

Exploration of the Application of Collective Information to Transportation Data for Safety White Paper



MIRE
MODEL INVENTORY OF ROADWAY ELEMENTS
Dictionary of Roadway Data Elements for Safety

FHWA Safety Program



U.S. Department of Transportation
Federal Highway Administration



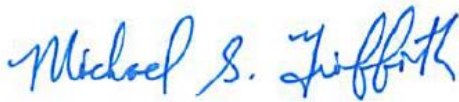
<http://safety.fhwa.dot.gov>

FOREWORD

The Federal Highway Administration's (FHWA) Highway Safety Improvement Program (HSIP) is a data driven program that relies on crash, roadway, and traffic data for States to conduct effective analyses for problem identification and evaluation. The FHWA developed the Model Inventory of Roadway Elements (MIRE) to provide a recommended listing and data dictionary of roadway and traffic data elements critical to supporting highway safety management programs. MIRE is intended to help support the States' HSIPs and other safety programs.

The MIRE Management Information System (MIRE MIS) was a project to explore better means of collection MIRE data elements, using and integrating MIRE data, and identifying optimal data file structures. The resulting products include a report documenting the MIRE MIS Lead Agency Program, a MIRE Guidebook on the collection of MIRE, a suggested MIRE data file structure report, and a report on Performance Measures to Assess Quality that will assist the States in conducting a more effective safety program. The intent of the MIRE MIS project is the integration of MIRE into States' safety management processes.

The Exploration of the Application of Collective Information to Transportation Data for Safety White Paper is one of the products of the MIRE MIS effort. This White Paper explores the relatively new technique of collective information as a means of gathering data needed for transportation safety. Collective information is a process of assembling or gathering information on a subject through the use of a large, disperse, and potentially uncontrolled group of people. This White Paper also examines current uses of collective information and potential next steps for this process to support the collection of transportation data.



Michael S. Griffith
Director, Office of Safety Technologies



Monique R. Evans
Director, Office of Safety Research and
Development

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document. The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this report only because they are considered essential to the objective of the document.

QUALITY ASSURANCE STATEMENT

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

TECHNICAL DOCUMENTATION PAGE

1. Report No. FHWA-SA-13-006	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Exploration of the Application of Collective Information to Transportation Data for Safety White Paper		5. Report Date January 2013	
		6. Performing Organization Code	
7. Author(s) Kim Eccles, Greg Bosworth, and Nancy Lefler		8. Performing Organization Report No.	
9. Performing Organization Name and Address Vanasse Hangen Brustlin, Inc. (VHB) 8300 Boone Blvd., Suite 700 Vienna, VA 22182-2626		10. Work Unit No.	
		11. Contract or Grant No. DTFH61-05-D-00024 (VHB)	
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Safety 1200 New Jersey Ave., SE Washington, DC 20590		13. Type of Report and Period Final Report, June 2012 - January 2013	
		14. Sponsoring Agency Code FHWA	
15. Supplementary Notes The contract managers for this report were Robert Pollack (HSA) and Dr. Carol Tan (HRDS-20).			
16. Abstract The Federal Highway Administration (FHWA) developed MIRE, the Model Inventory of Roadway Elements, as a recommended listing of roadway inventory and traffic elements critical to safety management. A critical step toward the acceptance and implementation of MIRE is the conversion of MIRE (which is now a listing of variables) into a management information system (MIS). FHWA undertook the MIRE MIS project to assist States in developing and integrating MIRE into an MIS structure that will provide greater utility in collecting, maintaining, and using MIRE data. This paper was developed as part of the MIRE MIS project to explore the potential mechanisms for MIRE data collection. The objective of this White Paper is to explore the application of collective information in transportation data, particularly in the collection and management of MIRE data. Collective information is the process of assembling data or collecting information on a subject using a large, disperse, and potentially uncontrolled group of people. The intent of the paper is to introduce this concept to the transportation community and provide a starting point for discussions on the application and feasibility of using collective information for transportation safety data, particularly for the MIRE elements. This paper also discusses the potential next steps and future applications of this concept.			
17. Key Words: Safety data, MIRE, collective data, roadway data collection		18. Distribution Statement No restrictions.	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 20	22. Price

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

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

TABLE OF CONTENTS

BACKGROUND 1

APPLICATION TO MIRE 2

OBJECTIVE 4

INTENDED AUDIENCE 4

DISCUSSION 5

Potential Applications for Transportation 5

Types of Data Collection 6

Data Contributor Considerations..... 7

Types of Contributors 7

Motivation for Collection..... 9

Training..... 10

Risk and Safety of the Individuals 10

Agency Considerations 11

Data Considerations..... 12

Data Accuracy 12

Quality Control 12

Analysis and Reporting 13

Application Considerations..... 14

FUTURE CONSIDERATIONS/NEXT STEPS 14

REFERENCES..... 16

APPENDIX A: AGENCY INTERVIEWS 18

APPENDIX B: NON-TRANSPORTATION EXAMPLES OF CROWD SOURCING..... 20

ACRONYMS

AASHTO	American Association of State Highway and Transportation Officials
ADA	Americans with Disabilities Act
BADCS	Buckeye Asset Data Collection System
CRM	Customer Relationship Management
DOT	Department of Transportation
EAM	Enterprise Asset Management
FHWA	Federal Highway Administration
GIS	Geographic Information System
GPS	Global Positioning System
HSIS	Highway Safety Information System
HSM	Highway Safety Manual
IHSDM	Interactive Highway Safety Design Model
IT	Information Technology
JMT	Johnson, Mirmiran & Thompson, Inc.
LIDAR	Light Detection and Ranging
MIRE	Model Inventory of Roadway Elements
MIS	Management Information System
MUTCD	Manual on Uniform Traffic Control Devices
NCHRP	National Cooperative Highway Research Program
NHTS	National Household Travel Survey
PTA	Parent Teacher Association
RITA	Research and Innovative Technology Administration
SFCTA	San Francisco County Transportation Authority
SMS	Short Message Service

BACKGROUND

The need for roadway data to support safety decisions is increasing due to the development of a new generation of safety analysis tools and methods. These include the 2010 Highway Safety Manual (HSM) (1), the Interactive Highway Safety Design Model (IHSDM) (2), SafetyAnalyst (3), and the American Association of State Highway and Transportation Officials (AASHTO) National Cooperative Highway Research Program (NCHRP) Series 500 Data and Analysis Guide (4), which all require crash, roadway, and traffic data to achieve the most accurate results. State and local Departments of Transportation (DOTs) need more detailed roadway and traffic data to support data driven decision-making that will have the greatest impact in safety on our nation's roadways.

Transportation agencies responsible for the collection, analysis, management, and interoperability of highway safety data have expressed that one of the largest hurdles to having a comprehensive data management information system for highway safety is the lack of roadway inventory data, particularly for local roads. More specifically, agencies identified a lack of resources (both personnel and monetary) to collect roadway data for local roads, or to collect detailed roadway data elements for specific types of locations such as intersections, ramps, curves, or pedestrian facilities. Similarly, the data are also difficult to maintain or update as agencies make changes to these facilities.

One potential solution is collective information. Collective information is the process of assembling data or collecting information on a subject using a large, disperse, and potentially uncontrolled group of people. Collective data is an emerging concept for highway safety data, although there have been some very powerful examples of success in other transportation fields. Agencies have the potential to significantly advance their data collection efforts with minimal investments by using applications that leverage data-enabled phones or a standard web browser for the collection, identification, or classification of data by the public.

The process of collecting the data could be described as crowdsourcing, but the results would be classified as collective information. The popular website *Wikipedia* is an excellent example of collective information. *Wikipedia* is a web-based, collaborative encyclopedia. Users of the website collectively populate the information in the online encyclopedia. This is an organic process as the users independently elect to populate the information, guided by some loose principles and forms of oversight. Most of the production of the information is done anonymously and essentially independent of one another (5). However, collectively the information presents a nearly complete description of many topics.

Collective information is starting to emerge as a method to collect data in the transportation world. Currently, two of the largest and most well-known examples of collective data applications in transportation are the City of Boston's Citizens Connect and Street Bump applications. The Citizens Connect application (described in more detail in Appendix A) allows residents and visitors in Boston to collect information on potholes, faulty street lights, and other

public works issues. Citizens report these issues to the City using their smart phones and populate a database that the City uses to prioritize public works efforts. A more detailed application by the City of Boston is the Street Bump application which uses a mobile device's accelerometer, or motion sensor, and global positioning system (GPS) to detect a distinct jolt when a motorist drives over a pothole or sunken manhole cover and pinpoint its location. The users of this application are collectively developing a pavements condition database for the City through their daily travel. While these applications utilize the general public, transportation agencies could rely on anyone from contractors, maintenance crews, volunteer organizations, to the general public for information.

Many of the other notable applications of this concept in transportation involve more planning level data to include applications that track travel behaviors for use in these models. Chapter 26 of *The On-line Travel Survey Manual* identifies five such examples, the most notable of which is San Francisco County's CycleTracks application (6). CycleTracks, a tool developed by the San Francisco County Transportation Authority (SFCTA), uses smart phone's GPS support to record citizen's bicycle trips for use in a route choice model.

Why are agencies using these collective data applications? One of the primary reasons is the applications present a cost-effective way to collect transportation data that agencies need for decision-making. The examples to date are only a handful; however, these pioneers provide a glimpse of the potential of the application this technology. The prevalence of smart phones and other mobile devices makes this type of data collection feasible. Distributed data collection applications may provide a tremendous opportunity for transportation agencies to cost-effectively collect the data needed to improve the transportation system or to develop a database of existing infrastructure or conditions.

APPLICATION TO MIRE

MIRE, the Model Inventory of Roadway Elements, is a recommended listing of roadway inventory and traffic elements critical to safety management. There are a total of 202 elements that comprise the MIRE listing. These elements are divided among three broad categories: roadway segments, roadway alignment, and roadway junctions. Examples of the MIRE data elements include:

- Roadway classification.
- Paved surface characteristics.
- Number and type of travel lanes.
- Shoulder, median, and roadside descriptors.
- Curve and grade information.
- Traffic control devices.
- Intersection features.

- Interchange and ramp descriptors.
- Pedestrian and bicyclist facilities.
- Traffic volumes.

MIRE is intended as a guideline to help transportation agencies improve their roadway and traffic data inventories. It provides a basis for a standard of what agencies can consider a good/robust data inventory, and helps them move towards the use of performance measures. Most States lack the critical elements in the MIRE listing on all public roads and also the funding to collect this information using traditional methods (7). More information on MIRE is available on the FHWA website, <http://safety.fhwa.dot.gov/rsdp/>.

A critical step toward acceptance and implementation of MIRE is the conversion of MIRE (which is now a listing of variables) into a management information system (MIS). The Federal Highway Administration (FHWA) has undertaken the MIRE MIS project to assist States in developing and integrating the MIRE into an MIS structure that will provide greater utility in collecting, maintaining, and using MIRE data. The proposed MIS design will include the exploration, development, and documentation of the following:

- Mechanisms for data collection.
- An efficient process for data handling and storage.
- Details of data file structure.
- Methods to assure (1) the integration of MIRE data with crash data and other data types, and (2) that access to these data can be accomplished through the MIRE MIS.
- The identification of performance metrics to assess and assure MIRE data quality and MIS performance.

The collective data concept falls under the first objective of the MIRE MIS project – to explore the mechanisms for MIRE data collection. A distributed data collection application, suitable for use on a mobile device such as an iPhone, iPad, Android phone or tablet, or a Windows-based phone, may provide an opportunity for an agency to cost-effectively collect many of the MIRE elements. The intent of this application would be to provide the public a way to assist in the collection of MIRE elements, in turn reducing costs of the DOT's data collection effort and allowing more funds to go towards safety improvement projects.

The data collected with this application would focus around traffic data elements and would integrate the functionality of these connected devices such as GPS, data connectivity, camera, video, accelerometer, and keypad data entry. If DOT's could harness just a portion of the mass of people that have connected smart phone devices, it would reduce their effort in collecting traffic safety data.

One of the largest challenges for State DOTs is collecting data on local roads. Collective data could provide a mechanism to cost-effectively populate local road or local intersection information.

OBJECTIVE

Collective data is an emerging concept in information technology (IT) and largely a new concept to transportation agencies. There are many issues that would need to be addressed before transportation agencies could envision the potential benefits of this approach. The objective of this White Paper is to explore the application of collective information in transportation data, particularly in the collection and management of MIRE data elements. The intent of this paper is to introduce this concept to the transportation community and provide a starting point for discussions on the application and feasibility of using collective information for transportation safety data, particularly for the MIRE elements. As previously identified, the *Travel Survey Manual* discusses the use of these applications for collecting travel behavior.

This paper presents the opinions and ideas of the authors. The information in this paper was developed by the authors in part based on information on transportation agencies that are currently using collective data applications. This includes reviews and in some cases interviews on the following applications:

- Citizens Connect App – Boston, Massachusetts.
- Street Bump – Boston, Massachusetts.
- CycleTracks – San Francisco, California.
- Spot Reporters – Baltimore, Maryland; Boston, Massachusetts; Brookline, Massachusetts; and Grand Rapids, Michigan.
- Buckeye Asset Data Collection System (BADCS) – Ohio DOT.

The detailed findings of these interviews are included in Appendix A. Note: this is not an exhaustive list of all the existing data collection applications; rather those that fit within the context of collective data and that the project team identified as being relevant and applicable to the collection of MIRE data.

INTENDED AUDIENCE

The intended audience for this paper is public agencies, transportation professionals, IT professionals, and researchers who are interested in finding ways to advance the collection and use of transportation data for use in highway safety analysis. There are potential opportunities for collective data applications in almost every field of transportation. The primary function for collective data applications that the authors describe in this paper are for use by public agencies, particularly local agencies, to cost-effectively collect roadway attribute data in support of their highway safety efforts.

DISCUSSION

Potential Applications for Transportation

The data needed to safely and efficiently operate the nation's transportation systems are immense and include data on:

- Demand, such as traffic volumes.
- System performance, such as queue lengths or crashes.
- Operating parameters, such as speed limits.
- Operating characteristics, such as travel times.
- Facilities, such as asset inventories.

Several of these types of data could benefit from collective information. Specific to the MIRE listing, potential applications include:

- Intersection inventory. This could include number of legs, intersection control (e.g., signalized, stop-controlled, yield-controlled), presence of turn lanes, left turn phasing, number of traffic control devices, etc.
- Signal inventory. This could include traditional red/yellow/green signals, pedestrian signals, and flashing beacons. The data collector could enter information on the presence of the signals or outages.
- Lighting. This could include the location and types of lighting.
- Pedestrian facilities inventory. This could include the presence and condition of sidewalks, crosswalks, pedestrian signals, push buttons, wayfinding signs, warning signs, etc.
- Bicycle facility inventory. This could include the presence and condition of bicycle paths, bicycle lanes, wayfinding signs, warning signs, etc.
- User reported walkability index. This could provide an opportunity for local citizens to rate facilities using existing walkability index assessments.
- Pavement surface characteristics. This could include the pavement condition and specific locations of deterioration.
- Sign inventory. This could include the presence, condition, and information conveyed for regulatory and warning signs.
- Curve and grade inventory. This could include the location and degree of curvature for horizontal curves or ramps. For vertical curves, it could include presence and grade. For both horizontal and vertical curves it could also include the presence and conditions of warning signs or the presence of features within the curve, such as driveways, that may impact safety.

The concept could also extend beyond the MIRE listing. Agencies could use these applications to collect other data to improve their transportation system including:

- Transit performance. This could include passenger reported arrival and departure times, and origin and destination information for the overall trip, not just the transit portions.
- Transit facilities. This could include the presence and condition of amenities such as benches and shelters at transit stops.
- Operating speed inventory. This could include passively reported operating speeds on roadways in a jurisdiction at peak and non-peak times.

The type of application used and the information collected would vary based on the needs of the agency. An agency may use one application to collect data across several categories. For example, an agency could use one application to collect inventory elements such as intersection, signal, sign, and pedestrian facility elements. Or, an agency could use an application to only collect elements of interest to one user group.

Once one agency develops an application, it may be able to be adapted to be beneficial to other agencies. The United States DOT Research and Innovative Technology Administration (RITA) tested San Francisco's CycleTracks application in Austin, Texas with positive results (8).

Types of Data Collection

The authors envision two types of data collection: mobile collection and remote collection. Mobile collection would occur in the field through the use of mobile devices such as smart phones or tablets. The user of the application would use GPS or maps to locate the data. The location of the person collecting the data would be important for the actual collection. CycleTracks is a good example of mobile collection. The GPS in the mobile devices of the data collectors tracks their location throughout their route.

The second type of data collection is remote collection. This type of collection could be done anywhere. The location of the person collecting the data is not important to the actual collection. The actual data collection is better described as data capturing. The agency in need of data would collect a large amount of data in the field, likely using a video camera or other device (e.g., light detection and ranging [LIDAR]). The remote data collector would review the video and process the captured data. For example, an agency might post a video of a corridor. Data collectors could process that video to develop a street sign condition inventory.

A good example of remote collection is a recent effort undertaken as part of the FHWA MIRE MIS project to develop an intersection inventory for New Hampshire DOT. Data entry clerks populated an intersection inventory using video logs and publically available photos in Google Maps Street View™ and Microsoft's Bing® Maps Bird's Eye. Remote data collectors, hundreds of miles away from the actual intersections in New Hampshire, reviewed the images and characterized attributes of each intersection in a database. Although in this example the data collectors were compensated, crowdsourcing could be used for similar efforts.

Data Contributor Considerations

For the purpose of this paper, the term *contributor* refers to a user of the application who would collect the data. These contributors are also likely users of the data in the more global sense in that they are users of the facility or users of the collected data. There are several types of potential contributors. They include: public contributor, qualified contributor, opportunistic contributor, and directed contributors, and are discussed in more detail in the following section.

Types of Contributors

The authors envision several types of contributors, or collective data imputers. The first type is **public contributor**. This type of contributor is used for applications that are open to any interested users. Public contributors are appropriate in the following circumstances:

- A large or representative sample of people is needed to accurately reflect operating conditions or to provide enough coverage for the application to reflect the transportation environment.
- Detailed training, beyond what can be conveyed directly by the application, is not needed.
- No equipment is needed, beyond the capabilities of the mobile device, to collect and transmit the information.
- The data are not needed urgently.

The previously described CycleTracks is an example of an application with public contributors. SFCTA made their application open to anyone who wanted to download it, and required the contributors to register using some basic contact information. SFCTA did not place any restrictions on who could register. This provided an opportunity for a diverse sample of bicyclists around the jurisdiction. Similarly, Boston's Street Bump and Citizen Connect applications all use public contributors.

The second type is **qualified contributor**. Civic organizations, Parent Teacher Association (PTA) members, or neighborhood watch members are examples of potential qualified contributors. Agencies would specifically identify these groups for the effort and would receive some form of training. Qualified contributors are appropriate in the following circumstances:

- Training, beyond what can be conveyed directly by the application, is needed.
- The needed reliability of the data is greater than what can be collected by public contributors.
- The collected data has some direct value to the qualified group.

An example of the type of application that may use qualified contributors includes school zone inventories or pedestrian facilities inventories.

The third type is **opportunistic contributor**. This could include agency employees whose primary responsibility is not data collection but their role with the agency provides an opportunity to collect data. Examples of this type of contributor include downtown

ambassadors, maintenance personnel, personnel responsible for collecting other data, such as traffic counters, or parking enforcement. Opportunistic contributors are appropriate in the following circumstances:

- The data collection is not time sensitive, since the collection of data by these contributors occurs only when the opportunity is present.
- Some training or equipment may be needed.
- The data elements to be collected are proximate to the general duties of the contributor, more likely in an urban environment.

An example is the Ohio BADCS. BADCS is a web-based application to view video of Ohio DOT's roadways to collect large-scale asset inventories. It is available to anyone within the DOT, although primarily used by Districts 1 and 2 (9).

An example of the type of application that may use opportunistic contributors includes signal inventories or Americans with Disabilities Act (ADA) compliancy in a central business district.

The fourth type is **directed contributors**. These are contributors that are specifically directed to collect data. This includes contractors retained to collect data, summer interns, or maintenance staff directed to collect data while performing their normal duties. In most cases, these contributors would be compensated for collecting the data. Directed contributors are appropriate in the following circumstances:

- The data are needed urgently, or must be collected over a short period of time.
- Some training, equipment, or specialized expertise is needed.
- The reliability of the data is very important.

ADAMobile developed by Johnson, Mirmiran & Thompson (JMT), Inc., is an example of an application collected by directed contributors (10). The Delaware DOT retained JMT to collect pedestrian facilities inventory data including ADA compliance data in the field and provide real-time updates to the office. JMT collected the data over several months with a high level of reliability.

The authors also envision other contributor groups forming beyond these four categories as the need arises.

An agency could construct an application for use by any of the four categories. The application could also recognize the type of contributor. Agencies could use this information to qualify the reliability of the data. For example, an agency could develop an intersection inventory as a combined effort of public contributors, qualified contributors from a local civic group, and directed contributors as part of a summer traffic counting program. The agency would characterize the reliability of the data collected for each intersection by the contributor group that collected it. Data from the public contributors would receive the lowest quality. In the case of redundant data (e.g., data for the same intersection collected by the local civic group and the

summer traffic counting program), the agency would use the data collected by the higher contributor group.

Motivation for Collection

What is the incentive of the general public to collect data? Jonathan Zittrain, a social theorist and professor of internet law at Harvard Law School, has provided some input on the subject (11). Mr. Zittrain theorizes that the motivation in related crowdsourcing examples includes a general desire to help others but could also include recognition, reward, and entertainment. One notable example is that the problems page on the popular website *Wikipedia* has more visitors to the page looking to solve problems than the number of visitors there to report problems (12). Other interesting (non-transportation) examples on people contributing for the greater good include XPrize, Facewatch ID, and Vertices, among several others. Appendix B provides more detail on these non-transportation examples.

The motivation for data collection could differ by the type of contributor and the type of data being collected. Public contributors and qualified contributors may be motivated by altruistic motives of contributing to the greater good; for example, collecting data that will be used to improve school zones. This is particularly true for qualified contributors as their existing affiliation with a community group, such as the PTA, signals their willingness to contribute to the community.

These same contributors may also be self-motivated or vested in the results. In the CycleTracks example, contributors were cyclists motivated to participate because the data were ultimately used to improve bicycle facilities. Contributors could also be motivated by the application itself. Using the CycleTracks example again, the application included some added functionality, beyond what was needed for the data collection, as a benefit to the contributors of the app. In addition to recording and sending GPS data to SFCTA servers, contributors could also see their recorded routes mapped with their distance and speed by trip. This added functionality was a motivation to use the application.

An application could also motivate contributors to use the app by being fun to use, or by introducing some elements of gaming. An application could include a scoring system so that contributors could be ranked against one another. This concept could be used in a sign inventory. Contributors would receive points both for the number of signs collected and the various categories of signs or diversity of collected signs. Each new type of sign would be like obtaining a tool or a randomized treasure in common adventure or role playing computer games. This concept could encourage contributors to collect data in a more distributed fashion. This may appeal to passengers on road trips. It may also appeal to players of the popular geocaching games or similar.

For directed contributors, and to some degree for opportunistic contributors, the motivation is compensation or negative consequences for not completing their assigned duties. Agencies could also use compensation to motivate public or qualified contributors. Unlike directed

contributors, all contributors would not be compensated. Instead, the agency could introduce some randomization of compensation. For example, contributors could be randomly selected to receive gift cards. The data collected could serve as a contributor's entry into the random drawing. That is, the more data a contributor collects, the more entries they have in the drawing. Similarly, compensation could also be incorporated into the previously described gaming approach.

Regardless of the type of contributor, the overwhelming take-away is that the contributors must understand the benefit of the effort to which they are contributing. Therefore, any successful application must identify the expected contribution to improved safety and communicate that to the contributors as part of the application.

Training

Most applications will require some form of training for contributors. This training could be very basic, such as information on how to use the application, or more detailed training such as classifying highway signs or assessing the condition of a crosswalk. The amount of training required would be dependent on the application. Agencies should consider the amount of training required when selecting the targeted contributor.

Agencies could deliver training through traditional methods for qualified, opportunistic, and directed contributors. However, for public contributors, web or application based training is more appropriate.

Several web-based collective data applications exist that could provide useful examples of training for classification data. The website Zooniverse uses public contributors to identify and classify galaxies. The contributors have to work through a self-paced training guide in which they essentially learn to classify galaxies by working example problems with some sample data (13). An agency could use this concept in the envisioned collective data applications. An agency could train contributors by working through the collection of some sample data on the device.

Risk and Safety of the Individuals

Any application used for collective data needs to prevent or discourage contributors from entering data while driving or from collecting data at a potentially hazardous location such as in the roadway. Agencies should consider this in the data input needs for specific types of locations. For example, an application should not require or allow the collection of video data on an Interstate.

There is always some risk that someone will use a mobile device in a situation that introduces a hazard to themselves or others such as driving while talking on the mobile device, or walking along a hazardous roadway. The application should launch with a disclaimer that warns the user of the hazards of using the application and instructions on appropriate use of the application. Specifically any active data collection should be limited to passengers in a vehicle or pedestrians

safely out of the roadway. (A driver's device could collect passive data such as pavement condition.) As part of the registration process for the application, the user could agree to the conditions of appropriate use.

Agency Considerations

As previously described, the primary function for collective data applications is for use by public agencies to collect transportation data in support of their highway safety efforts. Cost-effectiveness is one of the primary motivations for agencies to seek this type of application. However, there are other agency considerations including the following:

- **Need for Access to a Larger or More Representative Data Set.** Collective data applications may appeal to a broader base of contributors than traditional survey or data collection methods as smart phones are almost ubiquitous. This is important if the collected data needs to reflect the diversity of the contributors or be conducted in diverse locations throughout a jurisdiction. SFCTA compared the demographics of their contributors to the National Household Travel Survey (NHTS) respondents and found their CycleTracks contributor to be a more diverse and representative sample of their jurisdiction. According to a 2012 study by the Pew Internet & American Life Project, nearly half of Americans (45 percent) use a smart phone or other mobile device, and the number is growing (14).
- **Agency Size.** The size of the agency and jurisdiction it manages would impact several of the considerations. Collective data efforts that would rely on a "community" of people would be most effective for local jurisdictions rather than statewide initiatives.
- **Ability to Administer the Application.** Agencies need to have the ability to oversee the collective data application. A contractor could develop or manage the application; however, the responsible agency has to be able to clearly identify their data needs, requirements, and intended use for the data.
- **Ability to Manage the Data.** Agencies would need the technology and knowledge on how to store and manage the additional data. This may require additional IT investments and training for the data managers.
- **Vision.** Although the process of collecting the data will be more cost-effective than the traditional methods, the initial development of the data collection application will be a cost or time investment. This cost may be larger than the initial start-up costs for some traditional methods. Therefore, the agency needs to have a clear vision of how they will use the data so they can develop the application appropriately. An agency should also develop the application with some flexibility as their needs may evolve.
- **Existing Applications.** Once an agency has established a clear vision, they should do some research to determine if another agency has already developed a similar application. This additional research could save valuable time and resources in the long-run by building off of existing applications rather than developing one from scratch.
- **Pilot the Effort.** As with any large data collection effort, the agency should consider a pilot effort before launching a large scale effort. This is related to the previous point

about flexibility in the application. It allows for the agency to identify and resolve problems or additional needs before wide-scale use.

- **Secondary Benefits of Data.** An agency may find that there are some secondary benefits of the application of the collected data. For example, inventory data collected for safety analysis may also help with asset management.
- **Secondary Benefits of Process.** Agencies that use collective data applications will gain experience, knowledge, and comfort/familiarity with collective data and mobile technologies. Agencies may view data collection and the actual data in a more dynamic, connected context. This process will fundamentally change how they view both the data and the process of data collection by introducing the agency to other possibilities and potentially generating ideas for future applications. This could be future crowdsourcing applications or future uses for mobile technologies in general.

Data Considerations

Data Accuracy

Accuracy concerns will certainly limit the potential applications, but in many cases, there can be tremendous value in collecting data through collective data applications, realizing that the accuracy of the data may be in question. Agencies might decide to use these applications as an interim measure, with a willingness not to allow the perfect to stand in the way of the good.

For the types of applications envisioned, data accuracy could include **location** accuracy (e.g., is the horizontal curve correctly placed), accuracy in **classification** of elements (e.g., is the sign a standard Manual of Uniform Traffic Control [MUTCD] R1-1 STOP sign (15)), accuracy in **characterization** of elements (e.g., does the intersection have a protected left turn on the northbound approach or a protected-permitted left turn), and accuracy in **measurement** (e.g., does the ramp actually have a 10% grade). Before using collective data for their efforts, the agency should realize that the accuracy of the data has some limitations when collected through this method.

Location accuracy for most data elements would be limited to the location accuracy of the mobile device or the contributor's ability to locate the position of the element or asset on a map. The application could increase the location accuracy by providing two measures of location. That is, the contributor's position as understood by the mobile device could be identified on a map and the contributor would be asked to verify that the position reflects their understanding of their location.

Classification accuracy and characterization accuracy would be dependent on the contributor's ability to understand the training or direction provided and apply that to their characterization or classification of elements. Therefore, the skill of the contributor is critical.

Quality Control

The type of contributors affects the expected quality and therefore the quality control required. Directed contributors should provide higher quality data and require less quality control. The

application could also be more complex and the data collected could be more in-depth. Public contributors are at the other extreme. There would be an increased priority to validate and normalize the data that are collected by these contributors.

An agency can employ several methods to increase the accuracy or at least understand the accuracy of the data through a quality control process. As previously discussed, there may be value to collecting the information, realizing the accuracy of the information may be in question. However, the agency needs to have an understanding of the accuracy of the data.

The first method is to provide some ground truthing. That is, a sample of data would be collected by, or on behalf of, the agency with more traditional data collection methods. The agency would consider these data the ground truth data. The agency would then compare the ground truth data to data collected through the application to determine the relative accuracy of the application data. This could be done for a sample of data (e.g., one percent of all observations) or in a structured fashion to provide a fixed reference point in the collected data that contributors could use to conduct their own quality control (e.g., one sign in every mile of segment is characterized).

The second method is through the use of redundancy. That is, multiple contributors collect information on the same element. The values of multiple contributors are compared for similar responses. In this case, the majority prevails. Galaxy Zoo, part of the previously described Zooniverse, uses this process to control the quality of their data, noting, *“Having multiple classifications of the same object is important, as it allows us to assess how reliable each one is. For some projects, we may only need a few thousand galaxies but want to be sure they're all spirals. No problem – just use those that 100% of classifiers agree on. For other projects we might want larger numbers of galaxies, so might use those that a majority say are spiral. (16)”*

The third method, ratings, incorporates elements of the first two methods. With this ratings method, contributors receive a rating on the quality of their data entry. This rating could be assigned by comparing the contributor’s individual collected data to the ground truthing data or to the data collected by other contributors. This rating would be similar to the seller ratings used in the popular online auction site, EBay. Similarly, the rating could be applied to individual elements instead of individual contributors. Agencies could tag those elements that show low consistency in contributed data based on the redundancy checks with a rating. For example, 75 percent of contributors agree on the left turn protection at this intersection. Agencies may decide that some elements or specific locations need to be collected with traditional methods to supplement the collected data.

Analysis and Reporting

Analysis and reporting are important data considerations for any data collection effort. Specific to collective data, the application should provide feedback to the contributors. This is important for at least two reasons. The first is related to quality control and providing feedback to the contributors. The second is related to the motivation to the contributors. As previously

described, the desire to contribute to the greater cause will be one of the primary motivations for the contributors using the application. Information as simple as summary statistics on the data collection effort can help to provide the necessary feedback to the contributors to keep them engaged in the effort.

Application Considerations

Just over five years ago the only feasible solutions for customized field data collection was through the use of hand-held data collectors, common with surveyors, or some type of ruggedized or semi-ruggedized laptop. If photographs or geo-location were required, the data collection team needed to bring a camera and/or a GPS device in the field. Connecting all these devices reliably in the field was sometimes a challenge. There were options of integrated hardware, but the cost of these solutions limited the use to professional services firms that could afford them, in turn limiting who could collect these data.

Today, there is a vast array of platforms available to agencies to collect data. The majority of these devices are consumer level devices. Connectivity and processing power on the mobile devices continues to improve. Built-in features on devices such as cameras, video recorders, GPS, and accelerometers enhance the ability of data collection.

There are many options for developing a data collection application for a laptop or a data collector, but for a smart phone, presently there are only two viable approaches for data collection – web based applications or native apps such as what you would find on the iTunes App Store or the Google Apps Marketplace. There are pros and cons for both options.

The potential benefit of developing an application on a web platform is that you can develop it once and it is usable on any device. There are many browsers and mobile development tool sets that make this mobile web development more streamlined, but the challenge is that the variety of browsers and hardware platforms will typically require a lot of compatibility testing. Native apps can leverage the built-in functionality of the device, but may require specialized development resources. Additionally, an agency will need to either select a platform to support, such as the iPhone, Android, or Windows, or invest in developing the application on multiple platforms to get as much penetration as possible.

FUTURE CONSIDERATIONS/NEXT STEPS

This paper introduces the concept of collective information to the transportation safety community. It is a starting point for discussions on the application and feasibility of using collective information for transportation safety data. This paper focuses on the use of these applications for transportation agencies in collection of MIRE data. The authors envision other uses beyond what is discussed here. For example, could collective data be used to supplement the data in the Highway Safety Information System (HSIS) database? What if researchers who use the HSIS data amend the database with additional data they collected as part of their research studies? In addition, there were aspects of collective information that are not addressed in this

White Paper such as how the applications are marketed to the public to solicit participation in the collection.

A logical next step for this concept is the development of a prototype. The prototype would provide an opportunity to explore some issues about the potential application of this concept. The prototype would not cover all situations and types of collection information for transportation data collection applications. Specifically, a good starting point would be a prototype for an intersection inventory. Many agencies are now aware of the importance and value of an intersection inventory. The prototype would demonstrate the concept while identifying additional issues or questions that need to be addressed.

Based on the initial exploration detailed in this paper, collective data collection applications may provide a tremendous opportunity for transportation agencies to cost-effectively collect the data agencies need for data driven decision-making to improve safety on the nation's roadways and should be explored further.

REFERENCES

1. American Associate of State Highway and Transportation Officials. *Highway Safety Manual*. Available online at: <http://www.highwaysafetymanual.org/>. Accessed September 2012.
2. Federal Highway Administration. *Interactive Highway Safety Design Model*. Available online at: <http://www.fhwa.dot.gov/research/tfhrc/projects/safety/comprehensive/ihsdm/index.cfm>. Accessed September 2012.
3. American Associate of State Highway and Transportation Officials. *SafetyAnalyst*. <http://www.safetyanalyst.org/>. Accessed September 2012.
4. Council, Forrest, et al. *NCHRP Report 500: Guidance for Implementation of the AASHTO Strategic Highway Safety Plan, Volume 21: Safety Data and Analysis in Developing Emphasis Area Plans*. Transportation Research Board, Washington D.C., 2008. Available online at: http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_500v21.pdf . Accessed September 2012.
5. Wikipedia. Available online at: <http://www.wikipedia.org/>. Accessed October, 2012.
6. *The On-line Travel Survey Manual: A Dynamic Document for Transportation Professionals*. Chapter 26.0 Using Smartphones for Travel Behavior Studies. Available online at: <http://www.travelsurveymanual.org/>. Accessed September 2012.
7. Lefler, Nancy X., F. Council, H. McGee, D. Harkey, D. Carter, M. Daul. *Model Inventory of Roadway Elements - MIRE, Version 1.0, Final Report*. FHWA-SA-10-018. Federal Highway Administration, Washington, D.C., June 2010.
8. Hudson, Joan G., et al. *Using Smartphones to Collect Bicycle Travel Data in Texas*. UTCM 11-35-69. Department of Transportation, Research and Innovative Technology Administration, Washington, DC, August 2012.
9. Federal Highway Administration. *A Look at Ohio DOT's Asset Management Applications*. GIS in Transportation Summer 2010 Newsletter. Available online at: http://www.gis.fhwa.dot.gov/documents/Newsletter_Summer2010.htm. Accessed October 2012.
10. Johnson, Mirmiran & Thompson, Inc. ADAMobile Solution for Pedestrian Inventory & ADA Compliance. Available on at: <http://www.jmt.com/project-portfolio/adamobile-solution-for-pedestrian-inventory-ada-compliance/>, Accessed October 2012.
11. Zittrain, Jonathan. *The Future of the Internet and How to Stop it*. March, 2009.
12. Zittrain, Jonathan: The Web as Random Acts of Kindness. TED Talks. Available at http://www.ted.com/talks/jonathan_zittrain_the_web_is_a_random_act_of_kindness.html. Accessed September 2012.
13. Zooniverse. Available online at: <https://www.zooniverse.org/>. Accessed October 2012.
14. Smartphone Ownership Update: September 2012, Pew Internet and American Life Project. Available online at: <http://pewinternet.org/Reports/2012/Smartphone-Update-Sept-2012.aspx>. Accessed October 2012.

15. Federal Highway Administration. *Manual on Uniform Traffic Control Devices*. Available on at: <http://mutcd.fhwa.dot.gov/>. Accessed October 2012.
16. Galaxy Zoo documentation. Available online at: <http://www.galaxyzoo.org/#/story>. Accessed September 2012.

APPENDIX A: AGENCY INTERVIEWS

Example A

Name: Citizens Connect App

Location: Boston, Massachusetts

Link: <http://www.cityofboston.gov/doi/apps/citizensconnect.asp>

Overview: This simple to use, award winning smart phone app enables residents and visitors of the City of Boston to send issues and problems directly from their iPhones or Android-based devices to the city's work order system. It makes services easily available to constituents, cutting down on the amount of time that elapses between when an issue is reported and sequentially resolved, reducing the City's cost to manage requests & transactions. Initially created to report potholes in the city, citizens can now report issues such as clogged storm drains, excessive jackhammer noise, illegal trash dumping, and faulty street lights, complete with pictures. Reported cases then go directly into the city's work order queue for resolution, and users are informed how quickly the case will be closed. When cases are resolved the date, time of the resolution is listed, providing users with the sense that the city is on the job. Within the application, users are able to publicly share reports they have made, see reports others have shared, and text or use Twitter to publish efforts they have made to improve their neighborhoods making it also a social networking app.

Example B

Name: Street Bump

Location: Boston, Massachusetts

Link: <http://www.cityofboston.gov/DoIT/apps/streetbump.asp>,
http://www.cityofboston.gov/news/uploads/46245_32_21_59.pdf

Overview: Street Bump is a mobile app that helps identify potholes by recording "bump" data, providing the City with real-time information to help improve roads. This app's roots came from Citizen Connect. The City was looking for a way to find potholes before they were reported. Initially the information that came from the smart phone devices had many false positives. The City chose to set up a contest to create an algorithm to analyze the information to get more accurate results. From this contest there were three winners; these algorithms are in place now. The app uses a mobile device's accelerometer, or motion sensor, and GPS to detect the tell-tale jolt when a motorist drives over a pothole or sunken manhole cover and pinpoint where it is. By collating the data from drivers and matching it against known speed humps, level crossings and roadworks, sections of road in need of maintenance can easily be identified.

Example C

Name: CycleTracks

Location: San Francisco, California

Link: <http://www.sfcta.org/content/category/4/67/145/>,
<http://www.sfcta.org/images/stories/IT/SFCHAMP/PDFs/BikeRouteChoiceModel.pdf>,
<http://itunes.apple.com/us/app/cycletracks/id338203326?mt=8>

Overview: CycleTracks, a tool developed by the San Francisco County Transportation Authority, uses your smartphone's GPS support to record your bicycle trips, display maps of your rides, and help transportation planners make San Francisco a better place to bike. At the end of each trip, real-time data representing the trip purpose, route, and the date and time are sent to the Transportation Authority. Planners use the data to improve the bicycle-use component of their computer model and better predict where cyclists will ride and how land development and transportation infrastructure will affect cycling. Users get to see maps and statistics of their rides.

RITA further explored the CycleTracks application in a 2012 study which tested the application in Austin, Texas (8). The study found that despite challenges, the amount of information that could be collected by CycleTracks exceeded what would be able to be collected using other data collection methodologies.

Example D

Name: Spot Reporters

Location: Baltimore, Maryland; Boston, Massachusetts; Brookline, Massachusetts; and Grand Rapids, Michigan

Link: <http://www.connectedbits.com/>

Overview: Spot Reporters was designed to integrate seamlessly with existing customer relationship management (CRM) systems, enterprise asset management (EAM) systems, and geographic information systems (GIS). It enables anyone with a mobile phone to report issues they encounter by sending photos, locations, and descriptions directly to Spot Reporters app server. Smartphone users can download apps or use Web, Twitter, or short message service (SMS) clients to report. Four cities have started to use the program: Baltimore, Maryland; Boston, Massachusetts; Brookline, Massachusetts; and Grand Rapids, Michigan.

APPENDIX B: NON-TRANSPORTATION EXAMPLES OF CROWD SOURCING

Some interesting examples of crowd sourcing/data collection that could provide some insight into potential applications for highway safety data include the following:

- XPrize.org – <http://www.xprize.org/>. Crowd sourcing innovative solutions to world challenges.
- Amazon Mechanical Turk – <https://www.mturk.com/mturk/welcome>. A crowdsourcing internet marketplace that enables what are described as *requesters* to identify human workers (referred to as *providers*) willing to perform tasks that computers are currently unable to do. For example, the workers might classify pictures. The workers are compensated at a very low level. This is referred to as Artificial Artificial Intelligence. (The duplication of the word “artificial” is intended.) This is an ideal solution for simple repetitive tasks and may have a use in transportation safety data for classifying collected elements.
- Facewatch ID – <http://facewatch.co.uk/cms/>. This is a crime prevention application in the United Kingdom. The public watches feeds from the vast number of closed circuit televisions in the UK.
- Zooniverse – <https://www.zooniverse.org/projects>. Zooniverse is a citizen science web portal that currently hosts eight collective data applications which allows users to participate in scientific research. As of March 2012, the project has over 600,000 human volunteers, collectively referred to as Zooites. Two examples of Zooniverse projects include:
 - Galaxy Zoo – <http://www.galaxyzoo.org>. Launched in 2007, Galaxy Zoo has over 250,000 people helping with the classification of galaxies, providing more than 60,000,000 classifications. This work is entirely uncompensated.
 - Ancient Lives – <http://ancientlives.org/>. Users transcribe ancient Egyptian papyri. The papyri belong to the [Egypt Exploration Society](#) and their texts will eventually be published and numbered in Society's Greco-Roman Memoirs series.
- Vertices – <http://www2.vertices.com/home/>. This firm specializes in community participatory mapping and has a few examples on their website on projects they have done with various communities, capturing things from walking routes to healthy restaurants. In these examples the community collected data using a variety of participants, from school kids to an identified task force. Additionally they have an interactive mapping tool called Mapper (<http://www.mappler.com/>) which provides the ability for users to go out and collect a variety of information and interactively map the information. One example they have is creating a road safety assessment.

For More Information:

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

FHWA, Office of Safety

Robert Pollack
Robert.Pollack@dot.gov
202-366-5019

Carol Tan, PhD
Carol.Tan@dot.gov
202-493-3315