

**BEFORE THE PUBLIC UTILITIES COMMISSION
OF THE STATE OF CALIFORNIA**

*Order Instituting Rulemaking on
Regulations Relating to Passenger
Carriers, Ridesharing, and New Online-
Enabled Transportation Services.*

Rulemaking 12-12-011
(Filed December 12, 2012)

**OPENING COMMENTS OF UC DAVIS POLICY INSTITUTE FOR
ENERGY, ENVIRONMENT, AND THE ECONOMY ON DATA
CONFIDENTIALITY, COLLECTION, AND SHARING ISSUES**

December 17, 2019

Austin Brown, Executive Director
Policy Institute for Energy, the Environment, and the Economy
1615 Tilia St, Davis, CA 95616
E-mail: dokbrown@ucdavis.edu

**BEFORE THE PUBLIC UTILITIES COMMISSION
OF THE STATE OF CALIFORNIA**

*Order Instituting Rulemaking on
Regulations Relating to Passenger
Carriers, Ridesharing, and New Online-
Enabled Transportation Services.*

Rulemaking 12-12-011
(Filed December 12, 2012)

**OPENING COMMENTS OF UC DAVIS POLICY INSTITUTE FOR
ENERGY, THE ENVIRONMENT, AND THE ECONOMY ON DATA
CONFIDENTIALITY, COLLECTION, AND SHARING ISSUES**

I. Introduction

The University of California – Institute of Transportation Studies is properly suited to store TNC data going forward. Academic universities are well situated to maintain the balance between the need that the data remain confidential among competitors and the need that it be utilized in critical policy research.

II. Comments

**a. The University of California – Institute of Transportation Studies
is an appropriate organization to host TNC data**

Within the Commission’s proceeding *Regulations Relating to Passenger Carriers, Ridesharing, and Online-Enabled Transportation Services* (R.12-12-011), the Policy Institute For Energy, Environment, and The Economy (the Policy Institute) seeks to respond to the request for comments in the October 25, 2019 *Amended Phase III. C. Scoping Memo and Rule of Assigned Commissioner*, specifically as to Issues 2.2.2(7), 2.2.3(5), and 2.5(8) regarding whether TNC information should be hosted by a third-party entity, such as an academic institution.

The Policy Institute would like to respectfully recommend that the University of California - [Institute of Transportation Studies](#) (“UC ITS”) be considered as the third-party that stores TNC data. UC ITS is a network of research faculty from across four of the UC campuses: Berkeley, Davis, Irvine, and Los Angeles. UC ITS was established by the legislature in 1947, and is dedicated to advancing the state of the art in transportation engineering, planning, and policy for the people of California.

The UC is the research body of the state of California, and the UC ITS is its transportation research arm. The CA Legislature allocates \$5 million annually to the UC ITS to support transportation research that is a benefit to the state and there are dozens of research projects that serve state objectives. Furthermore, there is a history of research partnerships between the UC ITS and Uber, Lyft, Ford, Toyota, and many other transportation companies.

Recently the UC ITS developed the Innovative Mobility Initiative (IMI) which will be directed by Professor Susan Shaheen, Co-director the Transportation Sustainability Research Center (TSRC) of the Institute of Transportation Studies at UC Berkeley. Professor Shaheen also serves on the Society of Automotive Engineer’s (SAE) [Mobility Data Collaborative](#). The IMI will leverage considerable intellectual capacity across the four campuses to serve as a direct partner to the State of California. IMI plans to develop a cohesive research plan across the campuses and provide policy guidance and direct technical assistance to constituent governments. (More information on projects conducted by the UC ITS research network is in Section V).

a. Challenges for Mobility Data Collection and Analysis

The goal of a TNC data repository will be to balance concerns about privacy, or proprietary interests, against the potential for community benefits. Academic research institutions, like the UC ITS and the IMI, are well situated to strike the appropriate balance. Academic research institutions can hold the data in trust, and determine when it is appropriate to disseminate raw, anonymized, or analyzed data to appropriate state and city governments.

i. Privacy Issues

Mobility data is challenging to anonymize. Fine-grained mobility data—such as trip-level data containing the longitude and latitude of trip origin and destination—is often personally identifiable, even when time and identity information is removed. This is especially true of data on origin-destination pairs that originate from granular trip traces, such as GPS traces of individual riders on mobility apps, or traces from cell-phone GPS data. Over multiple trips, granular trace data can easily be linked to an individual person by evaluation of common travel patterns (e.g., home-to-work routes). This raises privacy concerns for individuals using mobility services, including ride providers (for ridehailing services) as well as riders. Anonymity is generally less of a concern for data that represent aggregated trips at the census-tract level or larger. However, these levels of aggregation can blunt data utility. There is a tension between achieving aggregation sufficient to protect privacy and preserving enough specificity that the data remain useful.

ii. TNC’s Propriety Concerns

Private companies have proprietary, financial, and fairness concerns about sharing data, especially data captured in highly competitive markets. Proprietary concerns could arise if, for example two companies are responsible for the bulk of ridehailing trips in a given region. Requiring these companies to share their trip data publicly could enable each to study their main competitor’s trips and possibly gain a competitive advantage from doing so. Financial concerns arise from the high costs of data collection and storage, which may make private organizations more reticent to share data without reasonable compensation. Fairness concerns arise from the possibility of increasing data-reporting requirements for mobility providers without imposing similarly stringent requirements on the automakers or private motorists who account for a larger share of travel. Unbalanced data-reporting requirements could also hinder development and adoption of new and sustainable mobility alternatives—either by discouraging automakers from venturing into new mobility markets, increasing costs of mobility services, and motivating consumers to stick to conventional transportation options for which privacy concerns do not appear to be as salient.

b. USDOT as Exemplar Third-Party Platform for Mobility Data

Several agencies of the U.S. Department of Transportation (USDOT) are already coordinating and securing large data reserves. The USDOT's Secure Data Commons (SDC) is an exemplary model of how to create a third-party platform for storing and analyzing sensitive data. The SDC engages two main types of users: *data providers* and *data analysts*. Data providers voluntarily add data however they deem appropriate. Providers can contribute data in real time, schedule regular batch uploads of historic data, or upload data on an occasional, ad-hoc basis. The SDC grants different levels of access to different types of data analysts based on criteria defined by data providers. Providers also set parameters for whether datasets (or data analyses conducted on the SDC platform) can be exported.

Data analysts must file an application and sign a data-sharing agreement to be granted access to the SDC. If approved, analysts will have access to a set of datasets tailored specifically to the level of access they are granted, as well as to statistical tools such as R, Python, and SQL. The SDC only includes two current mobility programs (each with several datasets): the Waze Connected Citizens Program (CCP) and the Connected Vehicle Pilot Deployment Program (CVPDP).

The goal of the Waze CCP is to analyze Maryland traffic data provided by Waze (a subsidiary of Google), alongside crash data provided by the State of Maryland, in order to evaluate when and where traffic accidents are most likely to occur. While the goal of this project is admirable, it is still relatively narrow in scope. There are undoubtedly myriad other lessons that could be learned from Waze datasets. But because SDC data providers are able to restrict data use to specific purposes, researchers have limited freedom to pursue new lines of inquiry.

The Connected Vehicle Pilot Deployment Program (CVPDP) houses data for connected-vehicle pilots occurring in New York City, Wyoming, and Tampa, FL. Like the Waze CCP, the CVPDP includes a very specific objective: "to support the independent evaluation of CV [connected vehicle] Pilot deployments." Independent evaluators who can receive access to the data include Noblis, the Texas A&M Transportation Institute, TTI and Volpe. Select datasets involved in these pilots (and a few other pilots including Virginia

and San Diego) are also available for public access on a different USDOT platform, the Intelligent Transportation Systems (ITS) Data Hub. The publicly available data includes safety message visualization data, road weather demonstration data, vehicle awareness device data, and several other datasets determined not to include personally or commercially sensitive information.

Determining which data should be made publicly available and which requires access restrictions to protect individual privacy and proprietary interests is at the core of the dialogue around mobility data sharing. Another example of a USDOT program that walks this line carefully and effectively is the Transportation Secure Data Center (TSDC), maintained by the National Renewable Energy Laboratory (NREL) through a partnership between the USDOT and the U.S. Department of Energy. The TSDC aggregates data from travel surveys and studies—including household travel surveys and data collected from GPS—into a single, publicly available repository. NREL converts the data into an anonymized and consistent format prior to publication. NREL also grants researchers and other stakeholders access to the datasets for specific purposes, such as research. This limited-access model has enabled TSDC data to study transportation-system emissions, examine travel patterns, and plan alternative fuel stations.

The USDOT is also developing another data portal, the Data for Automated Vehicle Integration (DAVI), as a “multimodal initiative to identify, prioritize, monitor, and – where necessary – address data exchange needs for automated vehicles (AV) integration across the modes of transportation.” The DAVI website emphasizes that for automated vehicles, data will enable safer and more efficient “integration of AVs into the transportation system.” A recent example of where USDOT is expanding the use of data repositories is the spring 2019 release of a notice for funding opportunity for Automated Driving Assistance (ADS) demonstrations. Applicants are required to share demonstration data via an existing USDOT data repository or via a “third-party system where USDOT analysts can conduct their work.”

c. Examples of UC Data Projects

The University of California system has several existing data repositories. UC Irvine has a machine learning repository that currently maintains 488 datasets through a searchable interface. Additionally, UC Merritt is a repository service from the University of California Curation Center (C3) that lets the UC community manage, archive, and share its digital content. The UC system would likely be able to create or use an existing data repository with adequate usability, features, and formats.

UC ITS itself is engaged in several existing mobility data projects. The most pertinent research project is SAE International's Mobility Data Collaborative, which UC Berkeley ITS is supporting. The Collaborative seeks to solve the problem of managing access to infrastructure by providing an open and impartial forum for sharing information and developing best practices. Membership of the collaborative currently includes Bird, Lyft, Uber, the City of Bellevue, the City of Tallahassee, Denver Regional Council of Governments, and the Transportation Sustainability Research Center at UC-Berkeley, among others. The Collaborative seeks to develop consistent mobility performance metrics and provide consistency and clarity to the area of data privacy.

Finally, UC ITS primarily conducts research. There are many [UC ITS research projects](#) that include partnerships with state and municipal agencies and ride-sharing companies. One example is a recent paper out of UC Irvine ITS *An Analysis of Travel Characteristics of Carless Households in California*.¹ This research paper, in partnership with Caltrans, Orange County Transportation Authority, City of Irvine Planning Department, and City of Santa Ana Planning Department analyzed trip data from the 2012 California Household Travel Survey (CHTS) to understand the travel behavior of households that voluntarily decided to forgo cars and of involuntary carless households, to inform policy and create greater awareness, as these households are at greater risk of physical isolation, poor access, and social exclusion.

III. Conclusion

¹ UC Irvine ITS *An Analysis of Travel Characteristics of Carless Households in California*, <https://www.ucits.org/research-project/carless-hh-travel/>.

New mobility data can offer much for city, regional, and state planners. But there remain significant challenges for the collection and sharing of mobility data among public and private partners. Much valuable data is held by the private sector, which may be reticent to share data with public actors, and for legitimate reasons. Meanwhile, many public officials without access to key data struggle to make informed decisions. However, oversharing and under-sharing mobility data are both problematic and stakeholders need a policy framework to find a middle-ground approach. UC ITS could aid the State of California in striking this middle-ground approach.

If the UC ITS is approved to be the third-party entity that stores TNC data, additional issues will need to be addressed, including with whom the UC ITS is permitted to share data, and at what level of granularity.

We look forward to working on the many important and complex issues that will be addressed in this proceeding.

Dated: December 17, 2019

Respectfully submitted,

Austin Brown

Executive Director

UC Davis Policy Institute for Energy, the Environment and the Economy

Tel: 650-906-5586

E-mail: dokbrown@ucdavis.edu