Title
A Database for Active Transportation Infrastructure and Volume

Authors
Proulx, Frank R.
Zhang, Yuanyuan, PhD
Grembek, Offer, PhD

Publication Date
2014-11-15

Peer reviewed

A DATABASE FOR ACTIVE TRANSPORTATION INFRASTRUCTURE AND VOLUME

Frank R. Proulx
University of California, Berkeley, Safe Transportation Research and Education Center
2614 Dwight Way #7374
Berkeley, CA 94720
Telephone: 510-642-4049, Fax: 510-643-9922
E-mail: fproulx@berkeley.edu

Yuanyuan Zhang, PhD*
University of California, Berkeley, Safe Transportation Research and Education Center
2614 Dwight Way #7374
Berkeley, CA 94720
Telephone: 510-457-5689, Fax: 510-643-9922
E-mail: yuanyuanzhang@berkeley.edu

Offer Grembek, PhD
University of California, Berkeley, Safe Transportation Research and Education Center
2614 Dwight Way #7374
Berkeley, CA 94720
Telephone: 510-642-5553, Fax: 510-643-9922
E-mail: grembek@berkeley.edu

*Corresponding Author

Submitted for Presentation and Publication at the Transportation Research Board's 94th Annual Meeting
Submission Date: August 1, 2014
Resubmission Date: November 15, 2014
Word Count: 4,643 text + 3 Tables and 1 Figure = 5,643 words

ABSTRACT
Information about pedestrian infrastructure and volume is indispensable to monitoring, evaluating, and improving the environment for comfortable and safe walking. However, determining and organizing the various types of data in a way that is easy to update and analyze can present challenges. This study designed and developed a relational database for pedestrian infrastructure and volume, and comprises two core components (node table and approach table) and several sub-components (tables for crosswalks, sidewalks, buffers, signs, transits, bikeways, bicycle parking, and volumes). Important measurements were proposed based on the literature and practice review and grouped into different component categories based on their attributes and relationships. To connect all the components, links were defined according to their relative locations. To prove the feasibility of the database, an infrastructure data collection pilot was conducted across 100 miles of California highways using computer imagery, and across seven miles of highways using field inventory. Time costs associated with collecting infrastructure data for the entire State Highway System were estimated to be 4,006 hours and 8,935 hours for using computer and field collection methods respectively. This study demonstrates that the database is easy to maintain, flexible to update, and feasible for data collection both via computer imagery and in the field. Although most of the data in the database is related to pedestrian, basic bicyclist related information is also included to demonstrate the transferability of the database to store bicyclist infrastructure and volume in the future.

KEYWORDS
Database, activity transportation, infrastructure, volume

INTRODUCTION

As we continue to move forward into the age of big data with research based on ubiquitous sensor networks, social media, and terabytes of computer imagery, there is a conspicuous gap in the transportation data landscape. In particular, many departments of transportation do not have central databases storing pedestrian and bicycle infrastructure and volume data, despite maintaining similar databases for motor vehicle-related data. At the same time, many agencies have stated goals of increasing the number of people walking and bicycling for transportation *(1, 2)*. There are many avenues being pursued towards this goal, including encouragement programs, enforcement activities, and infrastructural modifications. However, in order to effectively monitor the successes being had and to plan for next steps, information systems are paramount.

Collecting bicycle and pedestrian infrastructure data introduces unique challenges to the extent that the networks do not perfectly overlap with motor vehicle networks. Without systematized data storage mechanisms in place, it is very difficult if not impossible to make empirically driven decisions. Given detailed information on pedestrian and bicycle infrastructure and volumes, the following applications could be improved upon:

- Performing systematic analysis of pedestrian and bicycle safety
- Calculating Multimodal Level of Service
- Monitoring pedestrian network completeness
- Ensuring compliance with the Americans with Disabilities Act
- Estimating walkability of a region
- Travel demand modelling

In this paper, a relational database format is proposed for storing these two sources of data conjointly. The framework for the database was developed in the context of the California statewide highway system, but the database is transferable to any other jurisdiction for replication. In addition to developing the database theoretically, an infrastructure data collection pilot was conducted across 100 miles of California highways using computer imagery, and across seven of those miles via field inventory. The data collection pilot was conducted both to ensure the feasibility of data collection and to estimate the time costs associated with collecting the infrastructure data.

While the data being described here can be collected and stored in Geographic Information Systems as individual layers, using a relational database as proposed herein could potentially have performance gains over storing the information in a format such as shapefiles, with little loss of analytical power. Additionally, the proposed database does not require the level of specialized knowledge that collecting fully detailed GIS data would entail, although the proposed database is able to handle storing geographic information on each of the recorded elements if the user chooses to do so.

The goals of this paper are threefold, namely to:

1. Present a flexible database framework for storing pedestrian and bicycle infrastructure and volume data,
2. Propose specific fields to be included in such a database, and
3. Document the approximate time requirements for collecting such data

Based on these various goals, general conclusions are drawn for the benefit of practitioners looking to develop similar databases in their jurisdictions.

PREVIOUS WORK

Infrastructure Inventories

Several successful pedestrian infrastructure inventories have been conducted by agencies at the state, county, and city levels. At the California Department of Transportation (Caltrans), although there is no state level database for pedestrian infrastructure, the Sacramento region (District 3) developed a Complete

Streets inventory database for all state highways *(3)*. Sample fields include the presence of sidewalks, pedestrian-scale street lighting, marked pedestrian crossings, median islands, and curb extensions. Another related Caltrans effort has taken place within the Americans with Disabilities Act (ADA) Infrastructure Program.

An inventory has been conducted of non-ADA compliant pedestrian infrastructure along the California State Highway System *(4)*. Non-compliant facilities are geo-referenced, and details pertaining to the particular features of the facility which are non-compliant and how far they are outside of the acceptable range are included in this inventory. Data was collected through field inventory using both paper spreadsheets and tablet computers which offer the benefit of automatically geo-referencing records. The data is used in prioritization of facility improvements to meet ADA requirements at the District level.

Washington State DOT records date-stamped video of approximately half of their state highways annually using a van with special video equipment *(5)*. To create an inventory of pedestrian infrastructure, this video footage was reviewed by analysts who recorded data on sidewalks, marked crosswalks, and other pedestrian and bicycle facilities. The inventory was then field-checked by driving along a subset of the highways. New bicycle and pedestrian infrastructure is monitored through project control forms, and subsequently added to the inventory. WSDOT anticipates renewing the inventory every three to four years. Total time costs of completing the 7,000-mile inventory (not including driving and video recording) are estimated at 700 hours for video analysis and 1,000 hours for field checking.

New Jersey DOT has constructed an inventory of pedestrian and bicycle infrastructure presence along all county roads in the state, a total of approximately 13,200 miles *(6)*. Data were collected using a vehicle equipped with GPS and four digital cameras. The imagery was then analyzed and compiled into a database, noting the presence of pedestrian and bicycle facilities. All data is available for download by county both as .pdf maps and in GIS data formats. The Maryland State Highway Administration performed a sidewalk inventory focusing on ADA compliance *(7)*. Data was collected in the field using GPS and covered 874 miles of sidewalk for ADA compliance of sidewalks, bus stops, curb ramps, driveway crossings, and median treatments.

A sample of cities that have conducted some form of pedestrian infrastructure inventory were also identified. In most cases, these are merely binary indicators of sidewalk presence along city streets. Data collection methods include review of video imagery, review of city records, field spot-checks, and comprehensive field inventories.

Some cities have conducted more elaborate inventories. For example, Rancho Cucamonga (California), Berkeley (California), and Alexandria (Virginia) all have sidewalk inventories that include priority ranking for missing segments *(8–10)*. Data in Berkeley was collected by review of video imagery, city records, and field spot-checks. Data in Alexandria, by contrast, was collected using mobile GPS units to pinpoint the locations of infrastructural elements such as curb ramps and sidewalk obstructions. In Rancho Cucamonga, the pedestrian infrastructure inventory also included a priority ranking of missing sidewalk segments.

Pedestrian/Bicycle Volume Databases

A number of efforts are currently underway to centralize storage of bicycle and pedestrian count data. At the federal level, the Traffic Monitoring Analysis System (TMAS) operated by the Federal Highway Administration is in the process of being updated to accommodate submission of bicycle and pedestrian volumes. This will allow for the collection and analysis of data following a common format nationwide. However, the TMAS system does not include linkages to substantial pedestrian or bicycle infrastructure data. Count sites are described by their geographic location, whether the count was conducted on a crosswalk, sidewalk, or exclusive facility, count direction, and the location of the count relative to the roadway orientation *(11)*. At the local level, multiple regions are currently developing or have recently developed public-facing webpages allowing access to their bicycle and pedestrian count data, including Los Angeles County (CA), Arlington County (VA), Delaware Valley Region (PA), Portland (OR), and Seattle (WA) *(12–16)*. All of these webpages allow access to geocoded count data that is displayed on a map interface and accessible for download.

DATABASE

Structure

The proposed database stores pedestrian/bicycle infrastructure and volume data in two parallel sub-databases. The infrastructure sub-database consists of two component types: core components and secondary components. The *core (or primary) components* form the skeleton of the data structure—consisting of approaches and nodes. *Approaches* are defined as unidirectional road segments demarcated by a road junction, midblock crosswalk, pedestrian overpass/underpass, or when the length of an individual segment exceeds one mile. Approaches represent the two sides of the roadway. *Nodes* consist of components that lie in between adjoining approaches. This includes intersections/junctions, midblock crosswalks, pedestrian overpass/underpass, and/or points where these other features have not been encountered for one mile. Nodes are named (based on the names of the intersecting roads or other characteristics) and uniquely numerically identified. If georeferencing of data is desired, the core components are likely the most appropriate place to make these connections.

The *secondary components* are the key part of this data collection effort. These components include sidewalks, crosswalks, bicycle facilities, and other pedestrian or bicycle related infrastructure elements (details can be found in the "Infrastructure" section of this paper). Every secondary component is linked to a set of primary components—acting as a subset of the primary components. For example, sidewalks are linked to a single approach, while crosswalks are linked to a node and two approaches. The links are developed following the logic detailed as below.

Figure 1 shows the main components comprising a typical street segment. The two core components are nodes and approaches, while the others are the secondary components which will link to the primary components. Separate tables are stored for each type of infrastructure.

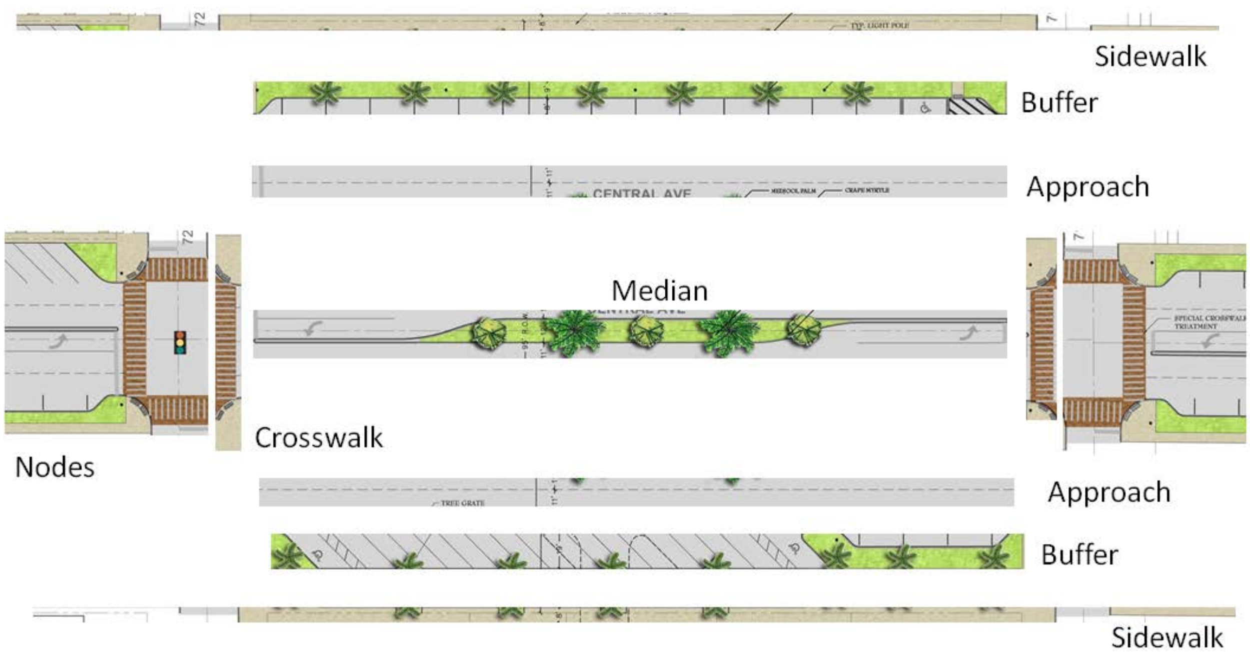

FIGURE 1 Components of a typical roadway segment.

The relationship between the components is developed based on their physical relationship to each other. There are two core components to which other sub-components connect. For example, sidewalks and buffers are connected to the approach with which they are associated. Signage, bicycle parking and transit are connected to the approach on which they are located. Following the same logic, crosswalks are associated with the intersections or mid-block breaks (two kinds of nodes), and approaches which advance toward or depart from the crosswalk.

The volume sub-database follows a similar structure. In this case, volume observations are secondary "components" linked to the core approach and node components. Volume observations are made on sidewalks and crosswalks. These values will then be linked to the core components in the volume sub-database in a manner parallel to that of linking sidewalks and crosswalks (respectively) in the infrastructure sub-database. The database is also flexible to accommodate recording mid-block crossing counts by omitting the "node" connection.

Content and Definition of Components

The infrastructure and volume components are organized in different sub-tables as detailed in Tables 1 and 2. Every table includes fields for primary key, foreign keys, and the date of data collection. The remaining fields store characteristics of the infrastructure as measured/observed during data collection. For example, the sidewalk table as proposed includes a linking field to the approach table, a width measurement taken at a data collector-defined "representative location", and a text field to describe any observed obstructions to the pedestrian right-of-way. While the fields included in this table were selected for the project under discussion, the database was designed with the intention of flexibly accommodating any other characteristics of the infrastructure. For example, some jurisdictions have collected information on the condition of their sidewalks. This could be included as a text description in the sidewalk table, a numeric rating in the sidewalk table, or as a separate table related on a many-to-one basis to the sidewalk table. It is also worth noting that infrastructure components pertaining specifically to motor vehicle traffic such as the number and width of lanes were excluded because Caltrans already has this information recorded statewide, but would be a worthwhile addition for any locations where these fields are not already available.

The motivations in selecting and defining the data elements to be collected in the database were multiple. Pedestrian and bicycle environment quality metrics were consulted for inspiration, including the Highway Capacity Manual's Multi-Modal Level of Service, the Pedestrian Environmental Quality Index (PEQI) and the Bicycle Environmental Quality Index (BEQI) *(17, 18)*. Additionally, one likely application of this database is in estimating pedestrian and bicycle volumes, so transportation infrastructure variables from a number of previously developed direct demand models were included *(19–21)*. However, not all of the variables identified in these various models were selected. A competing interest in development was developing a tractable data collection process using computer imagery, which would become far less feasible if minute details such as the number of driveway curb cuts (as is present in the PEQI) were collected. Future implementations of the database could use these additional variables if a primary desired application were to calculate a prescribed metric.

TABLE 1 Infrastructure Data Definitions in Current Database Implementation

Table	Field	Data Type	Definition/Possible Values
Node	Name	Character	Descriptive title for node (e.g. names of intersecting streets)
	Type	Categorical	3-way, 4-way, >4-way, midblock crossing, over/underpass, segment break
Approach	From Node ID	Key	Linking field
	To Node ID	Key	Linking field
	Informal Crossing	Categorical	Demonstrated, passible, not passible, no median
	Parking Lane Width	Numeric	<8', 8'-10', 10'-12', >12', none
Sidewalk	Approach ID	Key	Linking field
	Width	Numeric	Width of sidewalk at a representative location
	Obstructions	Character	Text description of obstructions
Buffer	Approach ID	Key	Linking field

	Width	Numeric	Width of buffer zone at a representative location
Transit Stops	Approach ID	Key	Linking field (defines presence of feature)
Bikeway	Approach ID	Key	Linking field
	Width	Numeric	Width of bikeway at a representative location
	Type	Categorical	Sharrow, Bike lane, buffered bike lane, cycletrack, separated path
	Color	Categorical	White, green, other
Bike Parking	Approach ID	Key	Linking field (defines presence of feature)
Signage	Approach ID	Key	Linking field
	Type	Categorical	Codes from California MUTCD
Crosswalk	Node ID	Key	Linking field
	Approach ID 1	Key	Linking field
	Approach ID 2	Key	Linking field
	Intersection Edge	Categorical	North, South, East, West
	Style	Categorical	Standard, continental, ladder, triple-4, solid, dashed, zebra, unmarked
	Curb Ramp Type	Categorical	Two-ramp, one-ramp, other, none
	Detectable Warning Surface Color	Categorical	Yellow, other, none
	Crossing Distance	Numeric	Distance measured along crosswalk edge away from intersection, curb-to-curb
	Color	Categorical	White, yellow, decorative pavers, other, N/A
	Pedestrian Signal Head	Binary	Y/N
	Pedestrian Call Button	Binary	Y/N
	Safety Island Width	Numeric	Width of pedestrian safety island, edge-to-edge. If none, "0"
	Advanced Stop	Categorical	Advanced stop bar, advanced yield warning, both, none
	Condition	Categorical	New, partially worn, faded, N/A

The volume characteristics defined in Table 2 give a location and directionality to pedestrian and bicycle counts, and are flexibly defined to allow for a variety of motions. Both approach and intersection/crosswalk volumes are accommodated. Every record in the database is a bidirectional count aggregated to a flexibly defined interval, which typically will take values of 15-minutes or 1-hour. For approach volumes, the directionality of volumes is defined as being either parallel or anti-parallel to the directionality of the approach, which in turn is defined based on the movement direction of motor vehicles along the approach. For crosswalk volumes, crosswalks are defined based upon the node which they abut and the two approaches joining that side of the node. Directional volumes within the crosswalk can thus be defined based upon which approach is crossed first. For bicycle intersection volumes, links are again defined based upon one node and two approaches. However, in this case the approaches are on different sides of the node. To record turning movement counts at an intersection for a given interval, three records would be created – one for straight-through bicyclists, one for left turning bicyclists and one for right turning bicyclists. Wrong-way riding can also be accommodated in this format.

TABLE 2 Fields Included in Volume Database

Table	Field	Data Type	Definition
Pedestrian Volume	Volume Type	Categorical	Approach, Crosswalk
	Count Type	Categorical	Manual, Automated

	Approach ID 1	Key	If approach volume, ID of linked approach. If crosswalk volume, same logic as crosswalk
	Approach ID 2	Key	Linking field for crosswalk volume
	Node ID	Key	Linking field for crosswalk volume
	Volume Direction 1	Numeric	Pedestrians counted during interval in "parallel" direction[1]
	Volume Direction 2	Numeric	Pedestrians counted during interval in "anti-parallel" direction[1]
	Bidirectional Volume	Numeric	Sum of Volume 1 and Volume 2
	Interval	Numeric	Duration of count
	Data Collection Start Time	Date/Time	Start date and time of count
	Weather	Character	Description of weather conditions
Bicycle volume	Volume Type	Categorical	Approach, Intersection (turning movement)
	Count Type	Categorical	Manual, Automated
	Approach ID 1	Key	If approach volume, ID of linked approach. If intersection volume, originating approach.
	Approach ID 2	Key	Linking field for intersection volume
	Node ID	Key	Linking field for intersection volume
	Volume Direction 1	Numeric	Bicyclists counted during interval in "parallel" direction
	Volume Direction 2	Numeric	Bicyclists counted during interval in "antiparallel" direction
	Bidirectional volume	Numeric	Sum of Volume 1 and Volume 2
	Interval	Numeric	Duration of count
	Data Collection Start Time	Date/Time	Start date and time of count
	Weather	Character	Description of weather conditions

[1] For approach volumes, "parallel" means "in the same direction as motorized traffic" and "anti-parallel" means in the direction opposite motorized traffic. For crosswalk (pedestrian) and intersection (bicycle) volumes, "parallel" means from approach #1 to approach #2, and "anti-parallel" means from approach #2 to approach #1.

Geometric Extension

The proposed database is designed to be flexible with regards to the degree of geographic information encapsulated. At the lowest level of detail, no geographic information would be recorded, and all elements would simply be related by relative locations to each other in the network. This approach requires the lowest level of knowledge on geospatial data, but does not allow for complicated spatial queries. It does, however, maintain the network structure. At the next level, the core components could be encoded geographically. Many modern database systems have extensions available for this functionality (e.g. PostGIS in the open source PostgreSQL system), and simply record each record's geometry as an additional field, with the coordinate reference system stored as table-level metadata. Using this method, the nodes would be recorded as points and the approaches would be recorded as lines. Because of the relation to the core components, each of the secondary components could be analyzed spatially in this method as well. Finally, the crosswalks could also be georeferenced. Crosswalks are the only element type that is not related on a to-one basis with either an approach or a node, and therefore cannot be easily represented in a graphical GIS if it is not itself georeferenced. However, for many analyses this should not be necessary as crosswalks are linked with the nodes, which should be a first priority for georeferencing.

PILOT DATA COLLECTION

The core research component of this project was populating the infrastructure database using a subset of the California State Highway System. Data was collected using computer-based data collection methods for 100 miles of highways, of which seven miles were duplicated using field data collection methods. Volume data collection was not included in the pilot data collection for a number of reasons. First, the costs associated with this type of data collection are better known and easier to anticipate. Second, the procedure for collecting volume data is fairly well established, both using manual and automated data collection techniques, and therefore was not the focus of this study. However, previously collected volume data from Caltrans was incorporated into the pilot database to ensure that the format worked as designed.

Computer Data Collection

The computer data collection involved using Google Maps Street View and satellite view, along with a spreadsheet for data entry. For each highway segment, data collectors first traversed the route in Google Maps and defined unique identifiers for every approach and node. These were stored in a layer with all nodes represented as pins and all approaches represented as line segments, labeled by ID. Next, the data collector traversed the highway segment imagery iteratively completing the data tables in groups. First, the node table was filled with all of the names and types of nodes. Next, the approach, sidewalk, buffer, transit, bikeway, and bike parking tables were populated—first for all values that could be captured using satellite imagery and next for all that require street level views. Finally, the crosswalk table was populated.

This iterative data collection approach was chosen as an effective workflow based on various data collectors' experiences. Using a dual monitor setup also helped with the workflow, as data collectors could simultaneously have the data collection form, map interface with labeled approaches and nodes, and map or Street View for data collection open on their screens. An Excel macro was developed within the data collection form to record the time taken to populate each cell of the data collection sheet. A timer ran continuously, recording the amount of time elapsed between each completed cell in a separate sheet. Most of the fields in the data collection form had a pull down menu of values defined to ensure uniformity of data coding and to expedite the data collection process, and thus minimizing typing.

Field Data Collection

The field data collection was conducted by a two-person team. The same approach and node IDs were used as in the computer data collection. A separate set of data collection forms was developed for use in the field with the same structure as the computer data collection forms. However, possible values for each field were abbreviated and printed onto the form so that data collectors could simply circle the value for categorical values. In the field, one data collector was in charge of completing the data collection forms, managing a stopwatch, and directing the overall process, while the other took measurements and made observations of the characteristics to be recorded. All distances were measured with a measuring wheel. Data collectors walked along one side of the highway recording characteristics for that side of the highway and for all crosswalk characteristics, and then recorded characteristics for the other side of the highway on the return trip. Caution should be taken if this process is implemented, as many highways do not offer adequate pedestrian facilities. For a number of highways, a simple "drive-through" observation may be sufficient to verify that no pedestrian facilities exist and to record the presence of any pedestrian-related signage or other such features.

Estimated Time-cost

The observed mileage and estimated time cost for collecting infrastructure data via computer imagery and in the field are shown in Table 3, distinguished by roadway functional classification. The estimated pace for collecting data using each method is calculated based on the total time required. In order to make sure that the data collection protocol accommodates the complex environment shared by non-motorized and motorized road users, urban arterials comprised a major component of the pilot. Intuitively, the time costs associated with collecting pedestrian and bicycle infrastructure data on urban roadways are substantially higher than on their rural counterparts, on account of how many more measurements have to be taken in urban scenarios where pedestrian amenities are more common. For example, collecting data along urban multilane divided non-freeways took approximately 49 min/mi, while collecting the same type of data alongside roadways in rural area took approximately35 min/mi.

By multiplying the total mileage of each class of highway in the entire State Highway System by the estimated pace, overall time estimates are derived. As shown in the last two columns, the total time cost for collecting data for the entire system using computer data collection would be approximately 4,000 hours, while field data collection would be approximately 9,000 hours. These figures do not take into account database creation and maintenance or access/egress time for field data collection. There is also likely a "learning curve" associated with either data collection approach which could result in lower overall costs once data collectors have learned how to optimize their workflow.

TABLE 3 Results of Data Collection Pilot

Roadway Class	Application results				Time cost estimation for entire system		
	Computer collection		Field collection		Mileage of entire system(mi)	Computer (hr)	Field (hr)
	Mileage (mi)	Time cost (min/mi)	Mileage (mi)	Time cost (min/mi)			
Urban freeways	13.17	3	0	1[d]	3,533	157	59
Urban freeways < 4 lanes	0	3[a]	0	1[d]	28	1	1
Urban two lane roads	1.64	40	3.3	50	868	579	723
Urban multilane divided non-freeways	47.1	49	4.0	270	1,081	883	4864
Urban multilane undivided non-freeways	11	29	0	270[e]	176	86	792
Rural freeways	3.5	2	0	1[d]	2,879	101	48
Rural freeways < 4 lanes	0	2[b]	0	1[d]	6	1	1
Rural two lane roads	16.6	7	0	7	12,422	1402	1470
Rural multilane divided non-freeways	4.3	35	0	38[f]	1,125	656	718
Rural multilane undivided non-freeways	0	21[c]	0	38[f]	407	141	260
Total	97.31	-	7.3	-	22,525	4006	8935

Notes:
[a] This value is estimated based on the urban freeways time cost assuming that urban freeways will have the same time cost no matter how many lanes there are.
[b] This value is estimated based on the rural freeways time cost with the same assumption as in note [a]
[c] This value is estimated based on the rural multilane divided non-freeways.
[d] The freeways in both urban and rural areas are all estimated by dividing 1 mile by speed limit (65 mph) because pedestrian facilities are very rare on freeways.
[e] This value is estimated by the assumption that the undivided and divided urban multilane non-freeways will have the same time cost in field.
[f] This value is estimated by the assumption that the time cost of field data collection for multilane non freeways will be 5.4 (time cost for urban multilane-divided non-freeways divided by time cost for urban two lane roads) times the cost for two lane roads.

As can be seen in Table 3, collecting data along urban non-freeway highways takes substantially longer than for other classes of roadway, due to the more complicated pedestrian environments that characterize these scenarios. These results are likely most applicable to cities seeking to collect this data, as most surface roads closely match the characteristics of these highways.

CONCLUSION

This paper introduces a database framework to store pedestrian and bicycle infrastructure and volume data. The database is centered on two core components, nodes and approaches, which are linked to sub-components such as sidewalks and crosswalks based on their physical relationship to each other. The database has been tested in a pilot data collection in which 100 miles of state highways were selected for computer-based data collection, while seven of those miles were also selected for field data collection. The volume database was also tested by importing outcomes from an automated counter. To demonstrate the feasibility of collecting infrastructure data, the time cost for completing the pilot was recorded and the total cost for collecting this data across the entire State Highway System was estimated. The following conclusions are based on the applications of the database.

The database was designed to allow easy updating and records maintenance. The sub-components, which may change frequently, were connected to the nodes and approaches, which are more stable. In other words, pedestrian amenities generally change more frequently than road rights of way. If pedestrian or bicycle infrastructure changes (such as a sidewalk being installed or expanded or a bicycle lane being striped), the relevant table can be updated without any modifications to the underlying approach and node tables.

Another benefit of the proposed database is that the data collection effort can be performed gradually, as opposed to collecting all data in a single project. Cities or regions who maintain large networks can start the data collection on the most important roadway sections and can populate the remainder of the database as time is available. The measurements suggested for inclusion in the database adequately cover pedestrian related facilities to assist agencies in tracking facility coverage. In addition, these measurements are useful for pedestrian related studies and can offer critical information for safety investigations, countermeasure selection, and level of service evaluations. The time estimates for collecting pedestrian infrastructure and volume data indicate that the computer data collection is more feasible than field data collection. Use of computer based imagery is much less costly (over 50% less) in terms of time than field data collection, even excluding site access time. Additionally, data entry and construction can be performed very quickly.

The database framework proposed in the paper offers a plan to store various types of pedestrian and bicycle related data in a central database. However, there are still several improvements that can be implemented for future study. First, the infrastructure and volume data for pedestrians and bicycles must be linked to the vehicle data. For example, in a causal factor analysis for pedestrian traffic safety at intersections, both non-motorized and motorized users' information needs to be attributed to the targeted intersection. Second, while the database has been tested both via computer and in the field, the accuracy of the data collected by the two methods has not been thoroughly compared. A preliminary finding indicates that although field data collection offers the advantage to the collectors of being able to see every corner of a street, the accuracy of the measurements based on computer data collection appears to be fairly high. Third, this database could potentially be implemented as a spatial database for Geographic Information System (GIS) application to support spatial analyses.

ACKNOWLEDGEMENTS
The authors would like to acknowledge the California Department of Transportation for funding this project under the California Strategic Highway Safety Program 08.09. In particular, meetings with the project's Technical Advisory Group led to many fruitful discussions over how to define infrastructural elements, and some members of the advisory team took it upon themselves to test the proposed data collection protocol. Meghan Mitman and Nikki Foletta at Fehr & Peers provided valuable perspective from the consultant point of view. Tony Dang volunteered on the team to help refine the database structure. Finally, a bevy of research assistants conducted the data collection. All views, conclusions, and mistakes in this paper are attributable to the research team and do not necessarily reflect the views of Caltrans.

REFERENCES

1. San Francisco Municipal Transportation Authority. *SFMTA Strategic Plan Fiscal Year 2013-Fiscal Year 2018*. San Francisco, 2013.

2. MassDOT Press Office. MassDOT Announces Mode Shift Goal to Triple the Share of Travel in Massachusetts by Bicycling, Transit, and Walking. http://www.massdot.state.ma.us/main/tabid/1085/ctl/detail/mid/2937/itemid/223/MassDOT-Announces-Mode-Shift-Goal-to-Triple-the-Share-of-Travel-in-Massachusetts-by-Bicycling--Transit-and-Walking-.aspx.

3. California Department of Transportation. District 3 Community Planning. http://www.dot.ca.gov/dist3/departments/planning/communityplanning1.htm. Accessed Aug. 1, 2014.

4. California Department of Transportation. ADA Transition Plan. http://www.adaac.org/secured/asset/caltrans/TransitionReport.aspx?mode=public.

5. Washington State Department of Transportation. *Statewide Bicycle and Pedestrian Facility Inventory*.

6. New Jersey Department of Transportation. County Road Sidewalk Inventory. http://www.state.nj.us/transportation/refdata/countysidewalks/. Accessed Aug. 1, 2014.

7. Maryland State Highway Administration. *Self Evaluation of Access to Public Right of Way Facilities*. 2009.

8. Vanderhawk Consulting LLC. Rancho Cucamonga Sidewalk Inventory Study. http://vanderhawk.net/2011/02/rancho-cucamonga-sidewalk-inventory/.

9. Alta Planning + Design. *Berkeley Pedestrian Master Plan*. 2010.

10. City of Alexandria. *City of Alexandria Comprehensive Transportation Master Plan*. 2008.

11. Federal Highway Administration. *Traffic Monitoring Guide*. US Department of Transportation, 2013.

12. Huff, H., and M. Brozen. Creating the Bicycle Count Data Clearinghouse for Los Angeles County, California. 2014.

13. Nordback, K. Online Non-motorized Traffic Count Archive. http://otrec.us/project/817.

14. Delaware Valley Regional Planning Commission. Pedestrian & Bicycle Counts. http://www.dvrpc.org/webmaps/pedbikecounts/. Accessed Aug. 1, 2014.

15. Bike Arlington. Bicycle & Pedestrian Counters. http://www.bikearlington.com/pages/biking-in-arlington/counter-dashboard/. Accessed Aug. 1, 2014.

16. City of Seattle. Downtown Seattle Pedestrian Counts Interactive Graph. http://www.downtownseattle.com/resources/pedestriancountgraph/. Accessed Aug. 1, 2014.

17. Transportation Research Board. *Highway Capacity Manual*. 2010.

18. Center for Occupational and Environmental Health. Pedestrian Environmental Quality Index. http://www.peqiwalkability.appspot.com/home.jsp. Accessed Aug. 1, 2014.

19. Schneider, R. J., T. Henry, M. F. Mitman, L. Stonehill, and J. Koehler. Development and Application of a Pedestrian Volume Model in San Francisco, CA. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2299, 2012, pp. 65–78.

20. Liu, X., and J. Griswold. Pedestrian Volume Modeling: A Case Study of San Francisco. *Yearbook of the Association of Pacific Coast Geographers*, Vol. 71, 2009, pp. 164–181.

21. Strauss, J., and L. F. Miranda-Moreno. Spatial modeling of bicycle activity at signalized intersections. *Journal of Transport and Land Use*, Vol. 6, No. 2, Aug. 2013, p. 47.