Exploring informants’ perspectives on the role of crowdsourced active travel data

Mohammad Anwar Alattar [a], Caitlin Cottrill [b], Mark Beecroft [b].

[a] *Department of Social Studies, College of Basic Education, Public Authority for Applied Education and Training, Kuwait City, Kuwait.*

[b] *Centre for Transport Research, School of Engineering, University of Aberdeen, Aberdeen Scotland, UK*

Mohammad Anwar Alattar

Department of Social Studies, College of Basic Education, Public Authority for Applied Education and Training, Kuwait City, Kuwait.

m.alattar@hotmail.co.uk

Abstract

In the era of ubiquitous technology, crowdsourced data is an emerging frontier for active travel (AT) studies. In this work, we utilize accrued knowledge from interviews and previous literature regarding crowdsourced data strengths, challenges, usefulness and reliability for future informants who seek to embrace crowdsourced data. We review four main types of crowdsourced data: social fitness networks, in-house developed apps, bike sharing systems and participatory mapping. The strengths of crowdsourced data include providing fine data coverage, precision, details, immediacy and empowering users to participate in decision-making. Potential challenges that might arise from adopting this data are related to technical, privacy, proprietorship, financial and data fragmentation factors. In terms of usefulness, crowdsourced data lend themselves to before and after analysis, assessing current infrastructure, and investment prioritization. Reliability issues that may undermine the credibility of crowdsourced data are also discussed, as well as remedies for these concerns.

Keywords: active travel; crowdsourced data; social fitness networks; bike-sharing; interview.

1. Introduction

Active travel (non-motorized travel [AT]) has recently received much attention due to its potential for conquering externalities from urbanization, including but not limited to, sedentary lifestyles (Townshend and Lake 2017), traffic congestion, and air pollution (Rissel 2009). More recently, during the COVID-19 outbreak, AT has been promoted as helping to maintain social distancing (Vos 2020). Thus, numerous cities around the world are currently promoting AT and undertaking activities to reduce reliance on motorized transport. However, a lack of sufficient AT data poses challenges in making well-informed decisions.

Previous research has employed traditional AT data sources to investigate various aspects of AT; for example, Raford et al. (2007) relied on cordon counts to model cyclists' route choices. Yet relying only on traditional AT data limits the available information on AT, as it reflects the spatiotemporal frame of the sensor used. Manual data collection methods (e.g. video recording, travel surveys, and handheld counters) require a low technological readiness level (AMEC E&I and Sprinkle Consulting 2011) and provide contextual data on AT users (e.g. helmet usage and gender) (National Academies of Sciences, Engineering 2017). However, these methods are deemed to be cumbersome and have a low spatiotemporal resolution (Day, Premachandra, and Bullock 2016). Automated methods (e.g., infrared sensors, magnetometers, and pressure pads) are more resilient to vagaries of weather yet are subject to faulty detection.

Advances in information and communication technology (ICT), particularly Web 2.0, have facilitated user-generated content, the proliferation of GPS-enabled devices, and the provision of crowdsourced data. Crowdsourced data leverages the input of users addressing the same issue or topic (Smith 2015). This data has a wide variety of applications. For example, Wikipedia allows users to share, confirm and edit content about any topic. OpenStreetMap operates in the same fashion, with a specialization in mapping locations (Barbier et al. 2012). The field of transportation is no exception, with a new stream of research and global applications embracing crowdsourced AT data (hereinafter referred to as 'crowdsourced data') to improve the understanding and monitoring of AT.

Benefits and applications of crowdsourced data (discussed in **Section 3**) can be seen by informants (e.g., scholars, transportation planners, advocates). Recently, accrued informants' expertise and knowledge from interviews and literature, have become a valid source for various data-related issues. For example, Griffin et al. (2020) examined biases in big data on transportation and mitigation approaches using semi-structured interviews (see also Griffin et al. (2018) and Lee & Sener (2020)). The current study deploys such a method to identify issues related to crowdsourced data strengths, challenges, reliability and usefulness, in an attempt to help those who are seeking to adopt this type of data in their research and to maximize its benefits.

1.1. Study rationale

The immense potential of crowdsourced data has been the focus of much attention. Accumulated knowledge from informants inherent in literature and their expertise provides valuable guidance for those who are seeking to embrace this type of data. Such guidance is supportive of a transportation paradigm shift towards a safe, pleasant and efficient AT environment. The guidance may also help with developing platforms that consider the perspective of informants. Through interviews and relevant literature, this work aims to document and bring greater clarity and visibility to informants' perspectives on the usage of crowdsourced data, with the corresponding objectives of identifying the following:

1- *Strengths: What are the strengths of crowdsourced data from an informant perspective?*
2- *Challenges: What are the challenges that might arise when adopting crowdsourced data?*
3- *Usefulness: What are the roles of crowdsourced data in investments and decision-making?*
4- *Reliability: Considering that reliability determines the success of crowdsourced data adaptation, what issues may arise?*

1.2. Paper outline

The rest of the paper is structured as follows: Section Error! Reference source not found. provides an overview of crowdsourced data sources and their applications in transport as well as the theory of change leading to the development of the research objectives. Section 3 discusses the methodology. The results are presented in Section **4** and discussed in the light of current literature in Section **5.** Finally, Section **6** presents the conclusions, implications and limitations of the paper.

2. Crowdsourced data in AT

2.1. Social Fitness Networks

Crowdsourced data in the transport sector can be acquired from various sources. Social fitness networks (SFNs) are networks that permit users to share, track, and analyze their physical activities (e.g. walking, cycling, kayaking). These networks often provide a gamified environment by allowing users to compete with their peers (Stragier, Evens, and Mechant 2015). For example, Strava Metro[1] is a prominent data service provided by the Strava SFN that offers cleaned, anonymized and aggregated data from Strava app users.

Lee & Sener (2020) presented a comprehensive literature review that discusses the applications, challenges and reliability of Strava data. In respect to Strava applications, Strava data has been implemented in route choice models (Orellana and Guerrero 2019), infrastructure appraisal (Hong, McArthur, and Livingston 2018), research on the conflict between recreational activities and nature (Jäger, Schirpke, and Tappeiner 2020; Thorsen et al. 2022; Venter et al. 2020), and determining cyclists' air pollution exposure (Lee and Sener 2019; Sun and Mobasheri 2017). The main challenge of Strava is the representativeness of the data, as it tends to over-represent certain segments such as affluent, youth and tech-savvy groups (Lee and Sener 2020). Validating the data using official ground-truth data is a common practice in order to ensure data reliability (Lee and Sener 2020).

In terms of cost, Ohlms *et al.* (2018) estimated the Strava license fees for Virginia to be $300,000 for a dataset consisting of 110,000 users with 2.5 million activities. Maus (2014) reported that the Oregon Department of Transportation paid $20,000 for 17,000 users with 400,000 activities. Though in the light of the COVID-19 pandemic, Strava data now are offered for free for certain organizations to accelerate the shift toward AT transportation (Strava Press 2020). Strava declared that they seek to partner with organizations that are working to enhance AT and do not partner with real estate investors, retailers or financial services companies (Strava Metro 2020a). More recently, no technical experience is required to use Strava data. Through the new Strava Metro platform, users can explore insights and statistics as well as visualize the data. Among other features, **Figure 1** shows the corridor feature which visualizes high volume corridors, the example used is in San Francisco.

---

[1] https://metro.strava.com

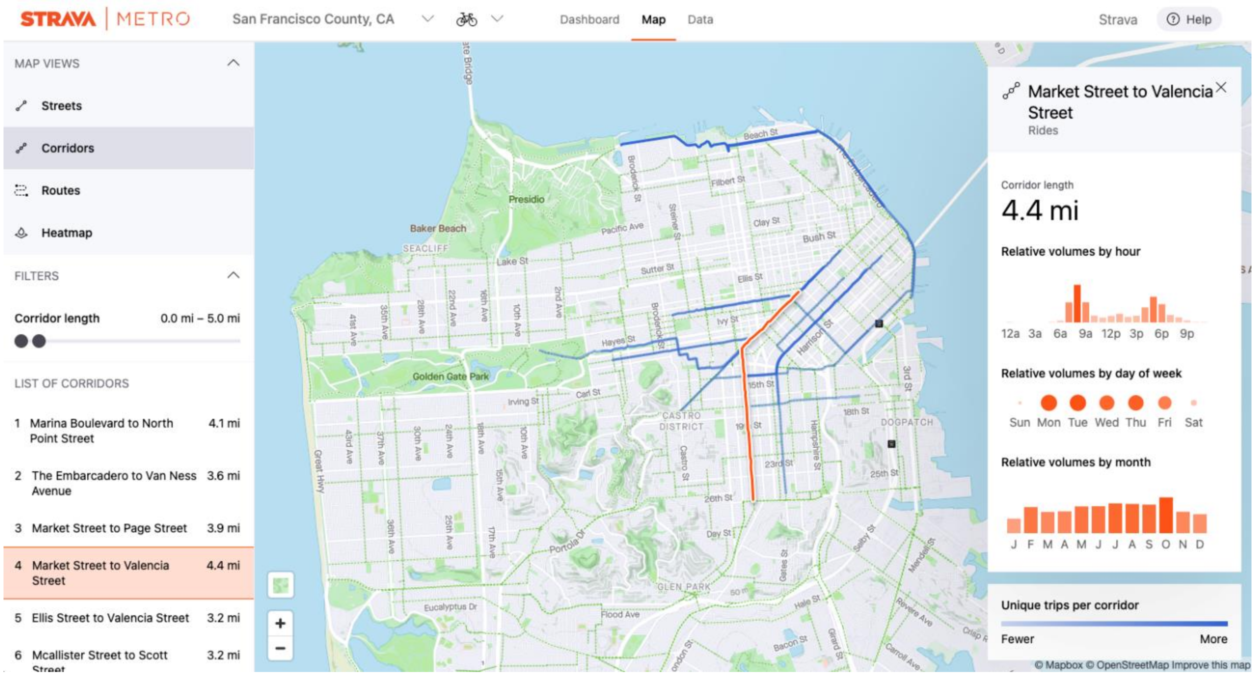

Figure 1. Strava Metro Corridors Feature (Bunn 2021).

2.2. In-house developed apps

Rather than using third party data, some agencies opt to develop their crowdsourced data collection platforms in-house. For example, the Cycletracks[2] app has been developed by the San Francisco County Transportation Authority (SFCTA) to collect data from cyclists about their route, trip purpose, data and time. The success of this app has led to its adoption in many cities along with rebranding in others (SFCTA n.d.). Hood et al. (2011) employed Cycletracks to model cyclists' route choices in San Francisco.

2.3. Bike-sharing systems

Bike-sharing systems (BSSs) operate by allowing users to check out/in bicycles in order to use them for a certain time period. Most metropolitan cities adopt these systems to facilitate short trips and address the first/last mile problem (Yang et al. 2019). A large amount of the data from these systems is openly available online, including CitiBike[3] (New York City, US), Metro Bike Share[4] (Los Angeles, CA), and Santander[5] (London, UK), amongst others. The datasets typically include the bicycles' origin and destination station and the check-in and -out time for each trip. Some systems also provide the demographic data of the trip user. Variants of these systems include dockless (or floating) frameworks where, unlike dock-based systems, bikes do not need a physical dock/station. An et al. (2019) investigated the effect of weather on cycling in New York using

---

[2] https://www.sfcta.org/tools-data/tools/cycletracks
[3] https://www.citibikenyc.com/system-data
[4] https://bikeshare.metro.net/about/data/
[5] https://cycling.data.tfl.gov.uk

CitiBike data, while Hamilton & Wichman (2018) examined the impact of Washington DC's BBS, Capital BikeShare, on traffic congestion. Qian and Jaller (2022) compared dockless to dock-based BSSs in San Francisco and Los Angeles. In both cities dockless BSSs were found to have greater area coverage, implying that dockless tend to provide more equitable access.

2.4. Participatory mapping

Participatory mapping (PM) uses community knowledge to obtain spatial information by involving users in mapping a given topic (Da Silva et al. 2020). PM is a survey-like instrument where participants are allowed to input spatial information (using points, lines, and polygons) and non-spatial information through filling in (usually online) forms. Its derivatives have an emerging presence in the collection of AT data. Platforms that are created by citizens to collect, assemble and share geographic information, known as volunteered geographic information (VGI), are often used to address cycling safety-related issues (Brown 2017). BikeLaneUpRising[6] allows users to report the location of bike lane obstructions to identify obstruction (e.g., vehicles and obstruction sites) hotspots, aiming to hold offenders accountable and to make cycling safer. Wheelmap[7] is a VGI platform used to determine and rate the extent of wheelchair accessibility, ranging from fully accessible to not accessible. The second derivative of PM is public participatory geographic information systems (PPGIS), which intend to involve the public in mapping activities with the aim of achieving more informed decision making (Brown 2017). Maptionnaire[8] enables users to easily design their PPGIS instrument allowing for modularity and scalability. Gerstenberg et al. (2020) deployed PPGIS to identify cycling activity hot spots in German urban forests by having cyclists sketch their routes.

2.5. Social media

Conventional social media platforms (e.g., Twitter, Facebook) have great potential in providing AT data given their popularity and ease of use, particularly with the advent of geo-tagged microblogging. These platforms can be mined to reveal useful insights relating to transportation. Evans-Cowley & Griffin (2012) conclude that social media platforms maximize community engagement in transportation planning. Gu et al. (2016) used Twitter to extract information on traffic incidents from highways and arterial roads in the Pittsburgh and Philadelphia metropolitan areas. Similarly, Kumar et al. (2014) detected road hazards using Twitter. Presumably, road safety impact all road users, among of which AT users Zeile et al (2016) and Hollander & Shen (2017) analyzed social media feeds to conduct cyclists' sentiment analysis; a technique used to detect users' sentiment (positive, negative, neutral) toward transportation. Rahyadi (2021) used 3805 posts on the social media platform Instagram, to understand how civil society is advocating for pedestrian rights.

2.6. Theory of change

To support the adoption of crowdsourced data, Adler et al. (2014) developed a Strengths, Weaknesses, Opportunities and Threats (SWOT) model to assess the practice of crowdsourced data in traffic management centers' operations. Primarily, SWOT analysis is a strategic tool to

[6] http://www.bikelaneuprising.com/
[7] https://wheelmap.org
[8] https://maptionnaire.com

assess internal capabilities (strengths and weaknesses) and external considerations (opportunities and threats) for businesses. Among other fields, SWOT analysis has been applied to ICT to assess crowdsourced data (Adler et al. 2014; Smith and Harris 2017). As presented in **Figure 2**, such a model helps agencies to delineate the readiness of adopting this type of data. Herein we used the following terms to develop our themes of interest 'challenges' instead of 'weaknesses', 'usefulness' instead of 'opportunities', 'reliability' instead of 'threats', while 'strengths' remain as is. Some of these themes have witnessed academic scrutiny. Strava reliability and challenges have been scrutinized by Lee & Sener (2020), whereas Lee and Sener (2020) puts forward the usefulness and challenges. Lee and Sener (2017) consolidated all mentioned crowdsourced data.

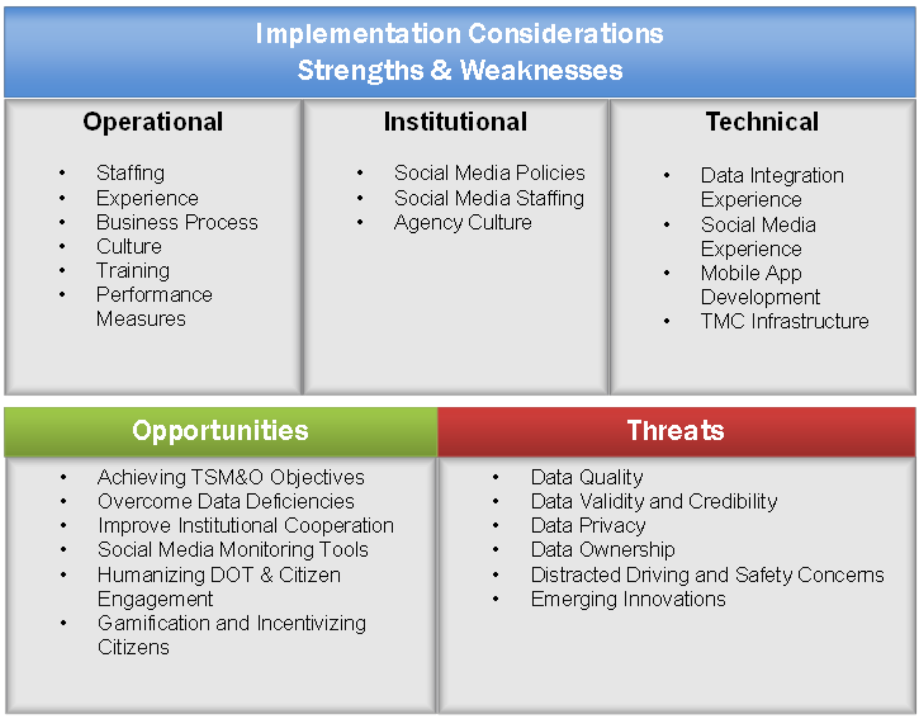

Figure 2. Adler et al (2014) crowdsourced data SWOT model.

In this work, we extend and improve the previously mentioned studies to deepen our knowledge on crowdsourced data strengths, weaknesses, usefulness, and reliability. Our research methodology has been derived from a Griffin et al. (2020) study on the biases on transportation big data using expert interviews. It is hypothesized that new insights on these topics will guide informants in adopting this type of data and further benefitting from its potential.

3. Methodology

Semi-structured interviews were undertaken with an aim to document informants' perspectives on crowdsourced AT data's strengths, challenges, usefulness and reliability (research objectives mentioned in **1.1**). Such interviews are loosely structured on the aforementioned research objectives themes, and their effectiveness in capturing such perspectives is two-fold. First, the interactive dialog encourages subjects to express their viewpoints, particularly in this work where our participants can contribute with detailed information given their 'special or authoritative' role. Second, this type of interview allows the interviewer to seek further clarification and more details where needed (Gillham 2000; Mason 2017). We developed a semi-structured interview guide to capture insights from crowdsourced data informants. This guide is equipped with follow-up questions (also known as probes) to encourage interviewees to express their thoughts.

3.1. Recruitment

An informant is defined here as any individual who has expertise in using crowdsourced data to conduct any type of analysis. Informants may belong to one or more of the following categories: (1) scholars using crowdsourced data for research purposes; (2) planners using crowdsourced data for monitoring and evaluating transport interventions and services; (3) developers creating a research instrument to collect crowdsourced data; and (4) advocates using crowdsourced data to improve AT (or a certain mode).

Interview candidates were identified in multiple rounds of screening. Google Scholar was used in the first round to search the keywords ‘crowdsourced’, ‘cycling’, ‘bicycle’, ‘cyclists’, ‘pedestrian’ and ‘Strava’, allowing for a focus on scholar informants. We contacted the authors of relevant research that has been extensively cited. The second round involved deploying the snowballing technique by asking interviewees to suggest other potential interviewees. Anonymity was offered to encourage interviewee responses.

After ethical clearance, interview requests were sent via email and included a consent form, participation sheet, and a link to an online meeting scheduling tool (Doodle). All interviews were conducted online using video conferencing software, with the exception of three interviewees who preferred to answer the questions by email. The advantages of flexibility and convenance offered by this interviewing protocol made this method to be appropriate, given that the interviewees spanned different countries (Weller 2015). **Table 1** lists the interview schedule deployed in this study. However, further questions related to each of the listed category might be asked to encourage the participants to provide more insights. For example, it is expected to mention the inherited bias in crowdsourced data. Nonetheless, if the participant did not mention any method to offset this bias, they will be asked ‘How can the application of crowdsourced data be optimized in order to reach reliable results?’

Table 1. Interview schedule. Questions in bold were intend only for informants who represent agencies (planners). Questions in italic were intended for other informants (scholars, developers, and AT advocates). Questions with normal formatting were intended for all informants

Category	Question
Data acquisition	**1- How does your department obtain data on AT ?** [If negative answer is provided go to 1a] **1a- Has** y**our** d**epartment considered using crowdsourced data in AT studies and monitoring?** [If negative answer is provided go to 1b] **1b- What are the main reasons to refrain from using crowdsourced data?** *2-* **How does your department encourage participation in AT crowdsourced data collection activities?**

	3- *Have you ever collected and/or used crowdsourced data? If so, what type of crowdsourced data did you use?* ***4-*** *How did you obtain this data?*
Reliability	Do you think crowdsourced data is reliable?
Usefulness	**How does your department use the crowdsourced data that you collect to inform planning and decision-making or monitor AT activities?**
	What is the role of crowdsourced data in investment prioritization?
	What do you think are the primary benefits crowdsourced data?
Challenges	What do you think are the primary challenges of crowdsourced data?
PM Awareness	Are you aware of the online map-based survey/public participation geographic information system (PPGIS)? Are you aware of other crowdsourced platforms related to AT such as BikeMaps, FixMyStreet and BikeLaneUpRising ?
	Is ***Your Department*** using PPGIS to collect data from travelers?
PM Usefulness	What do you think are the primary benefits of these platforms?
PM Challenges	What do you think are the primary challenges of platforms?
Potential Interviewees	Is there any other person, department or agency you would recommend to be interviewed?

3.2. Coding and analysis

NVivo (ver.12, QSR International) was used to transcribe the recordings and to code the transcripts for deductive thematic analysis with predefined codes (**Table 2**).

Table 2. Used codes and their descriptions.

Code	**Description**
Strengths	Key benefits of crowdsourced data.
Challenges	Potential challenges associated with adopting crowdsourced data.
Usefulness	Highlights how crowdsourced data can be utilized to improve AT.
Reliability	Crowdsourced data reliability related issues, including validation exercises.

4. Results

As listed in **Table 3**, a total of 18 persons were interviewed, spanning six countries. Compared to a related study (n = 10) (Griffin et al. 2020), our sample size is believed to be adequate for the purpose of the study. Our sample included interviewees from diverse groups in terms of their informant role and the corresponding crowdsourced data sources with which they are familiar or are actively using. Roles included scholars, transport planners, cyclist advocates and crowdsourced data platform founders, with some interviewees having experience of multiple roles as shown in **Table 3**, while data sources covered SFNs, BSSs, VGI and PPGIS platforms and in-house developed apps. Most sources were primarily concerned with cyclists. Nonetheless, some interviewees incorporate other travel modes in their data (e.g., walking) as well as winter sports (e.g., skiing and snowboarding).

Table 3. Interviewee characteristics.

Interviewee #	**Role**	**Number of publications±**	**Location**
#1	Scholar and planner	48	USA
#2	Scholar.	100	USA
#3	Scholar and former planner	12	USA
#4	Scholar	2	Germany
#5	Scholar	2	Germany
#6	Scholar	27	UK
#7	Scholar	--	USA
#8	Scholar	25	UK
#9	Scholar and platform founder	29	Canada
#10	Scholar	7	USA
#11	Scholar	24	USA
#12	Scholar	25	Norway
#13	Planner	--	USA
#14	Scholar	75	UK
#15	Cycling advocate	--	USA
#16	Scholar	12	New Zealand
#17	Scholar, former planner	10	USA
#18	Scholar and platform founder	20	Germany

± Number of overall publications until Feb 2021

4.1. Strengths

Topics related to the strength of crowdsourced data were coded in 41 instances. The data provided by crowdsourced data has unprecedented resolution in terms of coverage, precision, details, immediacy and empowering users to participate in decision-making. These qualities are more cost-effective and feasible compared to traditional counts. Interviewee #2 compared the data coverage (explained in 4.1.1) between crowdsourced and traditional data, saying:

'I don't want to give you the impression that I don't think it is a financial challenge. But I do think that the data is getting cheaper, right, and in relative sense the cost of going out and installing counters on every single segment is going to be cost prohibitive'.

4.1.1. Data Coverage

In virtue of crowdsourced data's grassroots, fine spatiotemporal granularity provides coverage that far exceeds traditional data. This means it is possible to collect data that covers the entire street network and extends temporally continuously. This is evident in the volume of data provided by SFNs and BSSs. Interviewee #17 said "*Maybe first and foremost thing is, unprecedented high spatiotemporal resolution data"*. Interviewee #11 added, '*Bike share [that we used] is different from other bike share in that, when the bike is in motion it collects GPS every six seconds*'. This feature is useful during the COVID19 pandemic. Interviewee #1 said:

'*I think the COVID pandemic is showing benefits to having crowdsourced and big data sources available at a global scale at high resolutions to be able to understand changes and behavior, and most of those are publicly available*". This is in alignment with the recent Strava announcement where the data are free of charge for certain organizations (mentioned earlier in 2.1). The change in AT ridership due the social distancing measures has been reported using Strava. Consequently, New York City dedicated 100 miles of streets to AT to match this increase

(Strava Press 2020). This was also confirmed by Interviewee #14, who mentioned, "*because you get a picture of where people are cycling, it tells you where it has been increasing in cycling'*.

4.1.2. Data Precision

When mapping single issues through PM platforms, such as safety (e.g., incidents, hazards and theft) or street problems (e.g., graffiti, potholes, lighting), usually the data feature greater location precision as users are asked to map the exact locations, unlike official data which tends to provide data aggregated to the nearest intersection or block centroid. As Interviewee #9 put it:

'*Traditional types of data get aggregated to say the nearest intersection or nearest midblock, so traditional data may lack precision in location. Whereas with crowdsourced data, people who are taking the time to report generally are very diligent and they make sure that is in the right spot'*.

4.1.3. Data Details

Crowdsourced data provide additional nuanced details which enrich the data that would not be obtained otherwise. For example, volume data provide demographic data and trip purpose. Interviewee #12 pointed out that, "*Strava data in general is divided into leisure and commuting trips. Also, more recently, they provided us the age brackets of the users and the gender*". This level of detail allows crowdsourced data to complement official data. Two interviewees suggested that official data sources tend to overlook certain occurrences (e.g., non-serious injuries in case of safety and minor incidents) whereas PM safety data cover these instances:

'*…bicyclists that they don't get injured or limited damages, they will not report it*', [Interviewee #3].

'*…official and more traditional data severely underreport crashes between bikes and motor vehicles and then it would not have collected near misses or any kind of incidents with non-motor vehicles*', [Interviewee #9].

4.1.4. Data Immediacy

The immediacy of the crowdsourced data allows it to be utilized as a surveillance tool, allowing decision-makers to monitor and intervene in a timely manner to any unexpected changes.

'*…when a city or jurisdiction puts a new infrastructure you can immediately keep an eye on some of the challenges with it*', [Interviewee #9].

'*…crowdsourced data can be collected quickly in real-time or near real-time and process potentially, with AI and computers also in near real-time. And get very quick picture that survey data cannot do because survey data you got to collect and analyze. So crowdsourced data really is great because it offers the potential to do data driven decision-making for public policy*.', [Interviewee #2].

4.1.5. Enabling Users

Generally, crowdsourced data enable users to be involved in decision-making processes through channeling their data to be used in data-driven decision-making. Interviewee #13 mentioned that, '*the ability to engage directly in decision making that opens up that avenue of saying "hey you can vote with your feet you just need to let us know where you are at and engage in this sort of process*"'. Given the PM modularity, which refers to the degree to which a PM administrator can structure a questionnaire using a wide range of both conventional questions (i.e., multiple choice, rating, matrix) and spatial questions (i.e., sketching points, lines and polygons); these platforms have a more direct influence on decision making and increase decision-making transparency, as users are allowed to annotate their participation, and in some cases view the participation of others.

'*It increases the transparency in the process'*, [Interviewee #1].

'*It enables a path of communication between citizens and the city or municipality or government so that we can better communicate to our government where the problem areas are*', [Interviewee #15].

'I think if you have the right type of research questions in mind, or the right types of courtesy, you can get a lot of insights about things that you don't just get from raw data ... So, to me, when it is in a situation where you are having a direct question about where do you see barriers or what are the issues or where do you like to ride. This is where the participants can input some more of that annotated or richer type of experiential information', [Interviewee #16].

4.2. Challenges

Topics related to challenges of crowdsourced data were coded 61 times. These challenges should be taken into account when using crowdsourced data for more accurate results. We categorized the associated challenges as follows:

4.2.1. Technical Challenges

Technical issues can at times result in challenges. Potential challenges with BSSs data result from misusing bikes or faulty equipment that may produce inaccurate records.

'*…somebody does not put the bike all the way, that might may misrepresent the trip length in terms of time … Or let's say somebody takes it out of the dock and then they realize that the seat is broken and then they re-dock it for example…it requires a pretty sophisticated level of cleaning to make sure that you are not getting false responses*', [Interviewee #2].

SFNs datasets that rely on GPS traces might be subjected to erroneous counting. There is a possibility of overcounting the number of users by double counting a GPS trace when using simple map-matching methods, which represents a single user. Interviewee #3 elaborated on this:

'*Each link [Streets] how many people cycle around that and they assign this number to the link. if you have two links and the buffers overlap with each other, that will double-count the number of cyclists. So, when this issue happens, we have to dig in the data and we clean this up. This is one of the issues you might face*'.

This issue has been further discussed in a Strava Support webpage as GPS errors and drifting might cause the appropriate matching and consequently counting. As a remedy, a Potential Segment Match Analysis tool has been introduced to allow user to interfere to match their activities with the appropriate street segment (G. 2022)

4.2.2. Privacy Challenges

There is a trade-off between the data granularity and maintaining privacy and anonymity. Interviewees have reported several challenges associated with this trade-off in Strava given its popularity, thus it is more susceptible to privacy incidents. For example, the inconsistences with the data due to changing the raw data specifications.

'*I think Strava was used to detect some military bases for the US because people left the phone on and mapped out where the troops go running. So they tightened down a bit and changed the data specification. They also changed how the raw data captured from the people's phones* *were put into the final data product, so you then have a problem of the method that they are using to adjust the data changed and the specifications change*', [Interviewee #6].

Another challenge is related to the number of counts. Strava bins counts to the nearest 5, and thus does not report the exact number of counts, causing some data gaps. This may limit the data applications where the number of users is expected to be below 5. Interviewee #12 mentioned:

'*That means if you have a remote trail segment in some protected areas or up mountains somewhere, and say only one or two users, that activity won't be registered on Strava. So you will have a bit of data gaps in more remote areas.*" Interviewee #14 witnessed changes in their results between the binned and raw data, suggesting that this measure threatens the replicability: "*we used their data, their un-binned data to replicate the binning process and look at the consequences to research that we have already published and already done. The results of the binning process change the results of that research and the conclusion of that research*'.

4.2.3. Proprietorship Challenges

Proprietorship challenges arise with third-party data. Changes in the terms and conditions may result in detrimental effects on data users and inconsistencies in methodologies, consequently impacting the availability of the data and its usage. Interviewee #14 mentioned that changes in their contract with Strava prevents them from validating the data. Similarly, Interviewee #6 complained about the need for approval prior to publishing any research involving Strava data.

'*They would not allow you to validate the data. So they specifically say "you cannot validate the data"*', [Interviewee #14].

'*New terms and conditions require you to run any output by the company for approval prior to publishing them. Which for academics is clearly a major problem*', [Interviewee #6].

4.2.4. Financial Challenges

Crowdsourced data is perceived as cost-effective owing to the advantages it confers (see Section 4.1), and some datasets are available for free. Interviewee #9, a PM platform developer, clarified that, '*We routinely share with people who reach out. We offer the chance to, if they are going to publish the material, offer the chance to weigh in on it*'. Nevertheless, financial challenges pose a threat to these platforms' sustainability.

'*We have been publicly funded by the public health agency. That funding has wrapped up but we can kind of keep the platform going just based on other research funding*', [Interviewee #9].

'*Another challenge is financial sustainability. So these things require money, right? Like, he [referring to PM founder] needs to pay for a server that is going to manage this data. He had to put man hours to build it, to make it functional, to do quality assurance, all that kind of stuff'*, [Interviewee #14].

All scholars who used Strava obtained it from funded entities (e.g., universities, research agencies, departments of transportation). However, acquiring data from third-party may impose a financial burden for entities with a limited budget and individual researchers. Interviewee #12 commented, '*It will be very difficult for individual researchers, for example, to enter into a contract like we have entered into*'.

Additionally, data acquiring expenditure may be a barrier for entities with limited financial resources. Interviewee #8 stated:

‘*If we are talking about some places that are big like San Francisco or New York or Phoenix or Glasgow or London, you know, these governments they have money to buy Strava data, and then we talk about places smaller, local authorities, they won't be able to afford it, most likely*’.

4.2.5. Fragmentation challenges

Multiple PM platforms may collect data on the same topic, causing data fragmentation. For example, both BikeLaneUpRising and SafeLanes (https://safelanes.org/) are concerned with bicycle obstructions. This fragmentation may be burdensome for data users who are inclined to obtain all relevant available data. Considering the variable platform visibility, this fragmentation may hinder comparability between different locations.

‘*So you have multiple platforms across multiple geographies and so the data is not necessarily comparable. So, like, I cannot necessarily compare SafeLane in SF to BikeLaneUpRising in Chicago. If it were all under one platform that would be a lot easier for people like me*’, [Interviewee #9].

4.3. Usefulness

The usefulness of crowdsourced data was coded at a less frequent level (27 times) compared to the other factors. Although not all interviewees were able to illustrate the usefulness of the data, we were able to identify the following subthemes.

4.3.1. Before and After Analysis

The high spatiotemporal resolution of crowdsourced data allows for its application in the comparison of AT user numbers. Such comparisons can be employed to evaluate and justify investments. Interviewee #14 suggests that monitoring changes without crowdsourced data may not be accurate, reporting:

‘*You put sensors up and you count those sensors before and then count those sensors afterwards, if there has been an increase in cycling. But that does not tell you that it encourages people to cycle. It just tells you that there are more people using that cycle lane. That could be because people are diverting, that might have no impact on increasing cyclists at all. It might just divert people from other places. So you cannot do that without data like Strava.*’

Interviewee #12 highlighted an important caveat regarding the comparison of numbers before and after introducing new infrastructure, emphasizing that controlling for usership is necessary:

‘*We have Strava data for 2016 but over time Strava users’ adoption increases. If you just look at the raw data, it looks like the recreational activity increased since 2016 onwards. That was confounded by the increase in the number of installed Strava apps on the phones. You have to correct for Strava usership increases over time and that is possible to do that.*’

4.3.2. Current Infrastructure Assessment

Current facilities and ongoing projects can be assessed and altered to accommodate the needs of potential users. Crowdsourced data can be used to identify where potential may occur. In the light of such findings, decision-makers could better cater for AT users.

‘*The direction of the Program has changed in response to feedback we heard, specifically that the program was not meeting the needs of lower income communities of color*’, [Interviewee #7].

‘*Maybe you do not have the room to have a segregated cycle path there for the whole time, but you can provide one so that it is for a given time, for the peak times where the cyclists are commuting and that is a way to use this kind of data to think about infrastructure and that is benefits not only for cyclists but also for drivers that won't be annoyed by cycle path*’, [Interviewee #8].

4.3.3. Investment Prioritization

Crowdsourced data can help in prioritizing and delineating the optimal location of future investments. This approach can be undertaken in a simple manner, where determining certain beneficiary criteria such as building a trail in a location used by a large number of people or serving vulnerable groups.

‘*If I want to promote active transportation or build a new trail like a bike trail. Where can I build and which route I can take that I can serve more people*’, [Interviewee #3].

An alternative approach, denoted as community engagement, engages users in the process of locating a new infrastructure. For example, CitiBike incorporates PPGIS to extend existing stations or to suggest a new station (**Figure 3)**. Furthermore, potential stations can be identified by equipping bikes with a GPS to signal the location of non-docked bikes.

‘*CitiBike they used what I would call crowdsourced community engagement so when they were designing the system they had not only in-person community meetings but an online platform you could go and say hey I want a kiosk or station here*’, [Interviewee #2].

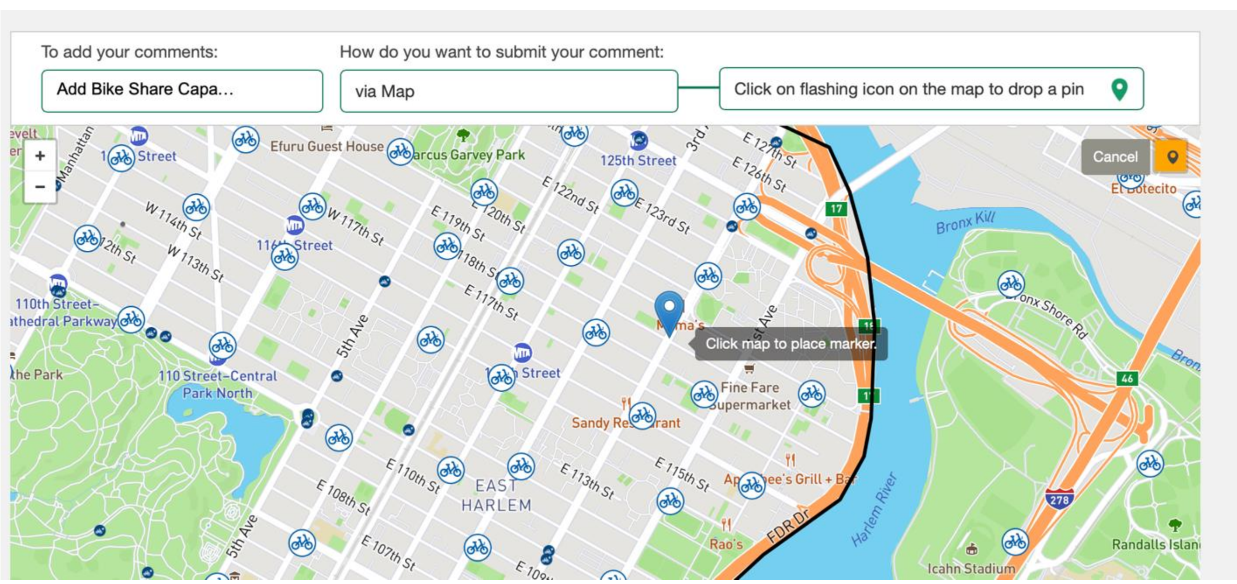

Figure 3. How to suggest the location of a new CitiBike station. The user can see existing stations and stations suggested by other users and can propose a new bike station (https://nycdotprojects.info/project-feedback-map/suggest-station-infill).

4.4. Reliability

Reliability issues were mentioned frequently by interviewees (40 times). In addition, interviewees discussed several remedies for reliability issues that may undermine the credibility of crowdsourced data.

4.4.1. Reliability of SFNs and in-house developed apps

Representativeness was a key topic for interviewees. SFNs, which typically collect volume data, were particularly criticized as being biased. Two principal biases were identified, namely, social desirability and self-selection bias. Social desirability bias occurs when participants share data to display a sense of accomplishment, Interviewee #1 remarked '*social desirability bias in Strava just like Instagram people only post things that they are proud of*'. Self-selection bias emerges when participants can include or exclude themselves from the data, as explained by Interviewee #13 as:

'*Amish population within the state, a religious sect that essentially, for the most part is staying from using modern technology. This population is not going to have a smartphone, and use it in the same way that would be necessary to capture their movement data in a platform like Strava*'.

Almost all interviewees who were SFN data users encouraged data cross-validation. Cross-validation is the practice of statistically comparing crowdsourced and ground-truth data.

'*I think validation is useful. Certainly, if you want to do any kind of generalization, you have to say something about how it relates to measures that we know a bit more about how it is generated. It is something difficult to get ground truth, you cannot always find something that matches the dataset well, but I think some sort of validation exercise is quite important to do*', [Interviewee #18].

This practice is typically conducted using data obtained from traditional data sources such as cordon counts. However, even traditional data sources are not necessarily representative for the population (e.g., counters), and should be distributed in a way that accounts for different AT users' profiles.

'*I think that if municipalities and local governments are actually thinking of using that sort of data for their transportation planning, they first need to think about implementing counts program and make sure they distribute their count so that they actually are representative of different ridership profiles and then validate*', [Interviewee #9].

4.4.2. PM Reliability

Unlike the majority of crowdsourced data sources, PM platforms are subject to further potential issues due to a higher level of interaction required from the user while reporting or donating information; for example, the erroneous data entry resulting from mistakenly inputting data. As Interviewee #10 suggests, '*someone could put the wrong street address. So that makes the data less reliable*'. As a countermeasure, several platforms require a photograph to submit the data, which verifies the submitted record by simply cross-checking the inputted data with the photo.

In the incident of mistakenly classifying a car obstructing a bike lane Interviewee #10 detailed that, '*the photograph has a way to check the quality of the record so they say it is a firetruck, you could get a picture, no it is a police car, that type of thing.*'. Additionally, the metadata from the photo (i.e., timestamp, GPS coordinates) can increase the data precision. Interviewee #10 continued, "*geotags stamp on the photograph, meaning the coordinates on the photograph are inputted or you can put it manually*".

The second issue with PM platform is data inconsistencies due to the variability of contributor levels in understanding the PM topic, as explained by Interviewee #10:

'*I believe SafeLanes has hundreds of people using it, all those people understand bike lane obstruction to be different things. So, does it mean a car parked in the bike lane? Does it mean a car parked near the entrance of the bike lane? Does it mean a car parked adjacent to the bike lanes?*'.

5. Discussion

Informants are becoming increasingly attuned to the impact of robust data in order to make informed decisions. Currently, crowdsourced data is shaping AT research. In this work, we interviewed knowledgeable informants identified from publications and snowballing. The majority of our sample used cycling crowdsourced data, whilst other AT modes remain understudied. This is also supported by Griffin et al. (2014), who claimed that pedestrian data is still primitive and limited to traditional sources. In the following subsections, we discuss each crowdsourced data sources' strengths, challenges, reliability and usefulness.

5.2. SFNs and In-House Developed Apps

SFNs are generally employed for volume data on each street segment. Bearing in mind that users of SFNs tend to be fitness-oriented, long and recreational trips are over-represented in these datasets. Although a great number of studies have used SFN data to study cyclists; several studies exist that examine non-cyclists (i.e. pedestrians and skiers). Data beneficiaries will receive cleaned and already processed datasets at the expense of relatively high overhead costs. In-house developed apps act similarly to SFNs, although they do not belong to commercial companies. However, unlike SFN data, raw data have to be cleaned (noise removal from the GPS signals) and subsequently go a through map-matching process where the cleaned GPS traces convert matched to street segments.

The strengths of crowdsourced data are related to its coverage, details, precision, immediacy and enabling users to contribute. Since AT is spatiotemporally-situated, the fine spatiotemporal resolution allows for a more detailed understanding of behavior. For example, cyclists tend to alter their routes in inclement weather conditions (Griffin et al., 2014). This change in cyclists' behavior will be captured through crowdsourced data. This is unlike traditional data sources, which have more limited spatiotemporal coverage; for example, cordon counts, which provide data on pedestrians and cyclists in a limited geographic area (Livingston et al. 2020). Bunn (2019) reported bike counts to underestimate the number of cyclists compared to Strava during the 'Five Boro Bike Tour' event. The underestimation was attributed to the use of lanes not dedicated to cyclists. Crowdsourced data additionally possess a wealth of attributes, such as AT user speed and direction (Desai et al. 2021).

These forms of crowdsourced data are, however, more susceptible to privacy challenges. Strava SFN data, for example, only offers consented trips, and includes limited information (e.g., trip purpose and demographic data) and aggregates counts in five-count buckets (Lee and Sener 2020).

Proprietorship challenges may yield data inconsistencies and deter data comparisons and method replications. Given the nature of SFNs, privacy is a key concern, leading to changes in the data specifications. In agreement with Interviewee #6, BBC News (2018) reported incidents where military bases and outpost locations were revealed unintentionally via Strava. Burgess (2018) reported an incident where a criminal targeted a cyclist by determining his location and stole his bike using Strava. These challenges are in line with Lee & Sener (2020), who indicated that for such privacy issues, Strava dataset specification changes; for example, Strava decided to stop providing minute-by-minute data for routes with fewer than 3 users. Additionally, the authors demonstrated that double-counting issues can be solved algorithmically. Another potential breach of privacy raised by Kazlouski et al. (2021) is that SFNs share personal information (e.g., location, phone model, and SIM carrier) with third parties. Further proprietorship challenges are expected by reviewing the Terms of Use of Strava. Licensee reports (i.e., summaries, research papers, studies, reports, charts, tables, graphs and other analyses that incorporate Strava data) are subjected to the following Terms of Use which might limit its applications.

i) *Data cannot be compared or benchmarked against third party data except for the validating purposes; though refuting Interviewee #14 claim in Section 4.2.3.*

ii) *Strava's review and approval is required before publishing a licensee report.*

In terms of the overhead cost of SFNs, Strava recently alleviated this challenge for organizations by providing their data for free to qualified groups. The qualification is determined by Strava following vetting and approval (Strava Metro 2020b). Alternatively, data beneficiaries may develop their apps in-house to mitigate proprietorship and financial challenges. The CycleTracks app and its variants have been adopted in over 10 cities and may provide a solution for such financial and proprietorship challenges. However, the data from this app requires processing as it provides raw datasets (SFCTA n.d.). Interviewee #18 suggested the application of open source initiatives rather than commercial products. Bike Data Project[9] is a platform where individuals can donate data from trips that have previously been recorded through Strava or other GPS tracking apps.

The main threat to SFNs' reliability was identified as representativeness. Cross-validation is a quasi-mandatory exercise to examine representativeness, whereby researchers can compare their dataset with a more robust dataset. However, several interviewees found that the unavailability of ground-truth data can hinder this examination. Kim (2020) and Serra et al. (2020) proposed the use of drone technology and CCTV, respectively, both of which were found to be highly correlated with manual counts. Data fusion is a promising technique applied to increase the representativeness of data, whereby two or more datasets are combined to yield a more accurate dataset. Some attempts have been made to increase the visibility of these apps to improve the representativeness. A cycling advocacy group in Ottawa and Gatineau encouraged commuter cyclists to used Strava to record their "mundane trips" such as library and grocery store journeys (Pritchard 2016). Similarly, in the Netherlands a diligent advertising campaign for the Bike PRINT app was able to reached representative samples (Garber, Watkins, and Kramer 2019). During the COVID-19 lockdowns, the bias in Strava representativeness (age-wise and gender-wise) decreased in Vancouver and Victoria, Canada (Fischer, Nelson, and Winters 2022). This is might be the case

[9] https://www.bikedataproject.org

in other locations as well as the pandemic encouraged more AT as a non-pharmaceutical intervention.

Previous studies have demonstrated the usefulness of SFNs and in-house apps. First, in before and after analysis, informants can detect changes resulting from certain types of infrastructure. Using Strava, Sunde (2017) conducted before and after analysis for the City of Seattle and Washington, whereas Hong et al. (2018) performed similar analysis for the City of Glasgow, Scotland. CycleTacks has been employed in before and after analysis to explore cyclists behavior in San Francisco, California. Ferster et al. (2021) and Boss et al. (2018) reinforced our finding related to the importance of correcting for usership by distinguishing between low risk locations (high number of cycling trips and low number of incidents) and potential high risk locations (low number of cycling trips and low number of incidents). Ferster et al. (2021) detected incident hotspots using this method. The second type of usefulness is associated with current infrastructure evaluations. Lee & Sener (2019) and Sun et al. (2017) assessed the current bicycle infrastructure in El Paso, Texas and Glasgow, Scotland, respectively, and the potential pollution exposure. Dhakal et al. (2018) evaluated street network design and cyclists' wrong-way riding in Philadelphia. Third is investment prioritization, where the optimal location for interventions is identified. Ferster et al. (2021) mapped bike incidents to prioritize locations for bicycle interventions.

5.3. Bike-Sharing System

BSS datasets, which represent the usage of rental bikes rather than personal bikes, are slightly sparse (Munkácsy and Monzón 2017). To reduce maintenance overheads, rental bikes usually feature one-gear and unpuncturable tiers, making them heavy and slow compared to personal bikes (DiDonato, Herbert, and Vachhani 2002). Thus, the produced data present short to moderate and utilitarian tips and are useful for the first mile/last mile problem and to explore tourist destinations and behavior. BSSs demonstrated more resilience in the COVID-19 pandemic than other transport systems (Teixeira and Lopes 2020). Although open access BSSs are widely available, some commercial products have also been developed such as Ito World[10]. Non-GPS equipped BSSs spatial granularity are restricted to check-in and check-out stations, limiting the usage of this data. The new paradigm BSSs, which are GPS equipped and dockless, provide greater spatial granularity via the geofence of this system (Cheng et al. 2019).

Privacy challenges are less severe in BSSs data. BSSs provide discrete trip details (e.g., trip duration, trip time, gender and year of birth). Nonetheless, the user identity might be exposed through delineating the exact time and location of the station (Aïvodji et al. 2016). Regarding data reliability, BSSs operators provide cleaned datasets. For example, CitiBike offers their bikesharing data after removing trips taken by staff and trips under 60 seconds (as they may indicate an attempt to re-dock a bike). Jiang et al. (2019) outlined possible noise data that may exist in BSSs datasets as: i) redundant records in which one user appears in multiple similar trips; and ii) incomplete records where one or more attributes are missing from a record.

Duran-Rodas et al. (2020) used BSS data to assess the fairness of bike-sharing infrastructure. A great number of studies have employed such data to rebalance their stations, which is arguably a form of investment prioritization (Costa Affonso, Couffin, and Leclaire 2021; Lu, Benlic, and Wu

[10] https://www.itoworld.com/ito-world-launches-bike-share/

2020; Tian et al. 2020). For dockless BSSs, Ji et al. (2020) proposed a user-based rebalancing approach that determines regions with a bike surplus and deficit.

5.4. PM

PM platforms have been applied to collect crowdsourced data related to topics such as bike lane obstructions and bike safety issues, though despite this extensive research on cycling, other AT modes are understudied. PM platforms such as BikeMaps, BikeLaneUpRising, SeeClickFix, WheelMap and FixMyStreet are believed to capture data that is not available through official traditional sources. Similar findings have been reported in Deparday (2011) and McCallum et al. (2016) in a non-transportation context. The involvement of users through crowdsourced data, and in particular, through PM, can influence decision making. Insua et al. (2008) declared such practices to enhance the legitimacy, acceptance and users' satisfaction of decision making.

Our results suggest that PM platforms are prone to inaccurate data entry. Olteanu-Raimond et al. (2017) adds the potential of vandalism and fraudulent data entry to this concern and suggested the restriction of contributions from non-registered users as a countermeasure. These findings were also in consistent with Kantola and Tuulentie (2020) findings, who articulated PM challenges to i) the possibility of overrepresenting certain groups resulting from circulating the PM platform through advertisement; ii) the vulnerability of PM abuse through participating several times.

Community engagement through PM helps planners to prioritize investments in an informed, legitimized and transparent manner. In addition to the CitiBike example detailed in 4.3.3, Spaces for People is a project across Scotland that utilizes PM to identify where temporary infrastructure can be offered in order to improve AT during the COVID-19 pandemic.

6. Conclusion, Implications and Limitations

In this paper, we sought to document the crowdsourced data strengths, challenges, reliability and usefulness from the perspective of informants. Our work is aimed at those who are seeking to embrace crowdsourced data for AT. Thematic analysis from interviews in conjunction with current literature were employed to obtain informants' perspectives on crowdsourced data. To summarize, crowdsourced data provide unprecedented data on AT. Potential challenges that informants may encounter were identified as technical, privacy, proprietorship, financial and fragmentation related issues. Three distinct types of usefulness were identified related to investments: i) before and after analysis; ii) current infrastructure assessments; and iii) investment prioritization. Reliability issues threaten SFNs (i.e., social desirability and self-selection) and PM (i.e., vandalism). Despite the related challenges, crowdsourced data is demonstrated to have many attributes that can help fill gaps left by traditional data sources.

6.1. Implications

Several implications can be drawn from this work. The value of traditional data is embedded in its role as ground-truth data to validate, adjust or complement data sources. The adoption of SFNs, particularly Strava, and other crowdsourced data is linked with several challenges and reliability concerns. Although Strava is deemed to be cost-effective, for some the costs hinder the adoption of such data. By providing free data, in-house developed apps can be alternatives. BSSs data can be optimized through serving deprived areas and reducing the cost of rebalancing user vehicles. PM data can be used to obtain information about specific issues such as safety and the allocation of infrastructure. Potential informants should be aware of challenges and reliability considerations. Furthermore, following this work, we encourage future researchers and crowdsourced data providers to address potential challenges and reliability issues by proposing novel approaches to alleviate their impact on the result accuracy and replicability.

6.2. Limitations

Our study sample covered a variety of informants, yet it is still relatively small and may not represent the population of crowdsourced data users. Furthermore, our sample did not include the perspective of social media platforms. Finally, given the dynamic nature of this field, some of the strengths and challenges are likely to change in the future.

7. References

Adler, Jeffrey, John Horner, Jeanette Dyer, Alan Toppen, Lisa Burgess, Greg Hatcher, and others. 2014. *Estimate Benefits of Crowdsourced Data from Social Media.*

Aïvodji, Ulrich Matchi, Sébastien Gambs, Marie-José Huguet, and Marc-Olivier Killijian. 2016. "Meeting Points in Ridesharing: A Privacy-Preserving Approach." *Transportation Research Part C: Emerging Technologies* 72:239–53. doi: 10.1016/j.trc.2016.09.017.

AMEC E&I, Inc., and Inc. Sprinkle Consulting. 2011. "Pedetrian and Bicycle Data Collection." Retrieved May 6, 2019 (http://ahtd.info/yahoo_site_admin/assets/docs/Pedestrian_Bicycle_Data_Collection_FHWA_Dec_2011.14274024).

An, Ran, Renee Zahnow, Dorina Pojani, and Jonathan Corcoran. 2019. "Weather and Cycling in New York: The Case of Citibike." *Journal of Transport Geography* 77:97–112.

Barbier, Geoffrey, Reza Zafarani, Huiji Gao, Gabriel Fung, and Huan Liu. 2012. "Maximizing Benefits from Crowdsourced Data." *Computational and Mathematical Organization Theory* 18(3):257–79.

BBC News. 2018. "Fitness App Strava Lights up Staff at Military Bases." Retrieved February 8, 2021 (Fitness app Strava lights up staff at military bases).

Boss, Darren, Trisalyn Nelson, and Meghan Winters. 2018. "Monitoring City Wide Patterns of Cycling Safety." *Accident Analysis & Prevention* 111:101–8.

Brown, Greg. 2017. "A Review of Sampling Effects and Response Bias in Internet Participatory Mapping (PPGIS/PGIS/VGI)." *Transactions in GIS* 21(1):39–56.

Bunn, Haynes. 2019. "The Power of Multiple Datasets and the Insights Hiding in Them." Retrieved August 15, 2020 (https://medium.com/strava-metro/the-power-of-multiple-datasets-and-the-insights-hiding-in-them-62aa2cb3eb40).

Bunn, Haynes. 2021. "Strava Metro Corridors." Retrieved May 9, 2022 (https://stravametro.zendesk.com/hc/en-us/articles/360053274534-Corridors).

Burgess, Kaya. 2018. "Thieves 'Followed Rider on His Strava App' to Make off with £12,500 in Bikes." Retrieved February 8, 2021 (https://www.thetimes.co.uk/article/thieves-followed-rider-on-his-strava-app-to-make-off-with-12-500-in-bikes-xqjrlgj0f).

Cheng, Geyao, Yang Guo, Yingwen Chen, and Yudong Qin. 2019. "Designating City-Wide Collaborative Geofence Sites for Renting and Returning Dock-Less Shared Bikes." *IEEE*

Access 7:35596–605.
Costa Affonso, Roberta, Florent Couffin, and Patrice Leclaire. 2021. “Modelling of User Behaviour for Static Rebalancing of Bike Sharing System: Transfer of Demand from Bike-Shortage Stations to Neighbouring Stations.” *Journal of Advanced Transportation* 2021.
Day, Christopher, Hiromal Premachandra, and Darcy Bullock. 2016. “Rate of Pedestrian Signal Phase Actuationas a Proxy Measurement of Pedestrian Demand.” Purdue University Press.
Deparday, Vivien. 2011. “Enhancing Volunteered Geographical Information (VGI) Visualization with Open Source Web-Based Software.” University of Waterloo.
Desai, Elvis, Philip Wang, Jeffrey Suway, and Krystina Engleman. 2021. *Bicycle GPS Systems Positional Accuracy*.
Dhakal, Nirbesh, Christopher R. Cherry, Ziwen Ling, and Mojdeh Azad. 2018. “Using CyclePhilly Data to Assess Wrong-Way Riding of Cyclists in Philadelphia.” *Journal of Safety Research* 67:145–53. doi: 10.1016/j.jsr.2018.10.004.
DiDonato, Michael, Stephen Herbert, and D. Vachhani. 2002. “City-Bike Maintenance and Availability.” *Project Report (Project No. 44-JSD-DPC3). Worcester Polytechnic Institute. Neuvedeno*.
Duran-Rodas, David, Dominic Villeneuve, Francisco C. Pereira, and Gebhard Wulfhorst. 2020. “How Fair Is the Allocation of Bike-Sharing Infrastructure? Framework for a Qualitative and Quantitative Spatial Fairness Assessment.” *Transportation Research Part A: Policy and Practice* 140:299–319. doi: 10.1016/j.tra.2020.08.007.
Evans-Cowley, Jennifer S., and Greg Griffin. 2012. “Microparticipation with Social Media for Community Engagement in Transportation Planning.” *Transportation Research Record* 2307(1):90–98.
Ferster, Colin, Trisalyn Nelson, Karen Laberee, and Meghan Winters. 2021. “Mapping Bicycling Exposure and Safety Risk Using Strava Metro.” *Applied Geography* 127:102388.
Fischer, Jaimy, Trisalyn Nelson, and Meghan Winters. 2022. “Changes in the Representativeness of Strava Bicycling Data during COVID-19.” *Findings* 33280.
G., Mat. 2022. “Using the Potential Segment Match Analysis Tool.” Retrieved May 15, 2022 (https://support.strava.com/hc/en-us/articles/216918247-Using-the-Potential-Segment-Match-Analysis-Tool).
Garber, Michael D., Kari E. Watkins, and Michael R. Kramer. 2019. “Comparing Bicyclists Who Use Smartphone Apps to Record Rides with Those Who Do Not: Implications for Representativeness and Selection Bias.” *Journal of Transport & Health* 15:100661. doi: 10.1016/j.jth.2019.100661.
Gerstenberg, Tina, Christoph F. Baumeister, Ulrich Schraml, and Tobias Plieninger. 2020. “Hot Routes in Urban Forests: The Impact of Multiple Landscape Features on Recreational Use Intensity.” *Landscape and Urban Planning* 203:103888. doi: 10.1016/j.landurbplan.2020.103888.
Gillham, Bill. 2000. *Research Interview*. A&C Black.
Griffin, Greg, Krista Nordback, Thomas Götschi, Elizabeth Stolz, and Sirisha Kothuri. 2014. “Monitoring Bicyclist and Pedestrian Travel and Behavior: Current Research and Practice.” *Transportation Research Circular* (E-C183).
Griffin, Greg P., Meg Mulhall, and Christopher L. Simek. 2018. “Sources and Mitigation of Bias in Big Data for Transportation Safety.”
Griffin, Greg P., Megan Mulhall, Chris Simek, and William W. Riggs. 2020. “Mitigating Bias in Big Data for Transportation.” *Journal of Big Data Analytics in Transportation* 1–11.

Gu, Yiming, Zhen (Sean) Qian, and Feng Chen. 2016. “From Twitter to Detector: Real-Time Traffic Incident Detection Using Social Media Data.” *Transportation Research Part C: Emerging Technologies* 67:321–42. doi: 10.1016/j.trc.2016.02.011.

Hamilton, Timothy L., and Casey J. Wichman. 2018. “Bicycle Infrastructure and Traffic Congestion: Evidence from DC’s Capital Bikeshare.” *Journal of Environmental Economics and Management* 87:72–93.

Hollander, Justin B., and Yaqi Shen. 2017. “Using Social Media Data to Infer Urban Attitudes about Bicycling: An Exploratory Case Study of Washington DC.” Pp. 79–97 in *City Networks*. Springer.

Hong, Jinhyun, David Philip McArthur, and Mark Livingston. 2018. *The Evaluation of Large Cycling Infrastructure Investments in the Glasgow Clyde Valley Planning Area Before, During, and After the Commonwealth Games: The Use of Crowdsourced Data (Strava)*.

Hood, Jeffrey, Elizabeth Sall, and Billy Charlton. 2011. “A GPS-Based Bicycle Route Choice Model for San Francisco, California.” *Transportation Letters* 3(1):63–75. doi: 10.3328/TL.2011.03.01.63-75.

Insua, David Rios, Gregory E. Kersten, Jesus Rios, and Carlos Grima. 2008. “Towards Decision Support for Participatory Democracy.” Pp. 651–85 in *Handbook on Decision Support Systems 2*. Springer.

Jäger, Hieronymus, Uta Schirpke, and Ulrike Tappeiner. 2020. “Assessing Conflicts between Winter Recreational Activities and Grouse Species.” *Journal of Environmental Management* 276:111194.

Ji, Yanjie, Xue Jin, Xinwei Ma, and Shuichao Zhang. 2020. “How Does Dockless Bike-Sharing System Behave by Incentivizing Users to Participate in Rebalancing?” *IEEE Access* 8:58889–97. doi: 10.1109/ACCESS.2020.2982686.

Jiang, Jian, Fei Lin, Jin Fan, Hang Lv, and Jia Wu. 2019. “A Destination Prediction Network Based on Spatiotemporal Data for Bike-Sharing.” *Complexity* 2019.

Kantola, Sini, and Seija Tuulentie. 2020. “Participation in a Large Arctic City--the Possibilities of PPGIS for Improving Interaction.” *Polar Geography* 43(4):295–312.

Kazlouski, Andrei, Thomas Marchioro, Harry Manifavas, and Evangelos Markatos. 2021. “Do Partner Apps Offer the Same Level of Privacy Protection? The Case of Wearable Applications.” Pp. 648–53 in *2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*.

Kim, Dohyung. 2020. “Pedestrian and Bicycle Volume Data Collection Using Drone Technology.” *Journal of Urban Technology* 27(2):45–60.

Kumar, Avinash, Miao Jiang, and Yi Fang. 2014. “Where Not to Go?” Pp. 1223–26 in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval - SIGIR ’14*. New York, New York, USA: ACM Press.

Lee, Kyuhyun, and Ipek N. Sener. 2017. *Emerging Data Mining for Pedestrian and Bicyclist Monitoring: A Literature Review Report*.

Lee, Kyuhyun, and Ipek N. Sener. 2019. “Understanding Potential Exposure of Bicyclists on Roadways to Traffic-Related Air Pollution: Findings from El Paso, Texas, Using Strava Metro Data.” *International Journal of Environmental Research and Public Health* 16(3):371.

Lee, Kyuhyun, and Ipek Nese Sener. 2020. “Strava Metro Data for Bicycle Monitoring: A Literature Review.” *Transport Reviews* 1–21. doi: 10.1080/01441647.2020.1798558.

Livingston, Mark, David McArthur, Jinhyun Hong, and Kirstie English. 2020. “Predicting

Cycling Volumes Using Crowdsourced Activity Data." *Environment and Planning B: Urban Analytics and City Science* 2399808320925822.
Lu, Yongliang, Una Benlic, and Qinghua Wu. 2020. "An Effective Memetic Algorithm for the Generalized Bike-Sharing Rebalancing Problem." *Engineering Applications of Artificial Intelligence* 95:103890.
Mason, Jennifer. 2017. *Qualitative Researching*. sage.
Maus, Jonathan. 2014. "ODOT Embarks on 'Big Data' Project with Purchase of Strava Dataset." Retrieved (https://bikeportland.org/2014/05/01/odot-embarks-on-big-data-project-with-purchase-of-strava-dataset-105375).
McCallum, Ian, Wei Liu, Linda See, Reinhard Mechler, Adriana Keating, Stefan Hochrainer-Stigler, Junko Mochizuki, Steffen Fritz, Sumit Dugar, Miguel Arestegui, and others. 2016. "Technologies to Support Community Flood Disaster Risk Reduction." *International Journal of Disaster Risk Science* 7(2):198–204.
Munkácsy, András, and Andrés Monzón. 2017. "Potential User Profiles of Innovative Bike-Sharing Systems: The Case of BiciMAD (Madrid, Spain)." *Asian Transport Studies* 4(3):621–38.
National Academies of Sciences, Engineering, and Medicine. 2017. *Bicycle and Pedestrian Data Collection Manual*. Minnesota Department of Transportation.
Ohlms, Peter B., Lance E. Dougald, Hannah E. MacKnight, and others. 2018. *Assessing the Feasibility of a Pedestrian and Bicycle Count Program in Virginia*.
Olteanu-Raimond, Ana-Maria, Mari Laakso, Vyron Antoniou, Cidália Costa Fonte, Alexandra Fonseca, Magdalena Grus, Jenny Harding, Tobias Kellenberger, Marco Minghini, and Andriani Skopeliti. 2017. "VGI in National Mapping Agencies: Experiences and Recommendations." *Mapping and the Citizen Sensor* 299–326.
Orellana, Daniel, and Maria Laura Guerrero. 2019. "The Influnce of Space Syntax on Cycling Movement in Manta, Ecuador." in *12th International Space Syntax Symposium*.
Pritchard, Trevor. 2016. "Ottawa-Gatineau Cyclists Urged to Map Their Journeys." Retrieved February 13, 2020 (https://www.cbc.ca/news/canada/ottawa/strava-app-ottawa-1.3546513).
Qian, Xiaodong, and Miguel Jaller. 2022. "Do Dock-Based and Dockless Bikesharing Systems Provide Equitable Access for Disadvantaged Communities?"
Raford, Noah, Alain J. Chiaradia, and Jorge Gil. 2007. "Space Syntax: The Role of Urban Form in Cyclist Route Choice in Central London." *SafeTrec Research Reports* 18.
Rahyadi, Irmawan. 2021. "Pedestrian Condition and The Social Marketing: Digital Marketing of Pedestrian in Social Media." *Review of International Geographical Education Online* 11(6):845–54.
Rissel, Chris E. 2009. "Active Travel: A Climate Change Mitigation Strategy With Co-Benefits for Health." *New South Wales Public Health Bulletin* 20(2):10–13.
Serra, Luis, David McArthur, Mark Livingston, Andrew McHugh, Kimberley Hose, Keith Scott, Kalim Uddin, and Bruce Whyte. 2020. "Creating Open Data Counts of Pedestrians and Vehicles Using CCTV Cameras." Retrieved December 29, 2020 (https://www.ubdc.ac.uk/news-media/2020/july/creating-open-data-counts-of-pedestrians-and-vehicles-using-cctv-cameras/).
SFCTA. n.d. "CyclingTraks for IPhone and Android." Retrieved July 28, 2019 (https://www.sfcta.org/tools-data/tools/cycletracks).
Da Silva, Christian Nunes, Vicka De Nazaré Magalhães Marinho, Gracilene De Castro Ferreira, Yasmin Alves Dos Santos, Roberto Magno Reis Netto, Aline Reis De Oliveira Araújo,

Rafael Dantas Dias, Adolfo Da Costa Oliveira Neto, Rodolpho Zahluth Bastos, João Marcio Palheta Da Silva, and others. 2020. "State's Technical Cartography and Social Cartography's Participatory Mapping: Contributions to a Contemporary Reflection on the Use of Maps." *Journal of Geographic Information System* 12(4):319–33.

Smith, Amy. 2015. "Crowdsourcing for Active Transportation." *Institute of Transportation Engineers. ITE Journal* 85(5):30.

Smith, Dave, and Dale Harris. 2017. "The Role of Real-Time Crowdsourced Information and Technology in Supporting Traveller Information." Pp. 1–15 in *Australasian Transport Research Forum*. Vol. 27.

Stragier, Jeroen, Tom Evens, and Peter Mechant. 2015. "Broadcast Yourself: An Exploratory Study of Sharing Physical Activity on Social Networking Sites." *Media International Australia* 155(1):120–29.

Strava Metro. 2020a. *Frequently Asked Questions*.

Strava Metro. 2020b. "The New Human-Powered Era." Retrieved December 14, 2020 (https://blog.strava.com/the-new-human-powered-era-20951/).

Strava Press. 2020. "Strava Announces Strava Metro, the Largest Active Travel Dataset on the Planet, Is Now Free and Available to Cities Everywhere." Retrieved May 13, 2022 (https://blog.strava.com/press/metro/?btn=4ngJq5kHx1SlNwIJ4IUG61&par=5WF6BYBmGyFYTCMihvXT4I).

Sun, Yeran, and Amin Mobasheri. 2017. "Utilizing Crowdsourced Data for Studies of Cycling and Air Pollution Exposure: A Case Study Using Strava Data." *International Journal of Environmental Research and Public Health* 14(3):274. doi: 10.3390/ijerph14030274.

Sun, Yeran, Yashar Moshfeghi, and Zhang Liu. 2017. "Exploiting Crowdsourced Geographic Information and GIS for Assessment of Air Pollution Exposure during Active Travel." *Journal of Transport & Health* 6:93–104. doi: 10.1016/j.jth.2017.06.004.

Sunde, Erik. 2017. "Analyzing and Improving Safety for Bicyclists in Seattle." Retrieved August 9, 2020 (https://medium.com/strava-metro/seattle-dot-case-study-994545fc8117).

Teixeira, João Filipe, and Miguel Lopes. 2020. "The Link between Bike Sharing and Subway Use during the COVID-19 Pandemic: The Case-Study of New York's Citi Bike." *Transportation Research Interdisciplinary Perspectives* 6:100166. doi: 10.1016/j.trip.2020.100166.

Thorsen, Neri H., Richard Bischof, Jenny Mattisson, Tim R. Hofmeester, John D. C. Linnell, and John Odden. 2022. "Smartphone App Reveals That Lynx Avoid Human Recreationists on Local Scale, but Not Home Range Scale." *Scientific Reports* 12(1):1–13.

Tian, Zihao, Jing Zhou, W. Y. Szeto, Lixin Tian, and Wenbin Zhang. 2020. "The Rebalancing of Bike-Sharing System under Flow-Type Task Window." *Transportation Research Part C: Emerging Technologies* 112:1–27.

Townshend, Tim, and Amelia Lake. 2017. "Obesogenic Environments: Current Evidence of the Built and Food Environments." *Perspectives in Public Health* 137(1):38–44.

Venter, Zander, David Barton, Vegard Gundersen, Helene Figari, and Megan Nowell. 2020. "Urban Nature in a Time of Crisis: Recreational Use of Green Space Increases during the COVID-19 Outbreak in Oslo, Norway." *Environmental Research Letters*.

Vos, Jonas De. 2020. "The Effect of COVID-19 and Subsequent Social Distancing on Travel Behavior." *Transportation Research Interdisciplinary Perspectives* 5. doi: 10.1016/j.trip.2020.100121.

Weller, Susie. 2015. "The Potentials and Pitfalls of Using Skype for Qualitative (Longitudinal)

Interviews.”
Yang, Yuanxuan, Alison Heppenstall, Andy Turner, and Alexis Comber. 2019. “A Spatiotemporal and Graph-Based Analysis of Dockless Bike Sharing Patterns to Understand Urban Flows over the Last Mile.” *Computers, Environment and Urban Systems* 77:101361.
Zeile, Peter, Bernd Resch, Martin Loidl, Andreas Petutschnig, and Linda Dörrzapf. 2016. “Urban Emotions and Cycling Experience--Enriching Traffic Planning for Cyclists with Human Sensor Data.” *GI_Forum* 1(2013):204–16.