NATIONAL ACADEMIES Sciences Engineering Medicine

Proceedings of a Workshop—in Brief

Location Data in the Context of Public Health, Research, and Law Enforcement: An Exploration of Governance Frameworks

Proceedings of a Workshop—in Brief

Location Data in the Context of Public Health, Research, and Law Enforcement: An Exploration of Governance Frameworks: Proceedings of a Workshop in Brief (2022)

DETAILS

12 pages | 8.5 x 11 | PDF
ISBN 978-0-309-69128-4 | DOI 10.17226/26645

CONTRIBUTORS

Steven Kendall and Dominic Lobuglio, Rapporteurs; Committee on Science, Technology, and Law; Policy and Global Affairs; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

National Academies of Sciences, Engineering, and Medicine 2022. *Location Data in the Context of Public Health, Research, and Law Enforcement: An Exploration of Governance Frameworks: Proceedings of a Workshop in Brief*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26645.

NATIONAL ACADEMIES Sciences Engineering Medicine

Location Data in the Context of Public Health, Research, and Law Enforcement: An Exploration of Governance Frameworks

Proceedings of a Workshop—in Brief

INTRODUCTION

On June 8-9, 2022, an ad hoc planning committee under the auspices of the National Academies of Sciences, Engineering, and Medicine's Committee on Science, Technology, and Law (CSTL) hosted a workshop, *Location Data in the Context of Public Health, Research, and Law Enforcement: An Exploration of Governance Frameworks.* The workshop examined the collection, interpretation, and use of location data by government, academia, and industry.[1]

During opening remarks, workshop planning committee Co-chair **Caroline Buckee** (Harvard T.H. Chan School of Public Health) noted that the impetus for the workshop was the deluge of location data from cell phones and other sources that flowed to policymakers and researchers seeking to contain the COVID-19 pandemic. Epidemiologists have been using location data in research for some time, but the pandemic vastly increased awareness of location data as a useful source of information for policymaking and law enforcement. With location data, Buckee said, there is a general lack of systematic frameworks for sharing and aggregating the data in a way that preserves individual privacy. She identified as risks related to the use of location data in the context of public policy risks to vulnerable populations and individuals and risks associated with corporate control of location data, suggesting that an understanding of risks must inform both the regulation of location data and decisions about how location data are aggregated and used.

Committee Co-chair **Paul Ohm** (Georgetown University Law Center) said that privacy scholars tend to discuss "omnibus" approaches to protecting privacy, but because the topic is very complex and surrounded by many uncertainties, a deep dive is merited. Privacy is contextual and giving attention to specific privacy contexts is very important, he said. Some focused "sectoral" privacy laws have had profound impact,[2]

[1] For this proceedings, location data is defined as information about the specific geographical whereabouts of a particular device. It can be collected and tracked by mobile phone operators, by devices through a global positioning system (GPS) satellite (e.g., when using an application), or by Wi-Fi access point. See, e.g., https://www.arm.com/glossary/location-data#:~:text=Location%20data%20is%20information%20about,such%20as%20a%20mapping%20application.

[2] See, e.g., the Wiretap Act of the Electronic Communications Privacy Act of 1986 (ECPA), 18 U.S.C. §§ 2510-2523 and the Illinois Biometric Information Privacy Act (BIPA), (740 ILCS 14/).

but Ohm noted that there are few laws, if any, focused specifically on location data. Ohm suggested that, if members of the location data ecosystem have not begun to agree on norms, the workshop might spur their development: in the absence of norms, regulators might fill an ethical void with rules that are not well calibrated to the needs of the community. We should, Ohm said, keep both the positive uses of location data for good and the negative harms that the spread of location data might cause in view, with balance and nuance as our watchwords.

THE COLLECTION OF LOCATION DATA

Navin Vembar (Atlassian) moderated the first panel.[3] He explained that the purpose of the session was to introduce core ideas about how location data are collected, distributed, and used.

Tom Lee (Mapbox) explained that his company is a mapping location services platform that provides modular software and Application Programming Interfaces (APIs) that allow others to build map experiences into their digital applications (apps). Mapbox mapping tools are built into tens of thousands of apps and collect large volumes of location data every day. These data are used to map traffic, inform the company's internal team mapping process, and shared as a dataset called Mapbox Movement.[4] Lee stated that Mapbox does not sell location data in raw form and that collected data are not used for advertising purposes, but instead to improve the company's products and services. Lee said that Mapbox only collects data that customer apps are collecting and that these data are heavily filtered to produce "relatively beautiful traces" of movement patterns. Lee raised the issue of user privacy and said that data minimization[5] is the most important principle for ensuring "maximal" privacy. What Mapbox is looking for, he said, is "the digital equivalent of stepping outside on the city street and having a very quick impression of how quickly traffic is moving, its volume, and the number of people that are passing by."

Eugenia Giraudy (Meta) then discussed Meta's Data for Good program.[6] The goal of the program is to leverage Meta or Facebook data and tools to help other organizations and researchers with topics such as climate change, migration, disaster response, economic opportunity, and public health. Giraudy discussed Meta's Maps for Good—which are constructed using geolocation data collected from Facebook users—and its Insights for Impact program, where, through the analysis of trends in public posts on Facebook, Meta helps organizations reach target populations through better informed outreach.[7] Using location data that Facebook users share with the company, Giraudy said, it was possible to help researchers, epidemiologists, and those working with governments better respond to the COVID-19 pandemic. Giraudy also discussed Meta's privacy protection mechanisms. Data are only shared in the aggregate and, in regions where there are low data counts, collected data are deleted from aggregate datasets to prevent the identification of individuals' data. Further, noise is added to the data to make demographics fuzzier. Differential privacy (DP) mechanisms are also applied.[8] In certain cases, access to data is limited to trusted partners. It is important that collected data are representative, and Meta employs weighting methodologies to help ensure the representativeness of data (with regard to gender, age, economic status, etc.). "I would encourage everybody that is using [...location] data to think about the biases the data may have," Giraudy said.

Julia Angwin (The Markup) is a journalist who runs a newsroom that often investigates issues related to privacy. "Never before in human history has it

[3] Dr. Vembar was a member of the workshop planning committee.

[4] "Mapbox Movement is a global privacy-forward dataset of anonymized and aggregated mobile device activity." See https://docs.mapbox.com/data/movement/guides/.

[5] The principle of data minimization "means that a data controller should limit the collection of personal information to what is directly relevant and necessary to accomplish a specified purpose. They should also retain the data only for as long as is necessary to fulfil that purpose." See https://edps.europa.eu/data-protection/data-protection/glossary/d__en#:~:text=Data%20minimization,necessary%20to%20fulfil%20that%20purpose.

[6] See https://dataforgood.facebook.com/.

[7] See https://dataforgood.facebook.com/dfg/tools/insights-for-impact.

[8] "'Differential privacy' describes a promise, made by a data holder, or curator, to a data subject: 'You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available. At their best, differentially private database mechanisms can make confidential data widely available for accurate data analysis, without resorting to data clean rooms, data usage agreements, data protection plans, or restricted views." See C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy'" at https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf.

been possible to track human movements so comprehensively," she said, "and all of this is because of the phone in your pocket." There is a huge industry that collects and monetizes location data collected from cell phones. Cell phone apps that collect location data sell that data to data brokers.[9] There are no real restrictions on the sale of location data, and data brokers sell to buyers that include the federal government, financial firms, and insurance, marketing, and transportation companies. Research has shown that 95 percent of people can be readily identified from just four distinct location points,[10] but many data brokers sell raw location data collected from apps. Such data makes it very easy to identify individuals. Angwin noted that the sale of raw data is permissible if the app discloses the practice in its privacy policy, emphasizing that privacy laws are insufficient to allow users to meaningfully consent to the sale of data.

Rebecca Williams (American Civil Liberties Union) discussed the risks of location data, noting that it is easy to identify individuals from location data, because, "We are not going to random homes and random workplaces and random friend hangouts—as we move through the world we're very predictable." It is therefore important to anticipate harmful secondary uses of location data. She suggested that current regulations offer inadequate protection against powerful location tracking tools.

Williams said that apps collect a vast amount of location data that can be useful to, among others, advertisers, members of law enforcement or the military that are interested in tracking specific groups, and those who might have an interest in, for example, those who went to a Planned Parenthood clinic. One branch of government might collect location data for public health purposes (e.g., COVID-19 contact tracing), but the same data may be repurposed by law enforcement for other uses. While some jurisdictions have enacted laws to prevent inter-governmental sharing of location data,[11] in general there are no prohibitions on secondary uses of data. The Fourth Amendment protects "against unreasonable searches and seizures" and states that no warrants shall be issued for searches without "probable cause, supported by Oath or affirmation,"[12] "but a lot of the big data collection…, including location data [collected] through cell phones or through automatic license plate readers or through" tile trackers, "can be purchased by police without a warrant." She asked the audience to consider how to prevent the exploitation of location data while simultaneously maintaining the ability to collect aggregate snapshots of location data that occasionally serve the public good.

AGGREGATION AND ANONYMIZATION OF LOCATION DATA: THE STATE OF THE ART

Session moderator **Yves-Alexandre de Montjoye** (Imperial College London) introduced the session by stating that the goal of the session was to discuss technical solutions that would permit the use of location data in ways that preserve privacy.[13] He said that the absence of a name, phone number, and address is insufficient to preserve the privacy of individuals and that the addition of noise[14] and de-identification, a traditional technique used to preserve privacy, are not effective in the case of location data. De Montjoye said discussion would focus on three tools that have been developed to use location data while preserving privacy: 1) query-based and open algorithm systems; 2) differential privacy; and 3) synthetic data.

Michael Platzer (MOSTLY.AI) discussed synthetic data as a mechanism to provide privacy in the case of granular data sharing. There has been a huge increase in the appetite for granular level data, but anonymization does not work in the case of high dimensional data.[15] There has also been a recent change in awareness regarding

[9] Angwin said her newsroom identified 47 data brokers based on their marketing material. As examples, she identified Near, a firm that "describes itself as the world's largest data set of people's behavior in the real world, with data representing 1.6 billion people across 44 countries"; Mobile Wallet, which "says it represents 40 countries, 1.9 billion plus devices, [and] 50 billion mobile signals daily"; and X-Mode, which "says its data covers 25 percent of the adult U.S. population monthly."

[10] See de Montjoye, Y.-A., C.A. Hidalgo, M. Verleysen, and V.D. Blondel. Unique in the crowd: the privacy bounds of human mobility. ***Scientific Reports*** 2013. (3):1376.

[11] See Massachusetts Information Privacy and Security Act (MIPSA) (S.2687).

[12] U.S. Constitution, amend. XIV.

[13] Dr. de Montjoye was a member of the workshop planning committee.

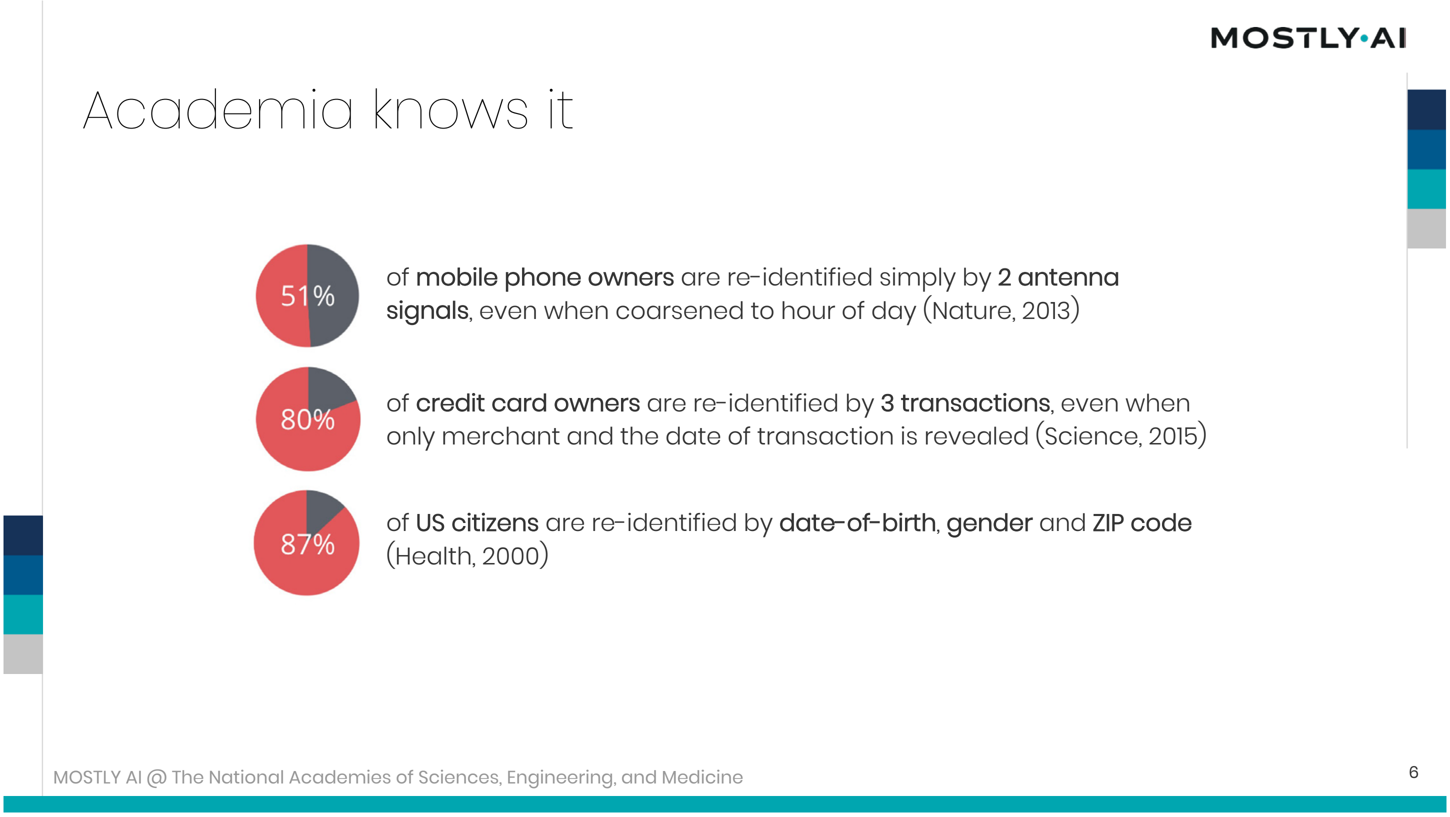
[14] For the purpose of this proceedings, noise is defined as any technique that aims to coarsen the spatial (e.g. GPS coordinates) or temporal (timestamp) properties of data.

[15] The dimension of a dataset corresponds to the number of attributes/features that exist in a dataset. High dimensional data exists within datasets with large numbers of attributes.

how easy it is to identify individuals based on simple socio-demographic attributes (see, for example, Figure 1). Encryption-based methods, query-based systems, and trusted environments may be used to protect privacy, but each has limitations. Platzer offered synthetic data as a potentially viable option.[16] Synthetic data "is based around the idea that the model learns what actual people look like, what they do, how they behave," he said. The use of synthetic data breaks the one to one relationship between data and user, but provides the same level of granularity as original data—this means that granular level data can be shared with a broad group of data users without compromising privacy.

Uttara Sivaram (Uber) spoke about the contentious topic of making location data available for government use. She identified 3 problems: (1) there seems to be no common agreement among governments, companies, and academics regarding how easily re-identifiable location data are; (2) this leads to a lack of acceptance for aggregated or otherwise anonymized data, which is exacerbated by the fact that; (3) privacy legislation rarely applies to the public sectors, and therefore fails to protect data when transferred from private organizations to public agencies.

Sivaram discussed the interest of governments in Uber trip data, which she defined as combinations of location data timestamps and vehicle information. Internal efforts to quantify the re-identifiability of Uber trips have shown that risk depends on the number of trips occurring at the same time in the same approximate area. It is important to understand which public projects or policies would benefit from location data and, from there, determine which data in what form would be appropriate to share. Decreasing the precision of fields that require less precision or accuracy in the context of a particular research question could allow for increased precision of fields where accuracy is more important, therefore



MOSTLY·AI

Academia knows it

51% of **mobile phone owners** are re-identified simply by **2 antenna signals**, even when coarsened to hour of day (Nature, 2013)

80% of **credit card owners** are re-identified by **3 transactions**, even when only merchant and the date of transaction is revealed (Science, 2015)

87% of **US citizens** are re-identified by **date-of-birth**, **gender** and **ZIP code** (Health, 2000)

MOSTLY AI @ The National Academies of Sciences, Engineering, and Medicine

6

FIGURE 1. The re-identification of individuals.
SOURCE: Michael Platzer, Workshop Presentation, June 8, 2022.

[16] Synthetic data are data coming from an AI model that has itself been trained on the original data. The generated data resemble the original data on average, e.g., the data would show people coming from and going to an airport, but each "person" is fictional. Importantly, while the generated data resemble the original data, there is no guarantee that the specific statistic of interest, e.g., the number of people actually going to the airport on Monday, July 2, is correct.

preserving data utility while reducing risks to the individual. We need more accountability for all entities, including government agencies, that collect and use and process location data, she said, as this is an important way of incentivizing responsible data practices from data providers and data recipients. Establishing a stronger legal framework that supports the flow of data between different entities could unlock data sharing opportunities that might be currently infeasible due to the imbalance of legal responsibility.

Gerome Miklau (Tumult Labs, Inc. and University of Massachusetts Amherst) focused his remarks on what he referred to as a data custodian—"the person who is in possession of sensitive location data and is responsible for allowing it to be used or shared appropriately." Data custodians need a "privacy filter" that allows insights about groups to pass the filter and to be shared externally by the organization that holds the sensitive location data. He said that reliable privacy filters are extremely difficult to design. Further, a class of re-identification attacks revealed that individual-level information can be recovered from de-identified data. Similarly, it was believed that if only aggregate statistics were released it would be possible to protect privacy. However, attacks have shown how aggregate query answers can be used to recover personal information. Even synthetic tables of structured synthetic records can be attacked. The foundation of reliable privacy filters is differential privacy (DP), a formal privacy model. DP performs standard computations on data through the application of a set of rules that limit, in a mathematically rigorous way, the personal information that could be revealed in the output. Output from a computation that satisfies a DP standard provides a rigorous guarantee that disclosure about individuals has been controlled. The guarantee of differential privacy holds regardless of the computational power of a potential adversary or the knowledge of the potential hacker. For that reason, Miklau said, DP resists attacks that we know about and see in the news but also future attacks using new methods.

LOCATION DATA USE CASES

Session moderator **Caroline Buckee** stated that the panel would focus on use cases for location data—specifically as related to disasters, crises, and urban resilience; public health; and law enforcement/national security. She asked panelists to reflect on privacy concerns and use case needs in relation to data resolution and scale.

Chao Fan (Texas A&M University) began his presentation by observing that advanced technologies like environmental sensors and smartphones have enabled the collection of data related to the human and built environment. Coupling this data with artificial intelligence (AI) technologies, it is possible to enhance community resilience in the event of a disaster. Fan discussed the utility of location data in the context of resilience measurement (data-driven insights about the spatial and temporal patterns of risk impacts); activity-based risk exposure (insights about the unfolding of a hazard event and condition of infrastructure); assessment of disaster preparedness (insights about the extent and spatial distribution of social, physical, and economic preparedness of a population); and recovery monitoring (rapid and continuous monitoring of the recovery patterns of communities). It is possible, he said, to build population-facility networks (systems in which human populations interact with urban facilities such as grocery stores and healthcare facilities) based on the demand of a population from a residential area for specific facilities. Such networks can assist in redesigning urban spaces to enhance equity and resilience.

Andrew Schroeder (Direct Relief) discussed the utility of location data in the context of humanitarian health dynamics; evacuation resource allocation; and refugees and displacement. During Hurricane Harvey in 2017, organizations had access, for the first time, to data on population density changes from Facebook's Data for Good Program. These data provided a form of "tactical remote sensing" that allowed researchers to understand change in a population over a particular, defined moment of crisis and helped relief organizations organize their thinking around medical resources and prioritize health care needs through an understanding of displacement patterns. In the case of the 2021 Marshall Fire in Colorado, over 1000 people were displaced. The insertion of location data into the emergency management workflow helped aid workers understand displacement during evacuation efforts. Similarly, in the aftermath of the Russian invasion of Ukraine, ready access to location data allowed for regular, mobility-based analysis on population concentrations compared to pre-war baselines. When cities and other areas were receiving refugees after the invasion, location data provided useful information to the public with regard to the provision of healthcare resources and access to hospitals and health centers, education, and housing. With regard to the utility of location data in emergency situations, Schroeder identified as key gaps: (1) temporal and spatial scale limits in app-based mobility data; (2) lack of a defined emergency access process from mobile network operators; (3) unevenness with regard to representativeness across different data providers and geographies; (4) lack of transparency across providers on sample size and representativeness; (5) lack of standards for emergency mobile data metrics; and (6) limitations on local capacity to perform emergency mobility data analysis. He suggested that a key next frontier will be building the ability to analyze and access location data at an appropriate level where events are taking place.

Amy Wesolowski (Johns Hopkins University School of Public Health) discussed epidemiologically relevant mobility patterns in the context of disease transmission. While there are many datasets available on different spatial and temporal scales, there are gaps that limit our understanding of disease transmission. Nevertheless, mobile phone data present an opportunity to quantify very fine spatial and temporal scales of mobility in many settings, and mobility can play an important role in understanding the spatial and temporal dynamics of pathogens. It is important, Wesolowski said, to be able to understand human behavior so that we can build better predictive models, improve the allocation of resources, and ultimately achieve disease control and elimination. The COVID-19 pandemic has spurred a great deal of research with (and access to) mobility data. Understanding how location data have been used can investigate the added value and ways for these data to be integrated into decision making.

Birgitte Freiesleben de Blasio [Norwegian Institute of Public Health (NIPH)] discussed COVID-19 modeling in Norway. Models based upon location data increased situational awareness, both nationally and regionally, of COVID-19 reproduction numbers, cumulative number of infections, and prognoses of hospital beds. Models also assisted with scenario analyses, planning, and response, particularly with regard to regional prioritization of vaccines, the understanding of the key parameters of the driving forces of the pandemic, and likely possible future disease trajectories. In many instances, the time between modeling and policy decisions was short (e.g., in the regional prioritization of vaccines). During the pandemic, NIPH has also used mobile data to investigate the effect of non-compulsory and follow-up mandatory COVID-19 non-pharmaceutical interventions in cities and rural areas. Freiesleben de Blasio suggested that more research is needed to understand the "value" of real-time mobility data.

The session's final panelist, **David Kris** (Culper Partners LLC), spoke about location data use by the law enforcement and national security communities. Location data are profoundly valuable to law enforcement officials, but the use of location data by these entities raises significant privacy concerns. In the law enforcement context, such data are useful for tracking the movement of suspects. In a national security context, the utility is similar, i.e., it might show that a suspect visited a foreign embassy. On the battlefield, location data might aid in weapons targeting or provide insights into the location and disposition of military forces. "I am not aware," Kris said, "of any serious argument that location data are not helpful to the authorities in solving crimes or protecting national security." Nonetheless, he continued, "The value of the data to the authorities in law enforcement and intelligence is directly related to the privacy invasion that these data represent."

Kris discussed the U.S. Supreme Court's decision in *Carpenter v. United States.* In that decision, the court ruled that, if authorities want to obtain persistent location data for an individual directly from a cell-phone provider, the Fourth Amendment generally requires that a warrant be obtained based on probable cause. He noted, however, that while "it may take probable cause and a search warrant to get location data for 7 days on one person, it takes just a little bit of money to get a massive amount of location data on a huge number of people for a very extended period of time."

LOCATION DATA: PRIVACY CONCERNS AND RISKS OF HARM

Session moderator **Paul Ohm** introduced the panel by stating that the panelists, experts in the field of information privacy, had been asked to consider location data in the context of privacy concerns and risks of harm.

Kirsten Martin (University of Notre Dame) stated that she views location data within a general framework of what would be called "privacy in public." Users have very specific privacy expectations and particular worries with regard to the collection, use, and sharing of location data. The technical details of location data collection does not matter much to consumers. Instead, the inferences drawn from location data are most important: if location data are gathered and aggregated to figure out things about individuals, this does not meet consumer privacy expectations. For Martin, this raises the important question: "What are reasonable secondary uses of location data?"

Danielle Citron (University of Virginia Law School) said that location data in the hands of third parties can endanger lives and economic opportunities, subject individuals to manipulation, and lead to criminalization. Apps can be downloaded onto phones to create a mirror of everything done with (and said on) them, inclusive of location. So-called "cyberstalking apps" have been installed on phones and used to track down people to murder them. Location data, she said, can reveal crucial aspects of our intimate lives, sexual activities, and sexual orientation or gender identities. Citing the case of a priest who was fired when his activity in the gay community was revealed from location data, Citron demonstrated that information revealed by location data can be job endangering. She noted that location data can also be used for exploitation, citing the case of a data broker in Massachusetts who provided location data on individuals visiting Planned Parenthood locations to a pro-life client. The client targeted these individuals with online advertising that condemned abortion as dangerous to women's lives—messages that were false, exploitative, and manipulative. Citron suggested that, if *Roe v. Wade* were to be overturned, law enforcement could use location data to identify women and girls visiting abortion clinics and prosecute them.[17]

Woodrow Hartzog (Northeastern University School of Law) discussed the concept of "waiver"—the idea that "when you knowingly expose yourself to others, you are consenting to being watched or somehow waving privacy interests." He said that even when we expose ourselves to some we do not expect to expose ourselves to *all*. Hartzog suggested that the creation of location data, in and of itself, is a moral act because it makes certain things easier to discover and makes certain kinds of activities (e.g., surveillance activities) easier to engage in. With location data, the reduction in transaction cost is truly remarkable and gives power to those that want to control the activity of others. Location data can be weaponized by those who seek to use that power against others (e.g., to track people everywhere they go and to intimidate them), and this can have chilling effects: "If you know that you are going to be monitored and tagged for every protest that you attend, for example, then you might be less willing to engage in fundamentally protected expressive activity," he said. Hartzog suggested that if location data are used persistently and pervasively to track individuals, this enables a "perfect enforcement of norms and rules" that were never meant to be perfectly enforced. The realization of perfect enforcement (e.g., of speed limits) feeds oppressive surveillance systems. Marginalized populations (e.g., people of color) suffer surveillance hardest.

LOCATION DATA AND COMMUNITY

Session moderator **Jasmine E. McNealy** (University of Florida)[18] began the session by stating that the panel would consider the intersection of location data and community. She asked the panelists to define community in the context of location data and to explore the implications of location data use and collection for privacy, democratic principles, and consent.

[17] The U.S. Supreme Court overturned *Roe v. Wade* on June 24, 2022, two weeks after the workshop.

[18] Dr. McNealy was a member of the workshop planning committee.

Megan Doerr (Sage Bionetworks) considered why the concept of community is so important in the context of location data. Because we are never truly alone, she said, location data about ourselves also reveals information about others. As a result, individual consent is insufficient in the case of location data. Location data affect communities. In the location data context, community is more than a group with a distinct history, culture, or tradition. Further, a location data community may be defined by transient attributes (e.g., education status, housing status, profession, use of mobility devices, or patterns of exercise) and a data community might be a group with a shared identity, such as religious affiliation, gender, or ethnicity. A community might even be defined by shared digital traces, such as by a common choice of an app (e.g., Grindr in the gay community) or a Bluetooth beacon. In this context, Doerr asked how we should engage with communities linked only by location data, suggesting that we must "build the social license for location data collection and use" and that we "need to dispense with the theater of anonymity or data neutrality" and "acknowledge that [location] data are identifying" and not neutral. She suggested that the governance of location data should be multi-directional and incorporate all sorts of stakeholders, including the community being surveyed, whose inputs can then be used to develop informed targets and specific policy recommendations.

Joon-Ho Yu (University of Washington) drew from his background in the nonprofit sector to explain that past and present interpersonal context within a community deeply affects the use and interpretation of location data, especially in "newcomer communities" of immigrants. He cited the example of South Korea's use of contact tracing during the COVID-19 pandemic, noting that, while South Korea's security concerns vis-à-vis North Korea have led to strict privacy protections, the South Korean government has carved out exceptions for emergencies when the imperatives of collective public health outweigh individuals' data privacy. "Prior experiences under the policies of one's country of origin shape" and "influence a newcomer community's" attitudes toward location data. Yu suggested that location data governance must engage often-overlooked newcomer communities and seek to understand their relationships to location data through time and across geography.

Dragana Kaurin (Localization Lab) spoke about how location data collection disproportionately affects marginalized communities, noting that the same technology that enables many of us to enjoy frictionless travel using conveniences like Uber and Google Maps may enable the surveillance and control of asylum seekers, refugees, and vulnerable populations affected by humanitarian crises. Location data have given government authorities unprecedented power to surveil migrant populations. She described how companies using surveillance data can detect patterns that allow them to predict who may be approaching borders with the intention of crossing. She noted that location data are often misleading because a party of migrants will often share a single cell phone. This can complicate asylum applications because it may suggest that individuals exchanged information with someone they had nothing to do with or were present at a location when they were not. Referencing the case of a Mexican father who told her of his discomfort at being surveilled when visiting his daughter, a U.S. citizen living in Los Angeles, Kaurin illustrated that knowledge that one is under surveillance can have a chilling effect on the movement of people of color.

Sabelo Mhlambi (Harvard University) discussed location data through the lens of colonialism. He cited the work of Sir Francis Galton, who promulgated the idea that you can take a small sample of a population and make extrapolations or predictions to better govern a larger population. This led to the development of a surveillance framework that allowed minority Whites to oppress the Black majority population of South Africa— and later inspired eugenics programs in California and Nazi Germany. Modern technology and location data have allowed methods of monitoring and control that defined historical colonialism to be applied to everyone. In order to use data for good, we must understand why we are collecting data, what we are trying to do with that data, who is benefitting from it, and how we can use such data to address existing inequalities within society.

LOCATION DATA GOVERNANCE

Session moderator **Stephanie Pell** (The Brookings Institution) said that the purpose of the location governance panel was to discuss laws and frameworks that govern location data; to illustrate how these frameworks might apply to the use of location data in the contexts of public health, safety, and law enforcement; and to identify gaps in governance frameworks.[19]

Albert Gidari (Affiliate, Center for Internet and Society, Stanford Law School) said that, from the earliest days of cell phones when location data first became available, law enforcement access has been an issue of concern— and so it is no surprise that these same questions now arise as to how these data might be accessed and disclosed in a public health setting. He discussed the federal Stored Communications Act (SCA)[20] of the Electronic Communications Privacy Act of 1986 (ECPA),[21] which sets out a detailed framework on how the government may acquire the content of stored electronic communications and associated transactional data. The statute restricts online platforms from disclosing content of electronic communications to any party except where a specific exemption applies. However, such providers may disclose transactional information (name, address, phone number, IP address, etc.) to anyone except the government even without notice to or without a user's consent (subject to any privacy policy promises). The government must use certain legal processes (e.g., warrants and subpoenas) to obtain transactional data. Under the SCA, the government can acquire historical location data from such providers by obtaining a court order based on a showing of specific, articulable facts that the information is necessary or material to the investigation of a crime, but needs a search warrant to obtain real time location data.[22] Nothing, however, prohibits the acquisition of location data from the commercial marketplace or the use of data obtained by one federal agency by another. Gidari concluded his remarks by stating that, with location data, there are many undecided issues with regard to when warrants are required for location tracking beyond real-time tracking.

Maneesha Mithal (Wilson Sonsini Goodrich & Rosati), who spent 20 years working for the Federal Trade Commission (FTC), spoke about location data from a regulator's perspective. She noted the potential benefits of the use of location data in the commercial sphere, harms animating regulators' concerns in this area, and regulatory frameworks that apply to the commercial collection and use of data. Location analytics are useful to commercial interests in making determinations about where to build a new store, to carmakers looking to enhance vehicle safety, and to public health officials. In terms of harms, Mithal said that regulators are concerned about safety (e.g., domestic violence victims), revelations of sensitive information about consumers (e.g., tracking of individuals attending protests), covert surveillance, the sale and sharing of data to law enforcement, and issues related to consumer autonomy and choice. Section 5 of the Federal Trade Commission Act, which prohibits "unfair or deceptive acts or practices in or affecting commerce,"[23] offers consumers protections against companies who make deceptive claims about their use of location data or who fail to provide notice and consent before using sensitive geolocation information. However, the FTC does not have the ability to impose penalties on first time violations, which may serve as an under-deterrence for companies—though the agency is in the process of developing a privacy rule that would make it possible for FTC to impose civil penalties. Further, the FTC has recently appreciated that the idea of notice and consent is insufficient and is instead focusing on practices related to data minimization, data security, and purpose limitations.

Margot Kaminski (University of Colorado Law School) discussed the regulation of location data in the European Union (EU). In the EU, regulations take two forms: (1) regulations that apply directly to member states; and (2) directives which act as soft law treaties between member states. Member states may implement directives with some variation. Kaminski identified the General Data Protection Regulation (GDPR)[24] as a regulation relating

[19] Ms. Pell was a member of the workshop planning committee.
[20] 18 U.S.C. §§ 2701–2712.
[21] 18 U.S.C. §§ 2511–2520; 2701–2712; 3121–3127.
[22] This standard pre-dates Carpenter, which requires law enforcement to obtain a warrant for at least seven days of historical location data.
[23] (FTC Act) (15 USC 45).
[24] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 (EU), hereinafter GDPR. See https://gdpr-info.eu/.

to location data and the Law Enforcement (LED)[25] and ePrivacy Directives as directives relating to location data. She also mentioned that the European Charter on Human Rights and European Convention enshrine certain privacy rights. The GDPR is specifically a data protection law that lays out the rules for government and private sector processing of personal data. It provides data protection by default and includes a design requirement that requires companies to consider data privacy principles before they "unleash technology on the public" and to build those principles into technology they design. The GDPR does not, however, cover law enforcement data that deals with criminal investigations (which is covered under the LED). Location data with respect to "electronic communications services" are covered under Article 9 of the ePrivacy Directive. In general, Kaminski said, location data other than what is necessary for communications or for billing can be shared with respect to value added services only when a consumer consents. Further, value added service providers (not the communications providers themselves) can only use that location data to the extent and for the duration necessary for the provision of a value added service.

NEW APPROACHES TO THE GOVERNANCE OF THE USE OF LOCATION DATA

Ohm moderated the workshop's final panel, which was organized to consider new approaches to the governance of the use of location data.

Nathan Wessler (American Civil Liberties Union) described his work as primarily on the dangers of unconstrained access to sensitive digital data and the use of surveillance technologies by law enforcement. The Fourth Amendment is the primary piece of the U.S. constitution that prevents police fishing expeditions and dragnet requests, and constrains abusive attempts by law enforcement and others in government to compile dossiers on past activities. While telecommunications companies are prohibited by law from selling cell phone location data without users' consent, most data brokers can sell location data to whomever they want, including government. While, Wessler said, a regulated service provider shall not knowingly divulge a customer subscriber record to any government agency without a warrant, without consent, etc., if an app company knowingly divulges location data as part of a sale to a data broker, and knows that the government is purchasing that information downstream, are they, he asked, breaking the law? Wessler supports legislation that would require the government to secure a warrant to acquire location data information from a data broker and said that purpose limitations and data minimization requirements should be clear and enforceable and include a private right of action that includes statutory damages.

Terrell McSweeny (Covington & Burling LLP) said that our understanding of the role of data brokers and large data holders in the data ecosystem is evolving. Legislative proposals regarding location data have typically focused on giving consumers more access to data held by data brokers, but this type of protection may be insufficient. Further, it is unclear whether consumers will interact with the data ecosystem in a meaningful way. There is a real need for consumers to understand where their data flows once it has been shared, she said. Some current legislative proposals call for more executive accountability and additional harm assessments. McSweeny reflected on the role of the FTC and suggested that the agency could expand the use of its unfairness authority and work to make sure that consumers have clear information, notice, and choices about how their location information is being collected and used.

Eleni Kosta (Tilburg University) focused her remarks on the regulation of location data in Europe. The most relevant piece of legislation is the EU's ePrivacy Directive, which says, "'location data' means any data processed in an electronic communications network, indicating the geographic position of the terminal equipment of a user of a publicly available electronic communications service."[26] This is a very narrow definition because it refers only to location data processed by the operator that are either made anonymous or have been collected with an individual's consent. Article 9 of the GDPR lists types

[25] See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L0680.

[26] See Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002, available at: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:en:HTML.

of sensitive data and enumerates the conditions under which these data may be processed, but there are many instances where location data reveal potentially sensitive information.[27] The European Data Protection Board and European Commission have issued numerous guidelines regarding the processing of health data for research purposes and these clarify issues related to consent and sources of data, how they are collected, and whether they are necessary for the performance of a task in the public interest. If it is not possible to obtain consent under the ePrivacy Directive and not possible for electronic communication service providers to process data that preserves anonymity, Article 15 of the directive permits, on an exceptional basis, the processing of location data. Further, the article allows EU member states to introduce legislative measures, for reasons of national and public security (etc.) that would allow the processing of location data.[28]

The workshop's final panelist, **Neil Richards** (Washington University in St. Louis), spoke about: (1) what privacy means; (2) the failure of existing ideas about privacy (and in particular location privacy); and (3) what to do. "We should," Richards said, "have substantive rules governing private gathering of location information to promote human values, and these rules should operate, either by direct regulation or by structuring the incentives for self-regulation in a manner similar to the way that the Common Rule does for university institutional review boards." To do this, we need to build trust in the digital environment in the information economy. Richards said that we need to think critically about the need for a consumer protection law for a digital age, noting that there is a lot of evidence that surveillance regimes, particularly location surveillance regimes, apply with the greatest frequency and most heavy-handed force to marginalized and disadvantaged communities. As such, we need to keep these communities in mind as we make rules for society.

CONCLUDING THOUGHTS

The workshop ended with the planning committee offering their reflections on what they had heard over the course of the workshop.[29]

Buckee remarked upon the fact that, though location data is quite niche in terms of the data itself and what people do with it, a rich and complex landscape of expertise is needed to provide a holistic understanding on a range of issues: "We need…conversations, where we have people from across the spectrum with expertise in different fields." She welcomed the workshop as a first step towards trying to make some headway on location data and as a template for how we can think about moving conversations about privacy forward in location data and other arenas.

Ohm said that there are significant upsides and worrisome downsides to the state of location data. He said there will be a need for balance, especially regarding public health. "We need to protect privacy ***and*** unleash the kind of data and research we've been talking to," he said. Ohm also noted that, while there are promising technologies that might allow us to do both, we need better laws for data protection. Interdisciplinary engagement is hard, he said, and prior to the workshop, there were not a lot of preexisting ties between disciplines. Part of the revelation of the workshop, he said, was that we each found another tribe on the other side with whom we want to continue having conversations.

Pell said that, because all these data sets are available and because law enforcement is buying data sets, pressure is being put on the Fourth Amendment. "We do not know," she said, to what extent the Fourth Amendment controls the buying of data by law enforcement or other parts of the federal government (and state and local governments)." Our regulatory mechanisms, she said, are lacking, and we have—notwithstanding the workshop's rich discussion—much work to do.

Vembar said that elements of technology need to merge with legal and regulatory frameworks. The implementation of powerful resources like differential privacy change the ways that we can think about how we apply privacy rights. The problems ahead require an incredible amount of technical, regulatory, and ethical consideration. We have to be able to use this data in the right way, he said. "We can't walk away from the power of it because it is scary and difficult."

[27] E.g., those that reveal racial ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership. See https://gdpr-info.eu/art-9-gdpr/.
[28] See https://gdpr-info.eu/art-15-gdpr/.
[29] Planning Committee Member Jasmine McNealy was unable to participate in this session.

De Montjoye expressed gratitude for meaningful discussion between data practitioners, privacy experts, and community engagement specialists. He said that he does not believe that we need to use the data less to get privacy. On the contrary, he suggested that, if we use the right tools, we can get more privacy and more research using this data. He suggested that standards for data release (as practitioners think about them, e.g., in terms of granularity) will not be possible from a privacy perspective because of the level of sophistication of attacks we need to prevent. De Montjoye noted that even perfect anonymity is insufficient to prevent all risk. As a result, he said, there is a real need for oversight of released data and a better understanding of how they could be misused.

COMMITTEE ON LOCATION DATA IN THE CONTEXT OF PUBLIC HEALTH, RESEARCH, AND LAW ENFORCEMENT: AN EXPLORATION OF GOVERNANCE FRAMEWORKS—A WORKSHOP **CAROLINE BUCKEE** (Co-chair) (Harvard T.H. Chan School of Public Health); **PAUL OHM** (Co-chair) (Georgetown University Law Center); **YVES-ALEXANDRE de MONTJOYE** (Imperial College London); **JASMINE E. MCNEALY** (University of Florida); **STEPHANIE PELL** (The Brookings Institution); **NAVIN VEMBAR** (Atlassian)

STAFF **STEVEN KENDALL,** Program Officer; **DOMINIC LOBUGLIO,** Senior Program Assistant

DISCLAIMER This Proceedings of a Workshop—in Brief was prepared by **STEVEN KENDALL** and **DOMINIC LOBUGLIO** as a factual summary of what occurred at the workshop. The statements made are those of the rapporteur(s) or individual workshop participants and do not necessarily represent the views of all workshop participants; the planning committee; the Robert Wood Johnson Foundation; or the National Academies of Sciences, Engineering, and Medicine.

REVIEWERS To ensure that it meets institutional standards for quality and objectivity, this Proceedings of a Workshop—in Brief was reviewed by **NAVIN VEMBAR,** Atlassian. **MARILYN BAKER,** National Academies of Sciences, Engineering, and Medicine, served as the review coordinator. We also thank staff member **LYNETTE MILLETT** for reading and providing helpful comments on this manuscript.

SPONSOR Support for this workshop was provided by the Robert Wood Johnson Foundation.

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2022. *Location Data in the Context of Public Health, Research, and Law Enforcement: An Exploration of Governance Frameworks, Proceedings of a Workshop—in Brief.* Washington, DC: The National Academies Press. https://doi.org/10.17226/26645.

For additional information about the workshop planning committee, visit https://www.nationalacademies.org/our-work/location-data-in-the-context-of-public-health-research-and-law-enforcement-an-exploration-of-governance-frameworks-a-workshop. For additional information about the Committee on Science, Technology, and Law (CSTL), visit https://www.nationalacademies.org/cstl/committee-on-science-technology-and-law.

Policy and Global Affairs
Committee on Science, Technology, and Law

NATIONAL ACADEMIES *Sciences Engineering Medicine*

The National Academies provide independent, trustworthy advice that advances solutions to society's most complex challenges.
www.nationalacademies.org