Pedestrian and Bicycle Information Center

White Paper Series

Crowdsourcing Pedestrian and Cyclist Activity Data

January 2015

Amy Smith, Fehr & Peers

For:
Federal Highway Administration
DTFHGI-11-H-00024

Pedestrian and Bicycle Information Center

This material is based upon work supported by the Federal Highway Administration under Cooperative Agreement No. DTFH610110H-00024. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the Author(s) and do not necessarily reflect the view of the Federal Highway Administration.

Executive Summary

Collecting cyclist and pedestrian activity data can be challenging due to data gaps and unique characteristics of active transportation that set it apart from other modes. Technological advances continue to improve data collection for demographic, infrastructure, and motorized travel data, much of which is made available online. However, active transportation data resources that can be used to better understand detailed spatio-temporal (location and time-based) travel patterns and personal experience may be absent or difficult to find. Crowdsourcing, or the process of obtaining information, insight, and knowledge from user-generated data provided through web and mobile applications, can help address these data gaps efficiently. In addition to increased data availability, crowdsourcing offers many additional benefits to active transportation planning, such as broad and diverse perspectives, local knowledge, data timeliness, and direct dialogue between planners and those affected by planning decisions. Existing crowdsourced datasets can serve as resources for many types of active transportation planning projects. New crowdsourcing applications developed for a specific use or planning implementation can produce high-quality, project-specific data about cyclists and pedestrians while at the same time opening lines of communication between planners and those contributing data. Data providers including public agencies, open source projects, and commercial developers are making data easier to find, and in some cases these resources are freely available. When used carefully and creatively, user-generated cyclist and pedestrian data has great potential to inform and add value to active transportation planning projects.

This paper considers how crowdsourcing applications and crowdsourced data are currently being applied, as well as potential new uses for active transportation research and planning efforts of various types. The objectives of this white paper are to review existing crowdsourced bicycle and pedestrian data resources and crowdsourcing tools; discuss potential planning implementations of crowdsourced data for a sample of bicycle and pedestrian project types; and provide examples of how crowdsourcing is currently being used to inform decision-making. Potential issues related to crowdsourced data are also considered (e.g., quality, privacy concerns, participation rates, bias). The research presented here highlights a decreasing skepticism over the quality of volunteered, user-generated data

provided by amateurs (as opposed to professionals) in light of a desire to open the lines of communication between the planning world and those affected by planning decisions, directly addressing (rather than being discouraged by) data limitations. The initiatives surrounding progressive data collection, management, and analysis are further reflected in the numerous conferences, meetups, and other events fostering collaboration between planners, developers, data scientists and others interested in applying critical thought and innovation in planning *(1)*. While this paper reviews existing crowdsourcing techniques and their current applications in planning, the pace of technological change and rate of adoption in planning indicates that planners will continue to develop and apply innovative approaches like crowdsourcing in response to continually changing community needs. This paper focuses on examples of current uses of crowdsourced data, crowdsourcing data suggestions, and data considerations.

Introduction

As active transportation research advances and grows, so does the need for high-quality active transportation data in greater quantities and at finer spatial resolutions to inform planners about the relationships between cyclists and pedestrians and their built environments *(2)*. Crowdsourcing tools and crowdsourced data are resources that can help planners understand travel behavior trends; inventory and assess the built environment; make modal and navigational recommendations; identify improvement opportunities; and forecast future needs and conditions. Spurred in large part by advances in social computing technologies *(3)*, many crowdsourcing tools are emerging from efforts led by tech-savvy planners to find creative ways to engage communities and gather information *(1)*. There are many sophisticated data collection and sharing efforts, synthesis techniques, and applications that incorporate crowdsourcing, such as open data, big data, and civic technologies (see "Popular Terms and Descriptions" table below for additional information). For example, web mapping applications are increasingly being used to solicit input from the online community about desired locations for new bicycle and pedestrian facilities, as well as needed improvements. These new developments in data collection are being harnessed and integrated into planning projects focusing on facilitating mobility, managing resources, improving health and safety, and meeting other objectives of active transportation planning.

Table 1. Popular Terms and Descriptions

Big Data	Large datasets that may require advanced processing, data synthesis techniques, and integration with other datasets to produce meaningful metrics and support decision making.
Civic Technologies	Web apps, mobile apps, and APIs developed to address community issues and improve communication between governments and their constituents, often by volunteers and non-profit organizations.
Collaborative Consumption/Sharing Economy	A term used to describe an observed consumer trends in renting, lending, and sharing goods and information, as well as the technologies and businesses that facilitate and reinforce these trends.
Crowdsourcing	The process of obtaining information, insight, and knowledge from user-generated data provided through web and mobile applications, often to address a specific issue or solve a problem.
Open Data	Public and private datasets made freely available for use without restrictions, such as datasets shared on www.data.gov.

Conferences, meetups, and other events are occurring in cities around the world to foster collaboration between planners, developers, data scientists, and others interested in applying critical thought and innovation in planning. Planners' strengths in research methods, citizen engagement, project management, and spatial analysis can help drive the development of successful civic applications *(1)*. At the same time that planners are beginning to embrace community-driven data, a recent growth of companies dedicated to facilitating the exchange of goods and services between individuals has been described in terms of an overall "spirit of sharing". In his *Wired Magazine* article "Trust in the Shared Economy," Jason Tanz describes an economic and cultural movement toward entrusting others with our possessions and experiences *(4)*. In her 2010 Ted Talk, author Rachel Botsman uses the term "Collaborative Consumption" to describe a broad cultural shift toward the adoption of crowdsourcing and peer-to-peer sharing systems enabled by real-time, location based applications *(5)*. This growing acceptance of crowdsourcing applications, open data, big data, civic technologies, and peer-to-peer sharing has many promising and yet-to-be-discovered implications for the field of active transportation planning.

This paper considers how crowdsourcing applications and crowdsourced data are currently being applied, as well as potential new uses for active transportation research and planning efforts of various types. The objectives of this white paper are to review crowdsourced bicycle and pedestrian data resources and crowdsourcing tools; discuss potential planning implementations of crowdsourced data for a variety of bicycle and pedestrian project types; and provide examples of how crowdsourcing is currently being used by the planning community. Due to software application turnover, many of the examples provided describe tools that may no longer be in use, have evolved significantly, or have been/will eventually be depreciated with the advance of new technologies. This paper is not intended to be a comprehensive outline of crowdsourcing applications in the transportation planning profession or a dictionary of crowdsourcing system types, but rather a resource for those interested in using crowdsourcing systems in active transportation planning and research.

Definitions of Crowdsourcing

The definition of crowdsourcing has evolved since its introduction as a term in a 2006 issue of *Wired Magazine (6)*. Broad usage of the term in a variety of fields has resulted in many flexible interpretations. Several variations on the term are described by Enrique Estellés-Arolas and Fernando González Ladrón-de-Guevara in their 2012 paper on the topic *(7)*. Many descriptions of crowdsourcing incorporate the value of knowledge compiled from a large group of diverse perspectives as compared to more traditional means of data collection, such as Brabham's description of crowdsourcing as "a strategic model to attract an interested, motivated crowd of individuals capable of providing solutions superior in quality and quantity to those that even traditional forms of business can" *(8)*. While the participation of the crowd plays a role in many definitions of crowdsourced data, recent advances in big data aggregation have provided new data access possibilities through implicit crowdsourcing techniques that repurpose large user-generated datasets collected for other intents. For example, many personal fitness mobile applications offer data through Application Programming Interfaces (APIs) or aggregated data products for purchase. While these datasets are generated by a large online community, end users might not always be aware of where, why, and how their personal data is being used. While some aspects of user data remain protected, issues surrounding location data privacy are emerging and beginning to be addressed *(9)*.

This paper uses a broad definition of crowdsourcing that includes: 1) actions explicitly taken to gather input from the online community, as well as 2) implicit re-use of data collected for other purposes. See the "Crowdsourcing Data Types" section for a more detailed breakdown of crowdsourcing categories discussed in this paper.

Benefits of Using Crowdsourced Data

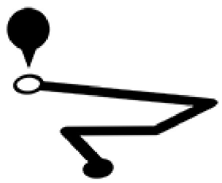
Spatial data related to active transportation is often limited, difficult to find, or in a less-than-ideal format for use in spatial analysis and other planning applications. Some studies may even be limited in scope and geographic reach due to sparseness or lack of data. Effective crowdsourcing data collection and aggregation efforts can be used to address the issue of data availability while providing perspectives from a variety of demographic and geographic regions on physical infrastructure, personal travel experiences, community priorities, and other topics. For example, while gathering, organizing, and visualizing base map data was once a time-consuming effort, base map services for use in crowdsourcing maps and mobile applications are easily accessible, and some are regularly updated by online volunteers, such as Open Street Map (OSM). A 2010 study comparing OSM data with Ordnance Survey data in England found a high degree of spatial accuracy in the London area *(10)*.

The advantages of a large pool of voices and perspectives in planning for and improving communities are not easily quantified but commonly sought and greatly valued once those perspectives have been collected. While crowdsourcing has limitations, it is one strategy for engaging a wide range of stakeholders in active transportation planning. Additionally, direct communication between planners and communities helps bring transparency to planning projects *(1)*. Through identification of potential participants and effective promotion, crowdsourcing efforts implemented for specific planning or research purposes can help inform a community and engage those involved in public and political dialogue *(2)*. Successful crowdsourcing implementations can also improve response rates as well as public perception of a planning project *(11)*.

Other benefits of using crowdsourcing in active transportation include the minimization of time and location constraints *(11)* and integration of crowdsourced geospatial data into advanced spatial analysis *(2)*. Crowdsourcing tools may be less expensive than alternative data collection methods such as surveys and bicycle and pedestrian counts, while producing valuable information that automated data collection methods might not capture *(12)*. The following sections outline various types of crowdsourced data with examples of their applications in active transportation projects.

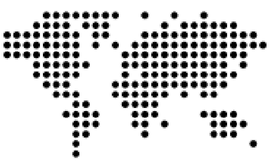
Crowdsourced Data Types

This paper examines four types of crowdsourced data with a focus on understanding how these data types apply when answering questions and solving problems in a planning context. The four types described below are in-situ data, thematic data, thumbtack data, and spatial inventory data. While these data types are not mutually exclusive, the typology provides a useful framework for discussing crowdsourcing options for planners.



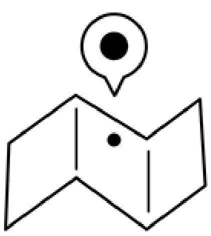
In-Situ Data

In-situ data sources include mobile applications that record real-time, geotagged data points representing the travel patterns of an individual. Examples include applications such as **Strava**, **MapMyFitness**, and **Moves**, as well as tools developed specifically for health and research purposes, such as a 2007 study on physical activity and neighborhood design in which subjects recorded their activity using accelerometer monitoring *(13)*.



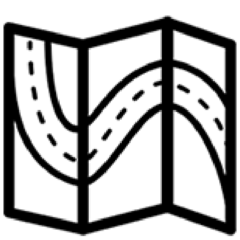
Thematic Data

Thematic data sources include data aggregated, categorized, and/or summarized within pre-defined geographic areas. These data sources may feature text-based data with active transportation characteristics linked to a particular place, such as city, county, Census tract, or transportation analysis zone. Decennial Census, American Community Survey, National Household Travel Survey, the EPA's Smart Location Database (SLD), and Census Transportation Planning Products (CTPP) are all examples of thematic data sources. While not typically classified as crowdsourced data, thematic data from these sources exhibit similar qualities, such as a large and diverse pool of contributors.



Thumbtack Data

Thumbtack data sources include point locations added to a map with associated attribute information. Examples include web mapping applications that serve as surveys to gather input from the online community, such as the suggestion map developed by Divvy, Chicago's bicycle sharing system, to solicit recommendations for new bicycle share station locations (http://suggest.divvybikes.com).



Spatial Inventory Data

Spatial inventory data includes digitized representations of ground features, often with associated attributes. Spatial inventory data can be greatly enhanced through crowdsourcing techniques that harness the knowledge of locals and enthusiasts. Examples include Open Street Map and Cyclopath.

Other Crowdsourcing Classifications

Data sources within the four crowdsourced data types can be further categorized into system types as described by Misra, Gooze, Watkins, Asad, and Le Dantec in their 2014 paper on crowdsourcing transportation data *(11)*. These systems include the following:

Explicit/Implicit Systems

While explicit systems are designed to gather data and feedback regarding a defined problem that is communicated directly to the participants, implicit systems often involve collecting data crowdsourced for other purposes and then re-used. Examples of explicit systems include web mapping applications designed for a specific purpose, wherein users are aware of why and how their contributions will be used. Re-use of crowdsourced data, in which the participants may be unaware of the secondary uses of their data, is characteristic of implicit systems such as in-situ GPS data accessed via personal fitness APIs.

General Purpose/Domain Specific Systems

General purpose crowdsourcing systems do not require specialized knowledge on the part of the participant. In contrast, domain specific crowdsourcing systems are designed to gather data from participants with an existing level of expertise. The use of general purpose vs. domain specific systems depends on the needs of the project. Some crowdsourcing efforts might benefit from a broad, general pool of users, whereas others might require their participants to have background knowledge or expertise.

Audience-Centric/Event-Centric/Geocentric/Global Systems

Audience-centric crowdsourcing systems require participants to respond in the same place at the same time. In event-centric systems, time constraints are applied; however participants can be in different places. If a time range is ongoing but the study is focused on a particular location, the system is known as a geo-centric system. A global crowdsourcing system is characterized as having no temporal or spatial limitations.

Planning Applications of Crowdsourced Data

Planning for Facility Improvements

Active transportation planning is informed by knowledge of existing conditions; identification of new or improved facilities and services; and the interests of the community. Location-based data generated from route tracking apps as well as data collected using map-based survey tools can provide helpful information about the interactions between cyclists and pedestrians and their built environments. These crowdsourcing systems can also be used to identify potential areas of concern. Whether applied in systematic spatial and statistical analyses or through interpretation of personal experience and anecdotal input from participants, crowdsourced data can be collected and assessed in a variety of ways to serve the objectives of the project or research topic at hand.

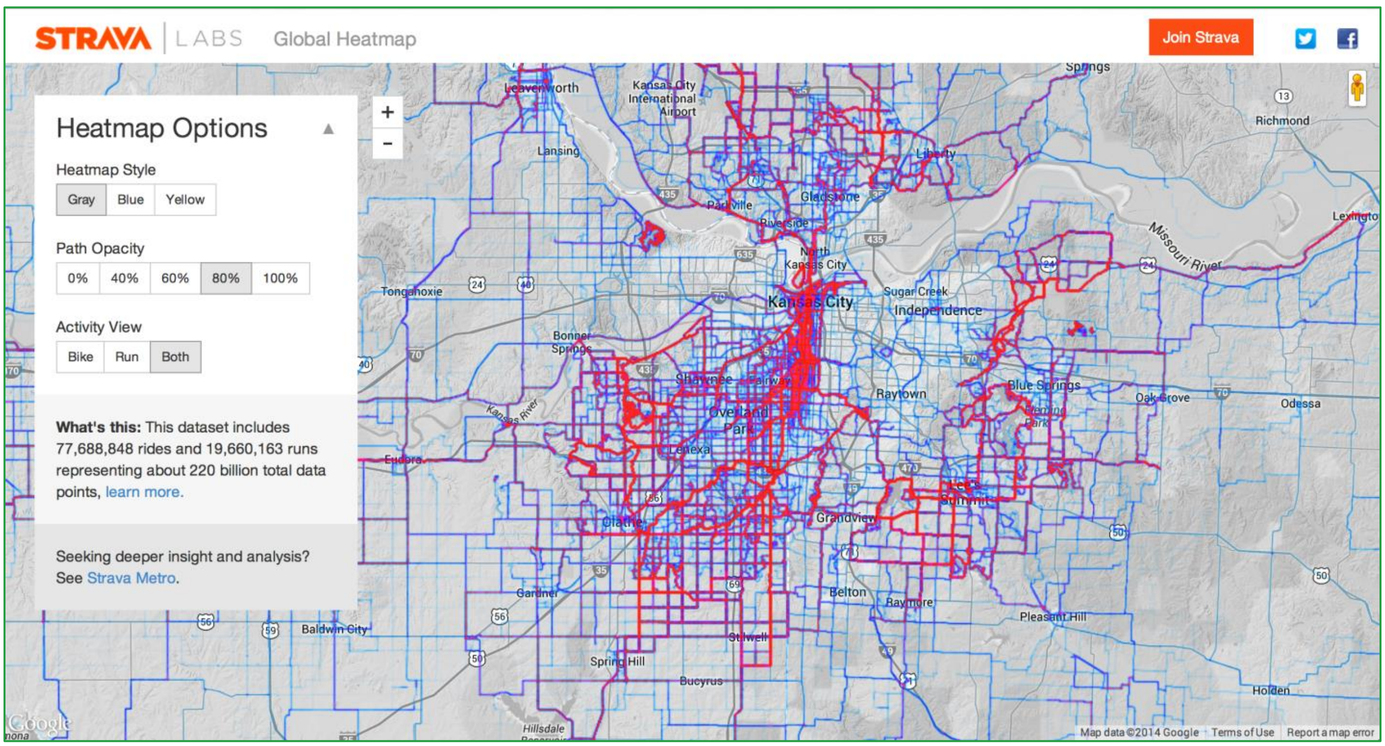
Example project types include:

- Bicycle / Bikeway Master Plans
- Bicycle Share Maintenance and Planning
- Pedestrian Master Plans
- General Plan / Mobility Element Performance Measures
- Non-motorized Access Improvements
- Bicycle and Pedestrian Circulation Studies
- Gap Closures
- Campus Plans
- Bicycle Storage
- Program / Project Evaluation: Before and After Studies
- Transportation / New Development Impact Assessments
- Striping
- Signage
- Pedestrian Overpasses / Underpasses
- Facility Sizing

Examples and Suggestions:

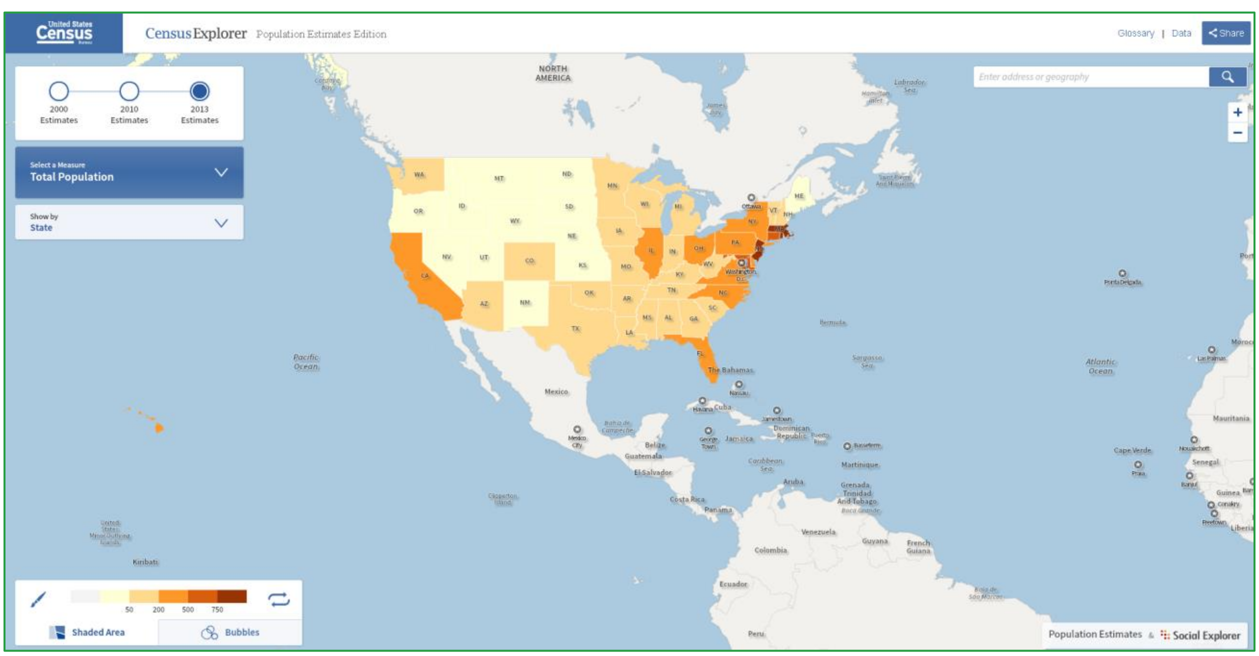
In-situ

Implicitly collected in-situ data can provide insight into facility usage and activity levels through aggregation of user data from mobile applications. For example, Strava, a San Francisco based mobile application company, has developed an application that lets users track their bicycle rides and runs using GPS. In addition to the mobile application that users can download to smart phones and other mobile devices, Strava offers a data product for purchase called "Strava Metro" that includes Shapefiles and database files (DBFs) containing aggregated user data. These data can be viewed and analyzed in desktop GIS environments. The public can explore visualizations of compiled datasets on the Strava Labs website (see screen capture below from http://labs.strava.com/heatmap).



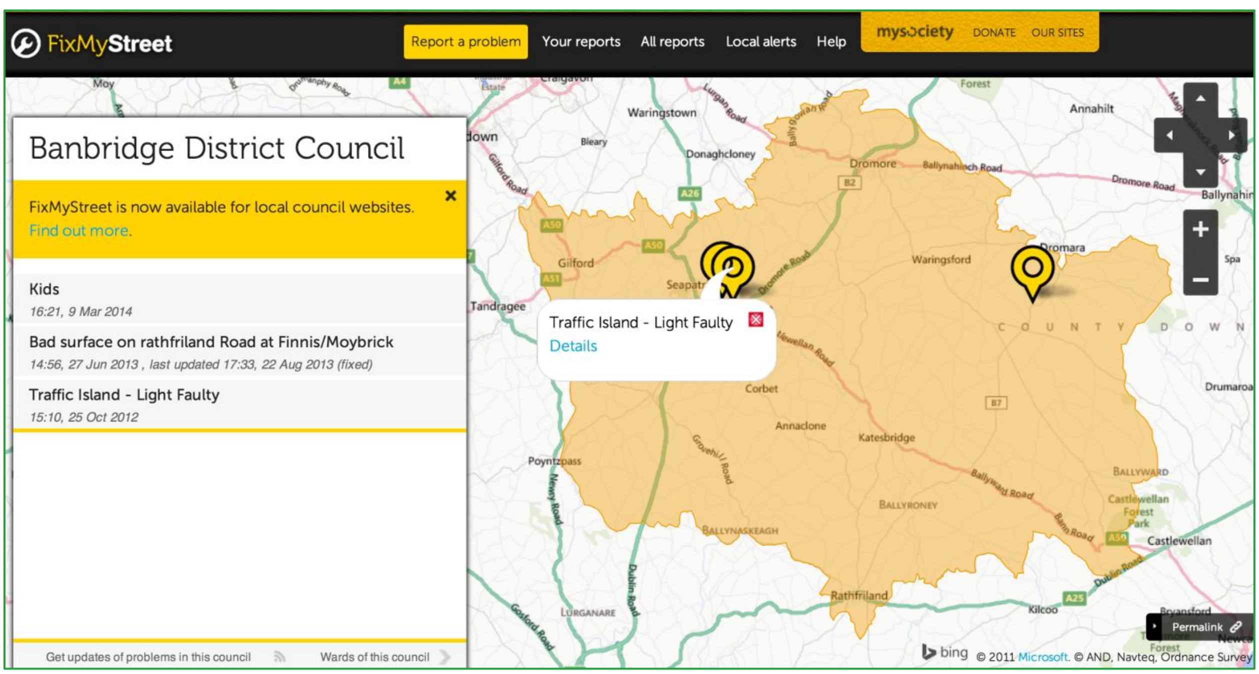
Thematic

Thematic datasets from the US Census, travel surveys, the Environmental Protection Agency (EPA), and other resources are not traditionally thought of as “crowdsourcing” but share some common characteristics, such as sourcing information from a broad and diverse range of contributors. Thematic data can be used in conjunction with other crowdsourced data types to paint a more complete picture of current travel behaviors, physical infrastructure, and anticipated future changes affecting transportation. For example, in-situ and thumbtack data can be overlaid with demographic data for visual inspection of the characteristics in the vicinity of logged travel patterns and community feedback points. Advanced spatial analysis techniques such as hot spot analysis and geographically weighted regression can be used in desktop GIS applications to examine the statistical relationships between crowdsourced data and geospatially aggregated characteristics. In addition to spatial analysis, thematic data from authoritative and comprehensive data sources such as the US Decennial Census can be used to validate travel behaviors observed in crowdsourced data. Web mapping services such as the Census Data Explorer (screen capture below from http://www.census.gov/censusexplorer/censusexplorer-popest.html) provide visualizations of thematic data for quick access to data. Thematic data from the US Census *(14)* and EPA *(15)* can also be downloaded for use in GIS desktop software.

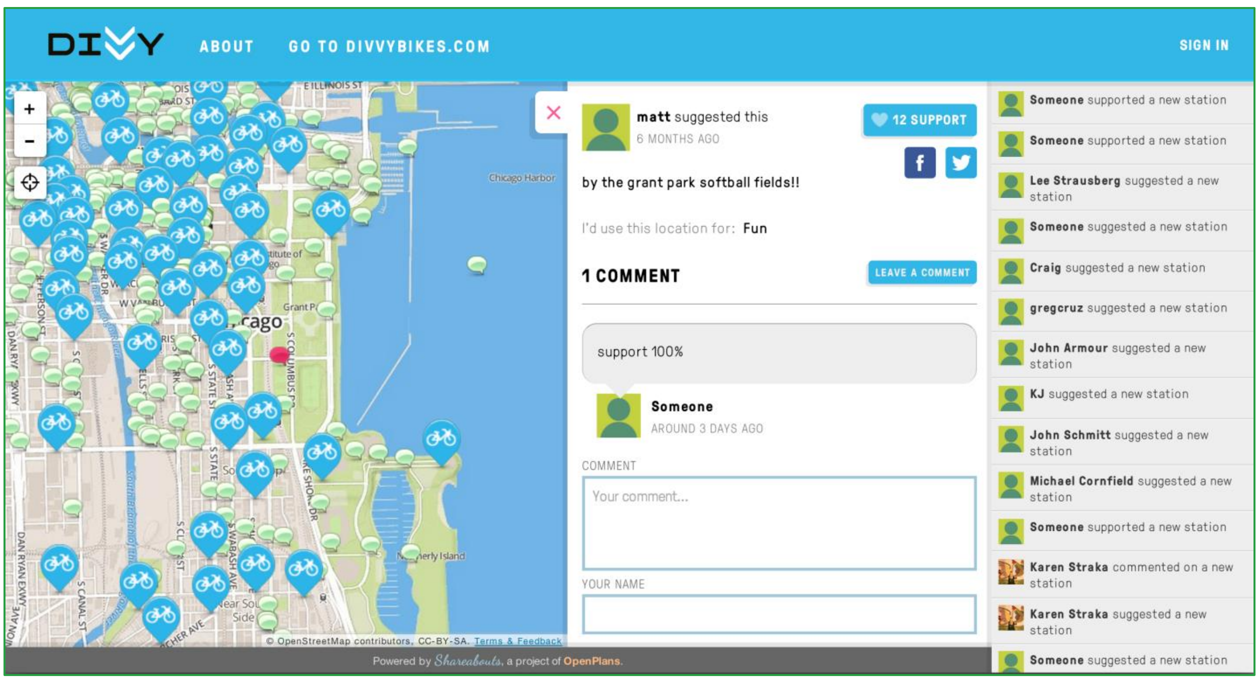


Thumbtack

Thumbtack data gathered using crowdsourcing techniques like web mapping applications can be used to identify needed improvements, current facility uses, and desired changes. For example, the FixMyStreet application developed in the United Kingdom has been used by citizens to report street-related problems. When users report problems through the application, notifications are sent to the local jurisdictions. Although many places implement automated monitoring systems, FixMyStreet has been used to identify problems that may not be immediately detectable through automated means. Similarly, the PEDS advocacy group in Atlanta hosts a website (www.peds.org) that serves as a venue for citizens to report issues such as faulty sidewalks and roadway crossings (16). While the concept is promising, in practice one concern with reporting apps such as FixMyStreet is that reports are not always directly addressed by responsible government bodies *(12)*. This illustrates the importance of designing a system that not only collects data but also engages those who are receptive and capable of making change. Below is a screen capture of the FixMyStreet user interface showing a report of a faulty street light at a traffic island from www.fixmystreet.com.



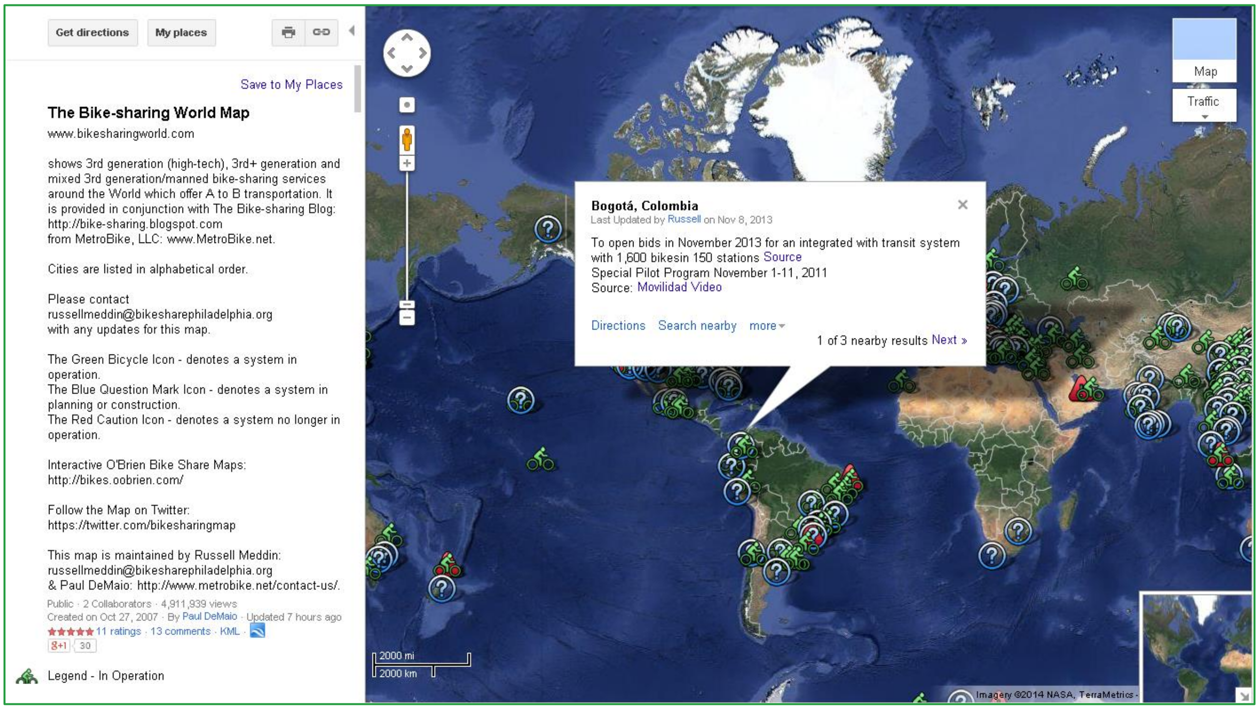
Another example of thumbtack data in active transportation crowdsourcing applications is the web mapping application developed by Divvy, Chicago's bicycle share system, to gather new bicycle share station location suggestions. The application, created using Shareabouts, a web-based tool designed to gather public input, allows users to either add thumbtack locations to the map or respond to locations added by others through comments or by clicking on a "support" button. Sean Wiedel, a transportation planner with the Chicago Department of Transportation involved in the crowdsourcing effort with Divvy, describes the process of gathering public feedback via the suggestions map as an effective ongoing planning tool for the city's bicycle share system. A screen capture below from http://suggest.divvybikes.com shows the web map of existing Divvy stations and a sample of the user-generated content. Used in conjunction with other public engagement strategies, the crowdsourcing map helped city planners identify key locations for new bicycle share stations, especially in areas that might not have been covered via the other strategies. Below is a screen capture of the suggestion maps website. See the project highlight "Divvy Bicycle Share: People Know Their Neighborhoods Best" for more details on how crowdsourcing and open data are being used in active transportation planning in Chicago *(17)*.



Spatial Inventory

Crowdsourced spatial inventory data such as Open Street Maps and the Bicycle Sharing World Map are sources of existing physical infrastructure and facilities. These resources are built from knowledge of existing conditions contributed by locals, travelers, and volunteers, often digitized using satellite and aerial imagery provided through publicly available base map services, such as Google Maps (www.maps.google.com), Esri (www.esri.com), and Mapbox (www.mapbox.com). The screen capture below from http://www.bikesharingworld.com shows the Bicycle Sharing World Map, a Google Map showing bicycle share stations around the world. Users can add stations by emailing Russell Meddin and Paul DeMaio who maintain the map *(18)*. If updated regularly by the moderators and volunteer contributors with the most current bicycle share locations, planners and bicycle share users alike can refer to this map for relevant information on bicycle share stations worldwide. The map was recently listed as a resource of bicycle share information in a white paper published by the University College London (UCL) Centre for Advanced Spatial Analysis titled "Bicycle sharing systems – Global trends in size" *(19)*.

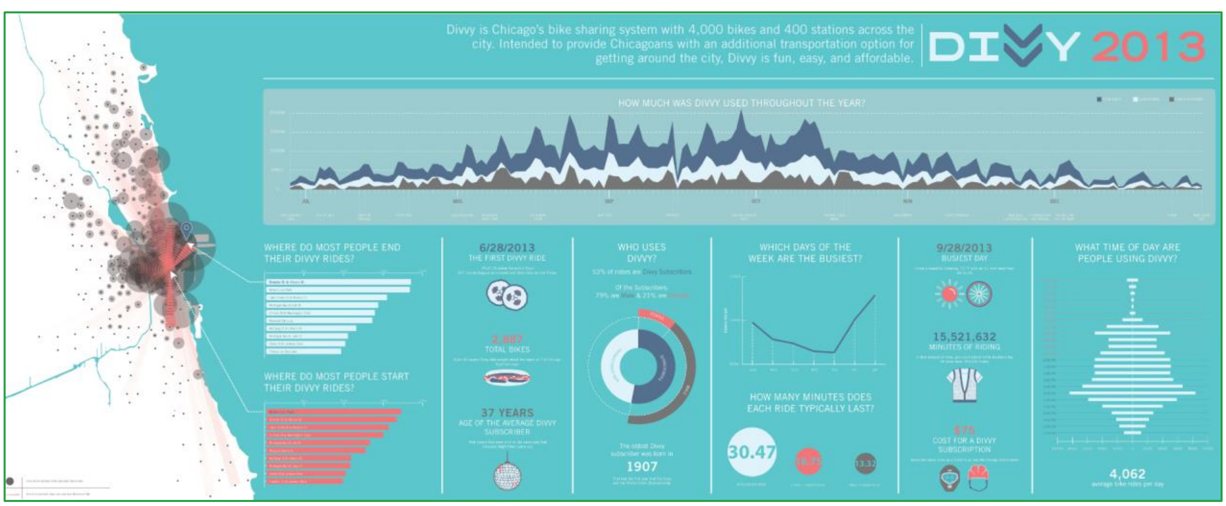
While very useful if up to date, crowdsourced spatial inventory datasets may quickly fall out of date if not well maintained and may be useful only as snapshots of conditions. This raises issues of participation rates and retention, which are major challenges with long-term crowdsourcing efforts and may eventually lead to the temporary suspension, termination, or abandonment of a crowdsourcing project. See the Considerations section of this paper for more on this and other crowdsourcing challenges.



Divvy Bicycle Share: People Know Their Neighborhoods Best

The Divvy suggestions map is a web mapping application developed by Divvy Bicycle Share in Chicago to gather input from the public on new bicycle share station locations. Developed to help address difficulties in gathering feedback about bicycle share via public meetings, which tended to be lightly attended, the suggestions map was one of many strategies Divvy used to address this issue. The suggestions website was used in conjunction with meetings with elected officials, user surveys, and popup meetings at functions such as farmers' markets, school board meetings, and other community meetings. Respected community organizations were also consulted in neighborhoods where Divvy expansions were planned. Divvy had previously launched a website based on a similar platform to the current version and received a strong response, which encouraged the Divvy planners to use this approach again. The suggestions map has been particularly helpful for planning new stations in less dense areas, where residents have been putting in requests for stations in their communities. This helped planners get a sense of the feasibility of new stations in locations both in and outside of the city.

In addition to their crowdsourcing efforts, Divvy also maintains a database of system information. Much of this data has been made available to the public through venues such as the Divvy Data Challenge (https://www.divvybikes.com/datachallenge). The challenge was designed to give people an opportunity to use the Divvy system data and present it in an interesting, creative and fun way. The challenge had about 100 entries, including Jessica Baker, who won the prize for best overall visualization (screen capture below). In addition to the data challenge, others have tapped into the challenge data as well as live data streams from Divvy to develop their own applications.



Forecasting Cyclist and Pedestrian Demand

Assessing geographic areas and facilities (i.e., street segments, intersections, and points of interest) in terms of their potential to attract and facilitate bicycle and pedestrian activity can be accomplished through forecasting techniques such as regional modeling with non-motorized travel components, as well as geospatial analyses and regression processes that estimate latent demand. High spatial resolution data about travel patterns and the built environment can improve forecasting of active travel modes. Furthermore, detailed walk and bicycle information can be used to model how factors such as neighborhood characteristics relate to traveler mode choices *(20)*. The following sections describe how crowdsourced data can be applied in forecasting with examples of projects that have successfully used crowdsourcing in this context.

Example project types include:

- Bike Share Ridership Forecasting
- Bicycle and Pedestrian Volume Forecasting
- Bicycle Facility Feasibility Studies

Examples and Suggestions:

In-situ

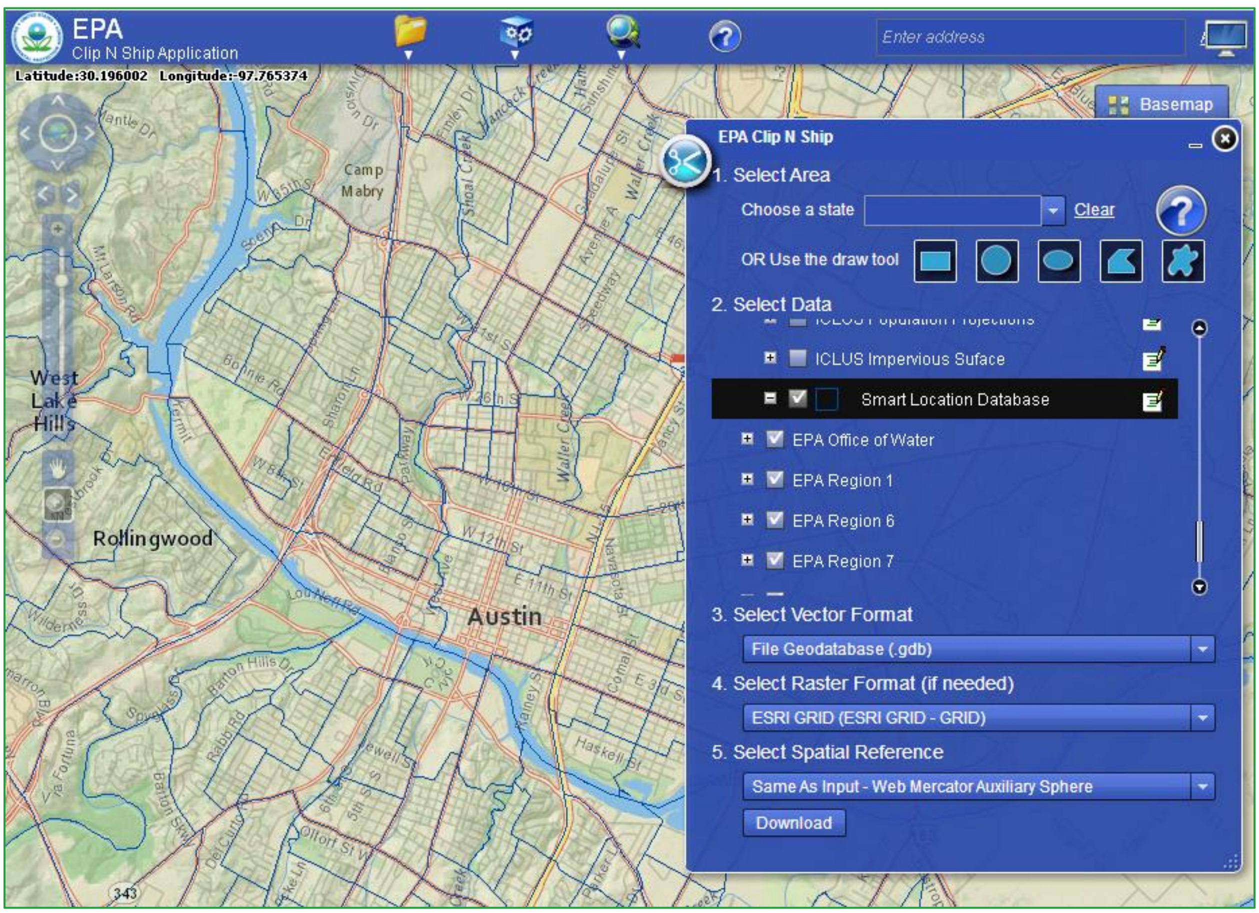
While in-situ crowdsourced data can provide large quantities of information (for example, data collected implicitly from mobile applications that offer APIs or aggregated data products), explicitly collected in-situ data contributed by a smaller but more informed audience can also be beneficial. For example, the San Francisco County Transportation Authority (SFCTA) reached out to smartphone users in San Francisco to provide their in-situ travel data via a mobile application called CycleTracks. Data collected by users was applied in the development of a route choice travel demand model in the city. Users can download the free, open-source application and record GPS data points along their routes of bicycle travel. Users also have the option of providing personal characteristics and trip purposes. SFCTA used this information in conjunction with travel records and built environment data as an input into the SFCTA's route choice model. The users' voluntary information helped in bias testing and expanded the responsiveness of their model to cyclists' needs for both utilitarian and recreational travel *(21)*. To address issues of sample size and bias, SFCTA used cyclist and pedestrian counts from a citywide network of sensors as well as on-the-ground counts to validate the model *(22)*. The CycleTracks app filled a gap in travel pattern data for cyclists for a specific location and purpose.

Since launching the program in 2010, numerous agencies and municipalities around the US and Canada have applied CycleTracks, or a rebranded version of the app, in their locales (e.g., Atlanta, GA; Austin, TX; Monterey, CA). The source code for CycleTracks is made available through GitHub, a web-based code repository. SFCTA also shares anonymized data with public agencies for further use *(22)*. Although this type of specialized crowdsourced data requires careful consideration when applied in other settings, finding new and creative ways to use this data to inform decisions is possible..

Thematic

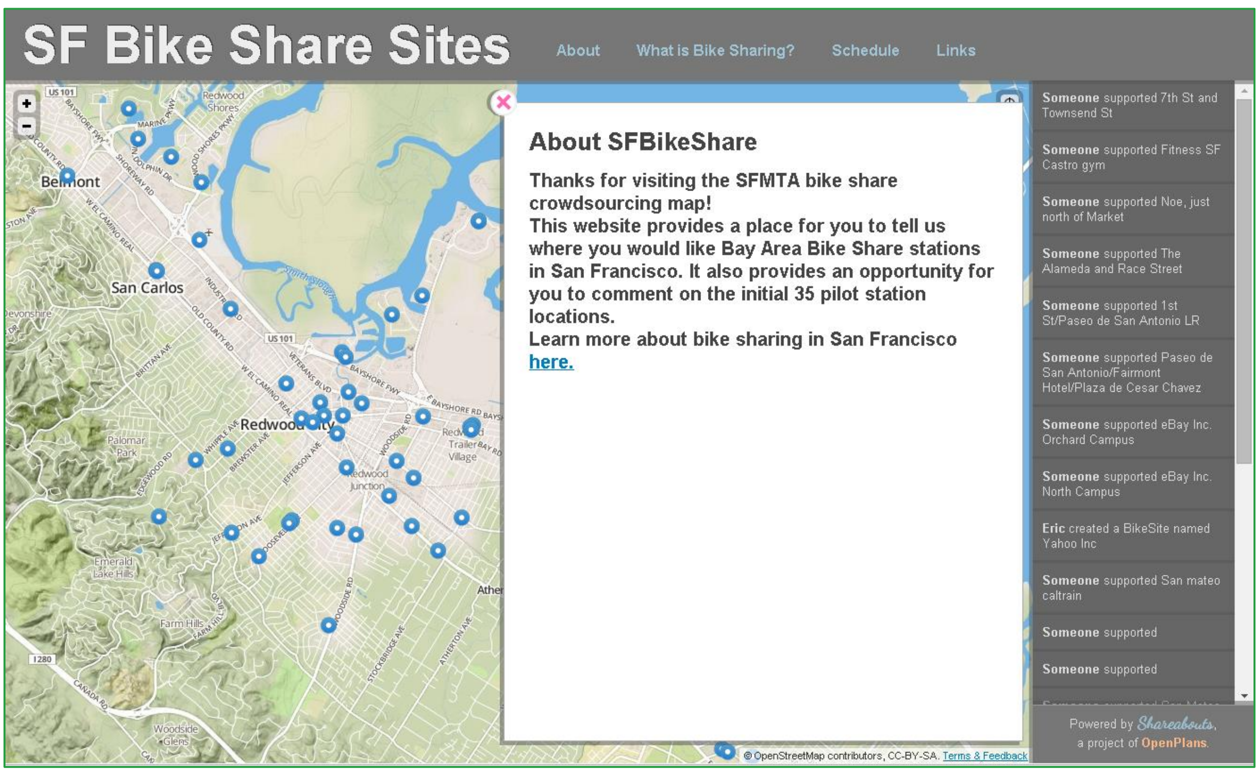
Thematic data from the US Census and travel surveys are commonly applied in forecasting models to develop relationships between existing demographics, land uses, and travel behaviors and estimate how travel patterns could change along with anticipated economic and demographic shifts, as well as planned and proposed land use and built environment changes. These characteristics, commonly referred to as the “D’s” (Density, Diversity, Design, Destination Accessibility, Demographics and Distance to Transit) are often compiled and associated with a variety of geographies depending on the available resources and the forecasting needs. For example, the EPA’s Smart Location Database (SLD) includes more than 90 attributes describing the D characteristics for US Census Block Groups nationwide. The SLD variables were developed from several sources, including the US Census, the Longitudinal Employer-Household Dynamics (LEHD) program, and the General Transit Feed Specification

(GTFS), among various individual agencies and jurisdictions *(24)*. The database is available for download on the EPA's website in multiple formats, including Shapefile, database format (DBF), and Esri Geodatabase. Web mapping applications hosted by the EPA can also be used to view and download the data. The screen capture below from https://edg.epa.gov/clipship shows the "Clip N Ship" application, a web mapping application that facilitates finding and downloading EPA data, including the SLD. Web services are available to the public for incorporation of the SLD variables into custom web mapping applications *(15)*. Thematic datasets such as the SLD can be used in conjunction with crowdsourced data to forecast bicycle and pedestrian activity, to validate active transportation trends derived using crowdsourcing techniques, and to visually communicate transportation data.



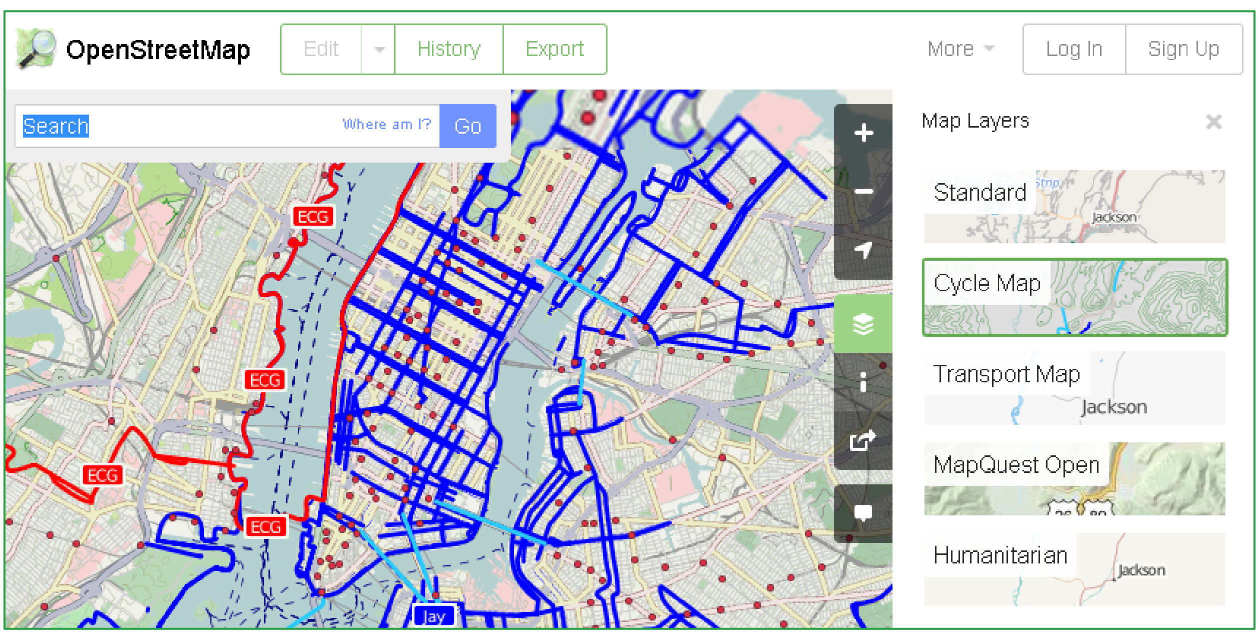
Thumbtack

Thumbtack data collection through web mapping applications can be used to help gauge the feasibility of new bicycle and pedestrian facilities. For example, a recent study in Redwood City, California, used a Bay Area bicycle share map hosted by the San Francisco Municipal Transportation Agency to consider feasible options for two new station locations *(25)*. The screen capture below shows the SF Bicycle Share Sites web map focused on Redwood City suggestions *(26)*. This web map is an example of how a crowdsourcing application developed to gather input for a specific purpose (in this case, gathering feedback on bicycle share station locations in the entire Bay Area) can serve as an effective open data resource for other projects. Although this type of implicit crowdsourcing is limited in terms of domain-specific responses (contributors of implicit crowdsourcing approaches are typically not aware of the secondary uses of their data), the data can still prove useful and cost-effective, especially when limitations are acknowledged and addressed through other means. See the Considerations section of this paper for more on this and other crowdsourcing challenges.



Spatial Inventory

Crowdsourced spatial inventory data can be applied in forecasting studies to incorporate classifications, distributions, densities, and other spatial frameworks into forecasting and regression models. Efforts to estimate demand for bicycle and pedestrian facilities can be enhanced by detailed and up-to-date information on roadways, paths, schools, parks, and other points of interest in the study area. Although applications such as Open Street Map (OSM) are highly dependent on the degree to which volunteers have digitized spatial content and related characteristics, assessments of this particular resource have shown that some regions contain highly accurate user-contributed data. A study reviewing the completeness of OSM data for bicycle facilities (i.e., trails and designated lanes) found that OSM data is relatively accurate and can supplement road network data that tends to lack cycling information, such as data provided by the US Census Bureau as well as commercial data providers such as NAVTEQ and TomTom *(10)*. The screen capture below from www.openstreetmap.org shows bicycle facilities include in the OSM "Cycle Map" view. OSM data can be freely downloaded from the web mapping interface or other websites that provide batch downloading capabilities, such as Geofabrik's OpenStreetMap Data Extracts download server *(27)*.



Planning for Healthy, Safe and Sustainable Communities

The relationship between health, safety, sustainability, and the built environment is a growing topic of interest for planners, engineers, health care professionals, advocacy groups, and research organizations. Understanding how cyclists and pedestrians interact with their environment, as well as community concerns, is integral to improving conditions and planning for the future. In 2010, the Centers for Disease Control and Prevention (CDC) released recommendations for improving health through transportation policy. These recommendations include supporting active transportation infrastructure with an emphasis on data and evaluation that informs planners about the impacts of transportation systems on health and a means to assess the effectiveness of planning implementations *(28)*. In addition to providing useful information for determining planning measures that encourage healthier communities, crowdsourcing approaches to data collection can be used to compile local knowledge of safe walking and biking routes, such as the CollabMap crowdsourcing tool developed at the School of Electronics and Computer Science at the University of Southampton in the UK to crowdsource the digitization of evacuation routes using satellite imagery and Google Street View *(29)*.

Crowdsourcing cyclist and pedestrian data for healthy, safe, and sustainable communities has much potential; however, examples of crowdsourcing efforts specifically tied to these topics are less common than in other realms of active transportation planning. For this reason, this section focuses on examples of crowdsourcing applications and crowdsourced data uses that, although not specifically linked to active transportation, are promising implementations of an evolving method of collecting and analyzing health, safety, and sustainability data.

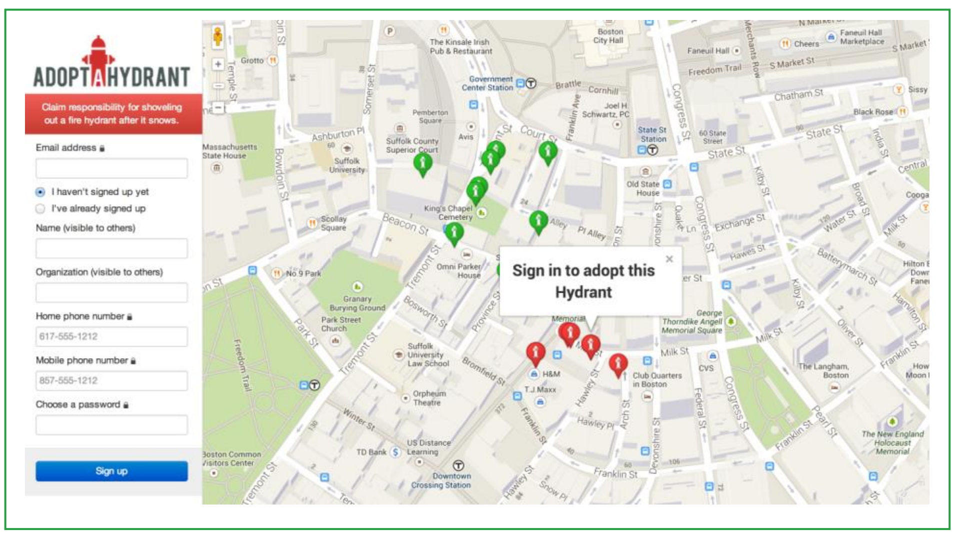
Example project types include:

- Intersection or Corridor Safety Studies / Pedestrian or Bicycle Safety Assessments
- Safe Routes to School
- Safe Routes for Seniors
- Safe Routes to Transit
- Health Impact Assessment
- Climate Change Studies

Examples and Suggestions

Crowdsourced data has the potential to be very helpful when identifying and communicating safety issues, assessing health impacts, delineating safe walking and biking routes, and developing sustainability strategies. In-situ data, whether collected implicitly through mobile applications or explicitly through research-specific data collection tools, can provide on-the-ground information about where people are choosing to travel. This information can be overlaid with other datasets such as collision records, air quality, roadway speeds, and points of interest to examine how the natural and built environments both influence and are affected by transportation patterns. Although crowdsourced in-situ datasets are not comprehensive representations of active transportation, crowdsourcing systems can be used to gather supplemental data implicitly as well as designed explicitly to gather domain-specific information from a defined user group that may otherwise be unrepresented or underrepresented *(11)*. For example, a 2007 study on neighborhood walkability and recreation environment variables as they relate to adolescent physical activity used accelerometer monitoring to estimate physical activity in a group of test subjects *(13)*. While not primarily focused on recording physical location, this study exemplifies the effective use of mobile applications to record personal metrics from an intentionally limited group of contributors.

Crowdsourcing can also be used by cities and regions as an outlet for citizens to influence infrastructure development and policy, as well as to encourage ownership and positive change at the street level. In 2012, the City of Boston implemented the crowdsourced data collection application Adopt-A-Hydrant. The application was developed by Code for America, an organization focused on improving communities through the use of innovative open-source applications. Using Adopt-a-Hydrant, Boston citizens can log in to a crowdsourcing web map and volunteer to shovel out hydrants covered in snow, which can cause delays for firefighters and add costs to city management if assigned to city employees. Users claimed ownership of hydrants by clicking on or adding thumbtack data to the map. Although initially introduced in Boston, Adopt-a-Hydrant is an open-source tool that can be used by other jurisdictions *(30)*. Citizen engagement applications such as Adopt-a-Hydrant can be an effective means to directly improve a community's infrastructure, engage citizens in the planning process, and reflect the concerns of the population.



Special Considerations for Crowdsourced Data

While there are many benefits to using crowdsourced data in active transportation planning, the complexities that could emerge through usage of multiple data sources, tools, and contributors may present some limitations and considerations. The section that follows describes some of these considerations that may need to be addressed when using crowdsourcing techniques and crowdsourced data.

Data Quality

Crowdsourced data quality can be difficult to manage and assess. For example, crowdsourcing applications that gather subjective user feedback and suggestions may result in multiple interpretations due to the complexity of insights offered toward real or perceived conditions *(31)*. Furthermore, crowdsourced data may lack documentation *(32)*. Evaluating the authenticity and accuracy of individual participants' responses is yet another difficult, if not impossible task. Even when evaluation techniques are included in the design of a crowdsourcing system (such as automated tests used to measure reliability of responses or clearly communicated punishments for malicious responses), some unanticipated and undesired contributions may go undetected, and conflicting responses may reduce transparency, making results difficult to decipher or use effectively *(33)*. For implicit systems that make use of data collected through other applications for secondary use, the quality assessment options may be limited, resulting in a need to develop additional methods of review or validation using other data sources, such as thematic data. Planning for and implementing quality control measures early on in a crowdsourcing project can save time and energy when collected data is ready to be applied or made available for additional uses. Awareness of potential data quality issues can also help prevent decision making based on inaccurate or incorrect reports.

Participation Rates, Sample Sizes, and Bias

Another challenge of working in a crowdsourcing data collection framework is recruiting and retaining participants. Recruiting and retention strategies depend largely on the nature of the collaboration, the type of data being collected, and the target problem *(33)*. For example, if the objectives of a crowdsourcing effort involve gathering input from a broad and diverse sample of volunteers, recruiting strategies may be more effective if broadcast from venues that attract a general audience, such as social media, podcasts, and radio. Participants with domain-specific knowledge may be more effectively recruited through special interest groups, conferences, and online forums. Geographic-specific crowdsourcing applications may consider strategies focused on particular locations. For example, in a recent study in Palo Alto, planners solicited input from cyclists in the community via a web mapping application that collected thumbtack data from participants (see Palo Alto Bicycle Boulevards Web Map feature below for more information). The web mapping application was promoted through the city's website (www.cityofpaloalto.org/bike).

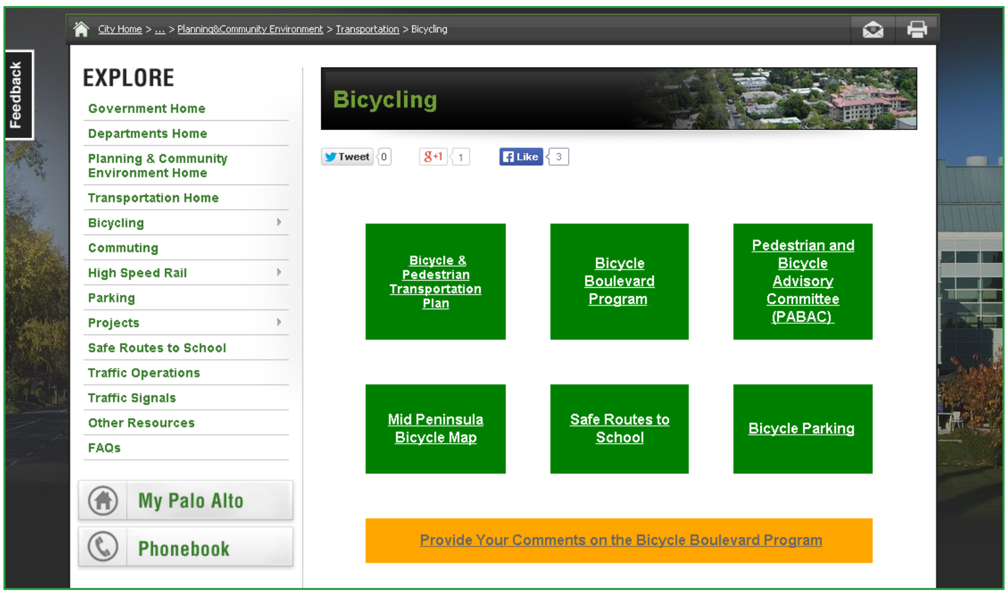
As with many data collection and survey methods, crowdsourcing efforts are susceptible to potential bias, whether using explicitly collected and available data or implicitly collected and repurposed data. Planning and decision making processes may be affected if data collection efforts do not receive the desired quantity of contributors or if the qualities of a group of contributors are either unknown or are not adequately representative. While social media, mobile APIs, and crowdsourcing apps can potentially be sources of large amounts of useful data, verifying that data samples are inclusive of all socioeconomic groups in a community may be difficult and a combination of datasets and outreach strategies may be needed. Crowdsourced data can be used in conjunction with other data sources or post-processed in order to address sample bias. For example, while SFCTA's CycleTracks app was limited to smartphone users and sample size was limited, data from CycleTracks, used with cyclist activity data from other data collection methods such as cyclist and pedestrian counts, was a valuable resource in SFCTA's travel forecasting model (23). Issues may also arise if questions of honesty and validity of user contributions significantly affect confidence in the outcomes of studies that use crowdsourced data. Implicit crowdsourcing such as the mining of in-situ data from mobile APIs may require additional filtering and validation to address these types of bias concerns. Data filters such as machine-based algorithms and human review have been used to improve the reliability of crowdsourced data, especially when data is re-used for other purposes (34).

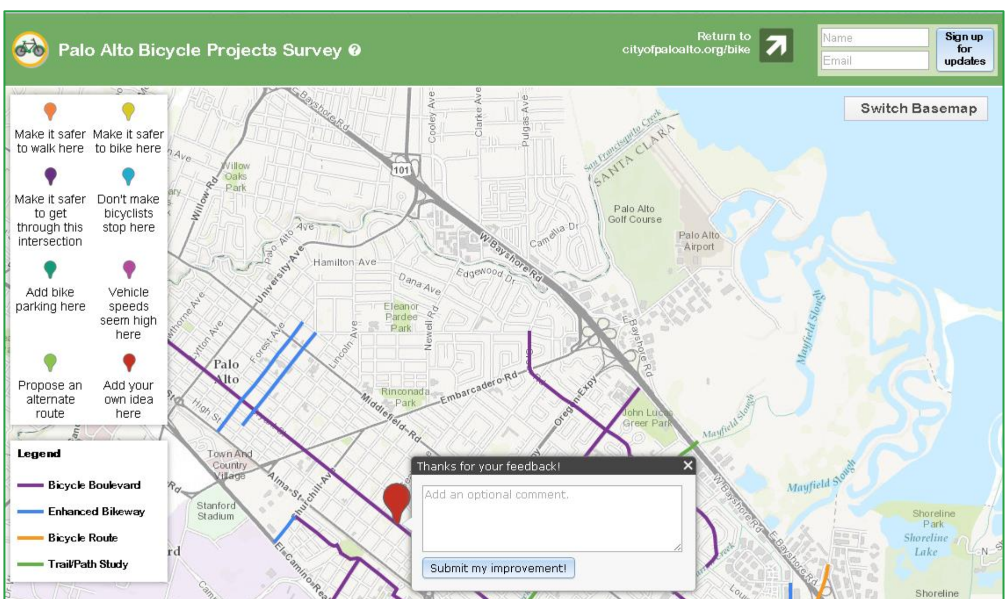
User Data and Privacy

User privacy and security are important considerations when crowdsourcing location information from GPS-enabled smartphones. Location and mobile privacy is the subject of growing concern, especially in light of reports of applications that may transmit user identification information without clear or direct notification to users. User data ownership is often either not specified in privacy agreements for mobile and web applications or specifically assigned to the collecting entity. While users continue to generate more location-based data, privacy policies are often absent, ambiguous, or simply not well understood *(9)*. Lack of participant knowledge or consent is an important consideration, especially when using implicit data *(11)*. Privacy policies must be consulted when applying implicitly crowdsourced data, especially if user identification information is included in datasets downloaded using mobile application APIs. While user information such as self-reported zip code, age and gender may enhance implicitly-collected datasets and help with bias assessments, proactive discussion and examination of potential conflicts and concerns can aid in the early identification of privacy and user protection issues that may arise. Privacy protection should also be addressed when designing crowdsourcing applications that gather personal information from contributors. Approaches to protecting users' privacy include increasing privacy awareness of users, maintaining user anonymity, and providing manual control over content sharing. While countermeasures for privacy threats have been researched and implemented, future research is needed to address many concerns related to privacy and mobile crowdsourcing (35).

Palo Alto Bicycle Boulevards Survey Web Map

The Palo Alto Bicycle Projects Survey is a web mapping application used to gather improvement suggestions from residents of Palo Alto and neighboring cities. Developed to gather feedback from local cyclists, the web map was advertised directly on the City of Palo Alto's website via a link to "Provide Your Comments on the Bicycle Boulevards Project". The efforts to recruit a geographically focused group of domain-specific contributors resulted in valuable input about needed improvements and desired changes. The screen captures below show the City of Palo Alto website (www.cityofpaloalto.org/bike) and the web mapping application to which users are directed (http://gis.fehrandpeers.com/Apps/PaloAlto). The web map was developed using ArcGIS for Server web mapping services and the ArcGIS JavaScript API *(36)*.





Summary

As described throughout this paper, crowdsourced data can be categorized into four data types: In-situ data, thematic data, thumbtack data, and spatial inventory data. The specifications of crowdsourcing applications depend largely on the goals and objectives of individual studies. Although a given crowdsourcing application may have been developed for a particular purpose, the resulting data can often be reapplied in other contexts. The table below summarizes the project types, crowdsourcing data suggestions, and data considerations presented in this paper.

Project Type	Data Suggestions	Notes and Considerations
Active Transportation Demand Forecasting	**In-Situ** • Explicitly collected crowdsourced data through custom application deployment (such as SFCTA's open source CycleTracks application) • Implicitly derived data from existing commercial applications or custom applications hosted by local public agencies • Commercial data aggregators such as Strava	In-situ data may be limited in sample size or biased toward specific types of users. Additionally, it's important to verify whether or not any data aggregation techniques applied are conducive to forecasting purposes. Privacy and data usage agreements should be consulted before using data.
	Thematic • "D's" data from the EPA's Smart Location Database (SLD) • Demographic characteristics from the US Census and American Community Survey (ACS) • Journey to Work data from the Census Transportation Planning Products (CTPP)	Some thematic datasets, such as the ACS, are developed using sample datasets and may require careful attention to potential sampling errors. Geographic scale differences may also pose challenges when working with multiple datasets.
	Spatial Inventory • Open Street Map (OSM) • Custom inventory web mapping applications	Quality and coverage of crowdsourced spatial inventory data depends on user contributions and may need to be compared with additional data sources for validation.
Facility Improvement Studies	**Thumbtack** • Custom web mapping applications designed explicitly for project-specific improvement suggestions and/or issue reporting • Implicit use of existing crowdsourced data relevant to project needs	Analyzing user feedback and applying user-generated data in conjunction with other datasets may be challenging due to subjectivity and context-specificity in responses. Aggregation and communication of responses may pose challenges if multiple collection methods are used. Increasing variety of response types can add complexity in data processing and understanding results.

Project Type	Data Suggestions	Notes and Considerations
Facility Improvement Studies (cont.)	**Thematic** • "D's" data from the EPA's Smart Location Database (SLD) • Demographic characteristics from the US Census and American Community Survey (ACS) Journey to work data from the Census Transportation Planning Products (CTPP)	Some thematic datasets, such as the ACS, are developed using sample datasets and may require careful attention to potential sampling errors. Geographic scale differences may also pose challenges when working with multiple datasets.
	Spatial Inventory • Open Street Map (OSM) • Custom inventory web mapping applications	Quality and coverage of crowdsourced spatial inventory data depends on user contributions and may need to be compared with additional data sources for validation.
Safe Routes Studies	**In-Situ** • Explicitly collected crowdsourced data through custom application deployment (such as SFCTA's open source CycleTracks application) • Implicitly derived data from existing commercial applications or custom applications hosted by local public agencies • Commercial data aggregators such as Strava	In-situ data may be limited in sample size or biased toward specific types of users. Privacy and data usage agreements should be consulted before using data. For safety studies, in-situ data can be used in conjunction with other safety-related datasets such as collision records to analyze spatial relationships between travel patterns and safety issues.
	Thematic • Demographic characteristics from the US Census and American Community Survey (ACS)	Some thematic datasets, such as the ACS, are developed using sample datasets and may require careful attention to potential sampling errors. Geographic scale differences may also pose challenges when working with multiple datasets. Demographic information can be used to identify regions with high populations of specific age and income groups, which may be helpful in safe route studies focusing on children and/or the elderly.

Project Type	Data Suggestions	Notes and Considerations
Safe Routes Studies (cont.)	**Thumbtack** • Custom web mapping applications designed explicitly for feedback on safety issues and improvement suggestions • Implicit use of existing crowdsourced data relevant to project needs	Analyzing user feedback and applying user-generated data in conjunction with other datasets may be challenging due to subjectivity and context-specificity in responses. Aggregation and communication of responses may pose challenges if multiple collection methods are being used. Increasing variety of response types can add complexity in data processing and understanding results.
	Spatial Inventory • Open Street Map (OSM) • Custom inventory web mapping applications	Quality and coverage of crowdsourced spatial inventory data depends on user contributions and may need to be compared with additional data sources for validation.

Conclusion

Crowdsourcing offers active transportation planning many benefits, such as increased data availability, broad and diverse perspectives, local knowledge, data timeliness, and direct dialogue between planners and those affected by planning decisions. Existing crowdsourced in-situ, thumbtack, thematic, and spatial inventory datasets can serve as resources for many types of planning projects. New crowdsourcing applications developed for a specific use or planning implementation can produce high-quality and project-specific data about cyclists and pedestrians while opening lines of communication between planners and those contributing the data. Data providers including public agencies, open source projects, and commercial developers are making data easier to find, and in some cases these resources are freely available. When used carefully and creatively, user-generated cyclist and pedestrian data has great potential to inform and add value to active transportation planning projects.

As with many data collection efforts, potential data quality issues must be considered, and specialized skills and expertise may be needed to manage, process, analyze, and visualize crowdsourced data. However, the skepticism over the quality of volunteered, user-generated data provided by amateurs (as opposed to professionals) *(37)* is increasingly being replaced by a desire to open the lines of communication between the planning world and those affected by planning decisions, directly addressing (rather than being discouraged by) data limitations. The initiatives surrounding progressive data collection, management, and analysis are further reflected in the numerous conferences, meetups, and other events fostering collaboration between planners, developers, data scientists, and others interested in applying critical thought and innovation in planning *(1)*. While this paper reviews existing crowdsourcing techniques and their current applications in planning, the pace of technological change and rate of adoption in planning indicates that planners will continue to develop and apply innovative approaches such as crowdsourcing in response to continually changing community needs.

Works Cited

1) Evans-Cowley, J. and B. Kubinski. We Built This Technology. *Planning*, June 2014.
2) Schlossberg, M. and C. Brehm. Participatory Geographic Information Systems and Active Transportation: Collecting Data and Creating Change. *Transportation Research Record: Journal of the Transportation Research Board. No. 2105*. Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 83-91.
3) Masli, M., Crowdsourcing Maps, *Computer*, Volume 44, Issue 11, 2011, pp. 90-93.
4) Tanz, J. Trust in the Share Economy. Wired. April 2014.
5) Botsman, R. The Case for Collaborative Consumption. TEDTalks, May 2010. Retrieved from http://www.ted.com/talks/rachel_botsman_the_case_for_collaborative_consumption, 2014.
6) Howe, J. The Rise of Crowdsourcing. Wired. June 2006.
7) Estelles-Arolas and E., F. Gonzalez-Ladron-de-Guevara. Towards an Integrated Crowdsourcing Definition. *Journal of Information Science (in press)*.
8) Brabham, D.C. Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence: The International Journal of Research into New Media Technologies*, Volume 14, Issue 1, 2008, pp. 75-90.
9) Cottrill, C.D., and P. Thakuriah. Privacy in Context: an Evaluation of Policy-Based Approaches to Location Privacy Protection. *International Journal of Law and Information Technology*, 2013, pp.1-30.
10) Hochmair, H.H., D. Zielstra, and P. Neis. Assessing the Completeness of Bicycle Trail and Designated Lane Features in OpenStreetMap for the United States and Europe. Presented at the 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.
11) Misra, A., A. Gooze, K. Watkins, M. Asad, and C.A. Le Dantec. Crowdsourcing and Its Application to Transportation Data Collection and Management. Presented at the 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.
12) Erickson, T., Geocentric Crowdsourcing and Smarter Cities: Enabling Urban Intelligence in Cities and Regions. *A Position Paper for the 1st International Workshop on Ubiquitous Crowdsourcing. UbiComp'10*, Copenhagen, Denmark, 2010.
13) Kligeman, M., J.F. Sallis, S. Ryan, L.D. Frank, and P.R. Nader. Association of neighborhood design and recreation environment variables with physical activity and body mass index in adolescents. *American Journal of Health Promotion*, Volume 21, Issue 4, 2007, pp. 274-277.
14) United Stated Census. TIGER/Line with Selected Demographic and Economic Data. Retrieved from https://www.census.gov/geo/maps-data/data/tiger-data.html, 2014.

15) EPA (Environmental Protection Agency). Smart Location Mapping. Retrieved from http://www.epa.gov/smartgrowth/smartlocationdatabase.htm, 2014.
16) Federal Highway Administration. A Guide for Maintaining Pedestrian Facilities for Enhanced Safety. Retrieved from http://safety.fhwa.dot.gov/ped_bike/tools_solve/fhwasa13037/chap4.cfm, 2014.
17) Wiedel, Sean. Personal interview. May 9, 2014.
18) Meddin, R, and P DeMaio (2014). The Bikesharing World Map. Retrieved from http://www.bikesharingworld.com/.
19) UCL Centre for Advanced Spatial Analysis. Bicycle sharing systems – Global Trends in Size. Retrieved from http://www.bartlett.ucl.ac.uk/casa/pdf/paper196.pdf, 2014.
20) Hwang, J., T. Reuscher, D. Wilson, and E. Murakami. Characterizing Walk Trips in Communities Using data from the 2009 NHTS, ACS, and Other Sources. Presented at the 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.
21) Hood, J., E. Sall, and B. Charlton, A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters: The International Journal of Transportation Research*, Volume 3, 2011, pp. 63-75.
22) Sall, Elizabeth. Personal interview. June 16, 2014.
23) SFCTA (San Francisco County Transportation Authority). CycleTracks for iPhone and Android. Retrieved from http://www.sfcta.org/modeling-and-travel-forecasting/cycletracks-iphone-and-android.
24) Ramsey, K. and A. Bell. Smart Location Database: Version 2.0 User Guide. Retrieved from http://www.epa.gov/smartgrowth/pdf/sld_userguide.pdf, 2014.
25) Pelon, M., M. Ridgway, and A. Rixey. Memo to San Mateo County Transit District, May 19, 2014.
26) SFMTA (San Francisco Municipal Transportation Agency). SF Bike Share Sites. Retrieved from http://sfbikeshare.sfmta.com/.
27) Geofabrik. OpenStreetMap Data Extracts. Retrieved from http://download.geofabrik.de/, 2014.
28) Centers for Disease Control and Prevention (2010). CDC Recommendations for Improving Health through Transportation Policy. Retrieved from http://www.cdc.gov/transportation/docs/final-cdc-transportation-recommendations-4-28-2010.pdf.
29) Stranders, R., S.D. Ramchurn, B. Shi, and N.R. Jennings. CollabMap: Augmenting Maps Using the Wisdom of Crowds (Extended Abstract). Association for the Advancement of Artificial Intelligence. 2011.
30) Code for America. Adopt-a-Hydrant. Retrieved from http://codeforamerica.org/apps/adopt-a-hydrant/.
31) Schlossberg, M., C. Evers, K. Kato, D. Maher, and C. Brehm. Active Transportation, Citizen Engagement and Livability: Coupling Citizens and Smartphones to Make the Change. Presented at the 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.
32) Estima, J., Using Volunteered Geographic Information to help Land Use/Land Cover mapping. 2012.

33) Doan, A., R. Ramakrishnon, and A.Y. Halevy. Crowdsourcing Systems on the World-Wide Web. *Communications of the ACM*. Volume 54, Issue 4, 2011, pp. 86-96.

34) Kamel Boulos, M.N., B Resch, D.N. Crowley, J.G. Breslin, G. Sohn, R. Burtner, W.A. Pike, E. Jezierski, and K.S. Chuang. Crowdsourcing, citizen sensing and sensor web technologies for public and environment health surveillance and crisis management: trends, OGC standards, and application examples. *International Journal of Health Geographics*. Volume 10, Issue 1, 2011, pp-67.

35) Wang, Y., Y. Huang, and C. Louis. Respecting User Privacy in Mobile Crowdsourcing. *SCIENCE 2*, Volume 2, 2013, pp-50.

36) Alexander, Charles. Personal interview. June 25, 2014.

37) Goodchild, M., Assertion and Authority: The Science of User-Generated Geographic Content. *Earth*, 2008, pp. 1-18.

Credits

All symbols used in this paper were downloaded from The Noun Project (www.thenounproject.com).

 In-situ symbol by Pedro Ramalho (https://www.behance.net/peramalho)

 Thematic symbol by Juan Pablo Bravo (http://juanpablobravo.blogspot.com/)

 Thumbtack symbol by Amy Lee (http://www.amyjlee.com/)

 Spatial Inventory symbol by Factor[e] Design Initiative (http://factore.ca/)

Acknowledgements

The following people were kind enough to provide interviews during the preparation of this paper: Sean Wiedel, City of Chicago; Elizabeth Sall, San Francisco County Transportation Authority; Eric Fischer, Mapbox. Many Fehr & Peers employees contributed insights and experiences. Meghan Mitman, with Fehr & Peers, provided direction and feedback throughout.

Pedestrian and Bicycle Information Center

730 Martin Luther King Jr. Blvd., Suite 300
Chapel Hill, North Carolina 27599-3430
pbic@pedbikeinfo.org
888-823-3977