Privacy in the Digital Age: Ethical data collection and storage systems for use in the public domain

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Zachery Key

Spring 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Rider W. Foley, Department of Engineering and Society

Introduction

Mobile technology is generating data at never-before-seen rates. A 2013 technological study estimated that the average smartphone user generates a staggering 60 gigabytes of data every year, a figure that has since been doubled (Talbot, 2013; Desjardins, 2019). Obvious data producing culprits include text messages, pictures, social media posts and internet queries, the latter two of which are commonly used as indicators of societal trends. Another valuable source of social insight are the mobility patterns contained in geographic data collected from cellular towers and location services. The analytic power contained in this geographic mobility data has been understood for some time, materializing itself in applications ranging from traffic forecasting to urban planning and epidemic control (Zhao, 2016). These benefits have a great potential to affect change in a number of public policy scenarios but often remain untapped sources of potential for government agencies lacking designated data and information management departments (Chawda, 2017).

High resolution, high frequency characteristics of mobility data present a barrier to entry for many organizations unequipped to deal with such sensitive personal information. The features of time and location that define a person's whereabouts are very unique, making it quite easy to trace their movements with minimal background knowledge (de Montjoye, 2013). Further complicating the design of this system is the antithetical relationship shared by privacy and utility in information systems (Chawla, 2005). In order to address these conflicting design principles, it is important to consider what level of risk an organization is willing to take (with respect to the privacy of its data providers) to gain a greater information potential from higher resolution, individual level data.

In my STS research paper, I investigate the cases of two organizations (LADOT and Cuebiq) which have taken different approaches to database design and development with a contrastive set of privacy protection principles. While Cuebiq is very careful to ensure the

anonymity and integrity of the mobility data products they sell to the public, LADOT's MDS mobility data specification has received a good deal of criticism from various academic and civil rights organizations arguing that the system does not provide an adequate level of protection against re-identification. This discrepancy has been a great source of unrest to many professional technological development analysts and civil liberty organizations prompting allegations of infringement on constituent's right to privacy. By analyzing the LADOT MDS case and comparing their database design decisions to those taken by Cubeiq, I provide clarification to public decision makers as how they can utilize constituents' data in a manner that respects and protects their privacy.

STS Case Context

MDS (mobility data specification) is a database architecture that was designed by LADOT (Los Angeles Department of Transportation) to create a universal data structuring regime for use by rideshare companies operating in large, metropolitan areas. Since its conception in 2019, MDS has been adopted by 130 cities around the world in 8 countries (OMF, 2022). MDS source code and API endpoints are all publicly available on GitHub, free to download and distribute. The project was intended to establish a baseline data format for the collection and storage of large amounts of mobility data from personal mobility devices in response to the growing number of personal transportation devices (e-bikes & scooters) in urban areas. This data system organizes individual level ridership information with details including the location of a device, the duration of a trip, start and end locations, the company name, and the associated trips of a device. For the purpose of this paper, when I refer to the mobility data system, it will be in regard to the specific instance of MDS as implemented by the city of Los Angeles.

Cubeiq is an international mobility data company that collects, stores and aggregates offline data to marketing companies to help them generate insight into their consumers

behaviors. Founded in 2015, it has since grown to 200 employees with offices in the US, Italy and China. Cuebiq collects data from 61 million monthly users in over 180 mobile applications, making it the largest mobility data collection firm in the US (Angelist, 2022). In exchange for access to location information, these apps are paid by Cuebiq when the user elects to participate in the app's services. In addition to their marketing business segment, Cuebiq also provides evacuation mobility data collected around an emergency event. This data is intended to be used to study the mobility patterns of citizens in a particular municipality and contains features such as location, time and device aggregated at the census block group level for any location in the US. For the purpose of this paper, when I refer to Cuebiq data, I will be referencing their mobility evacuation data.

This case study aims to compare and contrast the securitization approaches taken by these two organizations to better understand the public's response to the data design decisions of both systems with regards to data privacy. Through scrutinous analysis of database design, I uncover the values embedded in the design of these database architectures and analyze their compatibility with the principle of privacy. Drawing from Star's Infrastructure theory of technology I will analyze the properties of transparency, embeddedness and embodiment of standards in relation to the design, storage and management of mobility data to determine the valuation each organization places on the data securitization.

STS Theory

What is privacy? Moreover, what does it mean to protect one's privacy? 16th century author Michel de Montaigne presented one of the first modern definitions for privacy by comparing the human experience to existing simultaneously on two different worlds - the frontward facing external world and the internally facing personal world (Stalder, 2002). Montaigne describes the inner world as a "back room" where one can "retreat to reaffirm

their strength and identity" and "reflect on their lived experience". Moreover, Montaigne states, we have the right and responsibility of protecting our inner world "in order to establish our true liberty". This rudimentary description of privacy demonstrates the human need for separation between the public and the private aspects of our lives.

In his historical analysis of privacy, Stalder goes a step further to claim "the notion of privacy … is an unintended consequence of the emergence of a new form of communication: print". He argues that the printing press established a "print-culture" that "favored a one-way communication" in which "the author reveals without being revealed and the reader learns without being learned about". This model of privacy presents societies evolution from an 'manuscript culture' to a 'print culture' as a cultural paradigm shift that established the concept of privacy as we know it today (Stalder, 2002). This radical movement toward individual-level informatics has been attributed to the success of the protestant reform, the onset of the scientific revolution, and the proliferation of scholarly publications in the academic community. The widespread social influence of the printing press can be related to the modern impact of the computers and the internet in an equal but opposite fashion (Dewar, 1998).

Whereas the printing press established a unidirectional, one-to-many relationship in the distribution of information, mobile access to the internet has established a bidirectional, many-to-many relationship offering content consumers the ability to directly engage with content producers. This relationship directly opposes the one-way flow of information created by the printing press and lends itself to a culture that makes it increasing difficult to "reveal without being revealed" or "learn without being learned", forcing us to choose between protecting our privacy or maintaining normalcy in our relationship with society (Stalder, 2002).

Recently, this decision is becoming increasing detached from our own locus of control. Applications and webpages choose what data they will collect from you and you in turn are forced into signing away your personal information rights in a covert clause hidden somewhere in the depths of an infinitely vast privacy agreement (Brown, 2020). Regardless of whether or not we actually agree with the privacy terms and conditions of an application, if we consent to its collection, we maintain the right to understand how our data will be used (Belanger, 2011). But do the same principles apply in the public domain? What responsibility does the government have to disclose this information to constituents?

To understand the values embedded in LADOT's mobility data specification system, I will be utilizing Star's theory of technological infrastructure. Following Star's definition of infrastructure as "a relational property, not something stripped of use" I characterize MDS as a relational construction between the government and their citizens containing the properties of embeddedness, transparency and linkage with conventions of practice (Star & Ruhleder, 1996, p. 113; Star 1999). In the context of my project, MDS can be likened to a socially constructed technology that is embedded in context of the LA transportation system. By applying the property of transparency to MDS, I can evaluate the degree to which the general public understands the value of a data driven optimization method to create a safer and more efficient transportation system while simultaneously addressing the corresponding privacy risks inherent in the collection and storage of this information. Finally, I look at the conventions of practice for data storage that guided the structural design decisions of MDS and compare them to that of the Cubeiq mobility data used in the technical portion of this project.

Research Question and Methods

Despite the rapid growth of data collection and storage systems over the past two decades, very little has changed in the world of data anonymization and obfuscation. This presents a very real threat to our personal security if the databases that hold our information are compromised. The problem only worsens as the level of detail in the data increases: a 2016 study demonstrated that it was possible to uniquely identify 95% of the citizens of New York City with only 4 high resolution spatiotemporal datapoints (de Montjoye, 2013). Moreover, if you can reverse engineer these data points to identify their source signal transmitter, the same points could be traced to reveal visitation patterns to locations such as political rallies, nightclubs, marijuana dispensaries, abortion clinics, LGBTQ centers or rehabilitation centers (Atockar, 2014). Without ever meeting you in real life, a stranger can uncover highly sensitive personal information such as your political affiliation, sexual orientation, substance consumption, and/or daily routines.

In order to prevent such situations from arising, I seek to answer the following research question: How can governments protect constituents' privacy while collecting, storing and analyzing their data?

To address this question, I draw from Star's theory of Infrastructure to uncover the values embedded in Los Angeles Department of Transportation's Mobility Data Specification. I then contrast LADOT's values to those found in Cubeiq's evacuation data, considering both the utility and security of the two datasets. Comparing the common attributes shared between the data tables, I look into the informational value, resolution, and frequency of the data to establish the information collection objectives motivating each project. This analysis will be used to provide evidence for how to anonymize, aggregate and structure mobility data to retain the privacy of data providers.

Subsequently, I explore the issues with the lack of transparency present in MDS drawing from empirical evidence collected from news articles, government publications, and websites. Court rulings and letters addressed to LADOT serve as a proxy for popular opinion to provide an understanding of public concerns with the implementation of a novel government-specified data collection system. This analysis will be used to provide evidence for how to involve communities in the data design and planning process for public data projects.

Finally, I assess the reach and scope of each data project in an effort to understand the potential consequences of various data structure designs. I then present a series of recommendations to policy makers listing a variety of methods that can be used to obtain public consent and protect individual privacy when designing and building novel data analytics platforms.

Results

To respect constituent's right to privacy, governing bodies must obtain consent prior to the collection of their data. Looking at the MDS case, we find evidence that the public values a high level of transparency in the database design build process. In order to increase the security of the databases, public organizations can obfuscate their data by aggregating their records, engineering lower resolution attributes and applying differential privacy methods to their data. Furthermore, database designers can design database systems that periodically delete older records and save data in offline distributed systems to reduce data accessibility in the event of a breach.

To assess the disparities between *organizational* values and *societal* values, we look at the values embedded in the structure of these databases and compare them to the public's reaction upon deployment of the database. Ideally, these two sets of values align with each

other and the values embedded in the database follow those shared by society. In evaluating the organizational structures and level of detail contained in both the MDS and Cuebiq databases we can construct a rough idea of the underlying values held by the two institutions.

MDS Trip Data Structure

Field	Type	Required/Optional	Description
company_name	String	Required	Lime, Bird, etc.
device_type	String	Required	Bike, Scooter, etc.
trip_id	UUID	Required	Index number
trip_duration	Integer	Required	Time, in seconds
trip_distance	Integer	Required	Trip distance, in meters
start_point	Point	Required	Lat/Long Coords
end_point	Point	Required	Lat/Long Coords
route	Line	Optional	Path of Lat/Long Coords
device_id	UUID	Required	Unique identifier for device
start_time	Unix Timestamp	Required	Seconds since Unix epoch
end_time	Unix Timestamp	Required	Seconds since Unix epoch

MDS Parking Data Structure

Field	Type	Required/Optional	Description
device_type	String	Required	Bike, Scooter, etc.
availability_start_time	Unix Timestamp	Required	Seconds since Unix epoch
availability_end_time	Unix Timestamp	Required	Seconds since Unix epoch
placement_reason	String	Required	Reason for replacement
pickup_reason	String	Required	Reason for removal
associated_trips	[UUID]	Optional	List of associated trip ids

In the MDS data schema seen above, we observe an organizational paradigm that follows that of most relational databases: there are a series of attributes that describe a particular aspect of the data (device_type, route, start_time, end_time, etc.), a primary key that uniquely identifies an observation (trip_id) and a set of foreign keys that link the data to other tables (trip_id, device_id, company_name). These links to conventional practice allows the tables to be integrated into existing data systems that also use different variations of SQL. Designing the mobility data tables in this manner creates interchangeability between servers operating on the MDS platform, encouraging collaboration between localities that have adopted the platform. Another benefit to this approach is that it is easy to relate tables to one

another by the process of performing a SQL-style JOIN that merges tables together by a set of one or more common keys. This procedure provides interconnectivity between tables within the database, allowing a user to perform complex queries and combine data from multiple sources such as historical ridership information, parking information and device activation records.

Digging deeper we notice that MDS data has attributes that provide precise information collected at a high frequency: each trip has an associated start time, end time and duration accurate to the nearest second in addition to a start location, end location and route path accurate to the nearest meter. This 'high-resolution' data presents the user an excellent basis for analysis: even if the data is not collected in real time, the brevity of the average rideshare trip means that that new location information is collected and stored in short intervals that seldom last longer than an hour, providing a good approximation for an individual's current location.

Cubeiq Destination Index Data Structure

Field	Type	Required/Optional	Description
local_date	String	Required	Month, Day, Year
start_census_geoID	Integer	Required	Census block source location
destination_index	Float	Required	% of block that went to dest
dest_country	String	Required	2 Letter Country Abbreviation
dest_state	String	Required	2 Letter State Abbreviation
dest_county_geoID	Integer	Required	County FIPS Number

Cubeiq Evacuation Index Data Structure

Field	Type	Required/Optional	Description
local_date	String	Required	Month, Day, Year
start_census_geoID	Integer	Required	Census block source location
evacuation_index	Float	Required	% of block that left
max_distance_from_start	Float	Required	Average distance, in miles

Similarly, to LADOT's MDS, Cubeiq's mobility data is structured as a relational database that is comprised of a set of tables: one for destination index (flow of individuals from a US census block group to a county anywhere in the US) and one for evacuation index (ratio of individuals spending the night outside their home census block groups). Unlike MDS, each Cubeiq data observation aggregates its statistics on a census block group level and records measurements daily, obscuring detail and increasing the difficulty of deidentifying the data and tracing it back to its source signal. Cuebiq takes additional steps to abstract away any auxiliary information about the individuals living in a particular census block: while sources are given at the census block group level, destinations are given on the county level. The mobility data is then further anonymized by applying a differential privacy algorithm which introduces random noise to a subsample of the dataset. This approach, combined with census block group aggregation and date-time binning produces a lower-resolution, noisier output when compared to MDS.

Per Star's theory of infrastructure, technological systems contain values embedded within the context in which they were created. In LADOT's case, the primary motivation for MDS is for system optimization which depends on the collection and analysis of high-resolution data. This information provides a very high level of detail that is uniquely traceable to an individual operating within the system, suggesting that LADOT has designed MDS to prioritize data utility. In a complementary fashion, Cuebiq data has been anonymized, aggregated and obfuscated with the addition of noise suggesting a higher valuation of privacy. Now that we understand the organizational values embedded in database design, we turn our attention towards the LADOT - MDS case and evaluate the public's response (to MDS) in an effort to understand the social value of data privacy in comparison to that of data utility.

In May of 2018, LADOT presented its dockless scooter mobility collection plan to the LA city council, calling for the expansion of the city's existing e-scooter/e-bike program and the universal adoption of their novel mobility data specification (MDS) from all personal mobility providers (Reynolds, 2019). This project was intended as a solution to address mobility issues around the city of Los Angeles by increasing accessibility and speeding up commute times. LADOT's new state-of-the-art program was to be the largest in the country, built on the promise of providing equitable access to transportation for low income and disadvantaged communities (LADOT, 2019). In February of 2019, after many months of planning, reviewing permits and obtaining permissions from the city council, the program received permits from Lime, Lyft, Ride and others to deploy 22,500 additional scooters to the city's sidewalks, bringing the total to 37,000.

Less than 2 months later the Electronic Frontier Foundation (EFF) published a 12-page letter addressed to LADOT detailing a lengthy list of privacy concerns about the deployment of the novel Mobility Data Specification. Their primary complaint cites the violation of citizen's privacy interests as a result of the collection of individual-level rider data (Williams, Cyphers, Sheard, 2019). According to EFF, the collection and storage of such data is in violation of California Consumer Privacy Act and directly contradicts the ruling of Carpenter vs. US (2018) which stated that government agencies must obtain a warrant prior to accessing historical location data. To add insult to injury, LADOT's violated their own data protection principles just weeks before they were instated by allowing third party organization Remix access to raw trip data (Grass, 2018).

Fast forward to March of 2020. After being threatened with suspension for their refusal to share rideshare mobility data, Uber filed a request for hearing against LADOT. Uber's case was quickly dismissed by the California Court System who claimed that there was insufficient evidence of LADOT inappropriately utilizing personal information. After a

failed attempt to appeal, Uber quietly accepted the terms and conditions of the agreement and within the month had sold the entirety of their rideshare platform to competitor Lime (Carey, 2020). Three days prior to the Uber lawsuit, the Center for Democracy in Technology (CDT) wrote a letter to the United States Department of Transportation in a desperate attempt to engage the federal government after having been largely ignored by LADOT after voicing concerns four months prior. Despite their efforts, the outcome was no different - the federal government remained silent.

Frustrated by the failure of the Uber lawsuit and the indifference of LADOT towards the mounting concerns of technology advocacy organizations, the American Civil Liberties Union filed their own suit against the organization in June of 2020. They argued that the collection of mobility data constituted an illegal search and seizure of personal information in direct violation of the 4th amendment to the Constitution (Sanchez v. LADOT, 2020). But upon reaching the hands of Central District Court of California, their case was also dismissed with the court claiming that there was “no reasonable expectation of privacy” of the plaintiff for the MDS system and that “the program’s data collection was reasonable in light of the cities goals” (Albert, 2021).

Despite numerous lawsuits, protests and letters of opposition, LADOT was able to carry out their plans to collect, distribute and analyze publicly collected rideshare mobility data. Yet, LADOT’s path to success has been fraught with challenges. From delaying production to narrowly dodging legal action, MDS has cost LADOT a lot of time and money and resulted in a decrease in trust among their constituents. The resulting controversy has caught the public’s eye and left a bitter taste in the mouths of many who do not consent to the collection of their personal information.

On the flip side of the coin, Cuebiq has been publicly applauded for their data anonymization and securitization efforts. In a 2019 business wire to their NYC office, UNICEF stated, “Cuebiq’s location data is a privacy-compliant data source that enables anonymous human mobility analysis, at scale, while protecting the privacy of vulnerable populations”. The organization’s privacy sensitive approach to data handling has awarded then membership to the National Advisory Initiative. By anticipating reidentification efforts and designing data structures to counteract adversarial attacks, Cuebiq has set a high standard that has resulted in them becoming the largest mobility data provider in the United States.

One of the simplest and most effective data securitization methods Cuebiq employs is the reduction of data resolution through aggregation, or combining data into larger units of analysis. Temporal data sources can be aggregated by reducing the frequency of collection and binning into larger time periods (i.e. summarizing second level data into hourly or daily measurements). In a similar manner, Cuebiq combines and condenses high resolution attributes into lower resolution attributes by feature engineering, or producing irreversible summary statistics about data. An example of feature engineering would be Cubeiq’s evacuation index, which is calculated as the ratio of people that left an area to the total number of people living in that area. This aggregate statistic can then be used in place of the other two (total number of people and number of people that left) to provide a proxy for evacuation behavior and increase the difficulty of reidentification.

Outside of data aggregation and feature engineering, Cubeiq protects users’ privacy by limiting the scope of access to the collected data. By storing their data in an offline database, Cuebiq restricts accessibility to only include devices operating in their local network. Finally, Cuebiq utilizes differential privacy methods to introduce random noise into the dataset. This process creates uncertainty in which records are actual observations and

which have been added in artificially by replacing a small subset of true values with pseudo data. By implementing this approach, we reduce the absolute information potential of any single record by introducing variability to the veracity of individual observations.

Discussion

Expanding on the research of Bertino and Sandhu (2005), this work relates to the broader issues of privacy and security in relation to database design. While prior work has diligently looked into the securitization of databases prior to an adversarial attack, I consider the reidentification potential of the raw data itself by analyzing the values embodied in the database design. To build a secure public database, we must ensure that the data cannot be reidentified by removing identifying attributes, aggregating statistics and introducing noise into the dataset. Beyond database securitization, we encourage governments to engage in open and transparent discussions with their constituents about the intended use of their data, especially when containing sensitive information such as location.

Project limitations are mainly directed at the functional differences between the two datasets used in analysis. Even though both are publicly available mobility datasets, they serve different purposes: Cuebiq for evacuation analysis and MDS for ridership analysis. While there are a number of common attributes (location, time, device counts, etc.), there are also many unique attributes which limit the extent to which we can perform a head-to-head comparison of the datasets. Furthermore, the critiques offered against the MDS data do not account for the high utility of the data as compared to Cuebiq's lower resolution, lower utility data.

The results of this project will be used to inform the manner in which I design data collection and storage systems. Specifically, I will ask myself how to design attributes that maximize data security while meeting the minimal data utility criteria. I shall be deliberate in

my design process and make sure to clearly communicate with stakeholders the benefits and risks of deploying a data analytics system. During the design and build phase of the database lifecycle, I will carefully consider the accessibility requirements of the data system and assess the practicality of offline storage or periodic deletion to minimize the risk and damages assumed in the event of a breach.

Conclusion

The broader significance of this report is to provide public data management figures with a framework with which they can analyze the values embodied in data systems they want to deploy. I urge data decision makers to be careful in how they design and construct databases and data software systems by recounting the tale of LADOT's MDS controversial data system. I then evaluate the government – constituent relationship with regards to data collection, recognizing the innate tension between the organizational value of acquisition of information and the social value of protection of privacy. Recognizing this value differential, we seek to bridge the gap by proposing a set of privacy enhancement techniques that allow for an appropriate level of detail in analysis.

Building upon the findings of this paper, I suggest future researchers look into quantitatively comparing the securitization measures presented in the results to determine the optimal set of methods to employ for deidentification. By evaluating the computational probability of reidentification, we can assign performance metrics to anonymization algorithms and rate them on a common set of criteria. It would also prove beneficial to look into the cases of other publicly deployed data collection systems to provide a better grasp on the standards of practice employed by other public institutions. Increasing the scope of analysis, we may find that there are a common set of privacy enhancement practices implemented in the public sector (MDS) which differ from the private sector (Cuebiq). Future

efforts should also involve a deeper investigation into the utility-privacy tradeoff to balance the singular emphasis on data privacy observed in this paper.

To ethically design public data systems, it is important that we must consider all system stakeholders. We must be open and transparent in how we collect people's data and what we intend to use it for. Neglect to involve constituents results in public distrust and may result in legal action. To avoid these negative consequences, I suggest conducting a thorough investigation of the intended use-case of the data system. In doing so you can design minimally invasive data structures that only contain absolutely essential attributes and reduce the risk of reidentification. Finally, I would recommend engaging citizens in the design build process by collecting their feedback at major stages of production to ensure alignment with the public agenda.

Sources Cited

Albert, K. (2021, August 3). *Clinic files technologists' brief in Ninth Circuit Scooter Privacy Case*. Cyberlaw Clinic. Retrieved February 16, 2022, from https://clinic.cyber.harvard.edu/2021/08/03/clinic-files-technologists-brief-in-ninth-circuit-scooter-privacy-case/

Altman, M., Wood, A., O'Brien, D., Vadhan, S., & Gasser, U. (2016, May 16). *Towards a modern approach to privacy-aware government data releases*. SSRN. Retrieved March 24, 2022, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2779266

Atockar. (2014, September 16). *Riding with the Stars: Passenger privacy in the NYC taxicab dataset*. Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset. Retrieved February 5, 2022, from https://agkn.wordpress.com/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/

Bélanger, F., & Crossler, R. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, *35*(4), 1017–1040. https://doi.org/10.2307/41409971

Bertino, E., & Sandhu, R. (2005). Database security - concepts, approaches, and challenges. *IEEE Transactions on Dependable and Secure Computing*, *2*(1), 2–19. https://doi.org/10.1109/tdsc.2005.9

Brown, S. (2020, June 29). *Privacy isn't a right you can click away*. Wired. Retrieved February 25, 2022, from https://www.wired.com/story/privacy-isnt-a-right-you-can-click-away/#:~:text=When%20you%20sign%20away%20your%20privacy%20rights%20to,it%20sold%20to%20law%20enforcement%20around%20the%20country.

Carey, C. (2020, July 1). *How Los Angeles took control of its mobility data*. Cities Today. Retrieved February 5, 2022, from https://cities-today.com/how-los-angeles-took-control-of-its-mobility-data/

Carpenter v. United States, (2018) 138 S. Ct. 2206, 201 L. Ed. 2d 507

Chawda, V. (2018, June 21). *How is the government using data? how should it?* Forbes. Retrieved March 4, 2022, from https://www.forbes.com/sites/kpmg/2017/06/12/how-is-the-government-using-data-how-should-it/?sh=4157d1cc64e6

Chawla, S., Dwork, C., McSherry, F., Smith, A., & Wee, H. (2005). Toward privacy in public databases. *Theory of Cryptography*, 363–385. https://doi.org/10.1007/978-3-540-30576-7_20

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, *3*(1). https://doi.org/10.1038/srep01376

Desjardins, J. (2019, April 17). *How much data is generated each day?* World Economic Forum. Retrieved March 5, 2022, from https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

Dewar, J. A. (1998, January 1). *The information age and the printing press: Looking backward to see ahead*. RAND Corporation. Retrieved February 23, 2022, from https://www.rand.org/pubs/papers/P8014.html

Gang-Hoon Kim Electronics and Telecommunications Research Institute, Kim, G.-H., Institute, E. and T. R., Nebraska-Lincoln, S. T. U. of, Trimi, S., Nebraska-Lincoln, U. of, Profile, U. of N.-L. V., Ji-Hyong Chung Electronics and Telecommunications Research Institute, Chung, J.-H., Machinery, A. for C., Contributor MetricsExpand All Ganghoon Kim Electronics Telecommunication Research Institute Publication Ye, & Ganghoon Kim Electronics Telecommunication Research Institute Publication Years2014 - 2014Publication counts1Available for Download1Citation count332Downloads (cumulative)14. (2014, March 1). *Big-data applications in the government sector*. Communications of the ACM. Retrieved March 24, 2022, from https://dl.acm.org/doi/pdf/10.1145/2500873

Grass, M. (2018, November 9). *Los Angeles inks new data-sharing agreement with Scooter and Bikeshare Companies*. Route Fifty. Retrieved February 16, 2022, from https://www.route-fifty.com/smart-cities/2018/11/ladot-data-sharing-agreement-scooters-bicycles/152727/

Justin Sanchez et al v. Los Angeles Department of Transportation et al, (2020) C.D.Cal.

LADOT. (2019, February 15). *LADOT receives permit applications for over 37000 dockless on demand scooters and bikes*. LADOT. Retrieved February 5, 2022, from https://ladot.lacity.org/sites/default/files/press-releases/ladot-receives-permit-applications-for-dockless-.pdf

LADOT. (2019, March 22). *Ladot expands dockless scooter and bicycle program to be ...* LADOT . Retrieved February 5, 2022, from https://ladot.lacity.org/sites/default/files/press-releases/press-release-ladot-expands-dockless-scooter-and-bicycle-program-to-be-largest-in-country.pdf

LADOT. (n.d.). *Mobility Data Specification (MDS)*. LA City - LADOT. Retrieved February 5, 2022, from https://ladot.lacity.org/sites/default/files/documents/what-is-mds-cities.pdf

Open Mobility Foundation (OMF). (2022, January 28). *Who is using MDS?: Open Mobility Foundation*. Open Mobility Foundation | OMF. Retrieved March 5, 2022, from https://www.openmobilityfoundation.org/mds-

Papathanasiou, J., & Kenward, R. (2014, February 7). *Design of a data-driven environmental decision support system and testing of stakeholder data-collection*. Environmental Modelling & Software. Retrieved March 24, 2022, from https://www.sciencedirect.com/science/article/pii/S1364815214000358

Reynolds, S. J. (2018, May 18). DOCKLESS BIKE/SCOOTER SHARE PILOT PROGRAM. Los Angeles, California. Retrieved February 16, 2022, from http://clkrep.lacity.org/onlinedocs/2017/17-1125_rpt_DOT_05-18-2018.pdf

Reynolds, S. J. (2019, March 22). *LADOT Data Protection Principles* . LADOT. Retrieved February 5, 2022, from https://ladot.io/wp-content/uploads/2019/03/LADOT_Data_Protection_Principles-1.pdf

Stalder, F. (2002). The failure of privacy enhancing technologies (pets) and the voiding of privacy. *Sociological Research Online*, *7*(2), 25–39. https://doi.org/10.5153/sro.718

Talbot, David. (2013). Big Data from Cheap Phones. *MIT Technology Review: June 2013.* MIT. Retrieved March 4, 2022 from https://www.technologyreview.com/featuredstory/513721/big-data-from-cheap-phones

Teale, C. (2020, February 13). *Ladot wins appeal in data-sharing dispute with uber*. Smart Cities Dive. Retrieved February 5, 2022, from https://www.smartcitiesdive.com/news/los-angeles-ban-jump-dockless-bikes-scooters-upheld/572266/

UNICEF. (2019, September 10). *Cuebiq's Data for Good Program Provides UNICEF with High-Precision Human Mobility Data for Real-Time Response to Humanitarian*. Bloomberg.com. Retrieved April 21, 2022, from https://www.bloomberg.com/press-releases/2019-09-10/cuebiq-s-data-for-good-program-provides-unicef-with-high-precision-human-mobility-data-for-real-time-response-to-humanitarian

Williams, J., Cyphers, B., Sheard, N., & Thompson, A. (2019, April 3). Urgent Concerns Regarding the Lack of Privacy Protections for Sensitive Personal Data Collected Via LADOT's Mobility Data Specification. Los Angeles, California.

Zhao, K., Tarkoma, S., Liu, S., & Vo, H. (2016). Urban human mobility data mining: An overview. *2016 IEEE International Conference on Big Data (Big Data)*. https://doi.org/10.1109/bigdata.2016.7840811