Memo

To: Cybersecurity and Long-Term Policy Subcommittee of AV Task Force
From: Carrie MacLaren, Deputy Director, DLCD, and Becky Steckler, University of Oregon
Date: July 18, 2018
RE: Desired Data to Guide AV Policy Development

The future of AVs is uncertain; but there is a consensus that large-scale deployment of AVs will have significant impacts on our communities. Having good information about early AV usage will help Oregon build well-informed policy for the long term.

Decision-makers are rightly aiming to shape policies to ensure AVs can improve traffic safety, decrease congestion, boost transportation choices, and support a strong economy. The private sector should be asked to share useful information to assist in that effort, while protecting consumer privacy and proprietary information.

This memo aims to outline what information will be most useful.

What information do state and local governments need?

Data most relevant to understanding the impacts of AVs on travel and maximizing their benefits are listed below. The list is consistent with the National Association of City Transportation Officials’ (NACTO) Data Sharing Principles.

- **Trip origins, destinations, types (passenger, goods delivery, or zero-occupancy/goods), and time of day**, to understand travel demand. NACTO calls for origin/destination data at the block face level (i.e., which side of a city block a trip starts or ends at). Cities such as Portland, New York and Boston collect or are looking to collect TNC data at the block face level or an even finer scale.

- **The number of vehicle occupants**, allowing Oregon to incentivize shared travel and capture value from zero-occupant vehicles. Cities have become more interested in occupancy data as the impacts of TNCs on congestion have become more apparent (New York and San Francisco).

- **Location and severity of collisions; location of instances of rapid acceleration and deceleration and sudden collision avoidance; and AV operation disengagement**. As Oregon’s vision is to eliminate deaths and serious injuries on its transportation system by 2035, maximizing the safety benefits of AVs is a key opportunity to reach that goal.

Transportation agencies use state and federal collision data to identify safety problems. Yet data on non-fatal collisions is not always available, and collisions are often under-reported. AVs can provide those data, including data on near-misses to help identify potentially dangerous locations before collisions occur.

Data on AV disengagement (where a human driver has to take control of the vehicle) will be critical to understanding how AVs are impacting safety during the early years of deployment. Most TNC regulations require collision reporting (for example, see Portland’s City Code, section
16.40.280), and NACTO calls for collecting data on collisions, acceleration/deceleration, and disengagements in anticipation of AV deployment. The State of California also collects data on disengagements through its AV testing program (Section 227.46).

- **Data on AV operator distraction.** Distracted operation has been noted as a challenge, particularly for Level 2, 3, and 4 technology. As AVs work to train operators to not have false sense of security, data on operator distraction levels will be key to ensure long-term safety.

- **Route traces and parking data** to understand how AVs are affecting travel patterns (e.g., whether vehicles cruise or park, whether AVs are rerouting onto local streets to avoid congestion). This is not reflected in NACTO’s principles, but it is critical to fully understanding the impacts of AVs.

- **Traffic volumes and length of trips (in minutes) and/or vehicle speeds** to identify congested trips and causes of delay.

- **Data on traffic violations by AVs.** While hopefully a small set of data, the underlying challenges of safely integrating law-constrained AVs with human-driven vehicles, bicycles and pedestrians, may result in helpful data.

Additional data from TNCs operating AVs would be helpful to help ensure shared fleets provide safe and equitable service.

- **Service provider (e.g., Uber, Lyft) and type (e.g., UberBLACK, UberPOOL).** This is a standard requirement in city-TNC data sharing agreements. Cities are increasingly interested in service type since the introduction of shared TNC services, but few collect it.

- **Booking type (advance/real-time); wait time; cost of trip; and location, date, and time of unfulfilled, declined, and cancelled rides.** These data help ensure shared fleets are meeting people’s needs throughout our communities.

Shared fleets might be able to provide travel options for those who need them the most. The evidence is mixed. Some studies of TNCs have found people of color, people in wheelchairs, and other marginalized groups face longer wait times and greater numbers of unfulfilled rider requests. Of particular concern is the potential impacts on transit, a low-cost transportation option whose ridership could be reduced by AVs, leading to cuts in service. AVs could also supplement, rather than supplant, transit service.

Overlaying data on wait times, costs, and cancellations with Census demographic data can help us understand whether Oregonians are receiving equitable service. This could help us meet our Title VI requirements and inform equity-related regulations. Portland collects these data with the exception of cost (see Portland City Code 16.40.200 et. seq.), and NACTO recommends collecting data on unfulfilled rides.

- **Number and type of passenger complaints,** which can be a valuable resource for understanding safety and equity. Portland collects TNC complaint data (see pages 19-20 of the Greyball Audit Report).


How should data be provided?

Data must be properly managed to avoid compromising privacy and proprietary information.

Aggregation is the most common method to protect sensitive data, and one of the simplest to execute. Aggregation can enable agencies to use data. We are not interested in individual trips; we are interested in travel patterns, and many transportation agencies lack capacity to manage large quantities of data. Data can be aggregated spatially, temporally, or both.

NACTO’s data sharing principles recommend aggregating different data to different scales, and our recommendations below, which are largely consistent with NACTO’s recommendations, follow suit.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Spatial aggregation</th>
<th>Temporal aggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trip origins and destinations</td>
<td>Block face</td>
<td>Hourly averages</td>
</tr>
<tr>
<td>Number of occupants</td>
<td>Street segment</td>
<td>Hourly averages</td>
</tr>
<tr>
<td>Safety (collisions, sudden acceleration/deceleration, AV operation</td>
<td>Point (disaggregate)</td>
<td>Real-time</td>
</tr>
<tr>
<td>disengagement)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Route traces</td>
<td>Street segment or by origin/destination pairs, or disaggregate with trip ends truncated</td>
<td>Hourly averages</td>
</tr>
<tr>
<td>Traffic volumes and speeds</td>
<td>Street segment</td>
<td>Hourly averages</td>
</tr>
<tr>
<td>Booking type; wait time; cost of trip; unfulfilled, declined/cancelled</td>
<td>Census tract</td>
<td>Hourly averages</td>
</tr>
<tr>
<td>rides</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passenger complaints</td>
<td>N/A (complaint data does not need to be spatial)</td>
<td>Real-time</td>
</tr>
</tbody>
</table>

The data should be updated monthly at a minimum in an analysis-ready origin-destination format. For shared fleets, it would be best for data to be further disaggregated by service provider and type of service for regulatory purposes.

Some organizations have proposed third-party repositories to collect and process increasing amounts of transportation data and share them with public agencies. Both NACTO and Seattle DOT (in cooperation with University of Washington) are exploring this approach, which also has precedents in federal efforts such as the National Household Travel Survey and Fatality Analysis Reporting System, both of which allow users to either download raw data (cleaned of personally identifiable information) or query databases to get aggregate results for the data that they need. Creating a similar repository for AV data would:

- **Better protect of travelers’ sensitive information**: A central repository could employ more advanced techniques to provide privacy than those discussed above, such as differential privacy, which technology companies use to gain insights into users’ habits while preserving confidentiality. Many transportation agencies do not have the capacity to apply advanced privatization techniques. Such techniques could even enable the inclusion of certain demographic data in aggregate to assist transportation agencies in understanding equity issues.
• **Boost convenience for companies:** Instead of providing data to each of the local, regional, and state jurisdictions in which their vehicles operate, companies could supply all data in a single consistent format to a single third party.

• **Provide more manageable data for agencies:** A repository similar to the two federal tools mentioned above could enable transportation agencies to query data for only their area of interest, get it aggregated to the scale that they need, and create custom cross-tabulations to answer key policy and planning questions.

• **Increase adaptability:** As increasing numbers of AVs are on the road, more data will be available, and it will be both possible and necessary to understand AV travel patterns in more detail. A central repository would make it possible to display data at finer scales as increasing amounts of data become available.