nature of data acquisition, data linking, data cleaning, and data analysis, combined with the random and often widely distributed nature of collisions with pedestrians (and bicyclists), can create challenges to adequately accounting for pedestrian and bicycle safety. Simply put, the demand for data is high, although the availability of data itself may be limited.

Further compounding this challenge is the rigor of data needed. In a 2013 paper, Bauer et al. noted that any pedestrian or bicycle collision data acquired must be capable of linkage with other relevant datasets; the researchers focused specifically on hospital data to provide a more accurate picture of severity for collisions involving vulnerable road users, but a wide range of possible datasets can and often should be linked to pedestrian and bicycle collision data to account for risks [9]. This same need for linkage was also expressed more recently by Morris and Wier in 2016 and Black et al. in 2018. In both of these studies, the researchers noted the importance of linking pedestrian and bicycle collision data to geospatial data, because collisions involving vulnerable road users are often products of inherent risk factors within the built environment. It can be difficult to assess proper risks to vulnerable road users and potential treatments to combat these risks without an adequate sense of where specifically collisions occur [3] [10]. For example, pedestrians crossing midblock may face different risks than bicyclists riding through intersections, so collision data must be linked as closely as possible to other data to allow proper modeling.

In addition to geospatial linkage, researchers have noted that other modes (e.g., motor-vehicle data) should also be linked to pedestrian and bicycle collision data. In an analysis of tram and streetcar crash reporting, Nazin and Currie noted that collisions with pedestrians are often underreported, so linking existing pedestrian collision data with transit stop data may more properly account for exposure and gaps in pedestrian collision data. Nazin and Currie noted that transit stop data are especially important for urban locations [11]. Zhang et al. framed this need for linkage to other modes differently in an earlier study. While developing a plan to collect critical, geospatially linked pedestrian data for inclusion in California Department of Transportation's (Caltrans's) Traffic Accident Surveillance and Analysis System database, the researchers noted that a true evaluation of a multimodal system is difficult without pedestrian and bicycle data, so data linkage is critical [12].

Pedestrian and bicycle data may also be required to accomplish other safety goals. For example, Zhang et al. noted in 2014 that agencies may need more robust pedestrian facility data to identify locations in need of ADA compliance; missing ADA-compliant facilities may also explain risks to pedestrians, so accommodation is another motive for linking pedestrian safety and count data geospatially [12]. Long-term planning efforts within cities may also need consistent and detailed collision data to produce adequate collision prediction models that can in turn be used to justify spending and infrastructure development, as noted by Eluru et al. in 2016 [13].

Ultimately, pedestrian and bicycle data can be used in a variety of ways and thus should be linked to other data to give a comprehensive and clear picture. Nordback et al. established a number of properties for any database to possess to meet these data demands. According to the study, archives of pedestrian and bicycle data must be centralized, national, and uploadable so that users across the country can access, adapt, and supply their own data to fit needs in their jurisdictions [6]. Put differently, the end users of pedestrian and bicycle safety data have different needs that must be accommodated by linked, quality checked and sufficiently detailed datasets [10].

Available Data and Common Variables

Common Analyses

The variables and data types used in pedestrian and bicycle safety analyses may vary widely depending on the end user of the study, and a substantial range of both analysis types and specific models within these analysis types are attested in the literature. Though analyses can be used for different purposes, including high crash location identification, counter measure selection, and facility planning, this report focuses specifically on crash modeling for the purpose of understanding pedestrian and bicyclist safety performance and establishing crash modification factors for potential safety treatments. These analyses and models vary both in complexity and scope, and the usefulness for countermeasure selection, and long-term infrastructure planning depends both on the quality of the analysis and the variables used. However, many pedestrian and bicycle safety studies are still hampered by a lack of available data or linked data, thus making more than cursory examinations more challenging.

Despite the demanding nature of pedestrian and bicycle safety modeling, researchers like Talbot et al. and Kullgren et al. have used simple examinations of collision data involving pedestrians and bicyclists to identify potential cases of distraction [14] or identify potential interventions for collisions involving vulnerable road users [15], respectively. More complex pedestrian and bicycle safety analyses, however, often follow the type of collision prediction and network screening approach suggested by the Highway Safety Manual (HSM) [16].

In 2016, Chimba and Musinguzi used pedestrian and bicycle collision data linked to population and other explanatory variables to develop Safety Performance Functions (SPFs) for predicting both pedestrian and bicycle collisions at the census block level [17]. Monsere et al. also discussed SPF development as one of several analysis types that can be beneficial for improving safety for vulnerable road users [18]. Crash Modification Factors (CMFs) are often developed alongside SPF models to describe the relationship between specific variables (e.g., roadway features) and collision outcomes. In 2014, Zhang et al. described the need for more data so that better pedestrian CMFs can be developed in California to lead to more effective pedestrian collision countermeasures [12]. More recently, Black et al. [3] presented a Vision Zero-based screening tool to allow agencies to measure collision reductions to assess the efficacy of pedestrian and bicycle countermeasures. Their screening process, though slightly different from the network screening procedure described in the HSM, follows the same basic formula of measuring excess collisions over time to evaluate the efficacy of a safety management program [3].

In addition to these more typical analyses, the literature includes more novel analysis and prediction types to meet specific needs. A recent trend in safety analysis for vulnerable road users is termed the “systemic safety” approach. It involves identification of common risk factors and a systemic approach to addressing these risk factors at “risky” sites, regardless of collision history. Although risk analysis-type evaluations are not always systemic in nature, the two concepts (risk analysis and systemic safety approach) are often linked. For example, Ariffin et al. conducted a study to identify risk factors for pedestrians using collision data and roadway/traffic data [19]. This study, however, used odds ratios to identify risks rather than negative binomial regression or SPFs, as commonly used in systemic analyses [20]. Monsere et al. also used odds ratios, developed through logistic regression models of pedestrian and bicycle collision data, to construct risk-scoring tools. These tools were to be used systemically to measure risk for pedestrians and bicycles at both intersections and segments based on a number of risk factors identified through the analysis [18]. Das et al. also explored risk analysis in their 2018 paper. This approach, however, was novel in that the researchers used a data mining procedure to identify common risk factors in pedestrian collision data [21]. A further exploration of risk analyses and the data requirements of a systemic screening for pedestrian safety are contained in NCHRP Report 893 [20].

Eluru et al. and Chimba and Musinguzi describe two novel pedestrian and bicycle safety studies in 2016. In their study to develop planning-level collision predictions, Eluru et al. modeled pedestrian and bicycle safety in relation to origin and destination data using negative binomial regression (the preferred regression method for mitigating regression-to-the-mean bias), ultimately producing simulations for varying policy conditions to model how changes in trip demands affect vulnerable road user safety [13]. Chimba and Musinguzi's study to develop pedestrian and bicycle SPFs deployed a cluster analysis with collision data linked to population as the exposure variable with the intent of identifying locations with higher proportions of pedestrian and bicycle collisions [17]. Locating risky locations is similar to the systemic safety approach mentioned earlier, but Chimba's and Musinguzi's study differs by focusing on clusters of collisions rather than sites where collisions are predicted to occur.

Common Safety Analysis Models

A variety of safety analysis modeling techniques are used in each of the analysis methods above for the purposes of modeling what factors are related to or predictive of pedestrian and bicycle collisions or risk. Monsere et al. catalogued these various statistical models for pedestrian and bicycle safety analysis [18]; Table 1 highlights these methods and includes a few more not mentioned by Monsere et al. As can be seen in Table 1, wide varieties of tools are available to researchers and practitioners to measure safety for pedestrians and bicyclists.

TABLE 1: Common Modeling Types in Pedestrian and Bicyclist Safety Analysis

Road User Type	Model Type	Applications
Pedestrian	Negative Binomial	Collision prediction [18]
Pedestrian	Linear Regression	Collision prediction [18]
Pedestrian	Probit	Severity estimate [18]
Pedestrian	Multivariate Regression	Risk analysis [22] Severity estimate [22]
Pedestrian	Association Rules Mining	Collision pattern analysis [21]
Pedestrian	Logistic Regression	Risk analysis [23]
Bicycle	Poisson and Negative Binomial	Collision Prediction [18] Risk analysis [18]
Bicycle	Linear Regression	Collision prediction [18]
Bicycle	Logit	Severity estimate [18] Risk analysis [18] Hierarchical analysis [18]
Bicycle	Probit	Severity estimate [18] Risk analysis [18]
Bicycle	Multivariate Regression	Factor analysis [18]
Bicycle	Spatial Bayesian	Risk analysis [18]
Bicycle	Quasi-induced exposure	Risk analysis [18]
Bicycle	Side-path Safety	Factor analysis [18]
Bicycle	Origin-Destination	Network analysis [18]
Bicycle	Logistic Regression	Risk analysis [23]

Common Variables

A vast number of potential variables that may help explain risks to bicyclists and pedestrians and provide predictive power for collision modeling are described in the literature. Although these variables may be used in a variety of ways and can often be found in different datasets, they tend to fall into general explanatory groupings. Here we collect an extensive sample of the variables used in the various modeling processes shown in Table 1 and group them as appropriately as possible. Where appropriate, each category of variable is subdivided into collected variables. Variables in Table 2 are not subdivided into pedestrian or bicycle categories because pedestrian and bicycle concerns often overlap, and specific variables may have more explanatory power based on location. The purpose of this table is ultimately to provide a comprehensive overview of all of the types of data we see collected in the literature and in what types of datasets those variables may exist so that analysts can select parameters for studies appropriately.

TABLE 2: Common Variable Types in Pedestrian and Bicyclist Safety Analysis

Data Type	Category	Specific Variable
Collision Data	Motor Vehicle type	Make, model [14] [24]
Collision Data	Environmental/weather conditions	Weather/climate zone [25], time of day [25], day of week [25], month of year [25], surface condition [25]
Collision Data	Driver characteristics/ demographics	Age, race, ethnicity, sex, collision behaviors [14] [3] [18]
Collision Data	Collision severity	Fatalities, injuries, etc. [22] [3] [25] [26] [19]
Collision Data	Collision type	Single-vehicle, multi-vehicle, vehicle-pedestrian, etc. [21]
Collision Data	Occupant data	Number of occupants [21]
Collision Data	Pedestrian-related factors	Age, race, ethnicity, sex, behavior [21] [24] [19] [18]
Collision Data	Cyclist-related factors	Age, race, ethnicity, sex, behavior [25] [18]
Collision Data	Impairment	Driver impairment, pedestrian impairment, alcohol impairment, drug impairment [21] [25]
Collision Data	Speed	Impact speed [24], posted speed limit [19] [18]
Exposure Data	Motor Vehicle Volume (Counts)	Major road AADT [27] [18] [13], Minor road AADT [18], Segment AADT [18] [12] [15] [13], Truck AADT [13]
Exposure Data	Bicycle Volume (Counts)	Intersection counts [18], segment counts [18], App-based trip estimates [18], AADB [6] [12] [20]
Exposure Data	Pedestrian Volume (Counts)	AADP [6] [12] [20]

Data Type	Category	Specific Variable
Exposure Data	Conflicts*	Time-to-collision [15], Turning counts that conflict [18] [10]
Exposure Data	Trips	Origin-destination data [13]
Spatial/Socio-Demographic/Socio-Economic Data	Exposure estimates	Census tract data [22], population density [23] [18] [13], population count [17]
Spatial/Socio-Demographic/Socio-Economic Data	Feature Density	3-Leg intersection density [23], 4-leg intersection density [18], features that affect level of stress [18]
Spatial/Socio-Demographic/Socio-Economic Data	Development Patterns	Commercial development in surrounding area [18], activity density [22]
Spatial/Socio-Demographic/Socio-Economic Data	Socio-Demographic Data	Proportion of male population [13], proportion of 22-29 aged population [13], proportion of 65+ aged population [13], total population by age distribution [17], median age by gender [17], total population by race [17], educational attainment [17], households at or below poverty level [17], median household income in the past 12 months [17]
Spatial/Socio-Demographic/Socio-Economic Data	Trip and Mode Choice	Mode of transport to work [17], travel time to work [17], housing unit car ownership [17], commute mode shares [22], transit supply [22]
Infrastructure/Land Use	Segment properties	Sidewalk presence/sidewalk length [22] [13] [12], number of lanes [23] [12] [18], two-way left-turn lanes [23] [18], marked midblock crossings [23] [18], on-street parking [23] [18], traffic direction [23], posted speed limit [23], sidewalk condition [12], sidewalk obstructions [12], buffer width [12] [18], buffer type [12], ped/bike signage/warning devices [12], presence of bike lanes [18] [13], width of bike lanes [18], one-way or two-way [18] [12], number of driveways [18]
Infrastructure/Land Use	Land use	Commercial development [22] [13], neighborhood concepts [18], retail density [18], household density [18], household size [18], area designation (urban, residential, recreational, mixed) [13], Schools/education centers in proximity [13] [18], financial centers in proximity [13], numbers of different

Data Type	Category	Specific Variable
		development types (parks, restaurants, hospitals) [13]
Infrastructure/Land Use	Speed	Zonal speed [13]
Infrastructure/Land Use	Road network properties	Proportion of arterial roads [13], proportion of collector roads [13], proportion of local roads [13], Number of flashing beacons signs [13], Number of school signals [13], street centerline [13] [3], alignment and geometry [21] [23] [15] [19] [18], lighting [21] [12] [19] [18], traffic signals [3], functional classification [18], median type [12], median width [12], traveled way width [12], channelized left-turn lanes [12], channelized right-turn lanes [12], ramp data (type, exit or entrance) [12]
Infrastructure/Land Use	Intersection Data	Number of intersections [18], number of legs [23], major road right-turn lanes [23], major road median [23] [12] [18], major road total lanes [18], major road functional classification [23], minor road right-turn lane [23], minor road total lanes [23] [18], crosswalk presence [12], crosswalk type [12], crosswalk color [12], crosswalk condition [12], crossing distance [12], presence of curb ramps [12], presence of truncated domes [12], number of lanes to cross [12] [18], pedestrian signal heads [12], pedestrian signal actuator buttons [12], number of thru lanes [18], 85^{th} percentile speed of street being crossed [18], presence of bike lane at intersection [18], number of lanes to cross to make a right turn [18], number of left-turn lanes [18], number of right-turn lanes [18], presence of green bicycle markings [18]
Infrastructure/Land Use	Transit Facilities	Length of bus lane [13], transit stops [12], number of transit lines [18]

*Conflicts are also used a surrogate for collisions.

Databases Identified in the Literature

Because the end goal of this project is the development of a comprehensive pedestrian and bicycle database, the available literature was scanned to identify any sources of data used for the modelling efforts described previously. The databases listed in Table 3 were found in the literature and are grouped based on whether they contain pedestrian data, bicycle data, or both. Additionally, data components and source locations/contexts are listed.

TABLE 3: Sample Databases in Pedestrian and Bicyclist Safety Analysis

Pedestrian or Bicycle	Database Name	Data Types	Source Locations
Pedestrian	Rutger's Plan 4 Safety [7]	Police-reported collision data	New Jersey
Pedestrian	LADTOD [21]	Police-reported collision data	Louisiana
Pedestrian	Gernab In-Depth Accident Study (GIDAS) [24]	Police-reported collision data	Germany (Hanover and Dresden)
Pedestrian	Pedestrian Crash Data Study (PCDS) [24]	Police-reported collision data	Six sites across United States
Pedestrian	State DOT infrastructure databases [12]	Infrastructure and roadway data	Washington, New Jersey, Maryland
Pedestrian	Local jurisdiction databases in California and other states [12]	Sidewalk inventories	Rancho Cucamonga, Berkeley, Sacramento County, Oakland, Marina, Rockville MD, Alexandria VA, Piedmont Triad Rural Counties NC, Tucson AZ, Asheville NC, Portland OR, Lexington MA
Pedestrian	Malaysian Institute of Road Safety Research (MIROS) Road Accident Analysis and Database System [19]	Police-reported collision data, traffic volume data	Malaysia
Pedestrian	Strategic Highway Research Program 2 (SHRP2) Database [28]	Naturalist driving data and road inventory data	Intersections in Washington and Florida
Bicycle	European Injury Database (IDB) [9]	Police-reported collision data, hospital injury surveillance data	Nine EU members
Bicycle	Road Traffic Safety Directorate Database [25]	Police-reported collision data, insurance data (for injury surveillance)	Latvia

Pedestrian or Bicycle	Database Name	Data Types	Source Locations
Pedestrian and Bicycle	SafetyNet Accident Causation Database [14]	Police-reported collision data	Germany, Italy, the Netherlands, Finland, Sweden, United Kingdom
Pedestrian and Bicycle	Crash Risk Scoring Tool [23] [18]	Combines roadway geometry data from Google Earth, built environment data from the Environmental Protection Agency (EPA) Smart Location Database, traffic volumes from the Oregon Department of Transportation (ODOT) ATR stations, and bicycle volume data from Strava	Random roadway sample throughout Oregon
Pedestrian and Bicycle	Community database on Accidents on the Roads in Europe (CARE) [29]	Police-reported collision data	EU member states
Pedestrian and Bicycle	Swedish Transport Administration (STA) Fatal Crash Database [15]	Police-reported collision data	Sweden
Pedestrian and Bicycle	Road Accident Sampling System-India (RASSI) [26]	Police-reported collision data	Sampled data from India similar to NASS-CDS or GIDAS
Pedestrian and Bicycle	Highway Safety Research Center Ped/Bike Cost Database [30]	Infrastructure cost data	Bidding sheets and cost summaries from state DOTs
Pedestrian and Bicycle	A Safety Perception Indicator for Vulnerable Road Users in Urban Environments (ASPIRE) [4]	Equity-based risk data	Based in Canada, but little detail is available
Pedestrian and Bicycle	ECO-Visio [6]	Count data for pedestrians and bicyclists	Counting locations across United States
Pedestrian and Bicycle	Bike-Ped Portal [6]	Count data, segment/roadway data, infrastructure data, directional flow data, and detector data	National clearinghouse

Pedestrian or Bicycle	Database Name	Data Types	Source Locations
Pedestrian and Bicycle	Various State, City, County, and MPO Datasets [6]	Intersection data and count data (many of these data were added to Bike-Ped Portal)	City of Portland/Metro, TriMet, Washington Park, Washington County, Mid-Willamette Valley Council of Governments, Lane Council of Governments, Deschutes County, Olympia, Seattle, Washington State DOT, Texas, Austin, Colorado, Boulder, Virginia, Arlington, Minnesota, University of Minnesota, Minnesota DOT, Minneapolis, California, San Diego
Pedestrian and Bicycle	Traffic Accident Surveillance and Analysis System – Transportation System Network (TASAS-TSN) [12] (Caltrans internal database – not publicly available)	Police-reported collision data, highway inventory database, traffic volumes, and a traffic investigation reporting tracking system	California highway system
Pedestrian and Bicycle	Pedestrian and Bicycle Safety Decision Support Tool [17]	Police-reported collision data, socioeconomic data, demographic data	Tennessee
Pedestrian and Bicycle	National Household Travel Survey [13]	Trip demand data combined with US Census Tiger/Line Data and an FDOT data repository (for [13])	National trip data (for NHTS) combined with Central Florida
Pedestrian and Bicycle	TransBASESF.org [10]	Geospatial data, collision data, injury data, infrastructure data, development type data, demographic data, zoning data	San Francisco
Pedestrian and Bicycle	Database for Active Transportation Infrastructure and Volume [31]	Infrastructure data, pedestrian volume data, bicycle volume data	Database framework developed in California but applicable to other sites
Pedestrian and Bicycle	Statewide Integrated Traffic Records System	Police-reported collision data	California

Pedestrian or Bicycle	Database Name	Data Types	Source Locations
Pedestrian and Bicycle	Fervor [32]	Uses transportation incident data from the Maryland State Highway Administration's CHART program	Maryland

Interviews

Eight leading pedestrian and bicycle safety researchers were interviewed in order to understand what data safety researchers need, where they find this data, what data problems they have and how the Clearinghouse might better meet their data needs. Those interviewed included Asad Khattak (UTK), Offer Grembek and Aditya Medury (UCB), Noreen McDonald (UNC), Chris Cherry (UTK), Robert Schneider (University of Wisconsin, Milwaukee), Chris Monsere (Portland State University), Daniel Carter (HSRC). The interviews asked each participant the same eight questions. Below is a summary of the information shared by the interviewees, organized by interview question.

Synthesized Interview Results

1. **What resources do you typically use when conducting a pedestrian and/or bicyclist safety analysis?**
   - Collision/injury data: HSIS, FARS, state police-reported collision data, police reports with narrative, state injury databases, EMS records, lighting/weather/roadway surface data
   - Contextual data: stress surveys, Walkscore, American Time Use Survey (ATUS), perception of risk
   - Exposure/operations data: NHTS, any existing volume data by mode (AADP, AADB, AADT), crossing counts/turning movements, trajectory data, video for conflict analysis
   - Infrastructure/built-environment data: Road inventory by county, roadway data from cities or MPOs, design schematics, geometric data, operational data, pedestrian and bicycle infrastructure data, speed limits, sidewalks, signage, treatment installation dates, lighting, land use data, landscaping data, street furniture installations
   - Behavior data: interaction types, actual speeds, compliance, communication between road users
   - Socio-demographics data: Census data, demographics, vehicle ownership, income, population, employment density
   - Supplemental/Proprietary data: Strava, raw video feeds (for individual spots), Big Data from apps and bike share programs

2. **What data do you need for pedestrian and/or bicyclist safety analysis?**
   - Collision/injury data: severity, collision report narratives/diagrams, collision data
   - Exposure/operations data: turning movement counts, better pedestrian and bicycle exposure, AADT linked to GIS, intersection counts
   - Infrastructure/built-environment data: crosswalk level volumes, roadway centerline in GIS, posted speed limit, curb radius, curb extension presence, land use data, signal timing/type data
   - Socio-demographics data: socio-economic census data
   - Transit data: transit stops with reference sites
   - Supplemental/Proprietary data: CAV data, app data
   - Comparative and quality data: needs for data to be standardized across city/county/state, reliable and consistent counts, reliable sidewalk data, time of installation and accurate facility data (e.g. bike lane width)

3. **What data are missing that you'd like to have?**
   - Collision/injury data: EMS/ER data for accurate injury data, collision typing
   - Contextual data: perception of risk
   - Exposure/operations data: exposure data, naturalistic data for specific pieces of infrastructure, video feeds of pedestrians, motor vehicle AADT, turning movements at all intersections
   - Infrastructure/built-environment data: sidewalk, shoulder width, signal plans (e.g. LPI locations), curb radius, medians, curb extensions, on-street parking, signage, traffic calming devices, types of bike and ped facilities, installation dates of treatments, land use context for collisions
   - Behavior data: interactions between cars and bike lanes, traffic speeds on all roads, citation level enforcement data
   - Socio-demographics data: bike use among different socio-economic strata, road user demographics, demographics by census tract
   - Transit data: precise location of bus stops

- Supplemental/Proprietary data: CAV – how often are they not detecting peds, and in what situations
- Comparative and quality data: all data needs more records and more detail, better police-reported collision data, better data linkage

4. What percent of your project budget goes to data collection?

- 10%-30% when data need to be collected
- 10% if relying on data already collected
- Nominal if relying on secondary data
- Less than 10% if exposure data are provided, up to 60% for a continuous video project
- 20%-30% for a survey study
- 50%-75% if relying on Google street view for data collection plus 25% for additional bike/ped counts
- 10% if hiring a firm to collect data
- Most clients don't budget for data processing

5. What percent of your project budget goes to cleaning or processing data for analysis? How long does it take?

- 10%-20% if team collected data; double whatever time is expected
- 5%-10% for FARS data
- 60%-70% when relying on secondary data
- If linking app data to GPS and other sources, months
- If video data, about 1 hour per day of data
- 5% of data collection budget
- 0 if data provided by others
- 25%-30% for data cleaning, coding, and compiling

6. What features do you want from a data clearinghouse?

- Data elements
    1. Collision/injury data: pedestrians and bike collisions with and without injury linked to hospital records, de-anonymized hospital data, health data
    2. Contextual data: weather, temporal data (years, hours, seasons, days collected)
    3. Exposure/operations data: exposure, travel surveys other than NHTS, count data
    4. Infrastructure/built-environment data: infrastructure needs/accessibility, roadway geometry, installation dates and ownership, lighting, curvature, number of legs, intersection control, curb cuts, crosswalks, bike approaches, sidewalks, on-street parking, land use, street buffering
    5. Behavior data: speed data, perception data
    6. Socio-demographics data: socio-demographics by region/area, census, TAZs, population, employment density, mode share
    7. Transit data: transit usage
- Features:
    1. Usage: simple cross tab analysis tools, download to CSV at resolution chosen by researcher, data visualization, glossary, uploading and sharing capabilities, mapping capabilities, analysis tools, categorization and classification schemes
    2. Data contents: availability information, timeframe of data, data availability at different geographic levels, ownership information for facilities, data features unique to bikes and pedestrians involved in collisions
    3. Data connections: links to publications that use the data, integration with other road users, links to Scalable Risk Assessment (ScRAM) tool [8], linkage to city files, linkage to CMF Clearinghouse
    4. Quality: consistency, standardization, appropriate scale, ability to tailor clearinghouse to user, high-quality, ease of navigation, high-quality, geometry should be easy to access,
    5. Administrative: documentation of changes/cleaning to data

7. How might a pedestrian and/or bicyclist safety data clearinghouse help you?

Interview participants highlighted the potential usefulness of a clearinghouse for comparisons across space and time. Including a variety of geographic scales would enable comparisons between regions or jurisdictions, and including data that extends for a long period would enable better before and after evaluations. The clearinghouse could also facilitate better cross-sectional studies. A data clearinghouse would be useful to participants in facilitating project planning and operations, ideally reducing time and budget spent on data collection and cleaning. Interviewees valued the potential of a clearinghouse as a central space for bike share and exposure data, as well as allowing researchers to share their own data. The clearinghouse could also offer data packages to researchers. Other interview participants expressed interest in the clearinghouse to facilitate surveys of facility owners. In addition, some attractive aspects of a clearinghouse are data standardization and the usage of big data for machine learning.

8. Do you know of online datasets we should be sure to link to the clearinghouse?

- Fatality Analysis Reporting System (FARS)
- Highway Safety Information System (HSIS)
- National Automotive Sampling System (NASS)
- State collision data
- National Household Travel Survey (NHTS)
- State hospital and injury records (CDC or NIH)
- [Second] Strategic Highway Research Program Naturalistic Driving Study (SHRP2 NDS) data
- Existing national/state bike/ped surveys
- University researcher datasets (e.g. UCONN)
- Google Earth image repositories
- Census data
- Bikeshare
- Metropolitan Travel Survey Archive
- Travel Monitoring Analysis System (TMAS)
- Bike-Ped Portal
- Traffic Incident Management Systems
- Strava and other bike trip apps
- Crash Modification Factors (CMF) Clearinghouse
- Open Street map

Inventory

During the summer of 2018, students at University of California, Berkeley conducted an exhaustive web search of pedestrian and bicycle safety related data. The search focused on publicly available data or data that could be accessed with a relatively easy sign-up process. In addition, national datasets of importance to pedestrian and bicycle safety were also included, even if access was not readily available on the web. The search was conducted nationwide, targeting the following geographies: national data, data for all 50 states, all metropolitan planning organizations (MPOs), all cities with population greater than 100,000, and a limited number of counties.

The search yielded 4,126 datasets. Here we use the term dataset, rather than database, since some of the datasets found were not databases, but various collections of online information from static maps to extensive searchable online databases. The final version of the dataset archived in the CSCRS Dataverse contains approximately 4,170 records, about 40 more records than in the data discussed in this chapter. This is because additional data cleaning was conducted, and additional records were added after the summary tables were created. The Clearinghouse is a dynamic list, meant to be updated and improved with time. The team expects further improvements to the data in the online Clearinghouse in the next phase of work.

After grouping the data by geographic location, we catalogued the data by a variety of data components to accurately portray the content and availability of each source. The data categories we used for categorization, as well as the general types under each category, include:

- Category
  - Collisions
  - Counts
  - Infrastructure
- Source Name
- Geographic Scale
  - City
  - County
  - Region
  - State
  - National
- MPO Name
- Agency/Owner
- Availability
  - Publicly available
  - By request
  - Account needed to access
  - Access restricted
- Format
  - Non-static
    - Excel spreadsheet
    - GIS tool
    - HTML site
    - Other
  - Static
    - PDF
    - Map
- Date Ranges of Availability
- Time Period of Data Collected

Each of the primary data categories is composed of distinct elements. While these are simple in some cases (i.e. we categorized the count data based on whether the data were bike counts, pedestrian counts, or traffic volumes), others were more complex. A wide variety of infrastructure data are publicly available, but these data range significantly in terminology and completeness. We generally grouped infrastructure data by bicycle facilities, pedestrian facilities, street network or centerline data, and other. However, within each of these subcategories is a wide range of variation. Because these items are so diverse (for example, one dataset was denoted as containing alley data while another categorized the data as "bicycle green wave," indicating that the dataset had data on signal progression timing set to bicyclist speed instead of motor vehicle speed), a full list of infrastructure types is beyond the scope of this section. Instead, we discuss this variety in more depth as a limitation of our inventory.

Tables 4-6 below list samples available in different data categories. As can be seen in Table 4, the number of state (and territory) datasets ranged from 2 (in Puerto Rico) to 430 in California. Table 5 shows that there are more non-static datasets (2,316) than static datasets (1,810) in the inventory. Table 6 shows that the majority of datasets (4,108) included in the inventory are publicly accessible.

TABLE 4: Number of Datasets per State in the Inventory

State	Number of Datasets
Alabama	24
Alaska	12
Arizona	276
Arkansas	231
California	430
Colorado	92
Connecticut	81
Delaware	66
District of Columbia	116
Florida	130
Georgia	31
Hawaii	38
Idaho	32
Illinois	81
Indiana	65
Iowa	64
Kansas	31
Kentucky	33
Louisiana	52
Maine	25
Maryland	47
Massachusetts	69
Michigan	70
Minnesota	44
Mississippi	18
Missouri	25
Montana	22
Nebraska	19
Nevada	41
New Hampshire	21
New Jersey	27
New Mexico	53
New York	143
North Carolina	119
North Dakota	13
Ohio	47
Oklahoma	12
Oregon	312
Pennsylvania	119
Puerto Rico	2
Rhode Island	15
South Carolina	44
South Dakota	8
Tennessee	24

State	Number of Datasets
Texas	457
Utah	107
Vermont	46
Virginia	43
Washington	82
West Virginia	15
Wisconsin	67
Wyoming	3
National	89
Multiple states	4
Total	**4,126**

TABLE 5: Number of Datasets per File Format

File Format	Number of Datasets
Non-static	**2,316**
Spreadsheet	1,214
GIS	1,076
Other	19
Html	7
Static	**1,810**
PDF	1,727
Map	83
Total	**4,126**

TABLE 6: Number of Datasets by Availability

Availability	Number of Datasets
Publicly available	4,108
By request	11
Account needed to access	4
Access restricted	3
Total	**4,126**

Figures 1 and 2 below provide more context for the data distribution in the inventory. Figure 1 shows the geographic composition of the datasets by different localities. The majority of datasets contained city data, while there were fewer national datasets than any other type. Note that because some datasets contain data for multiple localities, the sum of the datasets shown in the figure is greater than the stated total of 4,126. Figure 2 shows the number of datasets in each of the main three data categories. There are more count datasets than any other type, but the sum of the datasets again does not match the stated total of 4,126 due to some overlap.

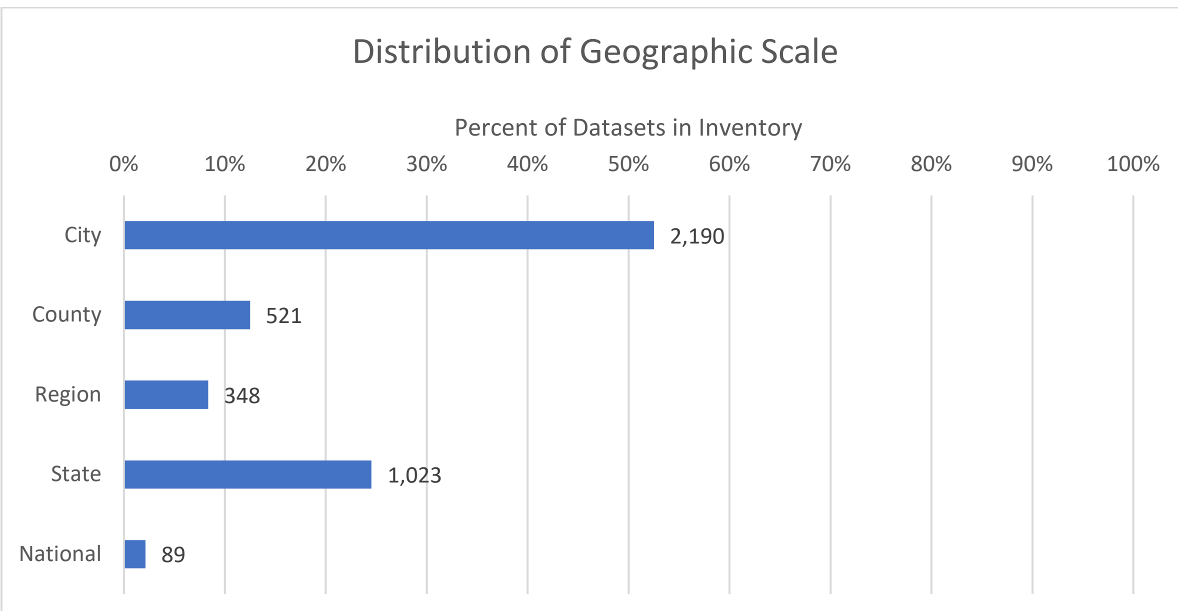

Figure 1: Distribution of datasets by geographic scale

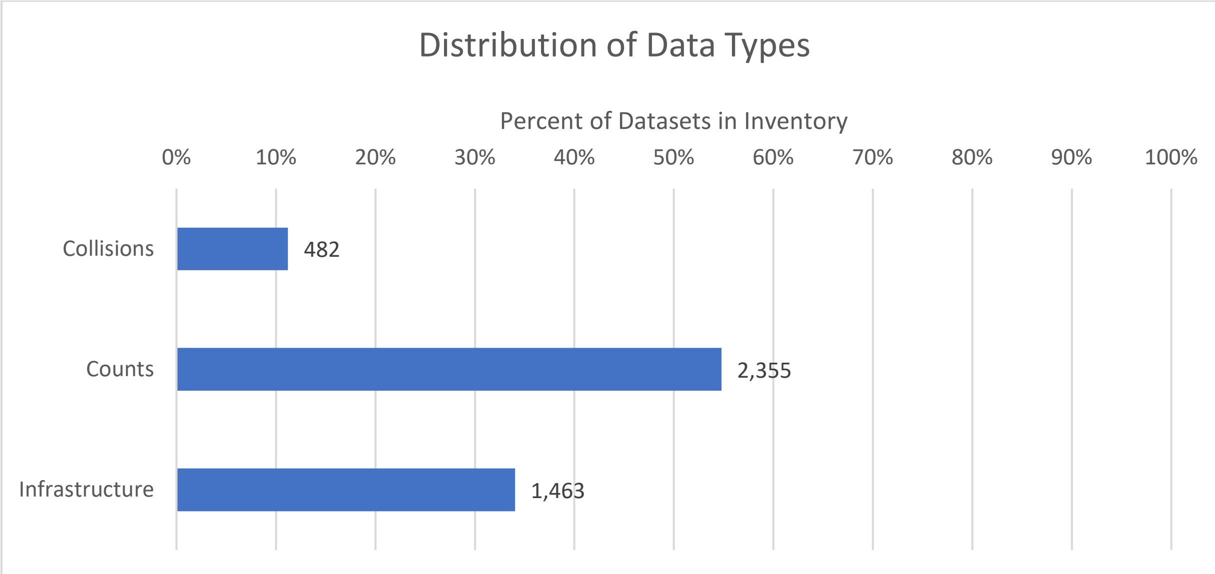

Figure 2: Distribution of datasets by primary data content

Rating of Datasets

Not all datasets are equal in quality. Some datasets have greater specificity; others have greater spatial coverage; and still others provide general information summarized at the annual or city level. Because the Clearinghouse is geared toward safety researchers, there is a need for correct and precise data with high specificity. With this in mind, the project team is proposing a five-star rating system described in this chapter. The rating system will help researchers quickly identify datasets with temporal completeness, geographic detail and ability to link to other datasets to be useful in their research.

Every dataset was classified as having information concerning collisions, counts (AADT counts, short duration counts, permanent counts), or infrastructure. Collision data are further divided into individual collision events or summarized collision data. Some datasets had multiple types of data: for example, some infrastructure data of road centerlines had attached AADT count value data. These types of datasets were evaluated separately for each data type. The team chose to exclusively rate datasets at the state level, and datasets in PDF formats were excluded, because such data are hard to use for analysis without spending time translating them into a different format.

The team rated each non-PDF, statewide dataset based on three components: temporal completeness, spatial completeness, and linkability. Temporal completeness was rated based on availability of time or date and the duration of data collection. Spatial completeness refers to the extent of the data, either based on elements per square mile or types of roadway covered. Linkability refers to the degree to which a dataset can be connected to geography, in formats of GIS systems or coordinate systems. Each type of dataset has qualifiers specific to the type of data's content.

The rating for each component (temporal completeness, spatial completeness, and linkability) were weighted and averaged to calculate the final score. For example, the temporal completeness of collision events averages together numeric scores for the availability of time (does the dataset include the exact time of day when each collision occurred?), availability of date (does the dataset include the date on which each collision?), and temporal duration of data (over what length of time does the dataset record crashes? Over one year? Five years?). For the final score, temporal completeness and spatial completeness were weighted with a value of 1.0, while linkability was weighted 0.8 as listed in Table 7. This weighting is based on the assumption that temporal and spatial completeness are more important than linkability, but different weights could easily be used in the future, if desired.

TABLE 7: Component Weightings

Component	Weighting
Temporal Completeness	1.0
Spatial Completeness	1.0
Linkability	0.8

The star value assigns a number of stars to a dataset based on its final score, with five stars being the best and one star being the worst. The star ratings are categorized in Table 8. While the final score is more accurate in analyzing the overall quality of a dataset, the star rating enables easier comparison across the various types of data.

TABLE 8: Score Ranges for Inventory Rankings

Final Score Range	Star Rating
0 - 0.19	1 Star
0.2 - 0.39	2 Stars
0.4 - 0.59	3 Stars
0.6 - 0.79	4 Stars
0.8 - 1.0	5 Stars

Collisions

Collision data were divided into two types: event data and summarized data. Each were rated separately due to their different characteristics.

Collision Events

Collision Event data concerns motor vehicle collisions with bicycles and pedestrians and have details about each collision. Collision severity levels recorded varies from dataset to dataset. Collision Events datasets were categorized by roadway type: “All” refers to federal interstates, state highways, and local roads; “All state and city” refers to state highways and local roads; and “Only state” refers to only state highways. Roadway type was categorized by mapping the collision data on a base map and seeing if collisions were on all three types of roadways. Some datasets explained the included roadway type. Intersection data had to be explicitly labeled as such, since collision events that list two streets do not necessarily mean that the collision occurred at an intersection. It may just indicate the nearest intersection to the collision. The rating system is summarized in Table 9.

TABLE 9: Rating Values Corresponding to Completeness and Linkability of Collision Event Data

Components	Criteria	Collision Event Rating Values
Temporal Completeness	Date	Date provided = 1.0 No date = 0.0
	Time	Time provided = 1.0 No time = 0.0
	Duration of data	>5 years data = 1 5-2 years data = 0.8 2-1 years data = 0.6 1 full year data =0.5 <1 year data = 0.2
Spatial Completeness	Roadway type	All = 1.0 All state and city = 0.8 Only state = 0.5
	Intersection vs non-intersection	Listed at intersection = 1.0 Not listed = 0.0
Linkability	Linkage code	Linkage code or data linked to roadway file = 1.0 No linkage code = 0.0
	Latitude & Longitude	Lat/Long provided = 1.0 No Lat/Long = 0.0
	Collision/hospital report	Collision report = 0.5 Hospital record = 0.5 Both collision and hospital report = 1.0 None = 0.0

Availability of a linkage code refers to whether the data was graphed in GIS or another interactive platform. If the count data was linked to a spatial file, it received a 1.0 for linkage code. If data has listed coordinate points or is attached to a spatial file, it received 1.0 for latitude & longitude.

Data connected to a collision report or hospital report was highly valued. Collision report narrative data could be connected with an ID, or the collision report may be included in the dataset itself.

One hundred datasets contained collision event data. One dataset was assigned two stars, 42 were assigned three stars, 40 were assigned four stars, and 17 were assigned five stars.

Collision Summaries

Collision Summary data has no individual collision data, but has overall data about collisions in a location. Some datasets include spatial data with the total number of collisions at various locations [33]. Others total the number of collisions based on counties or highways [34]. The inconsistencies of these datasets reduced the number of rating aspects, but temporal duration of data, roadway types, and coordinate linkability are included. Some datasets had related shapefiles, but this aspect was not rated. The rating system is summarized in Table 10.

TABLE 10: Rating Values Corresponding to Completeness and Linkability of Collision Summaries Data

Components	Criteria	Collision Summaries Rating Values
Temporal Completeness	Duration of data	>5 years data = 1.0 5-2 years data = 0.8 2-1 years data = 0.6 1 full year data = 0.5 <1 year data = 0.2
Spatial Completeness	Roadway types	All = 1.0 All state and city = 0.8 Only state = 0.5
Linkability	Latitude & Longitude	Lat/Long provided = 1.0 No Lat/Long = 0.0

Of the 13 datasets that had collision summary data, 3 datasets were assigned three stars, 7 were assigned four stars, and 3 were assigned five stars.

Counts

Count data includes traffic volume for motor vehicles, pedestrians, and bicycles. These data were classified into three types: Annual Average Daily Traffic (AADT), short duration counts (pedestrians and bike only), and permanent counts (pedestrians and bike only). In regard to spatial completeness, square mileage of each state includes solely land area (not water area) and was gathered from the United States Census. The number of count locations was estimated by examining the dataset itself, if the description of the dataset did not include this information. Availability of a linkage code refers to whether the data was graphed in GIS or another interactive platform. If the count data was linked to a spatial file, it received a 1.0 for linkage code. If data has listed coordinate points or is attached to a spatial file, it received 1.0 for latitude & longitude.

AADT Counts

The Annual Average Daily Traffic (AADT) Counts encompass both automotive and bicycle traffic. The data are primarily connected with roadway segments. The duration of data for AADT counts assumes that, unless specified, each year has a whole 365 days' worth of data. Assuming that AADT values should be collected everywhere, the location per square mile value has a higher threshold than that of the permanent count data. Some counts are only Average Daily Traffic (not annualized). Table 11 summarizes the rating system.

TABLE 11: Rating Values Corresponding to AADT Count Data

Components	Criteria	AADT Count Rating Values
Temporal Completeness	Duration of data	>5 years data = 1.0 5-2 years data = 0.8 2-1 years data = 0.6 1 full year data = 0.5 <1 year data = 0.2
Spatial Completeness	Count locations per square mile	>1.0 loc./sq. mi = 1.0 1.0 - 0.2 loc./sq. mi = 0.66 <0.2 loc./sq. mi = 0.33
Linkability	Linkage code	Linkage code or data linked to roadway file = 1.0 No linkage code = 0.0
	Latitude & Longitude	Lat/Long provided = 1.0 No Lat/Long = 0.0

We ranked 225 state-level datasets that included AADT data. Fifteen datasets were assigned two stars, 78 were assigned three stars, 100 were assigned four stars, and 62 were assigned five stars.

Short Duration Counts

Short Duration Counts include both pedestrian and bicycle traffic. The resolution specifies the time period into which total counts were aggregated. The duration per count site per year lists the length of time that data was collected for each site annually, though the overall temporal duration of data was not rated. Because it is more important for short duration counts to be spatially diverse, we decided to qualify a smaller count site per square mile as to have a higher rank. The rating system is summarized in Table 12.

TABLE 12: Rating Values Corresponding to Completeness and Linkability of Short Duration Counts

Components	Criteria	Short Duration Count Rating Values
Temporal Completeness	Resolution	15 min or per vehicle record = 1.0 1 hr =0.8 2 hr = 0.6 ≥3 hrs = 0.4
	Duration per Count Site per Year	≥1 week = 1.0 2 to 3 days = 0.8 23-25 hrs = 0.6 9-22 hrs = 0.5 2-8 hrs= 0.4 ≤2 hrs = 0.2
Spatial Completeness	Count locations per square mile	>1.0 loc./sq. mi = 1.0 1.0 - 0.2 loc./sq. mi = 0.66 <0.2 loc./sq. mi = 0.33
Linkability	Linkage code	Linkage code or data linked to roadway file = 1.0 No linkage code = 0.0
	Latitude & Longitude	Lat/Long provided = 1.0 No Lat/Long = 0.0

In total, 16 datasets contained short duration counts. One dataset was assigned two stars, 8 were assigned three stars, 6 were assigned 4 stars, and one was assigned five stars.

Permanent Counts

Permanent counts constantly measure pedestrian and cyclist traffic for an extended period of time. These data are often collected from video or automated counters, such as in-ground loop detectors. The resolution dictates the time period into which total counts were aggregated. The duration of data was based on the time period over which the data are available. Count locations per square mile need not be very high because the purpose of the data is to track change overtime, not space. For this reason, the threshold to meet a "medium" location per square mile ranking is relatively lower than that of short duration counts. The rating system is summarized in Table 13.

TABLE 13: Rating Values Corresponding to Completeness and Linkability of Permanent Counts

Components	Criteria	Permanent Counts Rating Values
Temporal Completeness	Resolution	15 min or per vehicle record = 1.0 1 hr = 0.8 one day = 0.5 one month = 0.2 one year = 0.1
	Duration of Data	>5 years = 1 5-2 years = 0.8 2-1 years = 0.6 1 full year =0.5 <1 year = 0.2
Spatial Completeness	Count locations per square mile	>0.1 = 1.0; 0.1 - 0.01 = 0.66 <0.01 = 0.33
Linkability	Linkage code	Linkage code or data linked to roadway file = 1.0 No linkage code = 0.0
	Latitude & Longitude	Lat/Long provided = 1.0 No Lat/Long = 0.0

In total, 10 datasets included permanent counts. Four datasets were assigned three stars, 4 were assigned four stars, and two were assigned five stars.

Infrastructure

Infrastructure refers to a variety of roadway elements, signals/signs, pedestrian facilities, and bike facilities. The installation date or time was relatively rare. Some datasets had interactive mapping components or raw data that listed the number of locations. Description of number of locations was included in the rating but other information, such as the type of roadway ("state highways"), ownership ("all devices that are owned by NYSDOT"), or specific location ("University of Illinois campus") was not included in the rating.

Four infrastructure element types are included in the basic facility completeness: centerlines, sidewalks, paths, and bike facilities. Centerlines refer to roadways. Sidewalks refer to footpaths for pedestrians along roadways. Paths refer to off-road trails often used by pedestrians, cyclists, horse riders, and skaters [35]. Bike facilities include bike lanes, bike routes, and shared use paths. Sidewalks and paths are prioritized because they are important for safety and less common in infrastructure datasets.

Availability of a linkage code refers to if the data were mapped in GIS or another interactive platform. If the count data was linked to a spatial file, it received a 1.0 for linkage code. Data was categorized as point or line data for future reference. If data has listed coordinate points or is attached to a spatial file, it received 1.0 for latitude & longitude. The rating system is summarized in Table 14.

TABLE 14 Rating Values Corresponding to Completeness and Linkability of Infrastructure

Components	Criteria	Infrastructure Rating Values
Temporal Completeness	Installation date	Has installation date = 1.0 Has installation year = 0.8 No data = 0.0
Spatial Completeness	Number of locations	Not rated
	Basic Facility Completeness	Centerline = 0.2 Sidewalks = 0.3 Paths = 0.3 Bike facilities = 0.2
Linkability	Linkage code	Linkage code or data linked to roadway file = 1.0 No linkage code = 0.0
	Point or line	Point Line Both None
	Latitude & Longitude	Lat/Long provided = 1.0 No Lat/Long = 0.0

Out of a total of 307 infrastructure datasets, 9 were assigned one star, 230 were assigned two stars, 41 were assigned three stars, 20 were assigned four stars, and 7 were assigned five stars.

Summary

Table 15 summarizes the rating values for the sample of evaluated datasets. Additional detailed breakdown of the star ratings by state for each category are contained in tables in the Appendix.

The data category with the highest number of datasets was Infrastructure, but this category also had the highest number of 2-Star datasets. The highest number of 5-Star datasets corresponded to the AADT Counts. The category with the lowest number of datasets was the Pedestrian and Bicycle Permanent Count Data category with only 10 datasets total most of which were rated 3 and 4 star. The Pedestrian and Bicycle Short Duration Count data category was similarly low. This shows need for more such data in collection in the future. The collision summary category was also lacking, but since the collision events were better documented with 100 datasets, most in the 3 and 4 star rating and because summary data can be created from these, lack of summary data should not be an issue.

Altogether, these rating summaries indicate gaps in count data and a general lack of quality and consistency for all data types. Next steps for this work may include rating the city, county, regional and national datasets.

TABLE 15: Star Rating Count by Type of Data

	1 Star	2 Stars	3 Stars	4 Stars	5 Stars	Total Number of Datasets
Collisions						
Collision Event	0	1	42	40	17	**100**

Collision Summary	0	0	3	7	3	**13**
Counts						
Ped/Bike Short Duration Counts	0	1	8	6	1	**16**
Ped/Bike Permanent Counts	0	0	4	4	2	**10**
AADT Counts	0	15	78	100	62	**255**
Infrastructure	9	230	41	20	7	**307**
Total	**9**	**247**	**176**	**177**	**92**	**701**

Data Gaps

One of the aims of this project was to identify gaps in available data that researchers seeking to conduct pedestrian and bicyclist safety projects use. To accomplish this goal, expert researchers were interviewed from a variety of fields, including civil engineering, planning, and public health, synthesized literature specific to pedestrian and bicyclist databases. The intent was to determine what was missing, inventory existing pedestrian and bicycle safety related datasets, and rate the state-wide datasets for quality. Specifically, the expert panelists were asked what data are missing that they would like to have for their projects and what data they would like to see in a data clearinghouse. General data types were grouped together which are commonly seen in the literature, which was compared against the expert responses with the goal of revealing missing data types. It was noted what was and was not available in the inventory.

Generally, the expert-identified missing data types can be categorized as:

- Detailed injury and severity data, vulnerable road user perceptual data (e.g. level of stress)
- Better exposure data for every mode, project-level infrastructure data
- Vulnerable road user behavior data (e.g. speed measurements and routing)
- Sociodemographic data beyond estimates from census blocks
- Transit data
- Data regarding emerging technologies and trip patterns

One consistent theme that ran through all these responses was the need for quality data in standardized formats. While many of these data do exist in piecemeal fashion, they are often based on model estimates or grafted together from Google Street View surveys. For example, HSRC researchers, in cooperation with one expert panel member and consultants, recently completed a systemic pedestrian safety analysis for NCHRP Report 893, Systemic Pedestrian Safety Analysis. This project leveraged collision data from the City of Seattle, using count data on a sample of the roads to estimate pedestrian, bicycle, and motor vehicle traffic volumes on each roadway segment. Additionally, Census Bureau estimates were used in conjunction with Google Lighting and Transit data to improve the estimate of pedestrian, bicycle, and motor vehicle traffic volumes. Although the amount of data collected was both substantial and comprehensive, many of these data elements lacked actual measurements that may have improved the quality of the collision models. Additionally, some desired elements, like speed measurements, were simply unavailable.

The literature analysis revealed that projects routinely use piecemeal data that is usually compiled on a project-to-project basis and rarely transferred from principle investigator to principle investigator. While researchers consistently collect the types of data identified by our expert panel, there are known limitations to what is available.

First, Wang et al. noted in 2018 that exposure data are a critical component to vulnerable road user safety studies that is unfortunately often missing. Some researchers have attempted to use alternatives or surrogate measures, such as census tract data [22], filmed interaction data [15], naturalistic driving data [28], population estimates [17], and trip demand [13], but these surrogate measures can also suffer from inaccuracy and must be collected on a project-to-project basis.

Second, the ad hoc approach to data collection makes linking data difficult [7]. For example, injury data may be isolated behind privacy agreements or other legal barriers [12], and exposure data may be retained by a State or City agency but not shared with other entities [10]. These problems can make it difficult to accurately capture the risk factors specific to a particular roadway.

Third, as mentioned in the expert panel, data, even when available, are often inaccurate [7] or inconsistent [6]. Agencies may not regularly update a dataset, or those data may lack sufficient metadata (data about the data) to describe functional applicability [10]. This problem is particularly pronounced for infrastructure data, because records are not well-kept of when projects and countermeasures were implemented [12].

Finally, the cost of data collection and existing funding structures may make it simply impractical to collect all of the potential explanatory data to identify risks. Federal funding typically favors motor vehicle infrastructure, so data on pedestrian and bicycle infrastructure is rarely prioritized. The other issues mentioned can also make it time-consuming and costly to clean and prepare data for use [12]. For these and other reasons, fundamental data like exposure, may simply be excluded from a pedestrian or bicycle safety project [21]. Furthermore, this exclusion only makes it more difficult to perform the rigorous type of evaluation necessary to measure safety for these road users and to determine the effect of post-collision interventions [15].

Many of these discussed issues and gaps were also evident in the datasets as the data inventory was compiled. It can be difficult to find count data for same location as the collision data in order to calculate relative risk and to identify countermeasures that work in multiple locations. Metadata relating to date and available time period for each dataset are frequently missing, and datasets are often difficult to link to other datasets.

Compounding these issues are the limitations revealed by the ratings discussion. In every category that were rated, there are more low to mid quality datasets than high quality datasets. There were less 5-star datasets than 3-star datasets for AADT Counts, Ped/Bike Short Duration Counts, Ped/Bike Permanent Counts, and Collision Events. Even more problematically, there were far more 2-star infrastructure datasets than any other rating, leading to a greater total number of 2-star datasets than any other rating. Considering the scarcity of pedestrian and bicycle counts (both short and permanent) and collision summaries, the low ranking of all datasets indicates a wide-ranging lack of quality and complete data. Strange fields in the data themselves, including inconsistent naming conventions for infrastructure, only makes comparing risks and identifying cross-jurisdictional treatments more difficult. Comprehensively the data inventory, though expansive, is generally incomplete and inadequate.

Taken altogether, the concerns identified in the literature verify the limitations expressed by the experts interviewed and explain the general lack of quality that was found in the compiled inventory. While Federal datasets do exist – namely FARS for fatal collision data, HPMS for motor vehicle estimates, and Census Bureau data for population estimates – there are few useful national sources of facility level data needed for pedestrian and bicycle safety analysis. Researchers need linked and accurate exposure and injury data at the State and City level, ideally with spatial data to contextualize risk factors and how countermeasures affect those risks. While many city datasets do exist, these data are not easily linkable and contain inconsistent data fields; most problematically, there is a significant gap in available count data. The expert panel of interviewees expressed that a data clearinghouse that overcomes some of these gaps would be extremely beneficial. A central data clearinghouse may enable researchers to quickly find exposure and collision data in jurisdictions they are analyzing in neatly downloadable packages. It could also serve as a central catalogue of known pedestrian and bicycle projects so that the true effects of interventions can be measured. By identifying these gaps and highlighting the potential of a data clearinghouse, we hope to work toward a resource that circumvents these limitations.

Below is a list of the data needs and gaps by priority:

1. Pedestrian and Bicycle Count Data provides a fundamental metric of exposure. Both high quality Short Duration and Permanent count data are needed, since so little is currently available on exposure.
2. The ability to link datasets needs to be improved.
3. The accuracy of data needs to be better assessed and improved.
4. Funding for data collection and management for pedestrian and bicycle safety related data needs to increase.

Framework and Website Development

In order to share the information gathered in this project, the team has made the inventory of online datasets available on a new website called the CSCRS National Pedestrian and Bicycle Safety Data Clearinghouse, which can be found at www.pedbikedata.org. This public website allows the user to search the list of pedestrian and bicycle safety data by data type, mode, availability, format and geographic scale. The user can then download the resulting filtered listing of datasets with their name, URL and other associated information.

The list of datasets can be filtered by a variety of categories, as shown in Figure 3. Users initially choose the type of data they are interested in: counts, collisions, or infrastructure. Counts can be narrowed by the type of mode: pedestrian, bicycle, or motor vehicle. Infrastructure can also be narrowed to four types: street network/centerlines, bicycle facilities, pedestrian facilities, and other. The variety of data in these categories is detailed in Table 16. Data can be filtered by availability (immediate or by request), though the majority is immediately publicly available. Data can also be filtered by format: GIS shapefile, static map, spreadsheet, PDF or other (which includes HTML, raster, JSON, VOL). Users can also filter the data by the geographic scale: national, state, regional, county, or city level.

TABLE 16: Infrastructure Facility Types

Infrastructure Type	Attributes
Street network/centerlines	Speed limit, signs, streetlights, guardrail, pavement quality, traffic cameras, traffic counters, medians, curbs, intersections, bridges, alleys, shoulder width
Bicycle facilities	bike lanes, bike routes, bikeways, path, trails, bike share locations, zoning, topography
Pedestrian facilities	Sidewalks, paths, trails, underpasses, curb ramps, greenways
Other	Mass transit, Railroads, railroad crossings

The output is a filtered list organized by geographic scale based on the search terms as shown in Figures 4 and 5. All relevant records will be returned with the following attributes (if available): source name, data type, (if counts selected, then modes), (if infrastructure selected, then infrastructure types), format, city, state, and date ranges. The listed name links to the URL where the data was originally found. The clearinghouse user can download an Excel spreadsheet of all records returned. The user can also click on a record listed in the search results to see additional details about that dataset such as the facility or treatment types, as shown in Figure 6.

As stated in the disclaimer statement found on the bottom of the home page, the links to the datasets are not maintained. Information may be old or out of date, and links may be broken. The user can report broken links to info@pedbikedata.org, but we do not plan to update links regularly. Instead, this is a directory of data sources. If the data have been there in the past, at least users know that it once was posted and can look for where it may be posted now, or if no longer posted, can contact the agency directly to find it.

Figure 3: National Pedestrian and Bicycle Safety Data Clearinghouse (pedbikedata.org) Search Page

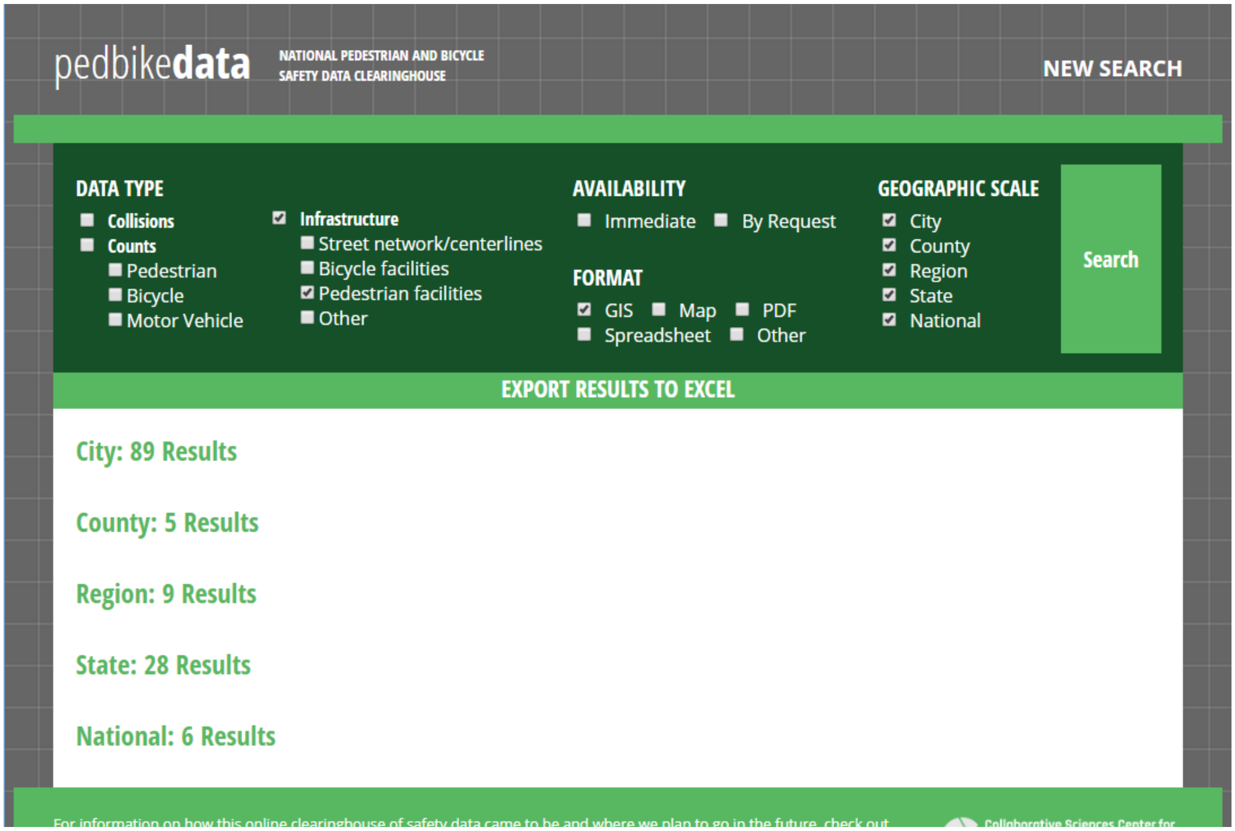

Figure 4: Example Results Page

Figure 5: Example Results Page Expanded

Figure 6: Example Details Page

Conclusions and Next Steps

The National Bicycle and Pedestrian Safety Data Clearinghouse is a work in progress, a first step toward making pedestrian and bicycle safety-related data more available to safety researchers and others. In this report, the team has:

- Summarized literature on data needs
- Interviewed leading pedestrian and bicycle safety researchers on their data needs
- Inventoried existing data sources, rated the state-wide data resources
- Identified gaps
- Explained how a searchable online data clearinghouse has been created to help researchers find data for pedestrian and bicycle safety studies in the U.S.

The next phase of this project will gather user input, prioritize improvements and implement the prioritized improvements in the Clearinghouse. Researchers and others are encouraged to visit the online pedestrian and bicycle safety data clearinghouse, pedbikedata.org, and provide feedback to the authors which can be incorporated into future updates. While keeping over 4,000 links up to date is not within the scope of this or future research projects, the team welcomes ideas and suggestions at info@pedbikedata.org.

References

[1] National HIghway Traffic Safety Administration (NHTSA), "Fatality Analysis Reporting System (FARS)," 2019. [Online]. Available: https://www-fars.nhtsa.dot.gov/Main/index.aspx. [Accessed 24 May 2019].

[2] U.S. Census Bureau, "American Fact Finder, Means of Transportation to Work, 2017 American Community Survey 1-Year Estimates, Table B08301," [Online]. Available: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_15_1YR_B08301&prodType=table. [Accessed 24 May 2019].

[3] T. Black, J. Swartz and T. Fremaux, "Vision Zero and Beyond: A Simple Yet Powerful Data Strategy for Evaluating Potential Engineering Solutions," in *Presented at the 96th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2017.

[4] N. Christie and A. Regan, "A Safety Perception Indicator for vulnerable Road users in urban Environments (ASPIRE)," in *Canadian Transportation Research Forum 51st Annual Conference*, Toronto, Ontario, 2016.

[5] L. Thomas, B. Lan, R. L. Sanders, A. Frackleton, S. Gardner and M. Hintze, "Predicting the Future? Development and Application of Pedestrian Safety Performance Functions to Prioritize Locations in Seattle, WA," in *Presented at the 98th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2017.

[6] K. Nordback, K. Tufte, N. McNeil, M. Harvey and M. Watkins, "Bike-Ped Portal, Development of an Online Nonmotorized Traffic Count Archive," National Institute for Transportation and Communities, Portland, OR, 2016.

[7] R. B. Noland, J. A. Sinclair, N. J. Klein and C. Brown, "How good is pedestrian fatality data?," *Journal of Transport & Health,* vol. 7, pp. 3-9, 2013.

[8] S. Turner, I. Sener, M. Martin, L. White, S. Das, S. Das, R. Hampshire, M. Colety, K. Fitzpatrick and R. Wijesundera, "Guide for Scalable Risk Assessment Methods for Pedestrians and Bicyclists," US Department of Transportation, Washington, DC, 2018.

[9] R. Bauer, A. Kuehnelt-Leddihn and C. Brandstaetter, "The Potential for the European Injury Database for traffic accident research," *Berichte der Bundesanstalt fuer Strassenwesen. Unterreihe Fahrzeugtechnik,* vol. 87, 2013.

[10] D. Morris and M. Wier, "Geospatially Enabled Database for Analyzing Traffic Injuries in San Francisco, California," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2595, pp. 40-49, 2016.

[11] F. Nazin and G. Currie, "Comparing Road vs Transit Authortiy Tram/Streetcar Crash Reporting Approaches and Data," in *Presented at the 98th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2018.

[12] Y. Zhang, F. R. Proulx, D. R. Ragland, R. J. Schneider and O. Grembek, "Develop a Plan to Collect Pedestrian Infrastructure and Volume Data for Future Incorporation into Caltrans Accident Surveillance and Analysis System Database," California Department of Transportation, Sacramento, CA, 2014.

[13] N. Eluru, S. Yasmin, T. Bhowmick and M. Rahman, "Enhancing Non-Motorized Safety by Simulating Non-Motorized Exposure Using a Transportation Planning Approach," SAFER-SIM, Iowa City, IA, 2016.

[14] R. Talbot, H. Fagerlind and A. Morris, "Exploring inattention and distraction in the SafetyNet Accident Causation," *Accident Analysis and Prevention,* vol. 60, pp. 445-455, 2013.

[15] A. Kullgren, M. Rizzi, H. Stigson, A. Ydenius and J. Strandroth, "The Potential of Vehicle and Road Infrastructure Interventions in Fatal Pedestrian and Bicyclist Accidents on Swedish Rural Roads - What Can Indepth Studies Tell Us?," in *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, Detroit, MI, 2017.

[16] American Association of State Highway and Transportation Officials, "Highway Safety Manual," AASHTO, Washington, DC, 2010.

[17] D. Chimba and A. Musinguzi, "Development of Decision Support Tools to Assess Pedestrian and Bicycle Safety: Focus on Population, Demographic and Socio-economic Spectra," Transportation Research Center for Livable Communities, Kalamazoo, MI, 2016.

[18] C. Monsere, H. Wang, Y. Wang and C. Chen, "Risk Factors for Pedestrian and Bicycle Crashes," Oregon Department of Transportation, Salem, OR, 2017.

[19] A. H. Ariffin, A. Hamzah, N. F. Paiman, M. S. Solah, S. F. M. Husin, Z. Zakaroa, M. R. Osman and W. S. Voon, "Risk Factor Identification and Issues Pertaining to Road Collisions Involving Pedestrian and Motorcylce," Malaysian Institute of Road Safety Research, Selangor Darul Ehsan, Malaysia, 2017.

[20] L. Thomas, L. Sandt, C. Zegeer, W. Kumfer, K. Lang, B. Lan, Z. Horowitz, A. Butsick, J. Toole and R. Schneider, "NCHRP Report 893: Systemic Pedestrian Safety Analysis," National Cooperative Highway Research Program, Washington, D.C., 2018.

[21] S. Das, R. E. Avelar, K. Dixon and X. Sun, "Pedestrian Crash Analysis Using Association Rules Mining," in *Presented at the 98th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2018.

[22] C. R. Bhat, S. Astroza and P. S. Lavieri, "A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level," *Analytic Methods in Accident Research,* vol. 16, pp. 1-22, 2017.

[23] Y. Wang, C. M. Monsere, C. Chen and H. Wang, "Development of a Crash Risk Scoring Tool for Pedestrian and Bicycle Projects in Oregon," in *Presented at the 98th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2018.

[24] G. Li, D. Otte, J. Yang and C. Simms, "Pedestrian Injury Trends Evaluated by Comparison of the PCDS and GIDAS Databases," in *2016 IRCOBI Asia Conference*, Seoul, South Korea, 2016.

[25] A. Bukova-Zideluna, A. Villerusa and A. Lama, "An overview of injured bicyclists in traffic accidents: analysis of traffic accident database in Latvia for the period 2010-2014," in *IRCOBI Conference 2016*, Malaga, Spain, 2016.

[26] N. Rameshkrishnan, A. Sathyakumar, S. Balakumar, A. M. Hassan, R. Rajaraman and J. Padmanaban, "The New In-Depth, At-the-Scene, Accident Investigation Database in India," in *2013 IRCOBI Conference*, Gothenburg, Sweden, 2013.

[27] Q. Wang, S. Tang, J. Hu, X. Chen and L. Wang, "Performance Measurement System for Assessing Transportation Sustainability and Community Livability," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2531, pp. 113-120, 2015.

[28] H. Xu, C. Reider, Z. Tian and H. Teng, "Influencing Factors on Conflicts of Turning Vehicles and Pedestrians at Intersections," Nevada Department of Transportation, Reno, NV, 2015.

[29] I. S. Han and W. Han, "Real-Time Pedestrian & Driver Analysis by Neuromorphic Visual Information Processing," in *24th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*, Gothenburg, Sweden, 2015.

[30] M. A. Bushell, B. W. Poole, C. V. Zegeer and D. A. Rodriguez, "Costs for Pedestrian and Bicyclist Infrastructure Improvements," UNC Highway Safety Research Center, Chapel Hill, NC, 2013.

[31] F. R. Proulx, Y. Zhang and O. Grembeck, "Database for Active Transportation Infrastructure and Volume," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2527, pp. 99-106, 2015.

[32] K. Wongsuphasawat, M. L. Pack, D. Filippova, M. VanDinker and A. Olea, "Visual Analytics for Transportation Incident Data Sets," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2138, pp. 135-145, 2009.

[33] New York State, "Motor Vehicle Crashes by Facility, Port Authority of NY NJ: Beginning 2000," Office of Information Technology Services, n.d.. [Online]. Available: https://data.ny.gov/Transportation/Motor-Vehicle-Crashes-by-Facility-Port-Authority-o/n4ys-au5c/data. [Accessed 16 10 2019].

[34] S. o. N. Jersey, "Crash Records, Crash Data," State of New Jersey, Department of Transportation, n.d.. [Online]. Available: https://www.state.nj.us/transportation/refdata/accident/crash_data.shtm. [Accessed 16 10 2019].

[35] N. Y. State, "Bicycle Routes Across New York State," Office of Information Technology Services, n.d.. [Online]. Available: https://data.ny.gov/Transportation/Bicycle-Routes-Across-New-York-State/7bg2-3faq. [Accessed 16 10 2019].

[36] National Center for Statistics and Analysis, "Bicyclists and other cyclists: 2015 data (Traffic Safety Facts/ Report No. DOT HS 812 382)," National Highway Traffic Safety Administration, Washington, D.C., 2017.

APPENDIX

TABLE A1: Count of Star Rated Datasets by State for Collision Events

	2 Stars	3 Stars	4 Stars	5 Stars	Number of Datasets
Arizona		12	8	2	22
California			1	1	2
District of Columbia	1		2	1	4
Hawaii			1		1
Idaho			2	1	3
Indiana		1	12		13
Iowa			1		1
Kentucky				1	1
Maine		2	2	1	5
Maryland		16	1		17
Massachusetts		4			4
Michigan		1		1	2
Minnesota				1	1
Montana				1	1
Nevada			1		1
New Hampshire				1	1
New Jersey			1		1
New York					0
North Carolina			1		1
Oregon		1			1
Pennsylvania				1	1
Rhode Island		1	3		4
South Dakota		1	1		2
Texas				3	3
Utah				1	1
Vermont		3	2	1	6
Washington			1		1
Total	**1**	**42**	**40**	**17**	**100**

TABLE A2: Count of Star Rated Datasets by State for Collision Summaries

	3 Stars	4 Stars	5 Stars	Number of Datasets
California	1			1
Connecticut			1	1
Iowa		1		1
Maine		2		2
Maryland		1		1
Montana		1		1
New Jersey		1		1
New Mexico			1	1
New York				0
North Carolina	2			2
Rhode Island		1		1
Vermont			1	1
Total	**3**	**7**	**3**	**13**

TABLE A3: Count of Star Rated Datasets by State for AADT

	2 Stars	3 Stars	4 Stars	5 Stars	Number of Datasets
Alaska			1		1
Arizona	3	1	1	1	6
Arkansas		2		1	3
California	5	8	1		14
Colorado		1	1		2
Connecticut		3	2	2	7
Delaware				16	16
District of Columbia				12	12
Florida		1	1		2
Georgia			1	1	2
Hawaii	4	1	5	1	11
Idaho		2	2	1	5
Illinois		1		1	2
Indiana		1	1		2
Iowa		22	20		42
Kansas		1			1
Kentucky		2			2
Louisiana	1	1	1		3
Maine		1	5		6
Maryland			1	2	3
Massachusetts				1	1
Michigan			23		23
Minnesota		1	2	1	4
Mississippi			8		8
Montana			2		2

	2 Stars	3 Stars	4 Stars	5 Stars	Number of Datasets
Nebraska		3			3
Nevada			1		1
New Jersey			2	1	3
New Mexico		1	1		2
New York	2			1	3
North Carolina			2	2	4
Ohio			1		1
Oklahoma			1		1
Oregon					0
Pennsylvania		1		1	2
Puerto Rico		1			1
South Carolina			1	16	17
Tennessee			1		1
Texas			3		3
Vermont		2	5	1	8
Virginia		18	1		19
Washington		2			2
West Virginia		1	3		4
Total	**15**	**78**	**100**	**62**	**255**

TABLE A4: Count of Star Rated Datasets by State for Short Duration Counts

	2 Stars	3 Stars	4 Stars	5 Stars	Number of Datasets
Arizona			1		1
California			1		1
Colorado		1	1		2
District of Columbia		1		1	2
Illinois	1		1		2
Kansas		1			1
New Mexico		1			1
Vermont		2			2
Virginia			2		2
Washington		2			2
Total	**1**	**8**	**6**	**1**	**16**

TABLE A5: Count of Star Rated Datasets by State for Permanent Counts

	3 Stars	**4 Stars**	**5 Stars**	**Number of Datasets**
Colorado	1	1		2
Massachusetts	1			1
New Mexico		1		1
Vermont		1		1
Virginia, District of Columbia			2	2
Washington	2	1		3
Total	**4**	**4**	**2**	**10**

TABLE A6: Count of Star Rated Datasets by State for Infrastructure

	1 Star	2 Stars	3 Stars	4 Stars	5 Stars	Number of Datasets
Arizona		10	6			16
Arkansas		2				2
California	1	1				2
Colorado				1		1
Delaware		27	1			28
District of Columbia		30	1	4		35
Florida		3		6		9
Georgia		1				1
Hawaii	2	9		5		16
Hawaii		2				2
Idaho		6				6
Illinois		1				1
Indiana		2				2
Iowa		2		1		3
Louisiana		1				1
Maine		3	1			4
Maryland		4				4
Massachusetts		2	4			6
Michigan		5	1			6
Minnesota	3	14	4			21
Mississippi		2				2
Missouri		1				1
New Hampshire		3				3
New Jersey		5				5
New Mexico		14	4	2	7	27
New York	2	6	2			10

	1 Star	2 Stars	3 Stars	4 Stars	5 Stars	Number of Datasets
North Carolina		6	9			15
North Dakota		1				1
Ohio		1	2			3
Pennsylvania		11				11
Rhode Island		3	1			4
South Carolina		2				2
Tennessee	1	2				3
Texas		4	1			5
Utah		1				1
Vermont		30	4	1		35
Virginia		2				2
West Virginia		10				10
Wisconsin						0
South Dakota		1				1
Total	**9**	**230**	**41**	**20**	**7**	**307**

Collaborative Sciences Center for
ROAD SAFETY

730 Martin Luther King Jr. Blvd.
Suite 300
Chapel Hill, NC 27599-3430
info@roadsafety.unc.edu

www.roadsafety.unc.edu