

AGRION INDUSTRY SNAPSHOTS

SMART CITIES SERIES

**Urban Mobility Workshop
Series #3: Public-Private
Urban Mobility Data
Sharing: Evaluating Security
and Privacy Solutions**

November 13th, 2014 in San Francisco

FEATURING:

Susan Shaheen (UC Berkeley)

Gerry Tierney (Perkins+Will)

Jeremy Donnell (PG&E)

Jason Lee (SFMTA)

David Ory (Metropolitan Transportation Commission)

Jan Urbahn (BMW Drive Now)

Co-Sponsored by:

PERKINS
+ WILL

AGRION
GLOBAL NETWORK FOR ENERGY
CLEANTECH, CORPORATE SUSTAINABILITY



UNIVERSITY OF CALIFORNIA *Berkeley*

Transportation Sustainability
RESEARCH CENTER

Industry Snapshot

Urban Mobility Workshop Series #3: Public-Private Urban Mobility Data Sharing: Evaluating Security and Privacy Solutions

November 13, 2014 | San Francisco | Smart Cities | Transportation

- Susan Shaheen, UC Berkeley Co-Director, TSRC & Adjunct Professor
- Gerry Tierney, Perkins+Will, Associate Principal
- Jeremy Donnell, Pacific Gas & Electric, Manager of Customer Energy Solutions
- Jason Lee, San Francisco Municipal Transportation Agency, Performance Manager
- David Ory, Metropolitan Transportation Commission, Principal, Planning
- Jan Urbahn, BMW Drive Now, Product Development & Strategy

This workshop was a result of the March 27, and June 19, 2014 workshops, held by AGRION, Perkins + Will and UC Berkeley's Transportation Sustainability Research Center, in which workshop attendees and panelists identified the need for commonality amongst data so that it can be effectively aggregated and repurposed by a range of entities in both the public and private sectors.

Speaker Presentations:

Gerry Tierney began Workshop 3 by describing where the idea for this workshop series began, citing that the team started imagining the ideal way to move around a city without needing to own your own car and not necessarily without using cars entirely. He described that such a system should be agnostic to the method, so long as it is safe, convenient, efficient, and can be integrated into other modes. From there, the team envisioned determining what the legal and technical hurdles and questions to such a world might be through a series of workshops.

Susan Shaheen then described the evolution of the previous workshops, culminating in Workshop 3, which aimed to examine best practices in data sharing—in and outside of the transportation sector—and how best to move forward. She described that we need to do to move the needle forward regarding how we handle data. Data sharing and the quality of that data are a critical part of a public-private mobility partnership.

Panel and attendee introductions helped elucidate the range of interests and involvement in data sharing. For Jason Lee, SFMTA Manager of Performance Unit, who works with data every day, data help to inform executives and staff member decision making, while for Ryan at SFCTA, data sharing affects a number of efforts including Transportation Demand Management projects, and it is used by their modeling group to examine regular travel demand models. At PG&E, Jeremy, senior manager and Garen Grigoryan, an expert on data, are collectively trying to figure out what the appropriate policies are regarding data sharing and privacy in the energy sector. Jan Urbahn, at BMW's DriveNow, noted that they have a long history of using data in the auto world and, for instance, want to use data to allow customers to be better able to find their carsharing vehicle, which may be parked at various locations within a given region. Others worked for data aggregators, like TransitS.

RidePal, a shared commute bus service, acquire a lot of data, which they use to understand the various means by which people get to and from work. Intuit is seeking the right balance of data transparency, along with privacy. SmartRide uses data to help commuters, while Navigant Research uses data to focus on economic forecasting and modeling. Carma Carpooling, a ridesharing service, is sharing non-personal information with the Metropolitan Transportation Commission (MTC) and would like to see standards for sharing real-time data (e.g., GTFS). At Velo Labs, a wireless bike lock company, open data will help to optimize their business; while Iteris, a transportation

analytics firm, has an intrinsic data-sharing core. An attendee from Freshwater explores data among water issues, and the California Public Utilities Commission is looking at data to examine how electric transportation such as electric vehicles (EVs) may impact the electricity grid and how they may be integrated into transportation systems more broadly.

Following the introductions, Jeremy Donnell presented on his work at PG&E involving customer data sharing. PG&E provides services to 15 million people in Northern and Central California and has been collecting reliable customer usage data for years. From a broad perspective, PG&E needs to understand the data governance framework to emphasize risk mitigation and the need to cover data security. There are three key overarching considerations:

- 1) Data management (needs to be high quality);
- 2) Data uses (what is the strategy for using it); and
- 3) Data value (what am I doing with this data).

Personally Identifiable Information (PII) is considered to be private under California state law; however, PG&E uses this information internally for its operational needs and does not share PII unless it meets these operational needs (such as vendors performing services or analysis on behalf of PG&E).

In other cases, it may be appropriate and beneficial to share data with other entities. Some questions to consider when dealing with PII:

- 1) Can you further anonymize the data for aggregation (putting groups together)?
- 2) Would this aggregation create enough security to allow other constituents to have access to the data without compromising the privacy of the individual?

PG&E always protects customer's privacy and personal interests, and is deploying data sharing platforms that address several needs.

Energy Data Request Program:

Is data request portal that allows eligible third parties (e.g. local governments, state and federal agencies) to request data in order to meet statutory obligations such climate action planning, or in the case of researchers for doing energy policy analysis.

Share My Data Program:

Is an electronic data sharing platform that allows customers to consent to sharing their usage data with select qualified third parties. The purpose of this data sharing is to enable third parties offering services to PG&E customers.

Green Button/Download My Data:

An online data access platform where customers can view, compare to their cohorts or download their usage information for the past thirteen months.

One of the major constants related to PII is the issue of consent. Without customer consent, PII is a no-go, unless there is a direct business need. Customer privacy is the number one concern at PG&E.

Jeremy then discussed PG&E launching a new data sharing platform (called Energy Data Request Program), which allows customers to pull down monthly data at the zip code level for both residential and non-residential customers. The monthly data, available on this portal, is aggregated (and anonymized) to the ZIP code level, by customer class, per CPUC aggregation rules and does not require customer consent.

In this scenario, data request options could also be made available within that portal for CPUC identified eligible users. For example, researchers may be able to request PII and anonymized data for research purposes; however, the utility would reserve the right to review the request and approve or deny it depending on its designated use.

Other data sharing observations from the utility perspective include:

- 1) It takes will power to get this done. It tends to be more effective to drive the decision to share data if sought collectively as a group.
- 2) Serving the best interests of your customers is a balancing act.

Key Point: Balancing privacy against all of the above interests

- Achieving a good balance is the best defense regardless of your position.
- Primary business purpose relationships, through contracts and strict non-disclosure agreements, are effective if allowed by law.
- Privacy advocates are powerful; green policy advocates may be even more powerful.
- Might need to look at governing legislation that limits data sharing in transportation; does this need to be changed?
- Providing scrubbed data or performing the analytics in-house are resource intensive but viable solutions.

Key Point: Ensure clear policies and effective processes to avoid unintended data leakage

Following Jeremy Donnell’s presentation, Jan Urbahn from BMW’s DriveNow presented from the private sector’s perspective on data sharing. DriveNow is a free-floating car sharing service with no late fee for use. They are continuously looking at new locations throughout the Bay Area and beyond, but they face regional limitations such as parking issues in San Francisco. The group plans on adding 80 all-electric BMW Active EVs to their fleet in the near-term.

Jan discussed how BMW’s cars collect several data points, having up to 138 sensors on some of their vehicles. As part of this data collection, they are often faced with the question of whether or not they would like to share this data. This data privacy concern was asked of OEMs in the mid-2000s in the context of Vehicle Infrastructure Integration (VII), a national privacy policy framework with respect to VII. A few questions that are critical to the success of data sharing with compromised privacy include: Who shares this data? Can there be a mandatory opt-in? What is shared: anonymous or personally identifiable data? Where are the data stored? How is it protected?

Key point: When are the data shared? In real time or with time delays (e.g., law enforcement question)?

Key point: Could you develop a working group to examine this issue (e.g., a VII working group)?

Following Jan’s presentation, Jason Lee from SFMTA gave a municipal perspective on data sharing. Proposition E was passed by the San Francisco voters in 1999 to merge the Department of Parking and Traffic with Muni to create San Francisco Metropolitan Transportation Commission (SFMTA). SFMTA also manages Transtat, which was launched by Ed Reiskin who came from the SF Dept. of Public Works, where he had launched “DTW Stat.” Transtat looks at data throughout the agency and helps visualize the data to assist SFMTA staff in making data-driven decisions.

One of Prop E’s goals in creating the SFMTA was to provide an 85% on-time performance standard of scheduled service, which to date has not yet been met. This goal forms the core of what is Transtat: to look at data and further analyze it and to make decisions about policies and how to design projects based on the data. Various data sources go into Transtat (e.g, NextBus, NTD, Clipper, Trapeze, SFMTA customer survey, automatic passengers counters, SFPD Compstat, 311). Transtat brings a framework around analyzing, organizing, and making better use of the data collected. SFMTA’s strategic plan also looks at what is needed to help improve the customer experience and employee workforce performance.

Jason touched on several potential opportunities that could improve service planning and enhance the overall customer experience:

- Clipper data does not have any way of telling ridership except when people “tag” their clipper cards (e.g., how to manage crowding and peak period demand).
- SFMTA cannot target people using their PII; this is a missed opportunity for SFMTA to suggest alternate travel options for specific riders.
- Bay Area Bike Share (BABS) and SFMTA have information when a customer checks out a bike and where the bike goes. Each generated trip is known, along with whether or not they are a casual user or a subscriber; however, we do not know if the user is using BABS with other transportation systems.

Another point to be aware of is that ridership data may not always be accurate due to the ability of sensors to accurately obtain and report data. SFMTA, however, has been looking into other technological solutions and are working on increasing the level of accuracy. This illustrates the importance of data quality, and this it needs to be looked at very carefully to ensure that there is not a false sense of precision.

Jason then brought up the significance of methodology and metrics. If you do not frame the metric right, you do not get the right data. Can SFMTA accommodate 85% on time service during a one-hour period? Can they accommodate enough people during a given day’s peak? People do not arrive at the same time nor do buses (e.g., bus bunching), which makes meeting this 85% target challenging. Alternative metrics may be needed to further evaluate service due to these realities.

Jason concluded by saying that SFMTA would like to house more data online and work on different ways to get it to the public. One of their main concerns is that mobile application software is not compliant with the Americans with Disabilities Act (ADA). As an agency, SFMTA is required to provide equal access to all demographics and possible users of their services.

Following Jason, David Ory provided the regional governmental perspective on data sharing. MTC was originally created as a planning organization to distribute funds, including federal dollars. The agency has facilitated many familiar services and has helped to design and implement several of the transportation services we use today, including Clipper, 511, and FasTrak.

MTC’s growth from a planning organization into a diverse planning and operations agency, has made it difficult to develop and implement a unified strategy for data sharing. MTC’s planning section is a consumer of data and is very interested in Clipper and Fastrak data. Currently, these data are difficult to access even by entities within MTC because both Clipper and Fastrak have the primary (and state-mandated) responsibility to collect fares; sharing data for planning purposes is a secondary objective.

Key point: Is giving out data for planning purposes in the business interest of Clipper? Purpose of Clipper is to collect fares. This is Clipper’s mission, at present.

The Clipper story is an interesting one, and it continues to develop. MTC is actually close to creating an anonymous Clipper data set. A few questions arise from this: What’s the right level of sharing this data? What is the best way to proceed? What is the purpose of giving out these data? One of the main takeaways here is that making predictions with data is now possible.

Key point: MTC has had good success working with private sector entities and are very happy to buy the data, if it adds value (e.g., fitness trackers, bike/pedestrian movements, cell phone tracking).

Key point: The “opt-in” idea, namely, giving riders the opportunity to opt-in to share their data in a more transparent form, is an opportunity

State law defines travel patterns as PII, but the law does not define what a “travel pattern” is. This lack of clarity makes it very difficult to know how far an agency can go with the data before it becomes a violation of the law. This is a major difficulty in Clipper and Fastrak data being shared for planning purposes.

Open Discussion

Following the panel presentations, Susan Shaheen led a discussion on the topics presented and other questions with the workshop attendees. The points of that discussion are summarized below:

It is very helpful when regulators believe sharing data is a good thing and act to create policies supporting data sharing. A focus on green policies, energy efficiency, renewable generations, and other sustainable motivations also create a viable case for gaining access to data as it serves a greater good. To enable all of this, PG&E currently shares data with third parties in both an opt-in format, as well as non-voluntary circumstances. In these non-voluntary instances, a case needs to be made that sharing the data is in the public and customer interest.

Public interest concept: Usually someone has defined what is in their interests and the definition is often unclear.

There is a section in the Public Utilities Code that defines public interest in the context of transportation: lowering energy costs, increasing use of alternative fuels, greenhouse gas emissions and air pollution. Even such definitions can counter one another.

Can we get EVs excluded from trip counts (e.g., case of trip caps)? Do not count EV trips in the trip caps because they provide a greater value to the environment.

Hackathons have proven to be an effective way of bringing innovative solutions to the forefront. SFMTA has held hackathons in the past, and may do more in the future. A lot of past efforts focused on data quality (e.g., NextBus, etc.).

Clipper contract system: Can there be a standard system to share data based on the benefits to all local agencies? Does it need to involve lawyers or can they talk directly to each other to share the data between MTC and the Bay Area Rapid Transit (BART) District? A lot can be decided among agencies themselves.

Planners need to engage in the “Clipper 2.0” development process. There are 30 odd entities running public transportation services in the Bay Area, and there are very few corridors in which public transportation is profitable, which now has competition from private corporate shuttles. A standardized approach toward designing the next generation of mobility could be beneficial for all constituents.

One main reason that MTC, SFMTA, and similar agencies cannot share their data is due to PII restrictions. Defining a “travel pattern” as PII in the legislation is an issue, and this needs to be more clearly defined. Public entities are doing one specific thing with Clipper data (i.e., money transaction), while it can be used effectively for many other purposes. There is a developing understanding among stakeholders that there is a data opportunity here.

Currently, agencies can see the movement of an individual across one day but not across a longer duration such as a week—for instance in the Clipper data. What is obscured by only looking at one day? Other pieces of data such as long-term patterns are needed for more thorough analysis.

There will be a procurement for Clipper 2.0 in the coming years. Ideally the revision will accomplish two things: 1) change the law around travel patterns (make the case that public transit agencies should have access to this data) and 2) have the capability to share this data, even if you do not act on it.

Transtat combines and provides access to huge sets of data (e.g., NextBus, individual Clipper transactions on the system, and more). This is a great software package to wrap up reports and visualizations. Most other public transit systems may be standalone or isolated transit systems. Currently, there is no access to individual transactions and a lot of limitations around direct communications with their customer (e.g., letting us help the customer find transportation alternatives).

Netherlands example: Getting access from the customer through an “opt in” clause is an effective way of mining data. Could a similar “opt-in” program be launched in the Bay Area?

Incentivizing opt-in data collection: Could we create a pilot program that rewards users who are willing to share their data (efforts that show behavioral change—are rewarded with cap and trade funds)?

MTC: Climate initiative plan for proposing pilot projects.

Planning agencies are partnering with entities that aggregate mobile GPS data. AirSage for example, collects data from cell phone towers and packages these data sets to sell to other entities. This could prove to be a far more efficient and easy way for agencies to have access to transportation data.

Sell Bluetooth reader: Buying these products (e.g., AirSage). It would be better for public sector to outsource more of its data needs to the private sector.

There is currently a unique value to the Clipper data because of its standardization. As the industry matures, more collaboration between the public-private sectors will grow.

The Bay Area has many public transit agencies, and many of them are large consumers of data, which creates a very unique situation. The data itself, however, may not be a completely accurate and representative of the user base. There are several nuances and monitoring these sources when they need further information could be a good approach on the agency side. Not all people have mobile phones either, which leads to equity considerations.

OEM data and PG&E data: Are there data that can tell us where we need new EV infrastructure? For example, choosing where to build out charging station infrastructure requires data and analysis to ensure that customers will use the systems. In BMW’s case, Jan mentioned that often executives will plan and determine the most optimal locations to site charging stations. Level 3 charging seems to be the next big thing, and BMW has recently started working with PG&E to install 50 new chargers, which will provide air quality and sustainability benefits by reducing the use of fossil fuels.

There needs to be a planning champion for a Clipper 2.0, which allows for greater data sharing. Planners need to be more vocal during the procurement process and voice their needs to regulators and the private sector. We all need to be more proactive at making these requests and seeing them through. In many cases, there is a lack of awareness that these needs even exist.

Another point made was that the MTC is currently looking at Clipper cards being linked to San Francisco parking meters. *SFpark* implemented sensors to track car movements; however, the sensors did not last very long. This is something that needs to be addressed and updated moving forward. Some technical limitations exist, but the goal is not impossible. Clipper could also be integrated with bike sharing, and the data could be sourced through the growing bike sharing network.