CHAPTER 6—DATA COLLECTION AND DATABASES

Chapter Summary

This chapter provides insight into database contents and data collection procedures for mobility analysis. The chapter begins with information on basic data sources and the role of estimation procedures. Data collection coverage is graphically illustrated for different data sources. Finally, insights into databases developed from real-time data sources and those developed with statewide estimation procedures are provided.

6.1 Basic Data Sources and Estimation

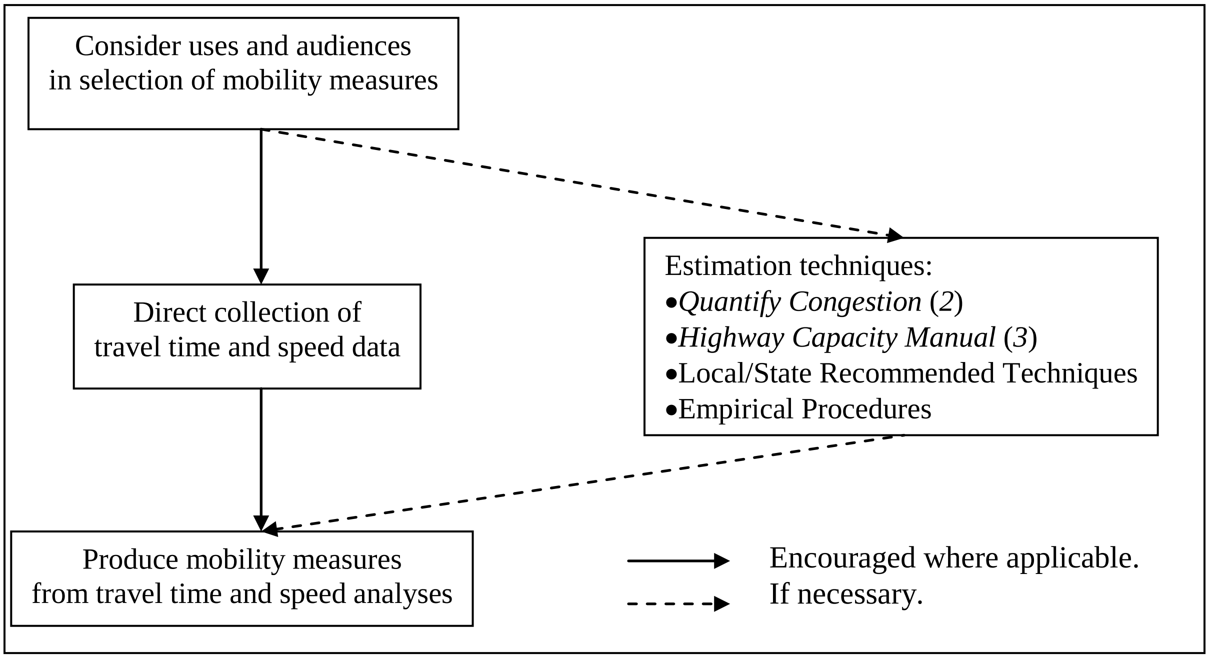
Concerns about the cost and feasibility of collecting travel time data are frequently the first issue mentioned in discussions of mobility measures. **There are many ways to collect or estimate travel time and speed quantities**; data collection should not be the determining factor about which measures are used. Exhibit 6-1 makes the point that while the direct collection of data is the desirable method of obtaining travel speed information, the selection of the proper measures should be the first step. As discussed in this report, while it is not always possible to separate data collection issues from measure selection, this should be the goal.

6.1.1 Sampling and Estimation Techniques

Sampling procedures and estimation techniques can provide useful travel time information with limited data collection budgets. Advanced technologies already provide a significant improvement in travel time data, and the number of transportation analyses that use this real-time information is growing. As these systems are installed in cities, travel time information will be more available in at least some corridors.

Travel time and speed data can be collected on a sample set of roads, routes, or modes in the analysis area. A strategic approach to sampling can be used—focus the travel time collection efforts on the problem or opportunity areas and estimate travel conditions on the rest of the system with a combination of limited data collection and estimating procedures. Techniques such as this allow mobility assessment programs to be more effective and affordable, especially for annual monitoring purposes or for complex study areas. Specific procedures and recommendations for data collection are included in the FHWA publication *Travel Time Data Collection Handbook* (*1*).

Exhibit 6-1. Desirable Role of Data Collection Issues in Measuring Mobility.



Source: NCHRP (*2*)

6.1.2 When and How to Estimate

The results of any estimate should be used with an allowance for the potential error that can be introduced when such estimates are derived. Travel time estimation procedures are most applicable for policy, programming, or planning purposes—situations where the future is not known with precision but it is important to select between alternative actions. This selection process often calls for mixing direct data and estimates. In these cases, a separate estimate of the speed must be made for existing and future conditions. The future rate should be calculated using Equation 6-1, which combines estimated travel rates for existing and future conditions with existing travel rates. This process reduces the error that would be induced by comparing actual rates to estimated rates—the difference may be related to the method used to obtain the estimate.

$$\begin{matrix}\textit{Future}\\ \textit{Travel Rate}\end{matrix} = \begin{matrix}\textit{Existing}\\ \textit{Travel Rate}\end{matrix} \times \frac{\textit{Estimate of Future Conditions}}{\textit{Estimate of Existing Conditions}} \qquad \text{(Eq. 6-1)}$$

Highway Capacity Manual (*3*) procedures are the basis for many national, state and local analytical processes. These count-based procedures are relatively detailed, with default factors provided when data are not available. The procedures and statistics have been developed for planning and operational analyses, and the products have not always been useful for communicating to audiences beyond transportation professionals. The *HCM* procedures have been developed from analysis of physical limitations of road systems at critical points. As such, the interaction between road sections that determines travel time along a congested road, as well as the spread of congested conditions beyond the peak hour, have not been a prominent aspect of *HCM*.

The incorporation of *HCM* procedures into computerized operations models has extended the usefulness of *HCM* to corridor and system analysis needs. Revisions to the basic products that are included in the 2000 edition of the *Highway Capacity Manual* are also developing travel speed and delay estimates for all the key analytical procedures and encouraging computer models for corridor analyses. *HCM*-based procedures will always have a role in producing mobility measures, but the direct collection of travel time data can assist in calibrating computer models and estimation techniques.

Vehicle occupancy data may be important for some analyses where modal, ridesharing, or other actions are being studied. The analysis may be able to use a set of regularly conducted studies in the urban area as a start for the analysis. Focused occupancy studies in locations where the average rate is likely to be different from the remainder of the urban area can be used to identify the effect of actions and assess locations where modal alternatives have been enacted. There may be many studies in an area where the use of general occupancy rates will be sufficient to adjust vehicle quantities to person values for economic analysis and presentation of results. Research performed in the State of New York provided a method for estimating areawide auto occupancy by county, year of occurrence, month of year, day of week, and time of day intervals (*4*).

The key to developing good mobility measures is to recognize the interaction between elements of the transportation system. Changes in one mode, operating system/procedure, or demand patterns can have effects that go beyond the original intent of the analysis. These potential effects should be considered in developing data collection plans.

6.2 Data Collection Coverage

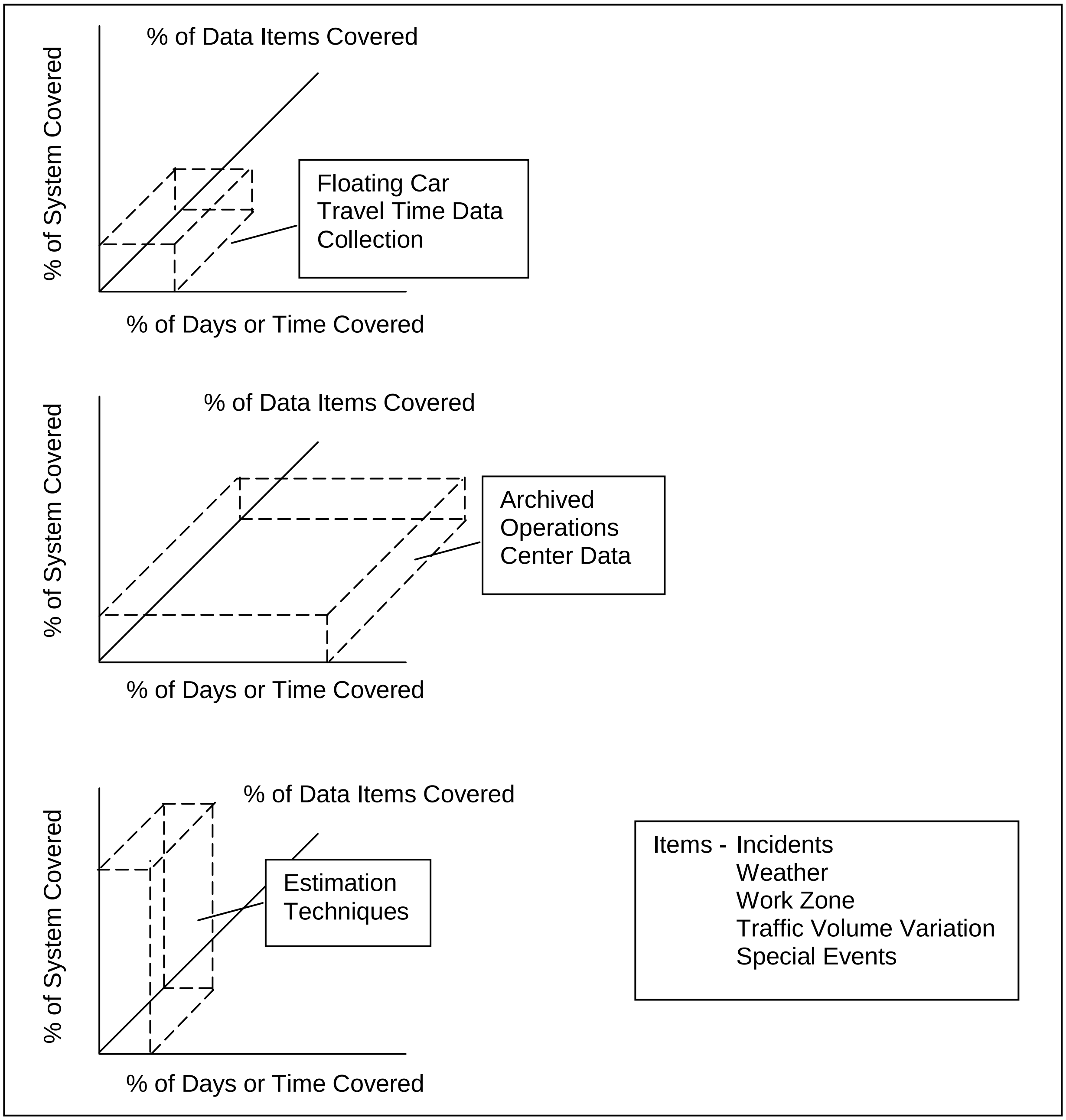
Supporting information for travel time analyses can be generated from three basic approaches—travel time data collection from floating car or other vehicle-based sampling procedures, data from traffic operations center archives, and estimation or modeling techniques. Each of these approaches has strengths and shortcomings. None of the approaches can be used for all analyses, and none of them include all the information required for a comprehensive assessment of congestion and reliability issues. Exhibit 6-2 compares the coverage of three data collection dimensions by the three approaches.

6.2.1 Vehicle-Based Travel Time Data

Floating car or probe vehicle travel time observations typically consist of a few trips on relatively few roads in a corridor or city. The observations are made on a few days and on a sample of roads or on only a few major roads. Data concerning some non-ideal conditions can be collected, but the sample size is typically small.

Some probe vehicle techniques provide more robust datasets. These include travel times obtained from toll tags or cellular telephones.

Exhibit 6-2. Summary of Data Collection Techniques.



Source: Lomax et al. (*5*)

6.2.2 Archived Traffic Operations Center Data

Traffic volume and speed data can be automatically collected and saved for each day of the year. These data will include many days when non-ideal conditions exist, which greatly improves the usefulness of the information. Unfortunately, in almost every city, freeways are the only roadway type where data are archived and usually for only a small portion of the freeway system.

6.2.3 Estimation Techniques

Equations, simulation models, and other estimation techniques are used when areawide or comprehensive network assessments are needed. Of necessity, these are simplifications of the day-to-day variation in conditions and travel patterns. They can be used to estimate the effect of some non-ideal conditions such as construction, maintenance, special events, weather, vehicle breakdowns, and collisions although these models cannot show the complexity of interactions or variations.

6.3 Creating an Archived Database

Archived data can provide much more information about the operation of freeway systems in normal time, during special or irregular events, or after the implementation of programs and techniques designed to improve operations. Until the technologies are more widely deployed, it will be difficult to compare performance characteristics from one area to another. The data collected from continuous monitoring systems result in different performance measure values than those calculated from estimation techniques for a variety of reasons. Any estimation program will have difficulty replicating actual conditions, but the current real-time data collection devices suffer from the lack of coverage of the travel system elements. When events cause travelers to leave the monitored portion of the roadway, the performance measure accuracy suffers.

The data concepts can be divided into four broad categories. Each represents a set of data needs, but fundamentally they are ways of thinking about the desired information and how to get it. Travelers and shippers want information about travel time—both the average and the variation—that indicates how much time should be planned for particular trips or sets of trips. There are a number of ways to obtain this kind of information and several ways to display or communicate it. The data collection concepts are discussed below in descending order of desirability (as ranked by the performance measures that might be developed by each type of collection technique). There are ranges even within each of these levels, but the following four concepts provide a good framework for the discussion of archived data collection and use.

1. **A computer chip on each person or each unit of cargo—**Massive personal privacy concerns are involved, but as a concept this would allow us to understand how people move around the city. It would give us door-to-door (D2D) travel time, allow us to monitor travel on all modes and all facilities, and provide us with information about route, departure time travel time, variation in travel time, and mode choice decisions.

2. **Travel time and volume detection over sections of road, transit systems, bike lanes, and sidewalks—**This would allow us to monitor travel on most facilities or modes and provide a significant portion of the D2D trip time data. Trip information could be estimated from the monitored data and supplemented with targeted surveys.

3. **Detectors of speed and volume at locations along the systems—**Estimating the volume and speed using point detectors is the practice of most archived data systems. The techniques are relatively well defined, but the level of detail does not provide

information at the trip level. Travel time must be estimated using the point speed to indicate the average speed over the adjacent road sections.

4. **Estimates based on volume and roadway inventory—**Equations or computer models that relate volume per lane and speed will always be needed for future condition mobility analyses or for evaluating improvements to existing conditions. The estimating procedures also provide information for portions of the system that are not continuously monitored. These procedures can be improved with information from the continuous monitors.

6.4 Real-time (ITS) Data Collection Practices

Data on operations (e.g., traffic volume, traffic density, speed, or travel time) are archived at some traffic management centers (TMCs). For most cities in the Mobility Monitoring Program (*6*), the data are collected at point locations using a variety of technologies including single and double inductance loops, radar, passive acoustic, and video image processing (some areas use multiple technologies; see Exhibit 6-3). These technologies establish a small, fixed "zone of detection," and the measurements are taken as vehicles pass through this zone. The Houston travel times are collected via their automatic vehicle identification (AVI) system. This system detects vehicles with toll tags and provides a direct measurement of travel time between sensors spaced at 2- to 5-mile intervals.

Data collection and processing procedures have been developed individually, and the details of the archiving vary from site to site. However, there are several procedures that are common to all sites. In general, the process works as follows for each city (with Houston being slightly different):

- Data are collected by field sensors and accumulated in roadside controllers. These field measurements are by individual lane of traffic. At 20-second to 2-minute intervals, the roadside controllers transmit the data to the TMC.
- Some areas perform quality control on original data, but this screening is typically simple and based on minimum and maximum value thresholds. These steps eliminate obviously incorrect data but do not identify all of the problems.
- Areas that use single inductance loops as sensors can only directly measure traffic volume and lane occupancy. In these cases, algorithms are used to estimate speeds for the combinations of volume and occupancy. The algorithms vary from site to site.
- Internal processes at the TMC aggregate the data to specified time intervals for archival purposes. These intervals vary from 20 seconds (no aggregation) to 15 minutes. In some cases, the data are also aggregated across all lanes in one direction at a sensor location.

- The aggregated data are then stored in text files or databases unique to each TMC. CDs are routinely created at the TMCs to reduce some of the storage burden and to satisfy outside requests for the data.

Exhibit 6-3. Summary of Data Collection Technologies and Data Level of Detail in 2003.

Participating City	Traffic Sensor Technology	Data Level of Detail: Time	Data Level of Detail: Space
Albany, NY	loop detectors	15 minutes	by lane
Atlanta, GA	video imaging, microwave radar	15 minutes	by lane
Austin, TX	loop detectors	1 minute	by lane
Baltimore, MD	microwave radar, loop detectors	5 minutes	by lane
Charlotte, NC	microwave radar	30 seconds	by lane
Cincinnati, OH-KY	loop detectors, microwave radar, video imaging	15 minutes	by direction
Dallas, TX	video imaging, loop detectors, microwave radar	5 minutes	by lane
Detroit, MI	loop detectors	1 minute	by lane
El Paso, TX	loop detectors, microwave radar	1 minute	by direction
Hampton Roads, VA	loop detectors, microwave radar	5 minutes	by direction
Houston, TX	probe vehicle (toll tags); also loop detectors, video imaging, microwave radar	vehicle-based link travel times	
Los Angeles, CA	loop detectors	5 minutes	by direction
Louisville, KY	microwave radar, video imaging	15 minutes	by direction
Milwaukee, WI	loop detectors, microwave radar, video imaging	5 minutes	by lane
Minneapolis-St. Paul, MN	loop detectors	30 seconds	by lane
Orange County, CA	loop detectors	5 minutes	by direction
Orlando, FL	loop detectors	1 minute	by lane
Philadelphia, PA	microwave radar, passive acoustic detectors	5 minutes	by lane
Phoenix, AZ	loop detectors, passive acoustic detectors	5 minutes	by lane
Pittsburgh, PA	microwave radar, passive acoustic sensors	5 minutes	by lane
Portland, OR	loop detectors	15 minutes	by lane
Riverside-San Bernardino, CA	loop detectors, microwave radar	5 minutes	by direction
Sacramento, CA	loop detectors, microwave radar	5 minutes	by direction
Salt Lake City, UT	loop detectors, acoustic detectors, microloops	15 minutes	by lane
San Antonio, TX	loop detectors, acoustic detectors	20 seconds	by lane
San Diego, CA	loop detectors	5 minutes	by direction
San Francisco, CA	loop detectors, microwave radar	5 minutes	by direction
Seattle, WA	loop detectors	5 minutes	by lane
Washington, DC			
- Maryland	microwave radar, loop detectors	5 minutes	by lane
- Virginia	loop detectors	1 minute	by lane

Source: FHWA (*7*)

The Importance of Maintaining an Accurate Real-time Data Collection System

An area of potential immediate benefit and continuing concern is maintenance of the data collection infrastructure. Funding limitations affect the ability to correct deficiencies even when devices are known to be producing erroneous or no data. The problem is exacerbated where sensors in the pavement are used because most agencies are reluctant to shut down traffic on heavily traveled freeways just to repair monitoring equipment. Maintenance is often postponed to coincide with other roadway activities, which helps spread the cost burden but may delay repairs.

Field checking of sensors is done periodically, but no standardized procedures are used across all areas. If a detector is producing values that are clearly out of range, inspection and maintenance are usually performed. However, calibration to a known standard is rarely, if ever, performed. This means that more subtle errors may go undetected. Bearing in mind that TMCs typically do not require highly accurate data for most of their operations, this approach is reasonable and practical. Work zones exacerbate these problems, and contractors often unknowingly sever communication lines or pave over inductance loops.

Calibration—at least to very tight tolerances—is not seen as a priority, given that operators focus on a broad range of operating conditions rather than precise speed/travel time estimates. This philosophy may be changing as a result of more stringent data requirements for traveler information purposes (e.g., TMC-based posting of expected travel times to destinations using variable message signs). However, the current data resolution used by TMCs is quite coarse; it supports their traditional operations activities, such as incident detection and ramp meter control.

6.5 Statewide Performance Measure Application with Estimation

Recent work was performed by the Texas Transportation Institute (TTI) for the Oregon Department of Transportation (ODOT) to identify and test operations performance measures for statewide application. The results include a methodology and test results that includes the use of real-time data and estimation procedures. The estimation procedures include the use of the Highway Economic Requirements System—State Version (HERS-ST) for obtaining statewide estimates of speed in a consistent manner (*8,9*). Observations relevant to statewide mobility measure estimation are described below.

6.5.1 Keeping a Consistent Speed in the Statewide Application

Any statewide performance measure analysis should keep a consistent source of speed data (and subsequently computed performance measures) because this value will change over time, and it is important to understand the extent that this measure is changing due to the measurement versus due to mobility improvements. For example, for the ODOT work (*8,9*), if HERS-ST is the source for the measures, then the "HERS-ST speed" should be kept from year to year as a data element. When supplemental speed information is available, it can be kept in the database next to the HERS-ST speed. There may even be speeds from more than one other source if different studies or local knowledge is available. Other speed sources might include real-time ITS data, floating car studies, automatic traffic recorder, etc. This would provide the opportunity to see trends not only in mobility performance from year to year, but to also see how these speed values may differ by data source (i.e., estimated versus direct measurement). This would allow for the calibration of the HERS-ST values with any other data sources that might be used. It should be noted that the geographic extent of data collection will be expanding each year. Consequently, it is important that the analyst only select the observations where prior year and current year data are available to derive the VMT-weighted areawide performance trend.

6.5.2 Possible "Beta" Version before Final Distribution

When implementing new mobility estimation procedures, occasionally there are concerns expressed that potentially unreliable data or errors in the new estimation procedure may indicate mobility problems when such problems do not really exist. To ultimately get a statewide methodology in place, identifying and fixing data issues are an inevitable part of the start-up process. Therefore, it might be prudent that the statewide implementation of the estimation procedures be performed over one to three years, and the results are identified as "beta" or "prototype" to allow a review of the process over the initial years. This could allow for calibration of the results across the initial years, and at different geographic locations, based on local knowledge or studies before the final "rollout."

6.6 References

1. Turner, S.M., Eisele, W.L., Benz, R.J., and Holdener, D.J. *Travel Time Data Collection Handbook*. Report No. FHWA-PL-98-035. , U.S. Department of Transportation, Federal Highway Administration, Washington, D.C., March 1998.

2. *NCHRP Report 398*. Quantifying Congestion—Final Report and User's Guide. National Cooperative Highway Research Program Project 7-13, National Research Council, 1997.

3. *Highway Capacity Manual*. Transportation Research Board, National Research Council, 2000.

4. Asante, S.A., Adams, L.H., Shufon, J.J., and McClean, J.P. "Estimating Average Automobile Occupancy from Accident Data in New York State." *Transportation Research Record 1553*. Transportation Research Board, Washington, D.C., 1996.

5. Lomax, T.J., Schrank, D.L., and Turner, S.M. "Selecting Travel Reliability Measures." May 2003. Available at: http://mobility.tamu.edu/resources.

6. *Monitoring Urban Freeways in 2003: Current Conditions and Trends from Archived Operations Data*. U.S. Department of Transportation, Federal Highway Administration, Report No. FHWA-HOP-05-018, December 2004. Available at: http://mobility.tamu.edu/mmp/.

7. "FHWA Mobility Monitoring Program: 2003 City Reports." U.S. Department of Transportation, Federal Highway Administration, February 2005. Available at: http://mobility.tamu.edu/mmp.

8. Eisele, W.L. and Lomax, T.J. *ODOT Operations Performance Measures: Final Report*. Sponsored by the Oregon Department of Transportation, June 2004. Available at: http://mobility.tamu.edu/resources.

9. Eisele, W.L., Lomax, T.J., Gregor, B.J., and Arnold, R.D. "Developing and Implementing Statewide Operations Performance Measures in the State of Oregon: Methodology and Application for Using HERS-ST and Archived Real-time Data." Paper prepared and presented at the Transportation Research Board's 84th Meeting, Washington, DC, January 2005. Available at: http://mobility.tamu.edu/resources.