Check for updates

Original Research Article

Philosophy, Theory, Models, Methods and Practice

EPF: Philosophy, Theory, Models, Methods and Practice
1–29
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/26349825251365637
journals.sagepub.com/home/epf
Sage

Navigating digital geographies: Black boxes, geospatial narratives, and the art of constructing location data

Thomas N Cooke[1], Dan Cohen[2] and Dillon Mahmoudi[1]

Abstract

Smartphone location data are often treated as objective and self-evident—but it is neither. This article opens the black box of how location is constructed on the phone and in the cloud, arguing that these processes are foundational to digital geography and central to how its infrastructures take shape. Drawing on an original experiment conducted in Kingston, Ontario and Baltimore, Maryland, we reverse-engineer and document the different methods of producing location data in Android smartphones. In doing so, we reveal three intertwined, overlapping, and contested geospatial narratives: raw GNSS location data, Google's computed location data, and the human narrative of embodied experiences. We analyze the frictions and contradictions among these narratives to demonstrate how location data are not simply measured, but actively produced through assemblages of surveillance, infrastructural power, and capitalist extraction. Against dominant portrayals of location as a neutral technical fact, our findings show that Google's location services depend on off-phone processing, structured by opaque systems designed for control and profit. We call for a critical reorientation in how digital geographers engage with location technologies—not as passive tools, but as politically charged systems that mediate and monetize everyday life.

Keywords

Digital geographies, surveillance studies, location data, geospatial narratives, urban geography

> "Intellectually, [urban location data] is a great challenging problem. But what about commercially—is this important? The answer is a resounding Yes!"
>
> Dr. Frank van Diggelen (2021), Principal Engineer at Google

[1]University of Maryland, Baltimore County, USA
[2]Queen's University, Canada

Corresponding author:
Dan Cohen, Department of Geography and Planning, Queen's University, Mackintosh-Corry Hall, E208, Kingston, ON K7L 3N9, Canada.
Email: dan.cohen@queensu.ca

In February 2020, we began a collaboration that brought together scholars from surveillance studies, computer science, and geography to focus on a fine-grained, but important question: how do smartphone-based technologies understand and model space, and for what purposes?

While in the popular imagination, and even in some scholarship, this is a relatively straightforward question: phones determine fixed locations based on connections between three or more satellites in a process called trilateration. In reality, the process is much more complex. As described by Google engineer, Dr. Frank van Diggelen (2021) for *Inside GNSS*, creating the discrete locations presented on a smartphone relies on factors such as the time delay of a satellite signal, the width of the satellite wave itself, atmospheric weather, the position of the planet, and the movement of the phone through space—all of which is complicated in urban environments where signals are blocked, or even altered, as they interact with "urban canyons" made up of signal reflecting buildings. The complexity of calculating an accurate location for a smartphone as it moves through urban space is so computationally and energy intensive that much of it occurs within Google servers, rather than within the phone itself.

In this respect, what seems simple—the dot on a smartphone screen that signifies a location—is the result of a complex calculation of heterogeneous elements that include spatial, temporal, and algorithmic elements across multiple devices. The goal of this multidisciplinary project, which we called "Big Data Exposed," has been to defetishize[1] the production of location data through tracking the process of the creation of a discrete location. To do this, we start from the very beginnings within the core processes of an Android smartphone and its interactions with applications, positioning satellites, and Google servers. As described in further detail below, this process involved the creation of a new Android app: a data parser, that extracts the location data from an Android smartphone at multiple stages. This process also depended on modifying an existing application—an open-source navigation app like Google Maps—to ensure that it contained location detection, processing, and data sharing algorithms known to be used by third parties to monitor how phone users move in real time. These applications were then put into use through experiments conducted in Kingston, Ontario and Baltimore, Maryland.

The process of unpacking how smartphone location data are produced has important findings for the philosophies and methods of geography as a whole, and of particular interest to the subfields of digital geography and critical GIS. We sought to collaborate with computer scientists and surveillance studies scholars, building on a methodological approaches that may be of interest to geographers (Ash et al., 2024; De, 2022; Fields et al., 2020; van Es and de Lange, 2020; Wilmott, 2020; Wilson, 2018). More specifically, revealing the particularities of how space is understood and modeled through smartphone-based technologies helps us understand how the production of location data is shaped by, and enmeshed with, projects of surveillance and profit-making. As we explore throughout this article, studying how space is modeled on smartphones—in combination with the complex chains of political economic factors that facilitate this production—highlights why Google and other technology firms are so focused on location accuracy and how both location and movement generate the construction of geospatial narratives by private actors.

The results of our experiment thus contribute to scholarship on the political economy of digital spatial technologies and knowledge production (Attoh et al., 2019; Bates et al., 2019; Cuppini et al., 2022; Hodson et al., 2020; Jefferson, 2018; Langley and Leyshon, 2017; Leszczynski, 2019; Stehlin and Payne, 2022; Thatcher et al., 2016; Wilmott, 2020) as well as to the methods used in digital geography and critical GIS (Ash et al., 2024). Most notably, this collaboration between computer and social scientists allowed us to identify how the production of "location" exceeds the smartphone itself, with companies such as Google and Uber acting as key nodes shaping how smartphone-based technologies understand and model space through calculations done on their servers. A key goal of our methodology is in revealing how space is differentially modeled by various techniques, each creating different geospatial narratives about location: a raw location story (produced through the smartphone's interactions with GNSS satellites), Google's story (produced through the co-opting of raw location data, combined with outsourced, third-party measurement points provided by cell towers and Wi-Fi routers), and the user's embodied experience. The story that gets told has considerable implications for (1) the phone user's authority over their location, and (2) how the phone user is represented not only as a traveler, but as a consumer, as a laborer, and as a private vs public human being.

The rest of the article proceeds as follows. First, we present a literature review of digital geographies and critical GIS with a close focus on issues of surveillance and the technical difficulties of generating location data in urban environments. Second, we outline the methodology of our study and highlight its interdisciplinary nature. Third, we provide an analysis of the results of the study. Finally, we conclude with a discussion of how our close attention to the production of location data may be of use to digital geographers in understanding the infrastructure that underpins the political economy of location.

Understanding location data

In 1989, engineers from Massachusetts Institute of Technology and the Electric Power Research Institute sought to improve energy usage awareness and efficiency. They developed and patented a metering device which monitored household electricity usage. The "Non-Intrusive Appliance Monitor Apparatus" recorded analog voltage and current signals which were then converted into a digital format and centrally processed off-site to identify individual appliances and the energy consumed by those appliances (Hart, 1989). The monitor could be placed outside the house and no in-home sensors were required for it to function properly.

In the same year, the lead engineer of the patent wrote the feature article in *IEEE Technology and Society Magazine* warning that the apparatus could also be used in surveillance with significant potential for abuse. As he put it, "[A] typical energy consumer would not understand the wealth of detail that can be extracted from the signals continually being transmitted out of the home over the power lines" (Hart, 1989: 15). In field results, the monitor was able to identify and differentiate between different light fixtures, and any number of other household appliances and devices, providing rich data about the household and the behaviors of household members. "[I]t is easy to tell when someone is in the shower, for example, based on the use of a water pump, water heater, bathroom light, and/or hair dryer" (Hart, 1989: 14). Comparing the monitor to a device used to tap a phone line, Hart suggested several surveillance scenarios including monitoring the activities of suspected criminals, monitoring or identifying political opponents, identifying vacancy status for break-ins or theft, or spying on foreign embassies. Hart also identified situations where such monitoring could be profitable; notably, the "junk mail" scenario, where "utility companies could sell advertisers and salespeople mailing lists of consumers lacking assorted consumer appliances" (Hart, 1989: 14).

More than 30 years later, the concerns of Hart are increasingly prescient. Similar to the non-intrusive monitors, smartphones collect data about individuals that are centrally processed both off-site and within the phone. These data are used to understand and model space as well as to analyze the behaviors of individuals. For users, these technologies enable the navigation of unfamiliar places as well as shape their spatial understandings of the places they occupy. Indeed, digital geographers have highlighted how applications like Google Maps influence how individuals understand their local geographies through providing contextual data as they move through the city and pick both routes and destinations (Dalton, 2018; Thatcher, 2013; Wilmott, 2016). However, as critical scholarship establishes, the role these data play goes well beyond its utility to its end users, as it also facilitates the production of datasets which are themselves valuable as a commodity (Beauvisage and Mellet, 2020). Furthermore, the representations of movement and location produce information useful for various surveillance infrastructures, from the carceral to consumer modeling (Kitchin, 2013).

In this section we explore, in turn, (1) the intersections of surveillance and location in digital geographies, (2) the place of location data in digital political economies, and (3) the particular challenges of calculating location in urban environments. Through doing so, we establish how our experiment illuminating the precise mechanisms through which spatial data are produced inside and outside of smartphones using both computational and geographic methods strengthens geographers' understanding of how (urban) digital geographies are produced and shaped through decisions made along the chain of production. As we will argue through our empirical case, such decisions shape the construction of movement and location narratives that result from the spatial data produced via smartphones.

Location data and surveillance

A key pillar of the literature on location data both inside and outside of geography has been on its role within surveillance infrastructures (Klauser, 2013; Leszczynski, 2015). If surveillance can be broadly understood as a set of related activities (such as looking, watching, staring, examining, etc.) in a technologically and digitally mediated context, those activities increasingly capture data for the purposes of monitoring (Amoore, 2014; Marx, 2015; Zuboff, 2019). Surveillance is instructive for thinking through the "why" of location data precisely because surveillance is a power mechanism designed to make that which is not normally seen more visible for the purposes of monitoring, judging, and regulating behavior (Foucault, 1977). It is through this scholarship that we theorize smartphone location tracking architecture and their data outputs not merely as a function of entertainment for the user but as a function of surveillance conducted by private and public sector actors who have vested interests in rendering human movement visible and legible.

Throughout the past two decades, multiple key events have transpired that demonstrate Haggerty and Ericson's (2000: 607–608) concept of surveillance-as-assemblage which posits that the practice of surveillance consists of a set of diverse, interconnected technologies and organizations that gain coherence by working in concert to monitor, control, and influence. Seldom do governments conduct their own surveillance on populations without leveraging and linking together otherwise disparate data collection and monitoring systems. For example, the Snowden Revelations revealed the extent to which the US government's National Security Agency tapped into existing communication networks and networked devices to conduct surveillance in the name of conducting the "war on terror" (Bauman et al., 2014). By tracking how users move and by monitoring the content of their digital communications, the NSA surveilled domestic and foreign populations allegedly in the name of national security. While corporate actors use telecommunications networks to monitor users for potentially different reasons and justifications (i.e. analyzing consumer behavior for sales, advertising, and marketing purposes), surveillance is invariably conducted along the same rationale: to render that what cannot be seen visible so that it can be monitored, judged, regulated, evaluated, and turned into profit.

Surveillance is thus what gives rise to location data narratives and storytelling power. By amalgamating location data together, public and private sector actors authorize themselves to construct truth claims such as, for example, explanations of potential consumer interest or users-as-security-risk in ways that cannot be self-evidently negotiated nor accessed by the subjects of these stories themselves (Mahmoudi et al., 2024). For example, in January 2020, a Florida resident named Zachary McCoy received an email from Google indicating that his local police department requested his location data history from his smartphone. Earlier that month, McCoy had ridden his bicycle past a home that was burglarized. The police department identified McCoy as a potential suspect of interest given the collection of location data available to them and their interpretation of its meaning, McCoy's location was used to create the police's narrative about what McCoy had been doing. Had the police department generated a fair and accurate narrative, they would have recognized that McCoy had never actually stopped nor dismounted from his bicycle (Bhuiyan, 2023). Instead, the police assumed he committed the crime because McCoy rode his bicycle past the victim's home around the same time that the crime took place. As McCoy's case illustrates, location and movement within the context of surveillance is thus more than a spot on a map, they are instead both produced and understood within a complex assemblage.

Tracking where citizens are, where they are heading, how long they take to arrive, how long they stay, and the routes they take along the way are also often motivated by the intersections of surveillance and profit-making. Location data in smartphones have been increasingly collected, interpreted, framed, and sold as stories that are interwoven with racialized logics of surveillance (Jefferson, 2018). In 2021, the data brokerage firm Predicio made global headlines after hoarding location data from smartphones containing the Muslim prayer app called Salaat First. Those data were sold to a US government contractor working with US Immigration and Customs Enforcement (ICE) and the Federal Bureau of Investigation (FBI) (Cox, 2021). Those data were not merely surrendered to the government. They were collected from a smartphone app, commodified, and sold for profit. The Predicio example is thus a salient one for not only thinking through the ways in which surveillance facilitates the ability for privileged actors and entities to generate narratives about location data. That is, surveillance by tech companies empowers them to narrate and commodify smartphone location data.

Location data and digital political economy

Location data's focus on the movement of people is revealing of both the evolving architecture of smartphone surveillance and the importance of social contexts, like urban boundaries, within which location data are understood. As hinted at in the Predicio case, smartphone architecture has evolved considerably not merely in the name of state surveillance, but as a central component of a multi-billion dollar global data marketplace with significant implications for urban spaces. Political economic factors play an important role in the growing surveillance of users' location and movement.

For technology companies, an accurate description of where a phone is located is important mechanically (e.g. knowing where a pick-up point for ridesharing is located) and for profitability (e.g. building accurate customer profiles). Both processes, however, require assumptions about what urban spaces are used for and by whom. As urban geographers have described, urban space is not a neutral category but ridden with a variety of associations that can produce both status and stigma (Jefferson, 2018; Kallin and Slater, 2014; Otero et al., 2022). Location data produced by smartphones exist within these dynamics, as digital information gathered as an individual moves through space draws on these associations in building profiles of smartphone users, their shopping habits, and their overall behaviors (Mahmoudi et al., 2024; Shelton et al., 2015). Location data thus become aggregated into a key aspect of consumer profiles that data companies then go on to sell.

In response to the COVID-19 pandemic, mobile location data from smartphones also became an important research tool for contact tracing and public health more broadly. Despite narratives of the positive public health applications of using these data, Human Rights Watch (2020) cautioned,

> that the use of incomplete and discriminatory datasets can misdirect public health efforts in ways that endanger the rights of the poorest and most vulnerable people. For example, stricter enforcement of social distancing measures in low-income counties could unduly penalize front line workers, people struggling to find shelter, or unemployed people traveling to food banks or welfare agencies because their movements may appear abnormal or in violation of social distancing norms when in fact they have to be more mobile to meet basic needs.

Indeed, during the pandemic, numerous geospatial data companies reached out to the authors to sell location data to glean insights from mobile location data. Such uses of data present two important problems. First, the companies' analysis and the surveilled data itself are presented as truth, rather than representations, agnostic of the calls by scholars to unpack the assumptions that go into creating and using data (see, for example, Dalton et al., 2016; Iliadis and Russo, 2016; Kitchin and Lauriault, 2014; Schuurman, 2006). Second, this reflects a broader trend where surveillance of users is commodified for surveillance advertising (Edelman, 2020). By tracking and profiling individuals' movements, these data are transformed into a data commodity that can be sold to advertisers under the premise that the data represent the best possible method for targeted advertising (Bodle, 2016; Elvy, 2018; Mahmoudi and Levenda, 2016; Thatcher, 2017; West, 2019). This commodification process underscores the economic incentives driving extensive location data collection and raises critical questions about privacy and the ethical use of personal data.

The challenge of location data in urban environments

As revealed through industry publications, the production of local data within urban environments specifically is essential to the political economy of digital geographies. Uber, for instance, requires precise location data in order to create the smooth operations of matching rider and driver in a seamless manner at the moment of pick up (Iland et al., 2018). As Attoh et al. (2019) also argue, this type of location data is then used to further accumulation strategies through the production of big data sets that can train new technologies (such as self-driving cars or machine learning) or which are sold to other actors. The "urban" is not merely where the processes of datafication and data extraction take place; rather, the city itself serves as a necessary precondition for both the collection of data and the generation of its value (Mahmoudi et al., 2024). From the combination of location data with its urban context, social interactions and consumptive behaviors can be inferred or enriched from other sources, providing vital sources of information for marketers to turn into targeted advertisements. These, in turn, produce new urban geographies of consumption and shape our understandings of space (Dalton, 2018; Mahmoudi and Levenda, 2016).

As described above, urban spaces represent an important site for location data that is fundamental to the profit proposition of technology companies and surveillance infrastructures. Yet at the same time, urban spaces are a challenge for modeling that the smartphone industry is perpetually trying to improve (as noted in the opening quote to this article). As Uber engineers argue on their company blog:

> Accurate estimation of rider and driver location is a crucial requirement for fulfilling Uber's mission of providing transportation as reliable as running water, everywhere, for everyone. . . To meet our mission, the Sensing, Intelligence, and Research team is working on a variety of approaches for improving location with creative use of sensors and computation on mobile devices, coupled with the computational power of our server infrastructure (Iland et al., 2018).

In the view of industry actors then, the production of location data in urban environments is particularly problematic (Iland et al., 2018; Van Diggelen, 2021). This is because the development of GNSS in its early stages assumed direct line of sight with a satellite. Urban spaces, often characterized by tall buildings made with reflective materials, either block line of sight signals or reflect them, creating a challenging environment for modeling. This can be compensated by tracking a phone's movement from an area where there is a clean line-of-sight, however, that approach requires that users start their journey in an optimal location. The production of location in urban environments is therefore never clean, but as Uber engineers highlight, requires workarounds like the "creative" use of sensors and computing power that exists outside of the smartphone itself (Iland et al., 2018).

Both Google and Uber's solution to this challenge include the creation of models of cities which includes building heights (Iland et al., 2018; Van Diggelen, 2021) and the known location of public Wi-Fi access points (Google Fused Location Provider, 2024). With the use of machine learning algorithms, Google, for example, spatially models how relative signal strength between satellites, signal strength of known public Wi-Fi access points, and GNSS signal bouncing can reveal the precise location of a user. In these algorithms it is assumed that where there is a strong signal strength from a satellite there is a clear line-of-sight, and when there is a weak signal strength, a building is interfering. Through doing so, the likelihood of a smartphone being on one side of a street can be modeled (GNSS bouncing) and predicted based on how the interference of line-of-sight to satellites by buildings of various heights (as determined using 3-dimensional (3D) models) is affecting signal strength (Iland et al., 2018; Van Diggelen, 2021).

Putting it all together: Thinking about location data

It is to the above literature on the digital geographies of surveillance, political economy, and urban space that our project contributes through the in-depth analysis of how smartphone location data are produced. While our intervention is primarily methodological, by going into the core processes of the construction of location within smartphones we hope to defetishize the production of location data (that is, we aim to reveal the underlying social and economic processes that construct meaning behind data), reveal how location data are used in the construction of narratives useful for surveillance and profit-making assemblages, and discuss new methodological tools that may be helpful to digital geographers.

For example, the 3D modeling of cities to improve location data, a computation process, masks the underlying assumptions and biases inherent in how these data are used to construct narratives about location. By scrutinizing these algorithms, we can better understand how they influence our understanding of how people use space, a topic that will be examined in the discussion section below. This examination is crucial, as it reveals the tensions between raw GPS data, computationally enhanced data, and the real-world movements and behaviors of individuals and how these shape people's understandings of urban space. In fact, recent developments, have highlighted how location, surveillance, and digital political economies are coming together through what Dayen (2024) calls "surveillance pricing," where apps offering goods and services (e.g. McDonald's app) gather data on users based on information like location and then use algorithms to create optimal prices that vary by consumer.

Yet, at the same time, digital geographers and other critical scholars have drawn on notions of embodiment to highlight that people's unique experiences cannot fully be captured by such data profiles (Kinsley, 2014; Smith, 2016). Instead, the body and its sensations (or "sensate") contains a multitude of factors including the sensory experience and social associations the shape how we react as we move through space in ways that are difficult to represent but which nevertheless guide how we negotiate our way through sites and, as Middleton (2010: 583) puts it, assess the "atmosphere of a place" (see also Harrison, 2000; Kinsley, 2014; Smith, 2016). Crucially, embodied experience is not merely individual, but social and hierarchical: how one moves through space is also shaped by how one is received within it. Social hierarchies of race, gender, ability, and class condition feelings of comfort, threat, and belonging, such that marginalized groups often encounter hostility, surveillance, or fear in ways that dominant spatial data cannot capture (Gieseking, 2020; Middleton, 2010). Such varied embodied experiences can even result in different awareness of what data one produces and strategies to limit one's data footprint, as what Smith (2016) calls the "embodiment-surveillance" nexus means that racialized groups (among others) are aware of how location data can be used against them and adjust their actions accordingly.

Combining these strands from the literature, and the ways corporations are combining surveillance and profit-making in new ways, raises important questions for digital geography: how can we verify that the location data that corporations and governments extract and analyze are the same as what we see, as users? Furthermore, what does it mean to rely on extensive training data and sophisticated algorithms to generate what might be deemed the most accurate conclusions about location and the downstream uses to which location data contribute? Addressing these questions is inherently challenging for two reasons. First, addressing these questions requires our research team to open-up the black box (Winner, 1993) of location-driven surveillance to both determine *how* location data are created and extracted as well as *what* location data embody. Second, coming to terms with the data, algorithms, and architecture inside the black box must also wrestle with the obfuscated nature of location data and the asymmetrical access to both its inner workings and to knowledge about how such data are used (Ball, 2019; Foster and McChesney, 2014).

Methods

This project adopts an interdisciplinary approach, leveraging insights from computer science, sociology, and geography to examine the production of location data via smartphones. Inspired by the method of datawalking (van Es and de Lange, 2020) our work is grounded in an embodied, situated, and generative practice. This approach not only allows us to reflect on our own experiences as we move through space, but to also capture and analyze the location data transmitted by an app and to document the data retention and transmission processes (whether to its own cloud-based storage or to third parties). We can then explore the tangible discrepancies between our own narrative of movement, location data visible to users, and the operations occurring “below the surface.” By integrating datawalking, our methodological aim seeks to connect embodied experiences with the underlying data infrastructures, addressing challenges related to data invisibility, context loss, and access.

This project began in 2019 under the name of A Day in the Life of Metadata (ADITLOM)—a multidisciplinary collaboration led by Dr. Tommy Cooke that brought together computer engineers and analytics developers together with sociologists, political scientists, geographers, and legal studies scholars to investigate precisely what location-based surveillance data looked like within smartphones. The premise of the project was that by “seeing” location surveillance data in a smartphone would enable investigators to understand their creation, transformation, and movement both within and beyond the smartphone itself. By working with the Centre for Advanced Computing at Queen’s University, the ADITLOM team designed a smartphone software intervention that would track when raw location data were generated, where those data were sent throughout an Android-based smartphone operating system, what apps received those data, and in what format those data were then sent out of the phone to third parties. Shortly after ADITLOM’s launch, a sub-project called “Big Data Exposed” was created along with Dr. Dan Cohen and Dr. David Lyonto test these methods in the field. Using ADITLOM’s primary software method, two additional, distinct software interventions were designed, which worked in concert with one another.

Technical framework for location data collection and analysis

The first intervention consisted of modifying an open-source mapping app called OsmAnd[2] to force it to commit its location data requests, receival, and transmission activities to a script stored on the smartphone (based on the phone’s own location data). These data then served as the primary source for our analysis. OsmAnd is an open-source[3] mapping and navigation app, like Google Maps. The open-source nature of the app was very important to our team because it allowed us to detect and monitor the creation of location data and movement activity. Data from the phone was sent to the app, passed through it, received modification by the app’s algorithms, and then left the app. The open-source design of OsmAnd also allowed us to include two additional sets of location data detection and monitoring mechanisms (described below).

To enable raw location data collection, we made a key modification to the OsmAnd app. First, we used the GnssMeasurement() function from the Android Application Programming Interface (API) to access raw satellite signal data, and the OnNmeaMessageListener() to detect when new measurements were recorded. These functions allowed us to observe and record raw GNSS output directly on the device. Second, we integrated GeoSpark’s (which became Roam) Software Development Kit (SDK), which collects location data by transmitting signals off the device for processing using Google's Fused Location services. These services combine GNSS data with information from nearby Wi-Fi access points, cell towers, etc.

Broadly speaking, both APIs and SDKs are methods of integrating third-party functionality. An API enables communication between one piece of software with another service, providing consistent and controlled access to features or data. An SDK includes code libraries that can be embedded into our app. This enables background processing or data collection. While the distinction between API and SDK is technical, the key here for the purposes of our experiment is understanding that they both allow data to be shared beyond the device. This dynamic raises concerns around surveillance. In the case of Predicio that we discussed earlier in the article, the company acquired mobile location data and sold it to a U.S. military contractor, which exemplifies how location data flow from consumer devices into opaque, obscure networks of data brokers and additional actors.

The purpose of the SDK in the context of our research was three-fold. First, it allowed Roam (formerly GeoSpark) to capture, transmit, store, and analyze any location data that passed through it. This is a standard method used by "location intelligence" firms, especially during the pandemic, to develop and sell location analytics services. For example, the French firm Predicio—unbeknownst to the app developer—extracted data from Muslim Americans as they traveled to mosques (Cox, 2021).

Second, the SDK's location data processing is powered in part by Google's Fused Location services. This cloud-based system aggregates data from GPS, Wi-Fi, and cell towers, then applies algorithmic smoothing to make the data more consistent. The specific workings of Fused Location are not disclosed, but the result is often a visually cleaner travel narrative—such as the straight lines seen in Google Maps—that may exclude anomalies or perceived errors. We intentionally included an SDK using Fused Location to observe how this black-box processing reshapes the raw GNSS data into more coherent, albeit altered, spatial histories.[4]

Third, by integrating an SDK reliant on Google's Fused Location, we were able to construct a counter-surveillance vantage point. The SDK made it possible to observe how a third-party analytics company collects, transmits, and receives location data, and how Google's smoothing returns affect the visibility of certain movements. This observation was critical to our research: it revealed how computational systems selectively erase or reshape location traces that appear erratic or deviant.

In summary, our software interventions allowed us to capture location data that the smartphone generated, passed to OsmAnd, or transmitted externally. This included all transformations that occurred through Google's Fused Location services. We also developed a data parser tool that enabled us to read and organize these data into timestamped scripts for analysis and visualized through GIS.

Embodied experience and datawalking

Our technical framework reveals two distinct location data narratives unfolding within the smartphone, both of which are fundamentally different from our embodied experiences as we conducted our research. As we navigated designed routes in Baltimore and Kingston, we intentionally traveled through high and low traffic throughout both urban environments. Our intention in doing so was to collect location data across a wide range of socioeconomic spaces as well as geographical areas that differed in terms of technical architecture. This included dense urban cores with significant building interference where we suspected satellite and cellular signals would be interrupted by buildings, but which were rich with public Wi-Fi routers, and conversely, more open, less networked spaces with minimal infrastructure. Adding intentional diversity to our datawalks positioned us to hypothesize on how technical, embodied, and urban factors, often invisible in computational measurements, shaped the production of geospatial narration.

As we navigated these predetermined routes it is important to emphasize that context-driven decision making emerged as a factor in shaping our embodied experiences—that is, the actual, physical walks that our teams took and not those represented in data points. As we highlight further in the Results section, we made decisions to walk under trees for shade or to pause briefly near landmarks. These decisions were dictated in real time and were often prompted in response to environmental and sensory stimuli, such as heat, fatigue, and even at some points a desire for an acoustically quieter path. These choices, though intuitive to most any traveler, are non-linear and may vary according to an individual or group's relative comfort in heat and/or desire for a sensory-rich or sensory-calm environment. As such, they are often at odds with the normative movement patterns assumed by location processing inside and beyond our smartphones which focuses on generalization rather than individualization. Datawalking, as described by van Es and de Lange (2020), embodies and situates practices that expose disconnections between lived experiences and digital representations. By sharing reflections on our datawalks below, we observe firsthand how digital processes often fail to capture the fluid, adaptive nature of human movement and cannot capture the embodied experiences which guide it.

The datawalk approach to analyzing and testing location data measurements is thus important because it reveals how digital systems prioritize and interpret certain pathways and behaviors over others. For example, computed narratives often default to predefined paths such as sidewalks and roads, while embodied experiences include spontaneous deviations based on sensory inputs like cutting across a section of grass rather than walking across an asphalt parking lot. These deviations reveal how location narratives are shaped not only by technical factors like signal interference or algorithmic smoothing but also the sociocultural, urban, and embodied contexts of mobility. Acts like pausing, seeking shade, or rerouting for convenience reflect the everyday, situated negotiations that shape mobility—subtleties that digital systems often flatten, oversimplify, or ignore altogether.

Moreover, integrating embodied experiences through datawalks offers a critical interrogation of the biases inherent in location data systems. Juxtaposing location measurements with lived experience allows us to identify the precise moments where computed representations significantly diverge from reality—divergences that may vary according to social difference and how individuals are perceived as they move through space. These divergences are not innocuous. They are more than technical errors; they are symptoms of a broader tendency in digital systems to prioritize coherence and predictability—along with assumptions about bodies—over the complexities of human decision-making. Human mobility reveals the assumptions embedded in location tracking technologies, highlighting their implications for urban planning, surveillance, and everyday navigation.

Route selection

To best achieve our goals for the data walk routes in Baltimore, Maryland, and Kingston, Ontario, we designed a strategic approach to capture diverse socioeconomic and infrastructural contexts. Routes included popular vehicular travel paths, downtown pedestrian corridors, and, in Kingston, a university campus. This selection ensured exposure to varied Wi-Fi router densities, cell tower coverage, and spatial characteristics critical for studying GNSS and digital connectivity.

In Kingston (see left portion of Figure 1), the route traversed the less dense urban core, a university campus with open spaces and institutional networks, and both high- and low-income neighborhoods. In Baltimore (see right portion of Figure 1), the route spanned from the city center to low-income areas with high vacancy rates, affluent waterfront zones, and established historic neighborhoods. By combining pedestrian and vehicular modes of travel in both cities, we examined the impact of velocity and travel mode on location data generation.

Figure 1. Routes for Kingston, Ontario (left) and Baltimore, Maryland (right). Basemap Copyright CARTO, OpenMapTiles, and OpenStreetMap contributors.

Table 1. Examples of parser data sentences.

-0.067131	61132.68435	-67.979666	23.975476	-1349374033384430000	36	18	501216000000	0	true	true	true	true	true	true	false	0.00	16
-0.05205	154.021211	-65.352778	28.871518	-1349374033384430000	36	18	502056000000	0	true	true	true	true	true	true	false	0.00	16
-0.577623	108.851964	-62.300981	42.389043	-1349374033384430000	36	18	503056000000	0	true	true	true	true	true	true	false	116.76	27
-0.328938	5036.731251	-13.355288	36.02374	-1349374033384280000	37	18	509632000000	0	true	true	true	true	true	true	false	0.00	16

This design captured a spectrum of urban density, spatial forms, and infrastructural conditions, ensuring robust, diverse datasets. It also accounted for technical variations in data construction and socio-spatial differences, minimizing the influence of location- or time-specific anomalies.

Project initiation

Our experiment involved using a mobile application to track movements in Kingston, Ontario, and Baltimore, Maryland. Collaborating with Dr. Dillon Mahmoudi, we investigated specific Application Programming Interface (API) methods commonly employed by Google Maps and similar navigation apps to collect, calculate, and transmit user location data. We focused on two types of location detection methods: one that uses GPS signals within the smartphone —the primary method for OsmAnd— and another that generates algorithmically smoothed location estimates via Google's servers, predominantly used by Roam. In essence, Roam operates within OsmAnd to produce alternate location estimates by leveraging external computational processes, without altering OsmAnd's internal reporting.

To analyze the data collected from both cities, we developed a Data Parsing desktop application that formatted location logs into Microsoft Excel Workbooks. These were imported into a Geographic Information System (GIS) for visualization and comparison of the location data from each walk. This analysis provided critical insights into how location data are generated and the methodological implications of combining independent and algorithmically augmented data sources, which we discuss in the results section.

Results

The parsed output from our desktop data parser app reports each single location recording in the format of a data "sentence." Each data sentence represents a different measurement that is made available to internal smartphone software and to external algorithms (i.e. Google Fused Location) to determine the device's location (see Table 1). There are over 35 variables in each sentence, ranging from satellite identification numbers and satellite constellation types (i.e. Galileo vs GPS vs BeiDou et al.) to the time recorded in nanoseconds. The log produced three sheets or three different categories of location-based recordings: National Marine Electronic Association 0183 (NMEA), Global Navigation Satellite System (GNSS) Raw Measurements, and Location Objects. NMEA is a standard format of data (called a data string), which is produced by the GNSS chip. Both NMEA strings and GNSS Raw Measurements are used by smartphones and third parties to calculate the device's location. Location Objects are those which are produced after the smartphone uses algorithms to smooth raw datasets to display more consistent location representations and movement pathways thereof within an app such as Google Maps (see Table 2).

Our analysis draws an important distinction between how OsmAnd and Roam produce location data. OsmAnd calculates location data using raw GNSS signals captured and processed entirely on the smartphone. Roam, on the other hand, transmits raw measurements to cloud-based servers where the data are combined with external sources such as Wi-Fi networks, cell towers, etc. Processed estimates are then returned to the device as computed locations. While this method is designed to improve consistency and precision in urban environments, it also introduces additional layers of abstraction.

0	-9223372036854770000	1600312448	0	0	41.60	3	0	336.84	0.86	76516529414060	7	0	49359	18	0	false	true	false	false	false
0	-9223372036854770000	1575420032	0	0	34.50	6	0	391.87	0.45	65735357368843	14	0	23567	33	0	false	true	false	false	false
0	-9223372036854770000	1599750016	0	0	32.30	3	0	210.55	0.25	76518372019974	17	0	49359	2	0	false	true	false	false	false
0	-9223372036854770000	1599750016	0	0	28.40	3	0	221.62	0.83	76524947872992	276	0	49359	2	0	false	true	false	false	false

Recognizing and understanding this difference is crucial because it shapes the kinds of location narratives each method generates.

While our method collected a wealth of data, found within GNSS data sentences for example, the goal of the method is to draw attention and awareness to the different kinds of location narratives that these data propagate. The location narratives, or stories about how we move, are constructed by the data and the processes responsible for creating, formatting, sharing, and transmitting them. As we discuss in more detail below, they are not always coherent narratives nor are they accurate representations of embodied experiences. Raw GPS, though granular, is prone to inaccuracies (Thin et al., 2016). While they are "smoothed" by Google's Fused Location processes, those inaccuracies are interpreted and recast as new data points that introduce their own biases and assumptions, for example, about whether or not our research team walked down the middle of a road or on a sidewalk. Our findings intend to defetishize and unmask these processes and data types to reveal how location data and the narratives they generate and share about travelers are not neutral or objective.

Table 2 represents the most important data to our analysis: Location Objects. This is for two reasons. First, the NMEA and GNSS datasets are "raw," and they tend to be inaccurate as such. Since the mid-2000s when location-tracking capabilities arrived on smartphones, Google has gradually developed a location calculating and smoothing architecture and installed them into the operating systems of Android devices. That architecture is referred to in the caption of Table 2 as the Location API index. They are, in essence, algorithms that accept raw location measurements and attempt to "smooth" them out so that they can be read in a more user-friendly fashion by apps that request them—and so that they can be displayed more cleanly, consistently, and coherently to the user. Simply, Location Objects are the data behind a pin drop or a travel path in a navigation app.

Of particular interest in the Location Objects data sentences (as depicted in Table 2) are two sentence types: Roam and OsmAnd. In the Method section of this article we shared that OsmAnd was the open-source navigation app, which we chose for this experiment. We outlined that OsmAnd calculates device location independently of third-party processing by using the Location APIs in the Android operating system. As such, OsmAnd is represented in the Location Objects dataset as its own data sentences that are calculated entirely within the phone itself.

The second data sentence, Roam (formerly GeoSpark), is constituted in a more complex way. The raw measurements that Roam's SDK collects are sent out of the phone to its own cloud-based servers for storage and to Google Play's servers where Google's Fused Location services exist. Google Fused Location uses cloud-based location calculating and smoothing algorithms that are more complex than the native Location APIs on the smartphone itself. As we indicated above, these cloud-based algorithms also benefit from modeling and calculating location accuracy by drawing upon existing networks and datapoints (i.e. public IoT devices such as Wi-Fi routers etc.; Google Android Developers, 2024b). This includes using publicly broadcasted Wi-Fi information from cell towers and routers in stores to assist in trilateration and thus smoothing and enhancing a device's accuracy (Bonnington, 2018), and it is from these external processes that Roam's data calculation benefits. It is these two data sentences—OsmAnd and Roam—that constitute the location narratives discussion below.

Viewing these data in GIS, we see the paths our teams followed across Kingston, Ontario and Baltimore, Maryland. We identify three distinct types of deviations evident in the data walks depicted on the maps. Analyzing the mapped data and contrasting it with our direct experiences can be beneficial for elucidating the methods by which location data are reported and processed. In the

Table 2. Examples of Osmand and Geospark data sentences.

Osmand	9.7423	-30.8338	0.0000	39.2835	-76.5864	gps	0	1665339318000	true	true	false	true	true
GeoSpark	9.1890	-29.6000	68.8439	39.2835	-76.5864	fused	0	1665339326000	true	true	true	true	true
Osmand	8.3970	-31.5387	0.0000	39.2835	-76.5864	gps	0	1665339327000	true	true	false	true	true
GeoSpark	8.7930	-29.6000	67.9861	39.2835	-76.5864	fused	0	1665339327000	true	true	true	true	true

accompanying figures, we illustrate the reported locations for OsmAnd (shown in orange) and Roam (shown in purple), along with their reported accuracies. The accuracy values, ranging from 3 to 18, indicate the radius in meters from the reported location to the actual position of the device.[5]

Accuracy deviation

Walking along sidewalks in both cities produced results where OsmAnd (orange) data reported high accuracy (represented by smaller dot sizes). Roam (purple), however, received this raw information but the additional computation performed by Roam introduced new accuracy issues (represented by the larger dots signaling a lower value for accuracy).

In some instances, the routes themselves were nearly identical between the two methods of calculating location, yet the Roam data reported resolution issues. While in Baltimore's Fells Point (see left portion of Figure 2), for example, walking along the water with a good view of GPS satellites, our path resembled what OsmAnd reported. Roam occasionally matches the accuracy reporting of OsmAnd but produces points with lower accuracy as it snapped us to the path we were already walking, sometimes reducing the typical GPS accuracy of 3 to 9 m. In the context of geospatial applications, the difference between 3 and 9 m accuracy can significantly impact location-based services and decision-making processes. For example, services like Uber depend on precise positioning to determine which side of the street a passenger is on, and a 9 m error can lead to incorrect assumptions about a user's exact location. This highlights how even seemingly minor inaccuracies can propagate through systems, affecting their functionality and user experience.

Figure 5 captures the Baltimore team walking across the bridge in a straight line, where neither Roam nor OsmAnd was able to capture our path but where OsmAnd reported higher accuracy than Roam. Kingston's downtown area provided similar results: high accuracy from OsmAnd and sometimes the computed or fused data introduces lower accuracy points again up to 9 m. These discrepancies highlight the black-boxed nature of Google's fused location, where the reasons for this perceived lack of accuracy may be related to factors such as 3D modeling or connections to Wi-Fi routers but which are hidden through the proprietary algorithm through which Google produces location.

There are two examples of accuracy deviation in Kingston. Understanding the deviation requires context that was critical to our team's embodied experience. The daytime temperatures exceeded 35° Celsius on the day of our experiment. Under direct sunlight (see Figure 3) for most of the day with minimal cloud cover while walking on concrete sidewalks and traversing asphalt crosswalks, staying cool and well-hydrated was a priority to the team. We often paused under trees to intentionally avoid direct sunlight.

At times throughout the campus and downtown core walks, our team would intentionally cut across grassy areas with tree coverage for a break. Our heat-avoiding behavior is rooted in our embodied experiences moving through the city and as such represents a notable deviation from the other linear, predictable paths often narrated by the privileged, computational model seen below. As van Es and de Lange (2020) described, there is importance in the embodied, situated practice of making visible the infrastructures that are often invisibly shaping digital geographies, especially as these can vary significantly across social differences (Middleton, 2010; Smith, 2016). The true path of our

Figure 2. Routes demonstrating accuracy deviation for Baltimore OsmAnd (top left), Kingston OsmAnd (top right), Baltimore GeoSpark (bottom left, now Roam), and Kingston GeoSpark (bottom right, now Roam). Basemap Copyright CARTO, OpenMapTiles, and OpenStreetMap contributors.

Figure 3. Images of the Kingston, ON project team under direct sunlight and tree cover

embodied experience is often different from the other two location data narratives generated by our devices and may have deviated in an entirely different manner depending on who was undertaking the walk. While the phone tracked a single path, our path was more complex as we traveled as a research group. In Figure 4, we have designated a portion of our embodied experience path as we intentionally navigated toward trees. The former is indicated by a light green line, while the latter is represented in dark green triangles. There are times when one of the smartphone-based location narratives reflects these decisions. Other times neither of our device-based data-driven narratives captured this decision making whatsoever.

Referencing the path of our embodied experience allows us to interrogate the accuracy of location data narratives. The decision to walk under trees for shade introduces subtle changes in trajectory that raw GPS data might capture but smoothed computational narratives failed to represent accurately. These deviations reveal gaps where algorithms prioritize coherence over precision, erasing the adaptive, non-linear ways our team navigated their urban surroundings. Such moments emphasize the limitations of computational systems in reliably modeling human mobility and embodied experiences, as they often misinterpret or overwrite the nuanced adjustments individuals make in response to environmental and social conditions.

Path deviation

As stated earlier, the hot weather prompted us to deviate from typical routes in both Kingston and Baltimore. We crossed the street or cut across the grass to enjoy the shade of buildings or trees. These deviations led to paths that differed from what one might expect a typical route to take. Such movements introduced differences in how the two location tracking methods, OsmAnd and Roam, placed our location data.

The OsmAnd trackers were notably less accurate when we followed predictable routes, such as sidewalks. There were instances where the location data placed us inside buildings while reporting high accuracy and then incorrectly placed us across the street while reporting low accuracy. In these scenarios, Roam was able to more accurately place us on the sidewalk and outside of buildings. However, Roam showed less flexibility when our routes diverged from expected paths. For example, in Kingston, moments when we cut across the grass to seek shade or move more quickly were more accurately modeled by OsmAnd compared to Roam. This suggests that while Roam may handle typical urban movements well, OsmAnd performs better in capturing non-traditional paths and deviations, highlighting the importance of context in evaluating the accuracy and usability of location data systems.

Another example can be seen in Figure 4. Near the end of the team's trip around downtown Kingston we stopped at the residence of one of the team members to hydrate and for a restroom break. The residence is circled in red. To the immediate northwest of the building, we can observe the device residing outside the residence before traveling inside the home. The device exits the home from the southwest corner before proceeding west-southwest across a road and then proceeding directly south. The cluster of data points to the northwest represents where our team resided while team members took turns going inside to refresh themselves. However, the device used in the experiment did not enter the house—despite the data points placing us inside, otherwise. Moreover, none of the team members exited the house from the southwest. Everyone waited outside on the driveway before some team members departed by car while the rest proceeded south for the remainder of the trip. The algorithms used to generate location narratives both contain the same discrepancies and errors. They represent false transitions and positional inaccuracies. The example underscores a challenge in interpreting micro-movements and the difficulty in reconciling raw and even smoothed data with human actions.

Missing data

In rare instances, OsmAnd's location was correct but reported low accuracy. This anomaly highlights the complexity of GPS-based tracking, where the raw data might be accurate, but the system's confidence in that data is low. For example, in our map of Baltimore (see Figure 5), a segment of travel along the I-83 freeway showed OsmAnd indicating low accuracy through large circles, despite correctly capturing the route. During this same period, Roam did not report any spatial data, resulting in an absence of location points. This is likely because the raw GNSS signals required for computation were unavailable. For example, if not enough state satellites were visible, OnNmeaMessageListener() may not have been triggered, leaving Roam's SDK with no data to process. Similarly, while walking on a sidewalk OsmAnd began to report low accuracy while Roam again failed to output any location data. These gaps reveal the technical limitations of cloud-dependent systems in constrained conditions. In addition, they highlight the fragility of computational services in the real-world. As importantly, they underscore a critical issue in location data systems: the potential for accurate data to be disregarded or underreported due to algorithmic confidence thresholds. These thresholds play an important role in balancing usability and reliability because low-confidence data could overwhelm applications or lead to errors in safety-critical contexts such as self-driving cars. Filtering these data can also discard valuable information. Due to the ways in which proprietary services tend to prioritize polished, high-confidence outputs to maintain trust, their process can introduce biases and limit transparency. We are particularly concerned about what these thresholds mean in larger volume over larger periods of time. For example, how are travel stories told and sold about refugee and human migration patterns when location data are missing? What interpretive liberties are taken to fill in those gaps in order to construct a story about how someone or a group has moved?

Similarly, while walking on a sidewalk in a gridded street layout, OsmAnd began to report low accuracy, while Roam again failed to provide any location data. This occurred despite there being no sudden change in our mode of transportation or direction. The failure of Roam to capture any data in these instances raises important questions about the robustness and reliability of computational location services under certain conditions. These gaps in data reporting suggest that both the physical environment and the specific algorithmic approaches of different services can impact the accuracy and completeness of location data. These findings emphasize the need for critical scrutiny of how location data are processed and reported, particularly in contexts where precise and reliable data are essential.

Discussion: Constructing geolocation, constructing geospatial narratives

This study's exploration of location data production through smartphones reveals critical insights into the complex interplay between raw GPS data, computational refinements, and embodied experiences. Our findings indicate variations in data accuracy and representation between the OsmAnd and Roam methodologies, particularly under different environmental conditions and user behaviors. These differences underscore the need for a deeper understanding of the mechanisms behind location data generation and its implications for digital geographies.

One of the most striking observations from our research is the contrast between the high accuracy reported by OsmAnd and the computational adjustments made by Roam, which often introduced new inaccuracies as the computational adjustments sought to correct travel that did not conform to, for example, the expected behavior of walking on sidewalks or pedestrian paths. This discrepancy raises important questions about the reliability and transparency of computational methods used to refine location data. Roam's approach, which relies on extensive training data to produce what it considers the most likely conclusions, exemplifies the challenges of interpreting and trusting black-box

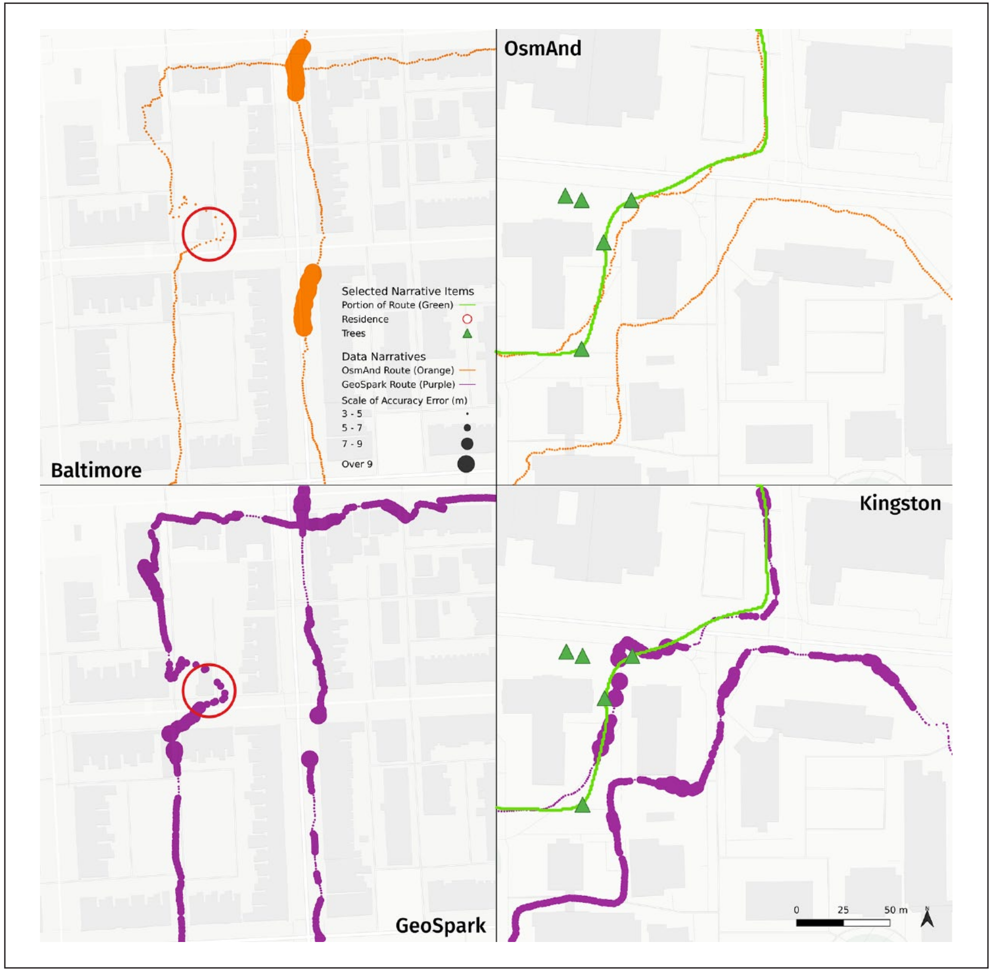

Figure 4. Routes demonstrating path deviation for Baltimore OsmAnd (top left), Kingston OsmAnd (top right), Baltimore GeoSpark (bottom left, now Roam), and Kingston GeoSpark (bottom right, now Roam). The red circle in the Baltimore maps denote the approximate location of residence we stopped at. The green line in the Kingston maps denote the path we walked seeking shade and the green triangles represent significant shade providing trees. Basemap Copyright CARTO, OpenMapTiles, and OpenStreetMap contributors.

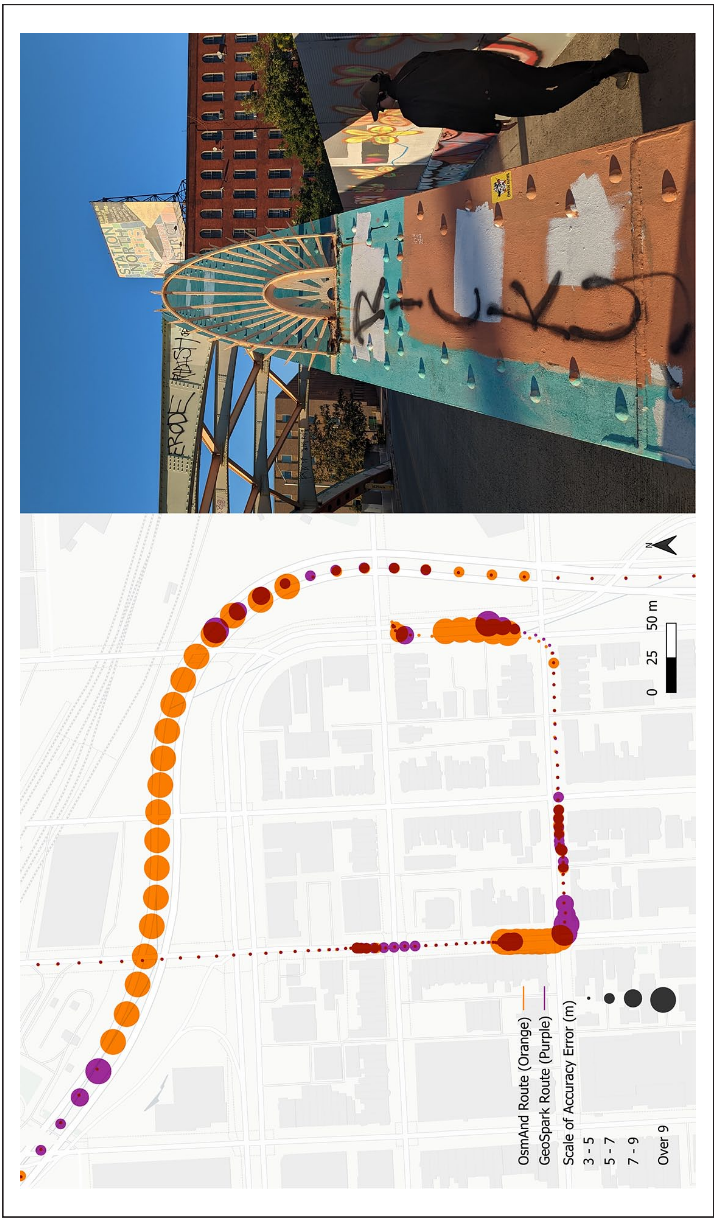

Figure 5. Geospark reported missing data as we walked (left), basemap Copyright CARTO, OpenMapTiles, and OpenStreetMap contributors. A team member walking across a bridge (right).

algorithms in location data services. This raises questions for digital geography and critical GIS in terms of the underlying data geographers are concerned with. If the urban geospatial narratives generated through how applications like Google Maps shape movement and how residents choose destinations (Dalton, 2018; Thatcher, 2013), what does it mean when locations themselves are shaped by decisions made by companies like Google? Or when that data are then used to feed into transportation planning and new technologies like self-driving cars (Attoh et al., 2019).

The contrast between OsmAnd's and Roam's location and movement are significant when considering how they aggregate over time. Though the inaccuracies and pathway differences between OsmAnd and Roam are minor in comparison to one another as represented in our figures, particularly in the context of our two data walk experiments, it is important to consider what these inaccuracies and differences mean as they compound over longer distances, repeated trips, and in historical databases used to sell narratives to third parties. As Kitchin (2013) notes, the more data that are collected over time, the greater potential for noise within them. That is, as more data are collected, more errors, redundancies, and irrelevant data increase within them. Each deviation in a dataset, no matter how minor, contributes a distorted narrative of movement and space over time. In our analysis, the micro-deviations we observe between the OsmAnd and Roam travel paths result in slightly different results in terms of our exact location. OsmAnd, relying solely on raw GNSS data, reports locations that were off from our embodied experience by a few meters. At times, OsmAnd placed us inside of buildings that we never entered. Roam, on the other hand, attempted to smooth inaccuracies through Google's Fused Location services. While the neater linearity visually represented in Roam's travel history appears more consistent at first glance, the smoothing process itself introduces distortions and erases travel behaviors. As is evident from the narratives, the smoothing process itself introduces distortions, erasing non-conforming travel behaviors such as cutting across parks or crossing streets to avoid direct sunlight, particularly when it prioritizes the normal distribution of expected sidewalk usage over the reality of spontaneous pedestrian choices.

The notion of data friction (Madsen et al., 2023)—the imperfections that arise as data from different sources is processed and reprocessed—reminds us that each time location data are smoothed or corrected, it moves further away from a traveler's embodied experience, removing critical factors such as the sensory experiences which shape the affective path-finding and the ways that varying levels of comfort in space along lines of social difference when constructing the story of a person's movement. As this process continues over time, it introduces ongoing tension between accuracy and representation; the goal of a company creating an accurate dataset is thus always elusive precisely because the very process of refining data introduces biases and assumptions about traveler consumption, preferences, and ideals. Within the location analytics or "location intelligence" industry, this issue recurs frequently because the industry itself is in its infancy, experiencing rapid growth especially since the pandemic. Companies in this industry, like Google and Roam, are still developing algorithms and models used to interpret and even predict location. The industry relies on coherence in its own storytelling and narrative capacities about human behavior when they are sold to clients like advertisers, urban planners, and law enforcement agencies. In our results, this is particularly self-evident.

How Roam handled deviations from well-defined paths—such as cutting across a lawn on Queen's University campus to pass under a tree to avoid direct sunlight on a hot day—Roam failed to capture these movements accurately. Instead, it defaulted to assumptions about where we "should" have been, exemplifying Kitchin's observation on the tendencies of noise emerging in a growing dataset. New noise is introduced as perceivably inadmissible or unwelcome data (or, what which deviates from a preferred narrative) are erased and replaced in the pursuit of a cleaner narrative. The long-term effect of this process is significant. As more data are collected, smoothed, and processed over time, the noise in the dataset grows, leading to inaccuracies that are not merely technical issues. They have real-world implications on how industries and actors understand and socialize norms and expectations of consumer behavior, urban design, and even location surveillance.

Moreover, situating this problem within the context of all three narratives contextualizes these kinds of issues. Each app that generates location narratives does it differently. In the case of OsmAnd, we see that it uses methods from Google's Location API and performs calculations about the user's position through processes of translating raw GNSS data. There are regimes of algorithmic interpretation already taking place here well before any external, cloud-based processing. For example, consider Google's Geocoder method, which converts geographical measurements (i.e. from GNSS raw measurements) into physical mailing addresses, and vice versa (Google Android Developers, Geocoder class, 2024a). The process utilizes location-based filters or bounding boxes to prioritize results (Google Maps Platform, 2024). When a user requests their location in an app that uses this API's method, the API may favor certain results that are biased toward popular or expected locations—even if those locations are not the most geographically relevant. Another potential bias can be seen within the way in which Geocoder displays results based on frequency of use or commercial priorities, meaning that known business or locations may appear more prominently even if they are not the most relevant to the user's experience. These biases distort how a space is represented as its privileges certain data types and data connections over others, thus further complexifying already complex narratives built from location data. More simply summarized, the computational processes involved in generating data narratives—whether OsmAnd or Roam—there is always a degree or interpretation and representation taking place that will always be different than a user's embodied experience. As this happens more and more over time, the more significant these differences become in terms of how they are translated and sold as stories about how we travel, how we consume, and how we interact with the urban spaces around us.

Furthermore, our study highlights the limitations of current location data methodologies in capturing the nuances of human movement in ways that further builds out our understanding of the production of digital geographies (Ash et al., 2016; Gregory and Maldonado, 2020; Huang et al., 2021; Leszczynski and Kong, 2023). The deviations from typical routes taken to avoid direct sunlight, for instance, revealed significant gaps in the ability of both OsmAnd and Roam to accurately model these behaviors. OsmAnd, while more accurate in unpredictable routes, sometimes incorrectly placed locations within buildings, whereas Roam struggled to adapt to unexpected paths, favoring predictable sidewalk routes. These findings suggest that existing location data systems may inadequately reflect the true complexity of urban mobility, particularly in diverse and dynamic environments. Given that these data are the building blocks for the production of digital understandings of urban space, with attendant feedback loops into how humans understand their spaces, where location data are accurate and inaccurate is important to understanding how space is modeled and then used. More concerning, as in the case of Zachary McCoy biking past the burglary site, that these locations can then be taken as truth within surveillance assemblages raises further concerns within the context of the carceral logics of urban space (Haggerty and Ericson, 2000; Jefferson, 2018).

In addition to these observations, the use of location data to understand urban space therefore raises three broad sets of critical questions about the underlying assumptions and implications of such practices. The first set of questions relates to how location data inform understandings of cities and their residents, potentially creating a feedback mechanism which reproduces the city in the image of the computed location data. For example, by privileging certain types of movements and spaces—such as well-trodden sidewalks over spontaneous detours—these systems may reinforce existing power dynamics and spatial inequalities. Thus, we ask, how do the computational methodologies employed by services like Roam and OsmAnd shape our understanding of urban environments (cf. Huang et al., 2021)? To what extent do these data-driven narratives reflect or distort the embodied experiences of urban inhabitants (cf. Dalton et al., 2019)? How do they shape the further uses of these data in surveillance and profit-making assemblages (c.f. Ash et al., 2016; Dayen, 2024)?

The second set of questions are predicated on location data-driven surveillance as a preoccupation that renders that which is invisible *visible* and that which is illegible *legible*—for purposes of judgment, categorization, and storytelling. We then must inquire into precisely *what* exactly is being made visible and legible (cf. Mahmoudi et al., 2022; Wilmott, 2019). What are the fundamental differences in the embodied experience of travel and the competing, non-negotiable, and often inaccurate stories told by the data (cf. Wilmott, 2016; Wilmott, 2020)? And how are those narratives shared, sold, or traded among corporations and governments (cf. Beauvisage and Mellet, 2020; Birch et al., 2021), and what impact might those differences have upon, for example, a user's locational privacy (cf. Fisher and Dobson, 2003)? What does it mean to enjoy relative degrees of obscurity or ambiguity around one's location as a means of mitigating a company's or government's desire to collect and analyze more accurate, more often, and more granular resolutions?

The third set of questions relates to the knowledge generated from data-driven surveillance and resulting competing truth claims, what we call in this article the geospatial narrative authority. The space for constructing narratives about movement and migration exists somewhere between the institutional pursuit of precise measurements and the inherently chaotic urban environment, where GPS satellite, wi-fi, and cellular signals are often disrupted by urban structures. This intermediate space, where stories are shaped, remains elusive and generally inaccessible—black boxed—to the average user, obscured by both technical complexities and institutional frameworks. What authority does the user have over the way their geospatial narratives are interpreted? Furthermore, the obscure content generated within these spaces must be recognized as truth claims about user location and movement: claims that are as much a reflection of measurable, scientific certainty as they are claims mired in error and ambiguity (cf. Wilmott, 2016). How much noise is normalized within these spaces and narratives? What degree of noisy location data is acceptable by these institutions, or even hidden from one another? What recourse might civil society have in terms of correcting, amending, or renegotiating narrative inaccuracy and exaggeration before they generate damaging profiles or catalyze data harms?

In addition, the opacity of the algorithms and the proprietary nature of the data sources used by tech companies obscure the processes behind data refinement, potentially leading to biased or incomplete representations of urban space. This critical perspective invites us to scrutinize not only the accuracy of location data but also the broader socio-political contexts in which these data are produced and utilized, challenging us to consider who benefits from these technologies and who might be marginalized by them (cf. boyd and Crawford, 2012; Dalton et al., 2016; Elvy, 2018; Wilmott, 2016).

The implications of these findings are profound for both researchers and practitioners in the field of digital geographies. By critically examining the production and refinement of location data, this study opens up avenues for more transparent and accountable methodologies. It highlights the importance of not only scrutinizing the accuracy of location data but also understanding the underlying algorithms and data sources that shape these outputs. This transparency is crucial for developing more reliable and inclusive location data practices that better serve diverse urban populations and for digital geographers seeking to understand the building blocks upon which the process they study are built.

Conclusion: Profit and the loss of geospatial narrative authority

In March 2013, Google introduced the Fused Location Provider as part of the Google Play services, which aimed to enhance location accuracy (Wilhelm, 2013). In 2020, Google enhanced the Fused Location Provider to improve location accuracy in urban environments specifically via the introduction of 3D mapping aided corrections, which utilizes building models to adjust for signal reflections and obstructions (Android Developers Blog, 2020). These technologies are used by Roam allegedly improving upon the raw GPS data reported by phones (like that data reported by OsmAnd). Given the widespread and global use of Android phones, which represent over 70% of the worldwide smartphone marketplace as of March 2024 (Sherif, 2024), this study makes a significant and timely methodological contribution by providing a nuanced understanding of how location data are produced and reproduced through smartphones. Our findings reveal three intertwined narratives: (1) the GPS location data with its inherent flaws, (2) Google's computed narrative, and (3) the human narrative of embodied experience. The interplay between these narratives underscores a tension, wherein each narrative depends on the context of "storytelling," ultimately shaping our identity as travelers, consumers, laborers, and residents.

The first narrative, the GPS location data (OsmAnd), represents the raw and initially compelling story of how smartphones determine and report user location through satellite trilateration. This narrative was once seen as providing users with control over their movements by allowing them to turn off GPS tracking. The second narrative, Google's computed location (Roam), illustrates how Google enhances and refines this data through additional sources such as local Wi-Fi routers and 3D models of cities, producing a more "precise" but less transparent story. This narrative is driven by commercial interests to sell location behavior, to advertise based on that behavior, and to optimize the information available to applications like Uber or in personalized pricing strategies like those used by McDonald's (Dayen, 2024). It reflects an on-going and continuous desire for ever-finer spatial and temporal measurements of users, transforming them into potential consumers. The third narrative, the human narrative, encompasses the embodied experiences and personal movements of users, often the most authentic yet the most variable and hardest to capture in digital location data as the multitude of factors which shape how a body experiences space, and is received by that space, cannot be easily smoothed data. This narrative captures the everyday decisions, detours, and deviations from the "norm" made by individuals that are not fully captured by computational models.

The contest over geospatial narrative authority reveals that the embodied experience, often the most authentic representation of movement and place, is paradoxically the one with the least agency in the digital age. Initially, GPS data empowered smartphones to narrate the user's journey, offering a semblance of control to the user who could ostensibly stop tracking by turning off GPS. However, our research highlights two critical evolutions. First, Google's location services have progressively undermined this autonomy. Despite users' attempts to disable GPS, Google's systems continue to collect location data, leveraging additional sources such as local Wi-Fi routers to enhance precision. Second, Google's integration of various data points into a cohesive location narrative illustrates a move toward computed storytelling that works backward into how location is calculated. These evolutions reflect a broader capitalist and data-driven imperative for increasingly granular and frequent measurements of human movement and behavior (Ash et al., 2024; Barns, 2016; Thatcher et al., 2016). This narrative is neither purely human nor raw; it is a product of sophisticated computational processes designed to serve commercial interests and, while technologically advanced, is not neutral. Instead, it carries implicit biases and motivations shaped by the interests of those who control the data and the algorithms. Roam's methodology involves producing location data conclusions that it deems most accurate, based on extensive training data and sophisticated algorithms. However, this process lacks transparency, or is deliberately obscured, raising critical questions about the implications of such data practices. What does it mean to open up the black box of location data for methodologies in digital geographies?

Opening the black box involves demystifying the computational processes and assumptions underlying location data production. This transparency is essential for several reasons. First, it allows researchers to critically assess the accuracy and biases of the data. Roam, like many advanced location data services, relies on complex algorithms trained on vast datasets to refine and predict location information. Understanding how these algorithms function and the nature of the training data they use is crucial for evaluating the reliability and representativeness of the produced data (Benjamin, 2019; Ettlinger, 2022; O'Neil, 2016).

Moreover, by unpacking these methodologies, researchers can identify discrepancies between raw GPS data and the refined outputs provided by services like Roam. This process can reveal how certain movements and locations are prioritized or marginalized within the data. For example, algorithms may favor frequently traveled urban areas while underrepresenting less common paths—as demonstrated in our results, where routes taken to avoid direct sunlight or follow less predictable trajectories were inconsistently captured. What is not yet known, and which requires further investigation, is the ways in which territorial stigma, which often involves the racialization of space in urban contexts, may also feature into how data are produced and modeled Jefferson, 2018; Kallin and Slater, 2014; Otero et al., 2022). For instance, if users avoid a space due to racial prejudice, does this then become hardwired into how space is modeled? And how is this then translated into economic processes like surveillance pricing that varies by how corporations model user behavior (Dayen, 2024).

Opening up the black box also has significant implications for how geographers might intervene in the study of digital geographies. It supports a more participatory approach to location data production and analysis. By making methodologies transparent, researchers and communities can engage more critically with the data—questioning, challenging, and potentially reshaping how it is collected and used. This participatory approach aligns with the principles of critical GIS, which advocate for more inclusive and socially responsive geographic information systems as a means of working toward just and equitable futures (Elwood, 2022; Mahmoudi and Shelton, 2022).

Furthermore, transparency in location data methodologies enhances accountability. As location data increasingly shape urban planning, policymaking, and commercial activity, it is vital that the processes generating these data are open to scrutiny. The widespread use of location data to monitor populations during the COVID-19 pandemic—often without consent—underscores the urgency of these concerns (Human Rights Watch, 2020). Methodological openness can help ensure that data-driven decisions are made more fairly and equitably, reducing the risk of reproducing existing social and spatial inequalities.

In conclusion, the transformation from GPS-based data to a fused, computational narrative represents an important shift in how location data are produced, interpreted, and utilized. This study highlights the importance of critically examining the sources and methods of location data production, recognizing the inherent power dynamics and the implications for individual agency and privacy. As we navigate an increasingly data-driven world, we argue that it is imperative to remain vigilant about who controls these geospatial narrative authorities and to advocate for transparency and agency in the representation of our movements and identities. Opening the black box of location data is a crucial step toward advancing methodologies in digital geographies. It allows for a critical examination of the accuracy and biases inherent in location data production, fosters participatory engagement with data practices, and enhances accountability in data-driven decision-making. As this research project evolves, embracing transparency will be key to understanding and improving the complex interplay between location data and urban processes.

Acknowledgements

This paper builds on the work done by the A Day in the Life of Metadata and Big Data Exposed projects which involved over 25 people from three universities. Most notable, thanks go to Fernando Hernandez Leiva for his leading of the software development and help with this article. Thanks also go to Aaron Iqbal, Alicia Sabatino, Benjamin Muller, Chris MacPhee, David Lyon, Dongmei Chen, Emily Smith, Evan Thomas, Francesco Virga, Grant Blank, Luci Bast, Nana Boateng, Noor El-Alfi, Peter Wright, Samir Toubache, Sanil Andapally, and Valerie Steeves for their work on the project. The authors would also like to thank Chu Chai for fueling work on this paper.

ORCID iDs

Dan Cohen https://orcid.org/0000-0002-9752-8586

Dillon Mahmoudi https://orcid.org/0000-0002-8816-1639

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by a Queen’s University Research Opportunities Fund Wicked Ideas Grant.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Notes

1. We use the term defetishize in reference to the Marxian concept of commodity fetishism which highlights how commodities (such as location data) present commodities as disembodied objects of exchange, obscuring the social relations needed to produce them (Marx, 1990[1867]).
2. The original app is available at https://osmand.net/
3. The term open-source refers to applications where the computational code and architecture of the app is visible and editable by anyone who wishes to modify the app.
4. While not explicitly advertised, Roam's SDK (accessible at https://docs.roam.ai/android/quickstart) uses the FusedLocationProviderClient (see details at https://developers.google.com/android/reference/com/google/android/gms/location/FusedLocationProviderClient) that is part of the Google Play services on Android.
5. While both OsmAnd and Roam report accuracy values, it is important to note that these values are calculated differently and may not be directly comparable. OsmAnd's value reflects the radius of a bounding circle based on trilateration of GNSS signals while Roam's value reflects a more complex calculation. A higher reported value for Roam does not necessarily indicate lower accuracy, but rather reflects differences in how the estimates are produced.

References

Amoore L (2014) Security and the claim to privacy. *International Political Sociology* 8(1): 108–112.

Android Developers Blog (2020) Improving urban GPS accuracy for your app. Available at: https://android-developers.googleblog.com/2020/12/improving-urban-gps-accuracy-for-your.html (accessed 21 November 2024).

Ash J, Kitchin R and Leszczynski A (2016) Digital turn, digital geographies? *Progress in Human Geography* 42: 25–43.

Ash J, Kitchin R and Leszczynski A (2024) *Researching Digital Life: Orientations, Methods and Practice*, 1st edn. Thousand Oaks, CA: Sage.

Attoh K, Wells K and Cullen D (2019) "We're building their data": Labor, alienation, and idiocy in the smart city. *Environment and Planning D: Society and Space* 37: 1007–1024.

Ball K (2019) Review of Zuboff's the age of surveillance capitalism. *Surveillance & Society* 17(1–2): 70–75.

Barns S (2016) Mine your data: Open data, digital strategies and entrepreneurial governance by code. *Urban Geography* 37(4): 554–571.

Bates E, by LK, Zwick A, et al. (2019) Gigs, side hustles, freelance: What work means in the platform Economy City/ Blight or Remedy: Understanding Ridehailing's role in the precarious "Gig Economy"/ labour, gender and making rent with Airbnb/ The gentrification of 'Sharing': From Bandit Cab to Ride Share Tech/ The 'Sharing economy'? Precarious labor in Neoliberal Cities/ Where is economic development in the platform city?/ Shared Economy: WeWork or We Work Together. *Planning Theory & Practice* 20: 423–446.

Bauman Z, Bigo D, Esteves P, et al. (2014) After Snowden: Rethinking the impact of surveillance. *International Political Sociology* 8: 121–144.

Beauvisage T and Mellet K (2020) Mobile consumers and the retail industry: The resistible advent of a new marketing scene. *Journal of Cultural Economy* 13: 25–41.

Benjamin R (2019) *Race after Technology: Abolitionist Tools for the New Jim Code*. 1st edn. Cambridge; Medford, MA: Polity Press.

Bhuiyan J (2023) TechScape: How police use location and search data to find suspects—and not always the right ones. *The Guardian*, 3 October.

Birch K, Cochrane DT and Ward C (2021) Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. *Big Data & Society* 8: 1–15.

Bodle R (2016) A critical theory of advertising as surveillance: Algorithms, big data, and power. In: Hamilton JF, Bodle R and Korin E (eds) *Explorations in Critical Studies of Advertising*. London: Routledge, pp.148–162. Available at: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315625768-18/critical-theory-advertising-surveillance-algorithms-big-data-power-robert-bodle (accessed 22 May 2024).

Bonnington C (2018) How Google uses Wi-Fi networks to figure out your exact location. *Slate*, 20 June.
boyd d and Crawford K (2012) Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.
Cox J (2021) Leaked location data shows another Muslim prayer app tracking users. *Vice*, 12 January. Available at: https://imrann40.sg-host.com/sfba/news/leaked-location-data-shows-another-muslim-prayer-app-tracking-users/
Cuppini N, Frapporti M and Pirone M (2022) When cities meet platforms: Towards a trans-urban approach. *Digital Geography and Society* 3(2): 100042.
Dalton C, Wilmott C, Fraser E, et al. (2019) "Smart" discourses, the limits of representation, and new regimes of spatial data. *Annals of the American Association of Geographers* 110: 485–496.
Dalton CM (2018) Big Data from the ground up: Mobile maps and geographic knowledge. *The Professional Geographer* 70(1): 157–164.
Dalton CM, Taylor L and Thatcher J (2016) Critical data studies: A dialog on data and space. *Big Data & Society* 3(1): 205395171664834.
Dayen D (2024) One person one price. *The American Prospect*, June. Available at: https://prospect.org/economy/2024-06-04-one-person-one-price/
De A (2022) Digital Geography and Its Methods. In: Mustafa FB (ed) *Methodological Approaches in Physical Geography* (Geography of the Physical Environment). Cham: Springer, pp.71–87. Available at: https://link.springer.com/10.1007/978-3-031-07113-3_5 (accessed 22 May 2024).
Edelman G (2020) Why don't we just ban targeted advertising? *Wired*, 22 March. Available at: https://www.wired.com/story/why-dont-we-just-ban-targeted-advertising/
Elvy S-A (2018) Commodifying consumer data in the era of the Internet of Things. *Boston College Law Review* 59: 423–522.
Elwood S (2022) Toward a fourth generation critical GIS: Extraordinary politics. *ACME* 21(4): 4436–4447.
Ettlinger N (2022) *Algorithms and the Assault on Critical Thought*. 1st edn. London; New York: Routledge.
Fields D, Bissell D and Macrorie R (2020) Platform methods: Studying platform urbanism outside the black box. *Urban Geography* 41: 462–468.
Fisher P and Dobson J (2003) Who knows where you are, and who should, in the era of mobile geography? *Geography* 88: 331–337.
Foster JB and McChesney RW (2014) Surveillance capitalism. *Monthly Review: An Independent Socialist Magazine* 66(3): 1–31.
Foucault M (1977) *Discipline and Punish: The Birth of the Prison*. New York: Pantheon Books.
Gieseking JJ (2020) Mapping lesbian and queer lines of desire: Constellations of queer urban space. *Environment and Planning D* 38(5): 941–960.
Google Android Developers (2024a) Geocoder. Available at: https://developer.android.com/reference/android/location/Geocoder?_gl=1*1pqe8gl*_up*MQ..*_ga*NzMwMTkyNjI2LjE3MjczNjI4MjU.*_ga_6HH9YJMN9M*MTcyNzM2MjgyNS4xLjAuMTcyNzM2MzEyMi4wLjAuOTIwNDM5ODEx
Google Android Developers (2024b) Location#getAccuracy(). Available https://developer.android.com/reference/android/location/Location#getAccuracy() (accessed 29 May 2024).
Google Fused Location Provider API (2024) Fused location provider API. Available at: https://developers.google.com/location-context/fused-location-provider (accessed 29 May 2024).
Google Maps Platform (2024) Geocoding Service. Available https://developers.google.com/maps/documentation/javascript/geocoding (accessed 22 September 2024).
Gregory K and Maldonado MP (2020) Delivering Edinburgh: Uncovering the digital geography of platform labour in the city. *Information, Communication & Society* 23(8): 1187–1202.
Haggerty K and Ericson R (2000) The surveillant assemblage. *The British Journal of Sociology* 51(4): 605–622.
Harrison P (2000) Making sense: Embodiment and the sensibilities of the everyday. *Environment and Planning D* 18(4): 497–517.
Hart G (1989) Surveillance studies. In: Wright JD (eds) *International Encyclopedia of the Social & Behavioural Sciences*. 2nd edn. Amsterdam: Elsevier, pp.733–741.
Hodson M, Kasmire J, McMeekin A, et al. (2020) *Urban Platforms and the Future City: Transformations in Infrastructure, Governance, Knowledge Production and Everyday Life*, 1st edn. New York: Routledge.
Huang H, Yao XA, Krisp JM, et al. (2021) Analytics of location-based big data for smart cities: Opportunities, challenges, and future directions. *Computers, Environment and Urban Systems* 90: 101712.

Human Rights Watch (2020) Mobile location data and Covid-19. Available at: https://www.hrw.org/news/2020/05/13/mobile-location-data-and-covid-19-qa (accessed 2 May 2024).

Iland D, Irish A and Sandler B (2018) Rethinking GPS: Engineering next-gen location at Uber. *Uber Blog: Engineering*. Available at: https://www.uber.com/en-UY/blog/rethinking-gps/ (accessed 2 May 2024).

Iliadis A and Russo F (2016) Critical data studies: An introduction. *Big Data & Society* 3(2): 205395171667423.

Jefferson BJ (2018) Computerizing carceral space: Coded geographies of criminalization and capture in New York City. *Environment and Planning A: Economy and Space* 50(5): 969–988.

Kallin H and Slater T (2014) Activating territorial Stigma: Gentrifying marginality on Edinburgh's periphery. *Environment and Planning A: Economy and Space* 46(6): 1351–1368.

Kinsley S (2014) The matter of "virtual" geographies. *Progress in Human Geography* 38(3): 364–384.

Kitchin R (2013) Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267.

Kitchin R and Lauriault T (2014) Towards critical data studies: Charting and unpacking data assemblages and their work. *The Programmable City Working Paper 2*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112

Klauser FR (2013) Political geographies of surveillance. *Geoforum Complete* 49: 275–278.

Langley P and Leyshon A (2017) Platform capitalism: The intermediation and capitalisation of digital economic circulation. *Finance and Society* 3: 11–31.

Leszczynski A (2015) Spatial big data and anxieties of control. *Environment and Planning D: Society and Space* 33(6): 965–984.

Leszczynski A (2019) Platform affects of geolocation. *Geoforum* 107: 207–215.

Leszczynski A and Kong V (2023) Walking (with) the platform: Bikesharing and the aesthetics of gentrification in Vancouver. *Urban Geography* 44: 773–795.

Madsen A, Munk A and Soltoft J (2023) Friction by machine: How to slow down reasoning with computational methods. *2023 EPIC Proceedings*, pp.82–105. Available at: https://www.epicpeople.org/friction-by-machine-and-computational-methods/

Mahmoudi D and Levenda A (2016) Beyond the screen: Uneven geographies, digital labour, and the city of cognitive-cultural capitalism. *Triplec: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 14(1): 99–120.

Mahmoudi D and Shelton T (2022) Critical GIS doing. *ACME* 214(4): 327–336.

Mahmoudi D, Hawn CL, Henry EH, et al. (2022) Mapping for Whom? Communities of color and the citizen science gap. *ACME* 21(4): 4372–4388.

Mahmoudi D, Levenda A and Sabatino A (2024) The urban-tech feedback loop: A surveillance and development data-walk in South Lake Union. *Digital Geography and Society* 7: 100106.

Marx GT (2015) Surveillance studies. In:Wright JD (ed.) *International Encyclopedia of the Social & Behavioural Sciences*, 2nd edn, vol. 23, pp.733–741.

Marx K (1990[1867]) *Capital: A Critique of Political Economy Volume 1*. Colchester: Penguin Classics

Middleton J (2010) Sense and the city: Exploring the embodied geographies of urban walking. *Social & Cultural Geography* 11(6): 575–596.

O'Neil K (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.

Otero G, Méndez ML and Link F (2022) Symbolic domination in the neoliberal city: Space, class, and residential stigma. *Urban Geography* 43(4): 632–658.

Schuurman N (2006) Formalization matters: Critical GIS and ontology research. *Annals of the Association of American Geographers* 96(4): 726–739.

Shelton T, Poorthuis A and Zook M (2015) Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning* 142: 198–211.

Sherif A (2024) Mobile OS market share worldwide 2009-2024. Available at: https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/ (accessed 16 May 2024).

Smith GJD (2016) Surveillance, data and embodiment: On the work of being watched. *Body & Society* 22(2): 108–139.

Stehlin J and Payne W (2022) Disposable infrastructures: "Micromobility" platforms and the political economy of transport disruption in Austin, Texas. *Urban Studies* 60(2): 274–291.

Thatcher J (2013) Avoiding the Ghetto through hope and fear: An analysis of immanent technology using ideal types. *Geojournal* 78(6): 967–980.
Thatcher J (2017) You are where you go, the commodification of daily life through "location." *Environment and Planning A: Economy and Space* 49(12): 2702–2717.
Thatcher J, O'Sullivan D and Mahmoudi D (2016) Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space* 34(6): 990–1006.
Thin L, Ting L, Husna N, et al. (2016) GPS systems literature: Inaccuracy factors and effective solutions. *International Journal of Computer Networks & Communications (IJCNC)* 8(2): 123–131.
Van Diggelen F (2021) End game for Urban GNSS: Google's use of 3D building models. *Inside GNSS*. Available at: https://insidegnss.com/end-game-for-urban-gnss-googles-use-of-3d-building-models/ (accessed 2 May 2024).
van Es K and de Lange M (2020) Data with its boots on the ground: Datawalking as research method. *European Journal of Communication* 35(3): 278–289.
West SM (2019) Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society* 58(1): 20–41.
Wilhelm A (2013) Inside Google's new location APIs for Android. Available at: https://thenextweb.com/news/inside-googles-new-location-apis-for-android (accessed 22 November 2024).
Wilmott C (2016) Small moments in Spatial Big Data: Calculability, authority and interoperability in everyday mobile mapping. *Big Data & Society* 3(2): 2053951716661364.
Wilmott C (2019) "Mapping-with": The politics of (Counter-)classification in OpenStreetMap. *Cartographic Perspectives* 92: 43–57.
Wilmott C (2020) *Mobile mapping: Space, cartography and the digital*. Amsterdam University Press. Available at: https://library.oapen.org/handle/20.500.12657/37349 (accessed 29 May 2024).
Wilson MW (2018) On being technopositional in digital geographies. *Cultural Geographies* 25(1): 7–21.
Winner L (1993) Upon opening the black box and finding it empty: Social constructivism and the philosophy of technology. *Science, Technology & Human Values* 18(3): 362–378.
Zuboff PS (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

Thomas N Cooke is an independent researcher of data ethics. He is also an AI ethics consultant. His work examines the politics of location data, the social implications of AI systems, and governance issues around emerging technologies. Since his SSHRC postdoctoral fellow at the Surveillance Studies Centre, he has focused specifically on surveillance and privacy issues surrounding user profiling through corporate collection of smartphone location metadata.

Dan Cohen is a Associate Professor in the Department of Geography and Planning at Queen's University. His research focuses on the power of finance and market systems in the creation of sociospatial inequalities.

Dillon Mahmoudi is a Associate Professor in the Department of Geography and Environmental Systems at the University of Maryland, Baltimore County. As an urban and economic geographer, his work maps the faultlines of digital capitalism, using heterodox political economics to understand how technology creates uneven development.