Crowdsourcing and Its Application to Transportation Data Collection and Management

Aditi Misra, Aaron Gooze, Kari Watkins, Mariam Asad, and Christopher A. Le Dantec

The participation of a large and varied group of people in the planning process has long been encouraged to increase the effectiveness and acceptability of plans. However, in practice, participation by affected stakeholders has often been limited to small groups, both because of the lack of reach on the part of planners and because of a sense of little or no ownership of the process on the part of citizens. Overcoming these challenges to stakeholder participation is particularly important for any transportation planning process because the success of the system depends primarily on its ability to cater to the requirements and preferences of the people whom the system serves. Crowdsourcing uses the collective wisdom of a crowd to achieve a solution to a problem that affects the crowd. This paper proposes the use of crowdsourcing as a possible mechanism to involve a large group of stakeholders in transportation planning and operations. Multiple case studies show that crowdsourcing was used to collect data from a wide range of stakeholders in transportation projects. Two distinct crowdsourcing usage types are identified: crowdsourcing for collecting normally sparse data on facilities such as bike routes and crowdsourcing for soliciting feedback on transit quality of service and real-time information quality. A final case study exemplifies the use of data quality auditors for ensuring the usability of crowdsourced data, one of many potential issues in crowdsourcing presented in the paper. These case studies show that crowdsourcing has immense potential to replace or augment traditional ways of collecting data and feedback from a wider group of a transportation system's users without creating an additional financial burden.

Researchers have long emphasized the importance of public participation in the planning process as a critical component to the successful implementation of any plan (*1–3*). Broad public participation leads to "greater legitimization and acceptance of public decisions, greater transparency, and efficiency in public expenditures, and greater citizens' satisfaction" (*4*). According to Burby, inclusion of stakeholders with varied interests and different backgrounds makes a plan comprehensive, acceptable, and more easily implementable (*2*). Moreover, a participatory planning process effectively recognizes that "society is pluralist and there are legitimate conflicts of interest that have to be addressed by the application of consensus building methods" (*5*). With these traits in mind, participatory planning has the potential to involve broader and more diverse groups of people into a planning dialogue and, hence, can bring in newer perspectives and ideas to the planning problem at hand (*6*).

Recent research, however, suggests that citizen involvement at different stages and levels of planning is steadily declining in the United States (*7–9*). This lack of involvement seems counterintuitive given the fact that over the past few decades, information accessibility and remote participation have been facilitated and made easier through the ubiquitous use of the internet and web-based social media. A wealth of emerging technologies has brought about significant new forms of communication and interaction, providing diverse new ways of documenting, sharing, and reflecting on the world at a truly global scale.

One possible reason for the apparent decrease in citizen involvement may be that planners and policy makers have yet to embrace technology-mediated forms of participation and instead still rely on methods that require the physical presence of participants. These methods limit the availability of the planning process for citizens by placing time and location constraints on participation and may also alienate or further disadvantage citizens for whom traveling to a planning meeting is neither physically nor financially viable.

One strategy for overcoming limited participation by interested stakeholders is to implement multiple methods of participation, which participants can choose from depending on their level of comfort and accessibility (*10*). Slotterback proposed that, along with the traditional methods of public hearings and open-house meetings, more accessible modes of communication such as project websites and web-based meetings and discussions may be adopted as a means of increasing public participation in the planning process (*3*). Toward that end, the purpose of this paper is to encourage the use of crowdsourcing platforms as a possible means of involving people from diverse walks of life to effectively participate in planning for transportation systems without putting additional financial burden on the transportation agency. The paper highlights the successful use of crowdsourcing in a few transportation projects, providing examples of projects that have overcome many of the initial challenges of adopting crowdsourcing in transportation planning and establishing a robust starting point for future work.

The paper is organized as follows: first, the concept of crowdsourcing is discussed along with a commentary on the existing platforms and types of crowdsourcing and the issues associated with crowdsourcing in general. Then, the crowdsourcing case studies in

A. Misra, A. Gooze, and K. Watkins, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA 30332. M. Asad and C. A. Le Dantec, Georgia Institute of Technology, Digital Media–Technology Square Research Building, 85 5th Street, NW, Atlanta, GA 30308. Corresponding author: A. Misra, amisra7@gatech.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2414, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 1–8.
DOI: 10.3141/2414-01

transportation planning are presented with reference to the different genres of crowdsourcing. The first group of case studies focuses on receiving feedback from transportation system users, while the second group focuses on the use of crowdsourcing for data collection. A standalone example is provided at the end of the case studies subsection, deserving special mention because of its use of data quality editors to ensure data usability and validity, thereby addressing one of the biggest issues of crowdsourced data collection.

CROWDSOURCING: CONCEPTS, PLATFORMS, AND ISSUES

At its conception, social computing focused mainly on building a network of collaborators and facilitating online communication between groups. This has eventually given rise to open source platforms and forums where people with similar motivation and outlook can come together to solve issues and find answers to problems that affect their community. Crowdsourcing is an example in which an organizer or an organization is able to use the network of collaborators to solve a problem that would otherwise be cost- or labor-intensive, or in which within a defined organization the expertise is unavailable or insufficient.

Crowdsourcing has been alternately defined as: the outsourcing of a job (typically performed by a designated agent) to a large undefined group in the form of an open call (*11*); a process that "enlists a crowd of humans to help solve a problem defined by the system owners" (*12*); or "a sourcing model in which organizations use predominantly advanced Internet technologies to harness the efforts of a virtual crowd to perform specific organizational tasks" (*13*). Common across these alternate definitions is the notion that crowdsourcing invites all interested people to form an open forum of ideas that can eventually lead to a solution of the assigned problem. As noted by Howe, crowdsourcing uses the "latent potential of crowd" to achieve a solution to a problem to which the crowd can relate (*11*).

According to Saxton et al., crowdsourcing systems are characterized by three main features: the process of outsourcing the problem, the crowd, and a web-based platform for collaboration (*13*). Outsourcing a problem generally implies getting a task done by outside sources even when it could have been performed by people within a system; in crowdsourcing, outsourcing is done in cases in which the in-house expertise has failed to produce a solution or is an expensive means to produce a solution, or in which there is no in-house expertise available for solving the issue. Crowdsourcing systems also rely primarily on an anonymous unidentified group of people (the "crowd") to come together willingly instead of using the business subcontract model of outsourcing where the task is performed by a previously identified and designated group of people or a company (*13*).

An important subset of the general crowdsourcing idea is the concept of citizen science, in which amateurs contribute to research projects in conjunction with professional scientists. Goodchild used the term "citizen science" in describing crowdsourced geomapping, referring to information generated through crowdsourcing as, although not of a professional level, helpful in expanding the reach of science (*14*). The nature of participation in citizen science projects takes different forms, depending on the type of project; it can range from data collection to data analysis and from instrument building to taking part in scientific expeditions. Recent citizen science projects tend to focus on using the ever-increasing reach and availability of electronic gadgets, particularly mobile phones and sensors, for data collection and monitoring purposes. In their experiments, Kuznetsov and Paulos (*15*) and Kuznetsov et al. (*16*) provided citizen scientists with sensors to monitor air and environmental quality, while the CycleTrack project in San Francisco, California, used GPS-enabled mobile devices to record cyclist trip data (*17*). Citizen science projects are gaining popularity as an alternative to cost-intensive data collection efforts, particularly in cases in which the information needed is global in character, and are thus being increasingly used for planning and monitoring purposes.

Existing Crowdsourcing Platforms and Systems

Despite the advantages discussed in the previous section, crowdsourcing can only be successful if a platform exists that can provide open access to incorporate, modify, and synthesize data. There are four versions of this shared platform: the wiki system, open source software, geocrowd mapping, and mash-ups using crowdsourcing data (*18*). Wiki systems are mainly centered on authoring information; open source software provides a platform to share and codevelop program source code; geocrowd mapping entails collecting, cleaning, and uploading GPS data; and mash-ups are combinations of some or all of these. While maintaining coordination between people coming from different backgrounds and with different motivations is a significant challenge, this voluntary coming together of a mass of people for a purpose is particularly useful in tackling problems that are large scale, e.g., mapping of a country.

Beyond the fundamental concept of providing an open access and participatory platform for a large group of people, crowdsourcing projects can be markedly different, depending on the purpose of the project, the nature of involvement required, or whether some special expertise is required for participation. Figure 1 schematically represents the different categorizations of crowdsourcing systems, which are further discussed in the paper. Based on the nature of involvement of the participants in solving the problem, Doan et al. classified crowdsourcing systems as either explicit or implicit systems (*12*) (Figure 1). Explicit systems are standalone systems in which users participate and collaborate in executing a stated problem such as answering questions via the web, testing software, and writing web content (e.g., Wikipedia). Within explicit systems, four types of tasks are generally performed by users:

1. Evaluating (e.g., book review),
2. Sharing (e.g., feedback on system performance),
3. Building artifacts (e.g., designing T-shirts at Threadless.com), and
4. Executing tasks (e.g., collaborating on finding gold-mining spots).

Implicit systems can be standalone or piggyback, depending on projects. In standalone implicit crowdsourcing systems, the system owners benefit from the indirect input provided by the users; the direct user input is used to solve a problem that is related to but not the same as the issue to which the users of the system respond. For example, although humans are more efficient at image recognition than computers, they are not necessarily willing to perform this task unless it is packaged in a form that attracts them. In the ESP game, the participants are shown images and asked to guess common words to describe those images as part of playing the game. Those words are then used to label the image (*12*). In piggyback

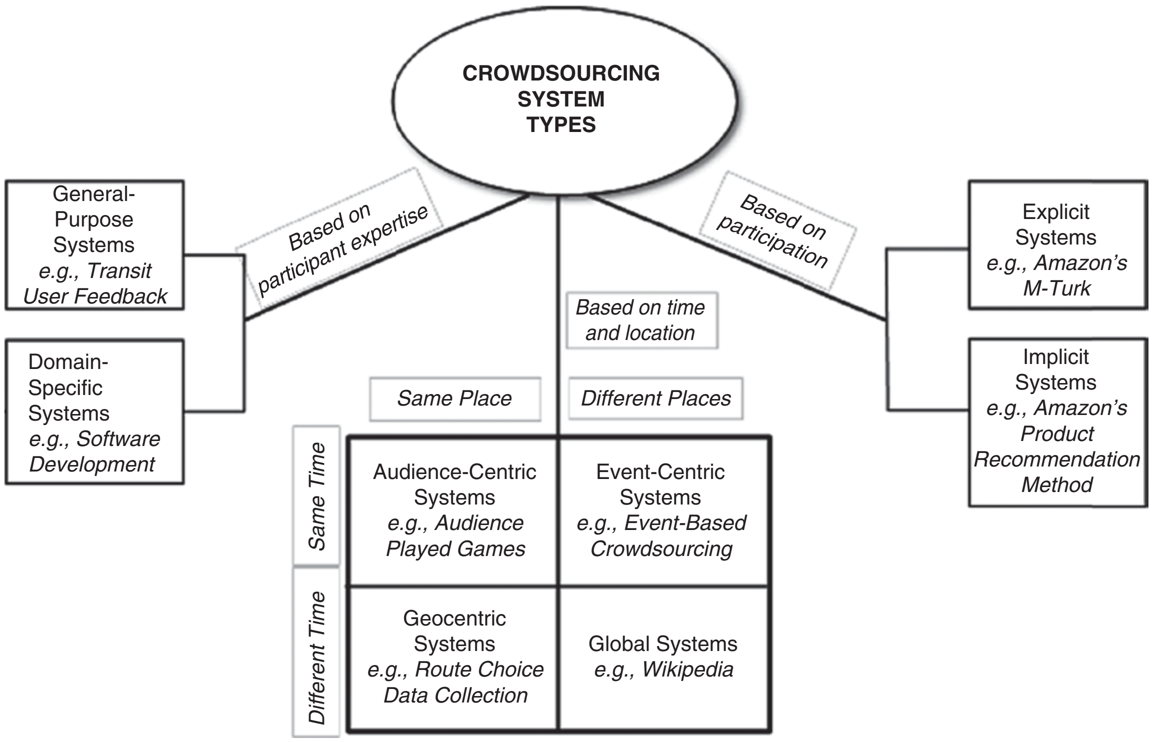

FIGURE 1 Classifications of crowdsourcing systems (*12, 19, 20*).

crowdsourcing systems, the traces of the users are collected from an entirely different system—ad keywords generated based on Google and Yahoo search traces are examples of piggyback implicit crowdsourcing systems.

Steinfeld et al. (*19*) categorized public participation as either general purpose or domain-specific systems. General purpose systems do not require special expertise from the contributors and are not targeted to any user group in particular, while domain-specific systems are designed for a special purpose user group (Figure 1). For example, most crowdsourced service quality feedback does not require special expertise on the part of the participants and is, hence, a general purpose system. Conversely, developing or beta testing open source software through crowdsourcing requires expertise in particular programming languages and platforms and is, hence, a domain-specific system.

Crowdsourcing systems are further classified based on whether the system is local or global in scope and whether the system is time bound or not (*20*) (Figure 1). For crowdsourcing systems in which the participants are at the same place at the same time, the system is termed audience centric (e.g., the use of clickers in class discussions). For systems in which participants can be at different places while the crowdsourced event is time bound (i.e., it has a start and end time between which the collaboration must happen), the system is termed event centric. An example of event-centric crowdsourcing is organized with online brainstorming sessions that are triggered by an event and span over a limited period of time. Systems in which collaboration can happen between people from different places and over an indefinite period of time are termed global crowdsourcing systems (e.g., Wikipedia). Finally, systems where people are at the same place and crowdsourcing is an ongoing process are termed geocentric crowdsourcing systems (an example is bicycle route-choice data collection for a city).

Crowdsourcing Issues

As crowdsourcing continues to evolve and gain in popularity, different and larger systems are being experimented with, and the issues uniquely associated with the characteristics of the systems are gradually surfacing. For example, domain-specific systems automatically reduce the crowd size by requiring some expertise from the participants, while implicit systems have the issue of not having explicit participant consent in using their contribution for the actual purpose of the project. A priori understanding of the project characteristics, and hence its category, can often largely help in setting up plans early to overcome such issues. The final case study presented in this paper is one such example of an expert group used as data quality auditor instead of the system being domain-specific. Use of an expert group helps in retaining a larger participant base and provides the necessary check on the usability of the data collected through a general crowdsourcing system. As crowdsourcing gets applied to different domains, and as the scale and scope of crowdsourcing systems increases, additional techniques for addressing these system-specific issues need to be developed based on the requirement of the projects.

In addition to the unique issues of the systems, operation and maintenance of crowdsourcing systems generally suffer from four major issues:

1. Recruiting and retaining the participant base,
2. Assessing user capabilities,
3. Aggregating the information provided by users, and
4. Evaluating the contributions of users (*12*).

The problem of recruiting and retaining participants is a major issue in adopting crowdsourcing for any project. Depending on the purpose of the project, it is often important that feedback be obtained from users with particular skills or expertise. Furthermore, retaining participants is often important for understanding a trend over time—to allow the crowd's understanding of the problem to evolve throughout the process. The use of recurring campaigns and marketing strategies at frequent intervals (along with new releases of apps) is suggested where applicable so that people remain curious about the project and the developers can help maintain a participant base over time (*21*). Using incentives in the form of material benefits as well as acknowledgement of contribution in the form of gratification announcements at project sites make people feel encouraged to participate in the project and can help recognize diverse kinds of contributions from the crowd (*12*).

Dealing with user capability is an important issue in citizen science projects and in problem solving projects where participants are required to have some background to appreciate the assigned task. While participatory planning may not generally require special skill sets, in cases in which the planning process targets a special group, it is important that the participants are aware of the specific problems of that group (e.g., planning for bicyclists' needs requires the presence of people who bike in that area so that the relevant problems and issues are brought up and placed on the table). In such cases, the crowdsourcing process may be most successful if it is designed as a domain-specific system—rather than a general purpose one—where specific tools and capabilities are made available to develop and maintain relevant user capabilities.

Problems with data quality and challenges with data aggregation are two important issues that often undermine the benefits of crowdsourcing systems. Regarding the importance of data quality, Heipke assessed that "quality issues have been a primary point of debate since crowdsourcing results started to appear" (*22*). From that perspective, a degree of loose hierarchical authority is needed to ensure that the data are useful for their intended purpose. Additionally, aggregation of the data from crowdsourcing is often a complicated task given the volume of responses received from a diverse pool of crowd participants. Coping with data issues is either often labor intensive as large data sets need to be manually cleaned, or more cost intensive as complex data management systems and processes need to be put into place in an attempt to reduce sources of human error.

Evaluating the contribution of the user is commonly accomplished by setting up an automatic screening program to evaluate the validity of user-submitted information based on predefined criteria. The screening program rejects any input that does not follow the set criteria, and thus only valid information is retained. However, this kind of automation is possible only in cases in which the input is sufficiently normalized to be evaluated programmatically; in cases in which the responses are descriptive or subjective, a manual evaluation stage is needed to evaluate each response based on its potential contribution to the project. Such manual processes are labor- and cost-intensive and prone to subjective biases of the evaluator but also much needed to ensure data quality for the project.

CROWDSOURCING AND ITS USE IN TRANSPORTATION

Crowdsourcing is particularly suitable and useful for transportation planning because it voluntarily brings together a large group of people on the same platform to address common issues that affect its members. Crowdsourcing works successfully for local purposes through localized knowledge and acquired experiences (*23*) because people in a region tend to identify themselves with the region where they live, work, and socialize, and are generally more interested in the systems that affect them (*20*).

A survey of transportation systems that use crowdsourcing reveals that the predominant purposes of using crowdsourcing are either for the collection of data or feedback from the transportation system's users. For example, one popular use of crowdsourcing is to collect route choice data from bicyclists using the GPS functionality of their cell phones; such data are not readily available through standard data collection procedures, and designing a separate survey for a small population of users is often not cost effective for regional planning agencies. Crowdsourcing in this case helps the geographically dispersed and diverse population of cyclists work together on a common interest without financially burdening the planning agencies. Similarly, crowdsourcing can also help in collecting feedback from a sociodemographically diverse range of users of any transit system, which can be immensely useful for improving transit service quality and standards.

Transportation related crowdsourcing systems designed to date can be implicit or explicit standalone systems as defined by Doan et al. and discussed in the previous section (*12*). These systems may also be either geocentric systems where only local users are engaged or global systems where any person can contribute to the system. Extending the categorization of public participation as defined by Steinfeld et al., transportation crowdsourcing systems may be further classified as either general purpose or domain specific systems (*19*). General purpose crowdsourcing systems do not require any special expertise from the contributors and are not targeted to any user group in particular, while domain-specific systems are designed for a special purpose user group.

Examples of transportation related crowdsourcing are presented below with reference to the above-mentioned classification systems: the first group of examples focuses on receiving feedback from users while the second group of examples focuses on use of crowdsourcing for data collection. A standalone example, provided at the end of the subsection, deserves special mention for its use of data quality editors to ensure data usability and validity and, at the same time, maintain a broad user base, thereby addressing one of the primary challenges of crowd-sourced data collection. The section is followed by a discussion on the advantages and disadvantages of crowdsourcing systems.

Crowdsourcing Case Studies

User Feedback–Based Crowdsourcing Systems

Three seminal examples of general purpose user feedback systems are SeeClickFix (http://seeclickfix.com), PublicStuff (http://www.publicstuff.com) and FixMyStreet (http://www.fixmystreet.com), all of which rely on public feedback about neighborhood issues and have been successful in mobilizing communities to take up the task voluntarily. While FixMyStreet is essentially for users to report road maintenance issues, the developers have a similar transit-based tool called FixMyTransport (http://www.fixmytransport.com). SeeClickFix and PublicStuff can be used to report "any nonemergency issue anywhere in the world that a user wants to be fixed" (*24*), be it infrastructural or governance related. In SeeClickFix, users can also set up neighborhood watches where they monitor and report local community issues which are then taken up by advocacy groups or elected officials, and solutions are proposed publicly. It is evident from the nature of the participation in these cases that no special expertise is expected from the users. The majority of the reported issues are local and community oriented in nature, reinforcing the concept that crowdsourcing can be successful in addressing local and regional issues, making it suitable for transportation planning.

Shareabouts is another example of a general purpose crowdsourcing system that uses an innovative approach. Shareabouts (http://www.shareabouts.org) is a web-based system that uses maps to generate user feedback on preferred location of facilities and amenities. A few ongoing projects that use Shareabouts are Chicago Bikeshare, Illinois, with people pinning preferred bikeshare locations on the map provided; North Carolina Alternative Bike Route Plan, with people voting for preferred alternatives as well as marking any segment that they think might be an inappropriate alternative; and Philadelphia Bike Parking Survey, Pennsylvania, with crowdsourced information collected for estimating the bike parking capacity of the existing stations and plan for future expansion. In Boston, Massachusetts, Street Bump (http://streetbump.org) is a mobile application that uses a smartphone's accelerometer to detect potholes and other street hazards as people drive around the city; the geolocated street quality data collected through crowdsourcing are automatically uploaded and integrated with the city's process for locating and fixing pavement quality issues.

A transit project using a general purpose crowdsourcing system, OneBusAway was created to address the reliability issues with on-time performance of transit systems in Seattle, Washington, and to expand upon existing transit tools in the region. OneBusAway provides several feedback mechanisms (email, Twitter, blog, bug tracker) that allow users to make comments or suggestions about the tools (*25*). The design of the various tools, along with development of new features, has been further shaped by feedback from users via several user studies and the IdeaScale feedback platform (another general use tool that can be applied to transportation). Because OneBusAway is open source software, users have also submitted improvements of their own to the code. Thus, users eventually become partners in development and design of the OneBusAway program, which promotes a sense of community among the transit riders in the region and a sense of ownership of the program. This ownership is an important factor in maintaining the user base for the program (*25*).

Another general purpose crowdsourcing project related to transit systems is Tiramisu transit (*26*), a user feedback–based real-time information system for public transportation in Pittsburgh, Pennsylvania. Tiramisu Transit, a crowd-powered transit information system uses riders as the human equivalent of automated vehicle location system, thereby providing an innovative alternative to more traditional cost-intensive data collection. Tiramisu Transit is a smartphone application (app) developed by researchers at Carnegie Mellon University to improve users' transit experiences and transit accessibility (*26*). Upon activation, the app shows a list of buses or light rail vehicles scheduled for arriving at the current time; the list is based on past arrival data as well as real-time data sent by riders on the vehicle. Tiramisu provides an option for the rider to indicate the level of fullness of the bus, which aids people with disabilities to choose the bus they want to access. Once aboard, the rider can use Tiramisu to find out which stop is next and to report problems, positive experiences, and suggestions. Use of Tiramisu is motivated by the riders' ability to use the same real-time arrival and fullness information they are reporting.

Crowdsourcing Systems for Data Collection

Issue-reporting crowdsourcing systems such as SeeClickFix and FixMyStreet do not call for specific expertise from the user, but there may often be systems in which data and information are needed from a group with specific expertise or purpose; these are termed domain-specific systems (*20*). Domain-specific systems may be nested under a general purpose system, such as the bike projects undertaken using Shareabouts. While all of these projects use the same crowdsourcing platform, the information is collected for one specific region, because it is more useful if it comes from the cyclists who use the facilities on a regular basis. Examples of standalone domain-specific systems are the crowdsourced bike route data collection projects undertaken in San Francisco; Minneapolis, Minnesota; Atlanta, Georgia; and Austin, Texas. These projects focus on developing smartphone apps and websites for cyclists to record their trips so that region-specific bikability maps can be created and facilities can be constructed on route segments as required.

CycleTracks (*17*) and Cycle Atlanta (*27*) are both projects for collecting bike route choice data through GPS-enabled smartphones. The creation of CycleTracks by the San Francisco County Transportation Authority in late 2009 was motivated by the lack of data on cyclists, cycling infrastructure, and eventually cyclist route choices. Traditionally, such data would be collected through public meetings because cyclists represent only 1% to 2% of commuters, making vehicle count methods less useful. CycleTracks made participation in data collection for cyclists more accessible by moving data collection to the increasingly common smartphone use. In CycleTracks, first-time users are asked optional information to determine cycling habits, such as riding frequency, age, gender, and zip codes for home, work, and school. Users record their trips by starting the app when they set out on a ride and then saving and uploading their data once they've reached their destination. The app records bicycle trip route, time, distance, and average speed, along with user-reported trip purpose and notes. The trip data are wirelessly uploaded for analysis of cyclist route choice and is later used for planning facilities along the predicted routes (*17*).

Cycle Atlanta, a similar smartphone app for collecting data about cyclists and their routes within the city of Atlanta, was built off the open source codebase of the CycleTracks app. Cycle Atlanta also uses the GPS capabilities of smartphones to save and upload routes to provide basic data on how cyclists navigate the city, but the project team added features to the app including the ability to note with photos and textual descriptions of specific locations as either issues (pavement issues, traffic signal, enforcement, etc.) or amenities (bike parking, public restrooms, water fountains, etc.). The app also includes the collection of additional demographic data, including cyclist ability and history as indicators of comfort level to allow analysis of route data around an established taxonomy of urban cyclists (*28*), and to enable correlation with existing cyclist count and census data. As a distinctly different approach from CycleTracks, Cycle Atlanta categorizes cyclists into groups based on their cycling comfort level. The categories include the strong and fearless, the enthused and confident, the comfortable but cautious, and the interested but concerned. This categorization helps to understand the preferences of different types of cyclists in choosing routes, and hence can be immensely informative in creating a tailored application such as bike maps for any particular group of users. Since the apps were launched in early October 2012, Cycle Atlanta has been used by more than 1,000 cyclists in Atlanta, who have recorded more than 10,000 rides—represented by more than 16.5 million individual data points. These data constitute the core piece of the City of Atlanta's effort to facilitate more streamlined communication between planners and cyclists.

A significant role of domain-specific crowdsourcing is its provision of information from an otherwise unrepresented or underrepresented community. For example, because of the small size of the cycling community, bicycle maps are not commercially attractive and, hence, are rare. Therefore, crowdsourced maps and geowikis are particularly suitable for understanding bicycle routes and for developing bicycle route maps (*29*). Also, cyclists can benefit from regularly updated information, which is easy to maintain through "delegated responsibility among a motivated community with common purpose" (*29*). Cyclopath (http://www.cyclopath.org), a crowdsourced geowiki-based bicycle map developed by researchers at the University of Minnesota provides an example of a domain-specific use of crowdsourcing in transportation. Cyclopath maintains an active database of user-contributed bicycle routes and trails within the Minneapolis–Saint Paul metropolitan area. The users of Cyclopath can add, modify, and delete roads and bike trails, road and trail segments, points of interest, and neighborhoods. In addition, Cyclopath allows users to add notes and tags describing any feature on the map, such as "bumpy" or "closed." Revisions are public and tagged to user logins for transparency and accountability. Cyclopath also has features that help the community moderate itself. A list of recent changes is also maintained so that other users can identify and undo malicious modifications to the geowiki. Finally, Cyclopath allows a user to rate bike routes on a five-point qualitative scale (excellent, good, fair, poor, and impassable) for their own use and for aggregation to enhance bikability ratings. The Cyclopath community has made more than 13,000 revisions since its release (*30*).

Standalone Crowdsourced Data Quality Auditor System

Along with generating data from underrepresented groups, domain-specific crowdsourcing also helps in data quality management, which is an issue with self-reported data in crowdsourced systems. As a study by Wiggins and Crowston revealed, most of the systems that use voluntary public participation include some form of expert control over the data (*31*). An expert user group can act as a bridge between general users and the system by filtering required information from general information and then translating the feedback from the system to the general users in a meaningful way. Use of such a group helps maintain a feedback loop that is important in retaining participants and also prevents losing the critical mass, which is often the case if the entire process is domain specific.

A standalone example of such an effort in transportation systems is the transit ambassador program initiated by the OneBusAway program in Seattle (*32*). The transit ambassadors are a super user group, with a solid understanding of the transit network and basic computational and analytical skills. Their role is to filter the incoming general purpose crowdsourced information and channel it to the respective departments within the transit agency for necessary action. Three core goals of the program development were addressing problem resolution, engaging the community, and improving agency-rider communication. Beginning in the fall of 2011, a number of errors with the real-time transit prediction data surfaced, affecting over 77% of a survey of riders (*33*). While the OneBusAway mobile application included an error reporting function to allow users to identify errors experienced, the amount and quality of the crowdsourced reports began to overwhelm the OneBusAway administrators. Oftentimes, reports were duplicates of previously reported errors or the information submitted was incomplete and required additional effort to utilize it. With upwards of 500 errors reported on a weekly basis, the time required to evaluate these reports and any attempt to leverage them in order to resolve underlying problems with the real-time system would have required an effort from a collection of individuals. In contrast to previously described crowdsourcing programs, this was not an issue of data collection, but rather a problem with information management. The management of the errors required the coordination between the agency, the OneBusAway administrator and the riding community, however, due to the constrained resources of each organization, there was no single contact to coordinate between these entities. This role fell to a collection of volunteer super users, or OneBusAway transit ambassadors. Figure 2 provides a visual summary of the flow of information established in the program as well as the role of the ambassadors in coordinating the process.

An initial group of three transit ambassadors were recruited from the rider community via blog solicitation and email outreach. The ambassadors were provided resources such as transit schedule data, agency alert information, and an error decision matrix to assist in categorizing the crowdsourced error reports. All error reports were collected into an online database that allowed the ambassadors to not just validate the error but to identify the nature and possible cause. This action of validation was a necessary and vital step in transforming the overwhelming amount of crowdsourced information from varying noise into usable knowledge. Finally, the ambassadors aggregated the information to forward onto the transit agency a clear and concise summary of notable issues reported by riders. For example, the summary of errors by vehicle and route provided the transit agency with valuable supporting information to help target potential actions to improve the real-time information system. The overarching role of the ambassadors provided a level of expertise that could accurately evaluate the incoming error reports and thus efficiently triage and divert any relevant issues to the appropriate organization.

Providing a behind-the-scenes look at the underlying issues confronting the transit agency allowed the ambassadors to relay that information to the rider community and to provide some context to the errors that everyone was experiencing. For example, a typical public relations response by the agency would have been interpreted far differently as compared with the ambassadors relaying this information to the community, thereby providing an enhanced level of trust. Although some underlying real-time issues could not be resolved by the agency, the ambassadors provided a means to explain to riders

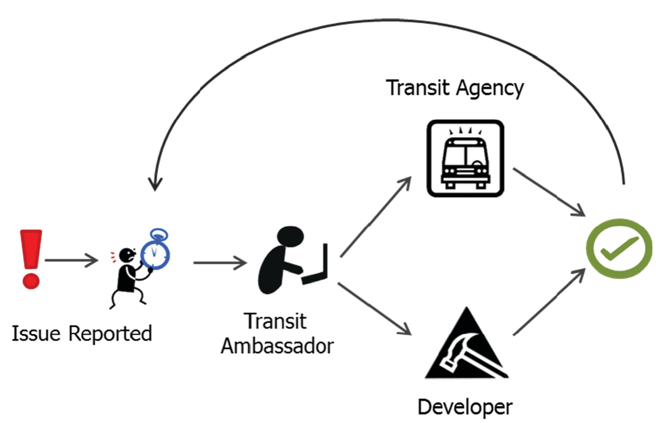

FIGURE 2 Information flow of transit ambassador program.

why an issue could not be fixed and how they could best adjust to the situation.

The success of the outreach exhibited by the ambassadors and their role in representing not just the agency but the riders themselves gave validity to the potential that a fully deployed ambassador program has within any real-time information system. With the proper adjustments to the available agency support and an expansion of the amount of ambassadors, a transit ambassador program can effectively accomplish the core objectives and serve as not only a means for improving the real-time information product but serve as a mechanism for an agency to fully engage its riding community in a method that improves the overall functionality and quality of the transit service provided.

Summing It Up

Despite the fact that crowdsourcing has been used in transportation planning only recently, it is evident from the case studies presented that it has immense potential in augmenting or replacing traditional survey methods, particularly for groups of stakeholders who have a small user base in the transportation system. As seen with all systems, crowdsourcing also has its own issues that need to be addressed through proper planning and understanding of the system. Although there are criticisms with respect to data quality and data management issues, it is undeniable that crowdsourcing has been successful in engaging groups of people in solving a problem that affects their community. Crowdsourcing for bike route choice data has successfully solved the issue of data aggregation, defining a role of the users and linking their contribution to the final goal of the project by developing facilities for the bicyclists in San Francisco, Minneapolis, Atlanta, and Austin. Meanwhile, transit information systems such as Tiramisu Transit and OneBusAway have been very successful in redefining the role of their users in monitoring service standards and quality. The OneBusAway transit ambassador program has the potential to address the data quality issues associated with crowdsourcing by filtering and validating the data received from participants before they reach the agency.

Most of the crowdsourcing systems use devices and technologies that are readily available and low cost; often crowdsourcing is based on devices that are owned by individuals (as in cycling data collection in CycleTracks and Cycle Atlanta), involving no major financial investment on the part of the system. In an exemplary case, the Tiramisu project described earlier uses crowdsourcing to actually replace the requirement of high-cost automated vehicle location systems. Tiramisu provides an example of ideal civic engagement in transit planning and operation where riders take care of other riders without the direct involvement of the transit agency and create an information sharing legacy that is beneficial to both the users and the agency. With current funding limitations, crowdsourcing can be a preferred alternative to involve the public despite limited resources.

However, CycleTracks and Cycle Atlanta are based on the widespread popularity and reach of the smartphone technology for crowdsourcing. Although smartphones are easy to carry and powerful devices that provide an inexpensive means of data collection, their use is not equally prevalent with all groups of people—thus, using smartphones for data collection comes with the issue of bias toward the input from populations of different ages, financial means, and ethnicity (*34*). Further research into possible biases arising from smartphone data collection is underway (*34*), and preliminary results show that age, income, and ethnicity are the major factors that should be considered in smartphone data collection. This, however, can be addressed using proper outreach efforts and using supportive traditional methods for people who are not currently smartphone users.

CONCLUSION

Crowdsourced transportation projects bear evidence that crowdsourcing has the potential to bring together a large group of people on the same platform when there is an issue that affects them all. Systematic use of information and feedback from users for the purpose of transportation planning or for improving service standards is receiving significant attention recently, and smart technology–based crowdsourcing provides an ideal platform for engaging a broad group of users with limited additional financial burden on the system or the agency—possibly even replacing costly equipment. Crowdsourcing for data collection is found to be financially most effective in cases where the user base is small but enthusiastic and motivated as in the case of bicyclists; in such cases crowdsourcing has a huge potential in augmenting the standard data collection procedures by including the requirements of otherwise marginalized groups of users. Examples of a few potential transportation-related cases where crowdsourcing can be used are traffic data collection, getting user feedback for different systems, monitoring pavement and sidewalk quality, and understanding group opinion in creating new facilities.

Crowdsourcing issues are mostly concentrated around problems with the quality, accuracy, and aggregation of data. However, these issues may be addressed through proper planning and with an understanding of the final goal of the crowdsourcing project. Further research and implementation of such strategies in real life projects are needed to establish a generic framework of crowdsourcing for transportation planning.

REFERENCES

1. Innes, J. E. Information in Communicative Planning. *Journal of the American Planning Association,* Vol. 64, No. 1, 1998, pp. 52–63.
2. Burby, R. J. Making Plans that Matter. *Journal of the American Planning Association,* Vol. 69, No. 1, 2003, pp. 33–49.
3. Slotterback, C. S. Public Involvement In Transportation Project Planning and Design. *Journal of Architectural and Planning Research,* Vol. 27, No. 2, 2010, p. 144.
4. Insua, R. D., E. G. Kersten, J. Rios, and C. Grima. Towards Decision Support for Participatory Democracy, *ISeB,* Vol. 6, 2008, pp. 161–191.
5. Hague, C., K. Kirk, M. Higgins, A. Prior, P. Jenkins, H. Smith, and W. Grimes. Participatory Planning for Sustainable Communities, 2003.
6. Rabinowitz, P. Participatory Approaches to Planning Community Interventions, http://ctb.ku.edu/en/tablecontents/sub_section_main_1143.aspx, 2013. Accessed May 2013.
7. Skocpol, T., and M. Fiorina (eds.). *Civic Engagement in American Democracy.* Brookings Institution Press, 1999.
8. Galston, W. A. Civic Education and Political Participation. *PS: Political Science and Politics,* Vol. 37, 2004, pp. 253–266.
9. Pew Research Center for the People and the Press. Cable and Internet Loom Large in Fragmented Political News Universe. http://www.people-press.org/2004/01/11/cable-and-internet-loom-large-in-fragmented-political-news-universe, 2004. Accessed May 2013.
10. Wagner, J. Measuring Performance of Public Engagement in Transportation Planning: Three Best Principles. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2397,* Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 38–44.
11. Howe, J. The Rise of Crowdsourcing. *Wired, Conde Nast Digital,* No. 14.06, June, 2006.

12. Doan, A., R. Ramakrishnan, and A. Y. Halevy. Crowdsouring Systems on the World-Wide Web. *Communications of the ACM,* Vol. 54, No. 4, 2011, pp. 86–96.
13. Saxton, G. D., O. Oh, and R. Kishore. Rules of Crowdsourcing: Models, Issues, and Systems of Control, http://www.acsu.buffalo.edu/~rkishore/papers/Saxton-et-al-Crowdsourcing-ISM-Forthcoming.pdf. Accessed May 2013.
14. Goodchild, M. Assertion and Authority: the Science of User-Generated Geographic Content. *Earth,* 2008, pp. 1–18.
15. Kuznetsov, S., and E. Paulos. Participatory Sensing in Public Spaces: Activating Urban Surfaces with Sensor Probes. *ACM Conference on Designing Interactive Systems,* Aarhus, Denmark, 2010.
16. Kuznetsov, S., G. N. Davis, J. C. Cheung, and E. Paulos. Ceci N'est Pas Une Pipe Bombe: Authoring Urban Landscapes with Air Quality Sensors, *ACM SIGCHI,* 2011.
17. Hood, J., E. Sall, and B. Charlton. A GPS-Based Bicycle Route Choice Model for San Francisco, California. *Transportation Letters: International Journal of Transportation Research,* Vol. 3, Jan. 2011, pp. 63–75.
18. Kitchin, R., and M. Dodge. Rethinking Maps. In *The Map Reader: Theories of Mapping Practice and Cartographic Representation* (M. Dodge, R. Kitchin, and C. Perkins, eds.), John Wiley & Sons, Ltd., Chichester, United Kingdom, June 2011.
19. Steinfeld, A., J. Zimmerman, and A. Tomasic. Bringing Customers Back into Transportation: Citizen-Driven Transit Service Innovation via Social Computing. In *Best Practices for Transportation Agency Use of Social Media* (S. Bregman and K. Watkins, eds.), CRC Press, Boca Raton, Florida, 2013.
20. 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 2010,* Copenhagen, Denmark, 2010.
21. Priedhorsky, R., B. Jordan, and L. Terveen. How a Personalized Geowiki Can Help Bicyclists Share Information More Effectively. *Proc., 2007 International Symposium on Wikis, ACM,* 2007, pp. 93–98.
22. Heipke, C. Crowdsourcing Geospatial Data. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 65, No. 6, 2010, pp. 550–557.
23. Brabham, D. C. Crowdsourcing the Public Participation Process for Planning Projects. *Planning Theory,* Vol. 8, No. 3, Jul. 2009, pp. 242–262.
24. SeeClickFix. http://seeclickfix.com. 2013. Accessed May 2013.
25. Ferris, B., K. Watkins, and A. Borning. OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit. *Proc., CHI 2010,* Atlanta, Georgia, April 10–15, 2010.
26. Zimmerman, J., A. Tomasic, C. Garrod, D. Yoo, C. Hiruncharoenvate, R. Aziz, N. R. Thiruvengadam, Y. Huang, and A. Steinfeld. Field Trial of Tiramisu: Crowd-Sourcing Bus Arrival Times to Spur Codesign. *Proc., CHI 2011,* New York, 2011, pp. 1677–1686.
27. Cycle Atlanta. http://www.cycleatlanta.org, 2013. Accessed May 2013.
28. Dill, J., and N. McNeil. Four Types of Cyclists? Examination of Typology for Better Understanding of Bicycling Behavior and Potential. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2387,* Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 129–138.
29. Masli, M. Crowdsourcing Maps. *Computer,* Vol. 44, No. 11, 2011, pp. 90–93.
30. Cyclopath. http://cyclopath.org, 2011. Accessed May 2013.
31. Wiggins, A., and K. Crowston. From Conservation to Crowdsourcing: A typology of Citizen Science. *Proc., Forty-Fourth Hawai'i International Conference on System Science,* Manoa, 2011.
32. Gooze, A. I. *Real-Time Transit Information Accuracy: Impacts and Proposed Solutions.* MS thesis, Georgia Institute of Technology, http://hdl.handle.net/1853/47638, 2013. Accessed May 2013.
33. Gooze, A, K. E. Watkins, and A. Borning. Benefits of Real-Time Transit Information and Impacts of Data Accuracy on Rider Experience. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2351,* Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 95–103.
34. Windmiller, S., T. Hennessy, and K. E. Watkins. Accessibility of Communication Technology Usage and the Rider Experience: A Case Study of St. Louis Metro. Presented at 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.

The Conduct of Research Committee peer-reviewed this paper.