Multimodal Data at Signalized Intersections: Strategies for Archiving Existing and New Data Streams to Support Operations and Planning & Fusion and Integration of Arterial Performance Data

Kristin A. Tufte
Portland State University

Christopher M. Monsere
Portland State University, monsere@pdx.edu

Sirisha Murthy Kothuri
Portland State University

Carl Scott Olson
Portland State University

OTREC

OREGON
TRANSPORTATION
RESEARCH AND
EDUCATION CONSORTIUM

FINAL REPORT

Multimodal Data at Signalized Intersections: Strategies for Archiving Existing and New Data Streams to Support Operations and Planning

OTREC-RR-430

September 2013

MULTIMODAL DATA AT SIGNALIZED INTERSECTIONS: STRATEGIES FOR ARCHIVING EXISTING AND NEW DATA STREAMS TO SUPPORT OPERATIONS AND PLANNING
&
FUSION AND INTEGRATION OF ARTERIAL PERFORMANCE DATA

Final Report

OTREC-RR-430

by

Kristin Tufte, Ph.D.
Christopher Monsere, Ph.D., P.E.
Sirisha Kothuri
Carl Olson

for

September 2013

Technical Report Documentation Page

1. Report No. OTREC-RR-430	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Multimodal Data at Signalized Intersections: Strategies for Archiving Existing and New Data Streams to Support Operations and Planning & Fusion and Integration of Arterial Performance Data		5. Report Date September 2013
		6. Performing Organization Code
7. Author(s) K. Tufte , C. Monsere, S. Kothuri, C. Olson,		8. Performing Organization Report No.
9. Performing Organization Name and Address Portland State University		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address Oregon Transportation Research and Education Consortium (OTREC) P.O. Box 751 Portland, Oregon 97207		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

There is a growing interest in arterial system management due to the increasing amount of travel on arterials and a growing emphasis on multimodal transportation. The benefits of archiving arterial-related data are numerous. This research report describes our efforts to assemble and develop a multimodal archive for the Portland-Vancouver region. There is coverage of data sources from all modes in the metropolitan region; however, the preliminary nature of the archiving process means that some of the data are incomplete and samples. The arterial data sources available in the Portland-Vancouver region and that are covered in this report include data for various local agencies (City of Portland, Clark County, WA, TriMet and C-TRAN) covering vehicle, transit, pedestrian, and bicycle modes. We provide detailed descriptions of each data source and a spatial and temporal classification. The report describes the conceptual framework for an archive and the data collection and archival process, including the process for extracting the data from the agency systems and transferring these data to our multimodal database. Data can be made more useful though the use of improved visualization techniques. Thus as part of the project, a number of novel, online visualizations were created and implemented. These graphs and displays are summarized in this report and example visualizations are shown. As with any automated sensor system, data quality and completeness is an important issue and the challenge of automating data quality is large. Preliminary efforts to validate and monitor data quality and automate data quality processing are explored. Finally, the report presents efforts to combine transit and travel time data and signal timing and vehicle count data to generate some sample congestion measures.

17. Key Words Multimodal, data archive, performance measures		18. Distribution Statement No restrictions. Copies available from OTREC: www.otrec.us	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 103	22. Price

ACKNOWLEDGEMENTS

DISCLAIMER

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the material and information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation University Transportation Centers Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof. The contents do not necessarily reflect the official views of the U.S. Government. This report does not constitute a standard, specification or regulation.

TABLE OF CONTENTS

APPENDICES

APPENDIX A: PORTAL DATA FEED SPECIFICATION

APPENDIX B: DATA DICTIONARIES

APPENDIX C: NON-FUNCTIONING DETECTORS REPORT

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

There is a growing interest in arterial system management due to the increasing amount of travel on arterials and a growing emphasis on multimodal transportation. The benefits of archiving arterial-related data are numerous. Long term, data can be used to calibrate regional travel demand models and to provide a more complete picture of the network. These data sets can also be used to activate incident management plans. Incidents on freeways often cause queue spillovers on arterials, Access to real-time arterial data and visualization of real-time performance can enable traffic engineers to respond effectively and reduce the cost of the incident. The incorporation of transit data into the arterial data archive allows the opportunity for interactions between transit and vehicle speeds. Transit performance measures and the effects of transit signal priority can be studied, leading to improvements in transit performance. These data can be used to study the effects of signal timing changes and other strategies to improve flow and efficiency. These data also helps in identifying and prioritizing areas along corridors that need improvement.

This research report describes our efforts to assemble and develop a multimodal archive for the Portland-Vancouver region. There is coverage of data sources from all modes in the metropolitan region; however, the preliminary nature of the archiving process means that some of the data are incomplete and are samples. The arterial data sources available in the Portland-Vancouver region and that are covered in this report are summarized in Table E.1. In the report, we discuss in detail these data sources, we describe the availability of data for the various local agencies – City of Portland, Clark County, WA, TriMet and C-TRAN, We provide detailed descriptions of each data source and a spatial and temporal classification. We describe the framework for an archive and the data collection and archival process, including the process for extracting the data from the agency systems and transferring these data to our multimodal database. Data can be made more useful though the use of improved visualization techniques. Thus as part of the project, a number of novel, online visualizations were created and implemented. These graphs and displays are summarized in this report and example visualizations are shown. As with any automated sensor system, data quality and completeness are important issues and the challenge of automating data quality is large. Preliminary efforts to validate and monitor data quality and automate data quality processing are explored. There are many possible uses for the data being archived on the arterial level. Finally, the use of the data in some sample data analysis is presented. In particular, we present efforts to combine transit and travel time data and efforts to combine signal timing and vehicle count data to generate some sample congestion measures.

Table E.1 Arterial Data Sources – Agency, Data Type and Detection Hardware

Data Type Name	Agency Responsible for Collection	Type of Data Collected	Detection Hardware
COP- Count Station	City of Portland, OR	Volume	Inductive Loop
COP-Travel Time	City of Portland, OR	Travel time	Bluetooth Stations – Bluetooth stations are built by City of Portland based on OSU specifications
COP-Signal Cycle Data	City of Portland, OR	Signal cycle changes	Signal cycle controllers.
Clark County – Count Station – Radar	Clark County, WA	Volume, speed, occupancy	High-definition radar
Clark County – Count Station – Loop	Clark County, WA	Volume	Loops
TriMet – Transit AVL/APC Data	TriMet, OR	Bus and MAX speed, passenger counts, on-time percentage	Bus/Light Rail AVL system
C-TRAN – Transit AVL/APC Data	C-TRAN, OR	Bus speeds	Bus AVL system
COP – Bicycle Data	City of Portland, OR	Bicycle counts	Bicycle inductive loops in bicycle lanes
COP – Pedestrian Actuations	City of Portland, OR	Pedestrian actuations	Detection programmed into controller hardware at intersections.

1.0   INTRODUCTION

Effective management and operation of the transportation system requires information. This information must be mined from an ever-increasing array of data sources. Despite an increasingly monitored and connected system, most transportation data are still siloed—collected primarily for one mode and sometimes only a subset of facilities serving that mode. Data that provides coverage over large areas of the network often only provide information for a sub-sample of users or vehicles. Integration of these large-scale, real-time data across modes is challenging. Archiving them in common spatial and temporal scales in a regional transportation data archive is even more difficult. To date, applications of Intelligent Transportation Systems technologies (such as freeway monitoring systems, transit management and signal systems) have produced much of these data stored in these archives. These data have successfully been used for performance measurement, operational analysis, and planning-level decision-making (Bertini, 2005; Chen, 2001; WSDOT, TTI 2001; Lund, 2010) (1-5).

The benefits of archiving arterial-related data are numerous. Long term, data can be used to calibrate regional travel demand models and to provide a more complete picture of the network. These data sets can also be used to activate incident management plans. Incidents on freeways often cause queue spillovers on arterials. Access to real-time arterial data and visualization of real-time performance can enable traffic engineers to respond effectively and reduce the cost of the incident. The incorporation of transit data into the arterial data archive allows the opportunity for interactions between transit and vehicle speeds. Transit performance measures and the effects of transit signal priority can be studied. Transit performance can be studied to identify areas of improvement. These data can be used to study the effects of signal timing changes and other strategies to improve flow and efficiency. This data also helps in identifying and prioritizing areas along corridors that need improvement.

1.1   BACKGROUND

There has been a growing interest in arterial system management and related performance measurement. This is due to the increasing amount of travel on arterials. It has been estimated that 40-50% of all vehicle miles traveled occur on arterials (Tarnoff, 2009). Unlike freeways, generating arterial performance measures is a challenging task, due to the complex nature of traffic flow on these facilities as well as the lack of high-resolution data. Unlike freeway networks where speed, flow, and occupancy measures are generally sufficient to quantify the performance of the link, the link-node combination of interrupted-flow arterial facilities requires additional information to accurately describe performance.

Few regions are actively developing integrated archived data systems to house, process and synthesize arterial data. One such effort, Arterial Performance Measurement System (A-PeMS), is underway in California to develop an extension to the freeway data archive (PeMS). The goal of A-PeMS is to integrate data from a variety of sensors – traffic data, signal data, bus Automatic Vehicle Location (AVL) data, and Bluetooth data to provide performance metrics on arterials (Kwon, 2006). A similar effort has been undertaken in Minnesota with the development of the

Systematic Monitoring of Arterial Road Traffic Signals (SMART) system. This system collects high-resolution traffic data from a number of intersections along an arterial corridor, and generates real-time performance measures including intersection queue lengths and arterial travel time (Liu). SMART system operates by storing all vehicle actuation and signal-phase change data, and estimates queue lengths using an algorithm based on shockwave theory. Travel times are estimated using a simulated virtual probe vehicle. The SMART signal system is also capable of optimizing traffic signal parameters by using archived traffic signal data and the estimated performance measures.

1.2   OBJECTIVES

This research report documents the expansion of PORTAL— the Portland Oregon Regional Transportation Archive Listing is the official regional data archive for the Portland-Vancouver metropolitan region— into a multimodal data archive, incorporating data from arterial systems. This report covers the efforts of two sequential projects that were funded by OTREC. The specific objectives of this research projects were:

- To develop an archiving framework to harness the existing arterial data;
- To assess the quality of existing traffic signal data and provide data quality feedback tools;
- To develop analysis and visualization tools to foster efficiencies and better decision-making;
- To explore the data quality of arterial data sources;
- To explore algorithms for fusing stationary (point-based) and mobile (link-based) arterial data sources for arterial performance metric calculations;
- To explore methods for long-term evaluation of an arterial corridor's performance; and
- To refine storage and reporting of arterial performance data in the PORTAL transportation archive.

1.3   ORGANIZATION OF THE REPORT

This report documents the Portland metropolitan area's efforts to archive data from arterial systems. Arterials are complex systems that have many conflicting movements separated by traffic signals. The traffic is multimodal in nature, with motor vehicles, transit, bicycles and pedestrians sharing the right-of-way and allocations of time.

This report is organized as follows: Chapter 2 is a literature review, and Chapter 3 provides an overview of the various data sources that are currently available for archiving. Chapter 4 describes our multimodal arterial archiving framework. Chapter 5 presents the archiving process, and Chapter 6 explores data quality. Chapter 7 presents the visualization tools that were developed to graphically view and analyze these data, and Chapter 8 provides sample data analysis. Conclusions of the research are provided in Chapter 9.

2.0   LITERATURE REVIEW

There has been a growing body of research towards developing techniques to measure performance on arterials. Arterial performance data sources range from probe vehicles and traffic signal information to transit travel time information and Bluetooth travel time collection. Work on arterial data ranges from the incorporation of arterial data into transportation data archives to work on fusing different types of arterial data together to produce performance metrics to the study of performance metric needs for arterials, particularly arterial travel time.

Several existing transportation data archives contain arterial performance data, including PeMS Iteris, RITIS and PORTAL. All three of these are general transportation data archives that store and produce metrics on a variety of transportation data sources, including non-arterial sources such as freeways and weather. They also feature arterial-related sources including transit data, arterial signal cycle and count data. Archiving arterial data and producing performance metrics on arterial data are different from archiving and producing metrics from non-arterial sources, such as freeway data, due to different characteristics of arterial data and different performance metric needs. Data archive structures that are specific to arterial data have been described by two groups. Olson et al. (2012) presented a framework for a multimodal arterial data archive. This paper describes the database schema, data sources and sample visualizations for a multimodal data archive in PORTAL. Data sources discussed include arterial count stations, pedestrian delay data, bicycle volumes, Bluetooth travel time data and traffic signal data. The data schema used to combine these sources into a coherent framework is included. Second, Liu et al. (2007) built a system called SMART-SIGNAL to archive event-based traffic signal data and to generate real-time performance measures. The system successfully collected real-time data and produced performance measure analysis using such metrics as queue length and travel time.

Several groups have investigated the use and fusion of arterial data, including probe data such as matched Bluetooth Media Access Control (MAC) address readings and traffic signal data. Original work on the concept of using matched Bluetooth MAC address readings to calculate travel times was done by Wasson et al. (2008). Using matched Bluetooth MAC readings has become a standard method of calculating arterial travel times. Remias et al. (2012) have described incorporating probe data such as Bluetooth MAC address matching, crowd-sourced data and agency-installed detection, into adaptive control of arterial traffic flow and performance monitoring of arterial flow. The paper concludes that such data fusion is feasible, but makes the case for richer probe data. Bhaskar et al. integrated probe vehicle data and cumulative plots to estimate travel time statistics – average and quartiles. Wolfe et al. (2013) estimated fusing arterial detector information with signal timing information, saturation headways and free-flow travel times can be used to provide accurate estimates. Temporal aggregation of detector data degraded the quality of the travel time estimates; however, if cycle-level signal data are available, the study determined that accurate estimates could be made.

A final way to estimate arterial performance metrics is to fuse transit data with arterial probe and/or signal data. Original work done on using transit buses as probes to estimate arterial travel times was done by Bertini et al. (2004). This work used the maximum speed between bus stops to estimate arterial travel times. Berkow et al. (2008, 2009) investigated the combination of traffic signal data and transit probe data for calculating arterial travel times. This study was

limited by the high granularity of the traffic signal data. Finer-grained traffic signal data are needed to effectively fuse signal and transit data for arterial performance management. Slavin et al. (2013) evaluate the SCATS adaptive traffic signal control system on transit performance in the context of transit signal priority (TSP). Metrics used in this study included traffic volume and speed, and transit performance metrics such as schedule and headway delay and travel time. The study concluded that TSP was not negatively affected by SCATS.

As for arterial performance measure needs, Li et al. (2013) have investigated performance measures for signal analysis. They focus on communication measures (making sure the signal is producing data and communicating) and on good allocation of green time. Their objectives include: "monitoring traffic signals often enough to identify when maintenance or operational intervention is warranted; good allocation of green times; and good progression on selected arterials at selected times of day." In terms of measures used to evaluate those objectives, they evaluated communications by examining the number of pings returned and data completeness was evaluated by examining the amount of data downloaded. The Portland Arterial Regional Concept of Operations project listed and classified feasible arterial performance measures (RCTO, 2013).

One of the key arterial performance metrics is travel time. Researchers have employed a variety of methods for travel time estimation. One set of methods utilizes loop detector data along with signal-cycle level data to build models that estimate travel times (Zhang, 1999; Skabardonis, 2005; Liu, 2007). Vehicle re-identification techniques have also been used to estimate travel times (Sun, 1999; Coifman, 1998; Ndoye, 2008; Oh, 2002; Sun, 2004; Kwong, 2008). Other techniques for matching toll tags, license plates, cell phones and MAC addresses have also been tested (Dion, 2006; Bertini, 2005; Wunnava and Wasson, 2008; Quayle, 2010; Sadabadi, 2010; Malinovskiy, 2010; Haghani, 2010). Some studies have used the AVL data from transit buses to generate arterial travel times by developing relationships between bus and automobile travel times (Uno, 2009; Hall, 2000; Elango, 2002; Cathey, 2002, 2003; Chakroborty, 2004; Bertini, 2004; Pu, 2009).

Finally, an emerging technique is to fuse data from multiple sources such as loop detectors, signal controllers, buses and MAC readers to generate estimates (Berkow, 2009; Choi, 2002). Some bus studies have documented additional applications of transit probe data, including use of data for analysis of travel time variability, congestion and incident detection (Uno, 2009; Hall, 2000). A few studies have compared transit speeds to loop detector speeds on freeways and used filtering and smoothing techniques to derive real-time trajectories (Hall, 2000; Elango, 2002; Cathey 2002, 2003). Finally, some studies have derived relationships between transit and automobile travel times for urban arterials (Chakroborty, 2004; Bertini, 2004; Pu, 2009).

3.0   DATA SOURCES

This chapter describes these data sources currently available for archiving on the arterial system. These data described are vehicle counts, bicycle counts, pedestrian actuations and delay, transit activity, intersection phase and timing data, and travel times.

3.1   OVERVIEW

The intent of this section is to provide a summary of the types of arterial data collected in the Portland-Vancouver region. This data includes vehicle counts, travel times, intersection phase and timing data, and transit activity. The sections below summarize the arterial data collected in the Portland-Vancouver region by agency, and describe the hardware and software used to collect it. A summary of this information is shown in Table 3.1.

3.1.1   City of Portland, OR

PORTAL collects the following types of data generated by the City of Portland traffic signal system: vehicle counts, travel time, signal phase and timings, pedestrian actuation counts and bicycle counts.

Arterial vehicle counts are collected using inductive loops placed upstream of intersections. These data are collected by the City of Portland in real time, and are periodically transferred to the PORTAL archive at Portland State University (PSU). The detectors are single loops and while the system does report volume and speed, only the volume numbers are usable. Minimal data quality checking has been done and the volume data are available in the PORTAL interface.

Travel times are collected using Bluetooth detectors at selected locations along US 26 (Powell Boulevard) and 82nd Avenue. The Bluetooth devices store data locally on the device, and scripts are used to periodically transfer data from the devices to servers at the City of Portland. The data are transferred to PSU where a matching process is implemented to calculate travel times from detector to detector. The matching process is described in Monsere (2013). The Bluetooth travel time data are available in the PORTAL interface.

Signal phase and timing data are collected and distributed through the Green Driver data feed. PSU subscribes to the Green Driver data feed, which contains information about phase changes for signals in the City of Portland. These data are archived in PORTAL.

Bicycle counts and pedestrian count data are collected by the City of Portland and are archived in PORTAL. The bicycle and pedestrian data are under active research by graduate student Sirisha Kothuri as described in her recent papers on the subject (2012, 2013).

3.1.2   Clark County, WA

PORTAL has samples of two types of arterial data from Clark County. Both types are arterial vehicle count data, but they come from different detectors and different systems. Clark County uses both inductive loops and high-definition radar (Wavetronix) for arterial detection. The Wavetronix detectors report speed and volume and are capable of classification. The loops report

volume data; loop speed data are reported but are not usable. Samples of both types of data are in PORTAL. The groundwork for automating these data transfers has been done as described in the tables below. The sample arterial data are available in the PORTAL interface.

3.1.3 TriMet

Samples of TriMet quarterly data are in the PORTAL database and are available in the PORTAL interface. The transit data are comprehensive covering all aspects of bus and light rail service. With respect to arterials, the two types of transit data that are useful are the bus speed data – maximum bus speeds between stops can potentially be used to estimate route travel times – and the ridership data. Ridership data could be used to compare bus rider volume versus volume of riders in vehicles. All data are high quality as it is cleaned by TriMet before being transferred to PSU.

3.1.4 C-TRAN

A small sample of C-TRAN transit data is available in PORTAL. C-TRAN is the Vancouver-area transit provider. This data was obtained from the C-TRAN Init system. Similar to the TriMet data, the C-TRAN data are comprehensive, with bus speed and ridership being the two pieces of this data most applicable to arterial performance.

Table 3.1 Arterial Data Sources – Agency, Data Type and Detection Hardware

Data Type Name	Agency Responsible for Collection	Type of Data Collected	Detection Hardware
COP- Count Station	City of Portland, OR	Volume	Inductive Loop
COP-Travel Time	City of Portland, OR	Travel time	Bluetooth Stations – Bluetooth stations are built by City of Portland based on OSU specifications
COP-Signal Cycle Data	City of Portland, OR	Signal cycle changes	Signal cycle controllers.
Clark County – Count Station – Radar	Clark County, WA	Volume, speed, occupancy	High-definition radar
Clark County – Count Station – Loop	Clark County, WA	Volume	Loops
TriMet – Transit AVL/APC Data	TriMet, OR	Bus and MAX speed, passenger counts, on-time percentage	Bus/Light Rail AVL system
C-TRAN – Transit AVL/APC Data	C-TRAN, OR	Bus speeds	Bus AVL system
COP – Bicycle Data	City of Portland, OR	Bicycle counts	Bicycle inductive loops in bicycle lanes
COP – Pedestrian Actuations	City of Portland, OR	Pedestrian actuations	Detection programmed into controller hardware at intersections.

3.2 DATA SOURCES AND SENSORS

This section provides detailed descriptions of the arterial data sources available in the Portland-Vancouver region. Data sources are described by type of data source and include vehicle counts, bicycle counts, pedestrian counts, transit activity, traffic signal phase, and timing and travel times.

3.2.1 Vehicle Counts

In the Portland metropolitan area, system detectors are installed along major arterials to report counts, speeds and occupancies on a continuous basis. In the City of Portland, these detectors are usually single loop induction detectors typically located 300-400 feet upstream of an intersection. Stations in Vancouver and Clark County are using radar technologies for these arterial counting stations. The locations of the currently archived system detectors are shown in Figure 3.1 (there are many other locations functioning but not being archived that are shown on the map).

In Oregon, a number of jurisdictions are sharing the central signal system software (TransSuite®, an ATMS software provided by Transcore). For each detector in the system, counts and occupancies are recorded. These data are stored in one-minute intervals. The file also contains the upload status, the duration of time during which these data were collected and the sequence number, which indicates whether the uploaded data are new, old, or whether an upload is missed. Prior to this research, these data were only temporarily archived within TransSuite®, and were discarded after three months. In the course of this project, data from all of these detectors has been gathered since August 2010. As of this report, the transfer is a manual process through remote access connection from the Intelligent Transportation Systems Lab at PSU to the TransSuite server that is housed in the City of Portland. Efforts are underway to automate this data transfer. Appendix B contains the metadata description of the table records and a sample of the raw detector data.
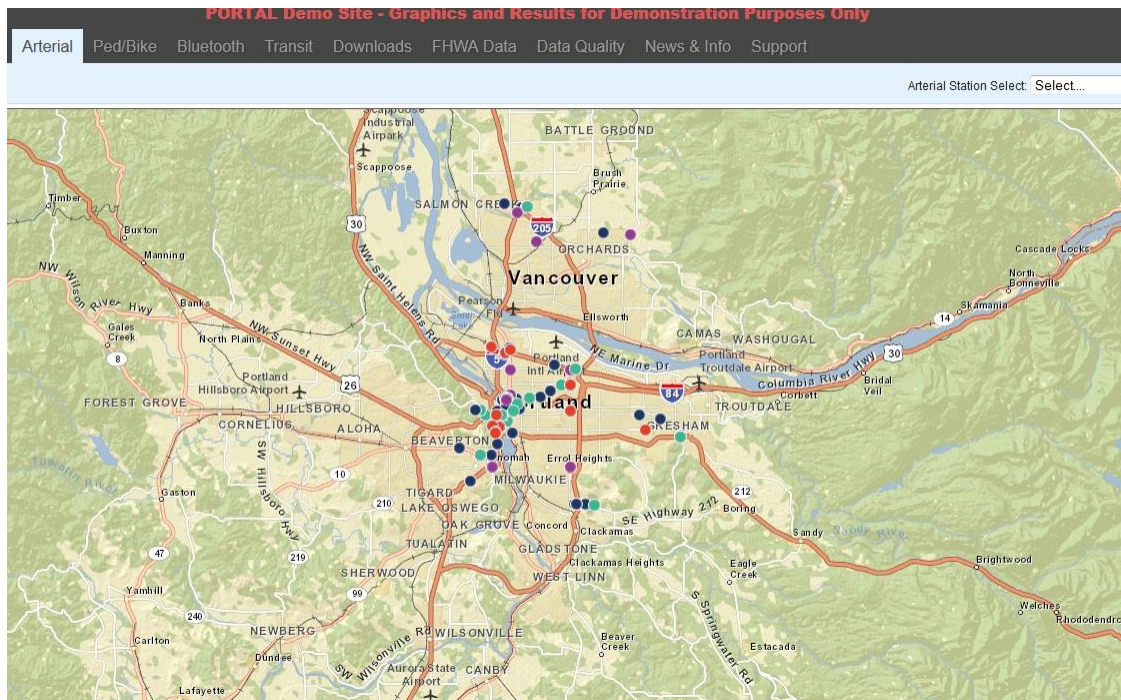
Figure 3.1 Screenshot of Arterial Count Stations Map

3.2.2 Bicycle Counts

The most common form of sensors used for detection and counting vehicles is inductive loop detectors. These devices work on the principle of electromagnetic induction and are typically embedded in the pavement. Detection and counts are recorded by measuring the change in inductance when a vehicle travels over them. Multiple loops are placed in each lane while approaching the stop bar at a signalized intersection to allow for vehicle detection for phase changes. Advance loops are typically installed 300-400 feet from the stop bar to gather vehicle counts; occupancy and speed and can also be used for yellow extensions. For bicycles, at a select number of intersections there are two sensors in the bicycle lane. The stop bar loop is typically used for phase actuation and commonly connected in series along with the vehicle loops. Therefore, counts cannot be obtained from the stop bar bicycle loops. These loops may also measure right-turning traffic that is near the loop due to the sensitivity settings of these sensors. Bicycle counts can be collected from advance loop detectors that are typically set back 40-60 feet from the stop bar.

There are specific criteria for obtaining bicycle counts from the advance loop detectors: the presence of a bicycle lane; the availability of an advance loop detector in the bicycle lane; the presence of an individual loop wire to the advance loops; and the availability of communication to the signal controller. The presence of a bicycle lane is important because it ensures that only bicycles travel in that lane and are counted by the loop detectors. The presence of advance loop detectors in bicycle lanes ensures that bicycles can be counted while in motion, which typically generate more accurate counts as compared to stop bar counts. The presence of an individual wire allows exclusive bicycle counts from the loop detectors, and the availability of

10

communication to the signal controller is necessary to view and retrieve counts. Counts are either available in 15-minute or hourly increments. Figure 3.2 shows a map of the locations that are currently configured to log bicycle counts.

In Portland, the counts from the advance bicycle loop detectors are routed to a detector channel and can be directly viewed using TransSuite® (the ATMS software used by the City of Portland). It is possible to obtain bicycle counts from loop detectors that are connected to either Type 170 or Type 2070 signal controllers. However, due to continued challenges with the legacy Type 170 controllers, all count stations for bicycles are now only obtained from Type 2070 controllers. This controller offers more functionality and channels that can be used for operations (detection) and keeping track of counts. There are 32 channels (inputs) available in 2070 controllers that can be used for counting purposes. These controllers have the ability to gather bicycle counts from multiple bicycle detectors (if available) at an intersection due to the availability of a larger number of inputs. Currently, bicycle counts are downloaded manually from TransSuite® (the central signal system software) and uploaded using scripts to PORTAL.

Appendix B contains the metadata description of the table records and a sample of the raw detector data.
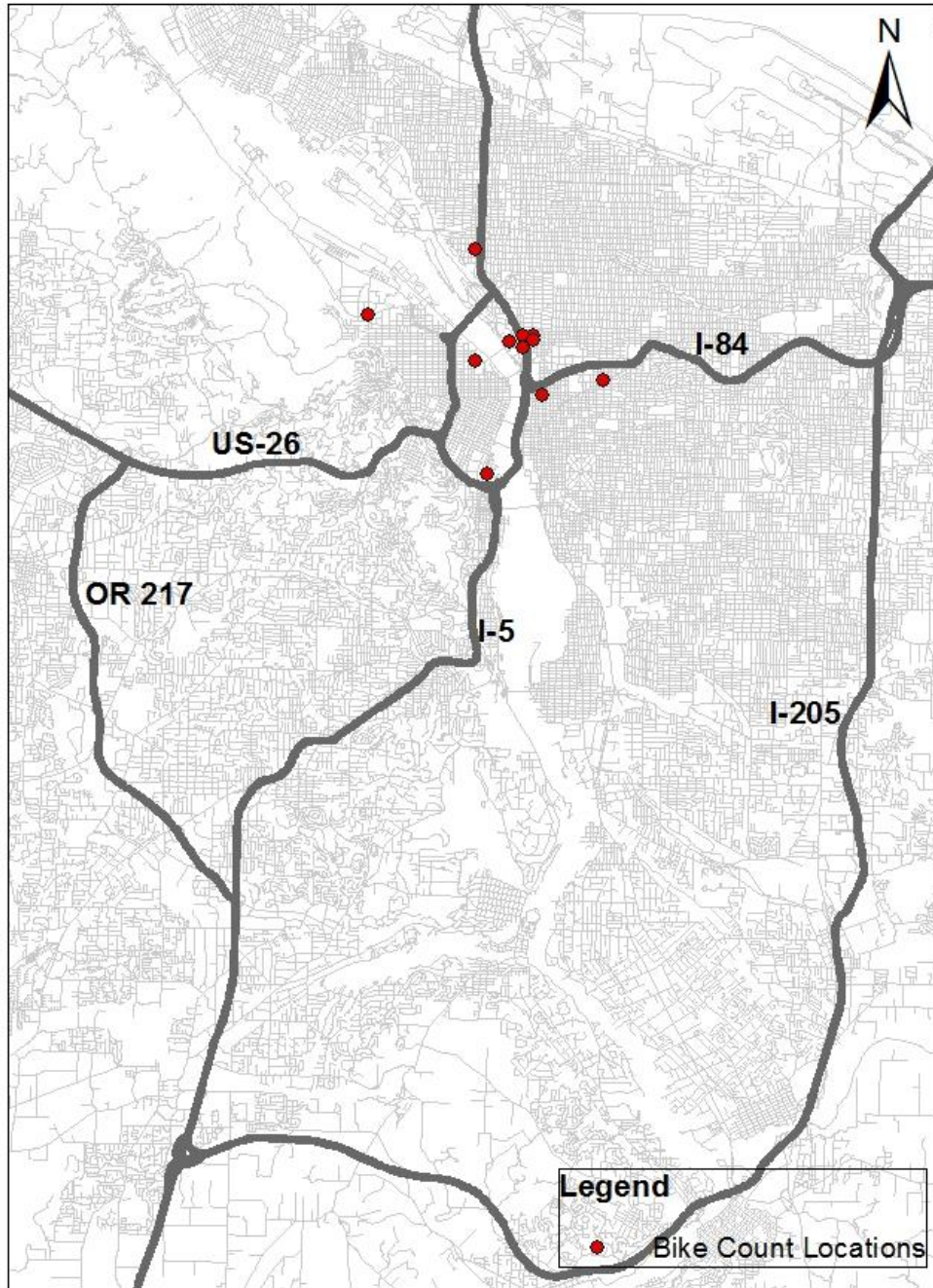
Figure 3.2 Locations of Bicycle Count System Detectors

3.2.3 **Pedestrian Counts**

The City of Portland operates more than 1,050 signalized intersections, with a mix of Type 170 and 2070 controllers. Type 2070 controllers are newer and are operated using Voyage controller software provided by Northwest Signal Supply, Inc. Voyage software can be used at isolated intersections or as a larger part of the central control system (Northwest Signal Supply). In

addition, another program called NWSCentral is also available to access the Voyage software from a Windows-based computer. Two methods were developed in this study to automate the process for pedestrian delay data collection using Voyage software. Both the methods relied on pedestrian push-button actuations to capture delay. In this research, we define pedestrian delay as the difference between the time when a pedestrian activates the push button and the time that the pedestrian phase is served. Previous research has shown that the delay calculated using *Highway Capacity Manual* (HCM) methods does not accurately reflect the delay observed in the field (Kothuri, 2013*).*

The pedestrian delay data in this study can only be collected for phases that are pedestrian actuated with push buttons. The National Electrical Manufacturers Association (NEMA) has adopted specific phase numbering standards (1-8) for defining phase movements. Through and left-turning movements are represented by even and odd numbers, respectively. Along major arterial corridors in the City of Portland, vehicle and pedestrian recall is implemented on the mainline phases (2 and 6), which ensures that these movements get served each cycle. Thus, the relative importance of measuring delay on these movements is less than the highly variable cross-street travel. Figure 3.3 shows the phase diagram for the intersection of Southeast 82nd Avenue and Tillamook Street in Portland. This intersection operates as a 6-phase intersection, with phases 2 and 6 serving the mainline through movements (Southeast 82nd Avenue south and north bound).
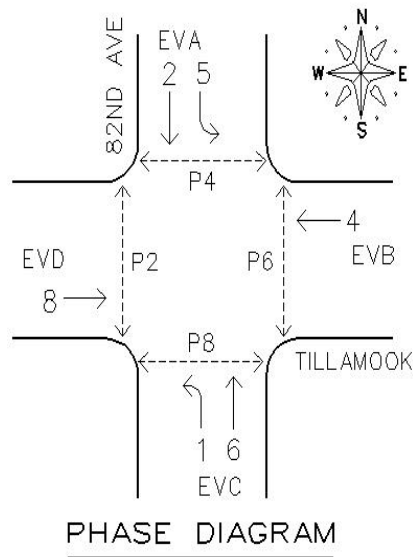
Figure 3.3 Phase Diagram

Pedestrian delay is captured for phases 4 and 8 (crossing the major arterial Southeast 82nd Avenue), which are not on recall and are equipped with push buttons. A description of the two methods used for collecting pedestrian waiting times is provided below.

13

3.2.3.1 Transit Priority Logging

Among the various advanced features present in the Voyage software is the ability to implement transit priority. At various signalized intersections, the City of Portland implements conditional transit priority if the bus is running late to either "extend the green phase" or "shorten the red phase," depending on the time in the cycle during which a bus arrives at an intersection. The transit priority log in the Voyage software records the time of the transit priority call and the time when the call was served for each event. Using the transit priority logging capability, logic was implemented in Voyage software to capture pedestrian delay (Kothuri, 2012). A pedestrian call is logged as a transit priority call. To record waiting time for pedestrians, we log the time when a pedestrian call is placed and when it gets served. At intersections where transit priority for buses is active, a sequence is needed to inform the controller regarding the order of priority for the two events. Currently priority is set equal, implying that the calls are handled on a first-come, first-served basis. The pedestrian push button is reassigned to a vehicle detector, and the logic is set up in a way such that a latch is on if the push button is activated and the walk phase is not active. The latch ensures that the transit priority input turns on and is only released when the walk is served. Additional pedestrian calls during the walk phase are ignored by the logging feature. This allows the system to keep track of the time the push button was activated and the time until the walk phase was served. One primary limitation of this method is that only one transit priority event can be active at a particular time. This limitation can hinder the actual transit priority operation if the priority sequence is not set correctly. In addition, at an intersection where there may be two pedestrian actuated phases and if the push buttons corresponding to both phases are actuated at the same time, only one pedestrian actuation event will be active and recorded at a particular time (meaning that the total delay is twice what is measured).

3.2.3.2 Volume Bin Logging

In addition to using transit priority logs, another method using internal timers in the controller was also implemented to record pedestrian delay, which is grouped into bins. Three bins are used for delay data collection: 0-20 seconds, 20-40 seconds and >40 seconds. When a pedestrian call is received, a latch is set using internal logic commands in Voyage software. Four timers per each pedestrian phase are activated when the latch is set. The first timer counts down from 20 seconds to 0 (timers 5/6); the second timer counts down from 40 seconds to 0 (timers 7/8); the third timer counts down from 0.1 second to 0 when the walk is served (timers 1/2); and the fourth timer counts down from 1 second to 0 when the walk ends and the clearance interval begins (timers 3/4). While the first and second timers classify the delay into the respective bins, the third and fourth timers keep track of beginning and end of the walk indication. The logic for classifying delay into bins is shown below: For each bin, the criteria listed below have to be satisfied in order for the delay value to be placed in that bin.

Bin 1: Delay between 0 - 20 seconds
   a. Latch 1 is set
   b. Walk indication is active
   c. Timer 1 $\neq 0$
   d. Timer 5 $\neq 0$

Bin 2: Delay between 20 – 40 seconds
    a. Latch 1 is set
    b. Walk indication is active
    c. Timer $1 \neq 0$
    d. Timer $7 \neq 0$
    e. Timer $5 = 0$

Bin 3: Delay greater than 40 seconds
    a. Latch 1 is set
    b. Walk indication is active
    c. Timer $1 \neq 0$
    d. Timer $7 = 0$

Each bin is reassigned to a vehicle detector, so that the counts (delay) from these bins can be obtained through the volume logs in TransSuite. For example, in Table 7 detectors 33 and 36 represent actuations in Bin 1, 34 and 37 represent Bin 2, and 35 and 38 represent Bin 3, respectively. While this method provides less detailed data (delays are binned rather than reporting actual time) than the transit priority log method, it is considered more accurate due to the inability of the transit priority log method to handle multiple pedestrian calls or the introduction of incorrect data when a bus transit priority event occurs at the same time. At this time, data from the Volume logs is manually downloaded in TransSuite® and uploaded to PORTAL. Efforts are underway to automate this data retrieval process. A sample of the pedestrian actuation and delay data from the volume log and the metadata description of the table records are shown in Appendix B.
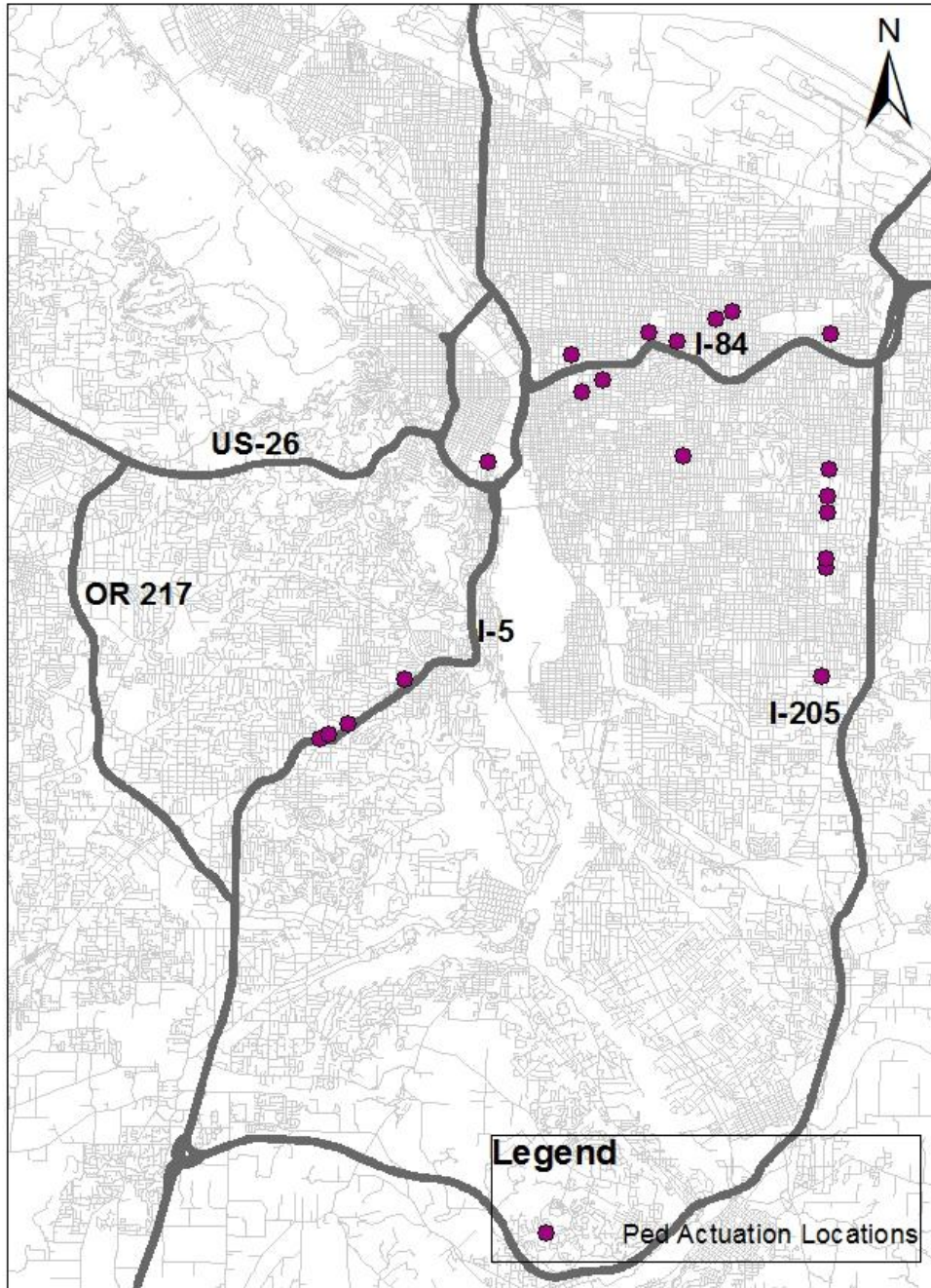
Figure 3.4 Locations of Bicycle Count System Detectors

3.2.4 **Transit Activity**

TriMet, Portland's regional transit provider, utilizes an automated Bus Dispatch System (BDS) for operational purposes. There are two components: AVL and Automatic Passenger Counter (APC) systems. The AVL system provides real-time location information that is used for scheduling and dispatch. These data collection system records stop-level information such as

arrival and departure times, dwell times, maximum speed achieved since prior stop, and number of boardings and alightings.

Bus data includes a schedule table and a table with stop event and automatic passenger counter (APC) data. The scheduled-stop timetable provides schedule information for the two-month time window. To facilitate ease of use, the schedule table includes an entry for each route and each day. This structure does replicate information, but makes it very clear which route schedule applies to which date. The stop event table provides information about bus activity including arrive and leave time from bus stops, passenger ons and offs, dwell time, lift usage and other detailed data. A stop record is created whenever a vehicle drives past a bus stop that it is scheduled to serve. In this case, the bus records the time that it arrived and left the stop's vicinity. The stop vicinity is currently defined as the area within a 30-meter (about 98 feet) radius of the recorded bus stop. A stop record is also created whenever a vehicle door is opened. Due to the age of the system, some buses have malfunctioning AVL units and thus a limited amount of stop data are missing. When the BDS unit fails to record times for one or two stops, the Katora program generates stop event rows with estimated times. These times are estimated by interpolating between the prior and following recorded times on the basis of the distance between the stops.

Appendix B contains the metadata description of the table records and a sample of the raw data for the transit activity.

3.2.5 **Traffic Signal Phase and Timing**

The phase time data contains information regarding the different phases at a particular signalized intersection. This archive is an important tool to assess the performance of the signal timing settings at each intersection. The types of data that are reported on a cycle-by-cycle basis are timestamp, id of the signal controller, and the current timing plan number in effect. A series of bit flag fields are used to indicate phase information such as the value of the local cycle timer, and phases that are displaying green and yellow at a particular time period. Phases that are displaying walk indication as well as active pedestrian and vehicle calls are also recorded. The mode of operation of the signal controller and whether or not the controller is online is also logged. Bit flag fields are a data compression technique that use the bits in an integer value to represent a series of yes/no flags. In computer hardware, an integer is often represented with 16 bits – each bit having a value of either 0 or 1. By setting each of the 16 bits to 0 or 1 as appropriate, those 16 bits can be used to represent 16 yes/no flags. Table 3.2 shows the values associated with each phase.

Table 3.2 Phases and Values

Phase Number	Value
1	1
2	2
3	4
4	8
5	16
6	32
7	64
8	128

To calculate the phases that are timing from the data, a matrix of possibilities was created, as shown in Table 3.3. For a particular phase, each column shows the possible values that could be recorded in the data file when that phase is active. This matrix was created using the standard NEMA phasing shown in Figure 3.3.

Table 3.3 Matrix of Phase Possibilities

Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8
1	2	4	8	16	32	64	128
3	3	12	12	17	33	68	132
17	18	68	72	18	34	72	136
33	34	132	136	48	48	192	192
19	19	76	76	19	35	76	140
35	35	140	140	49	49	196	196
49	50	196	200	50	50	200	200
51	51	204	204	51	51	204	204

For example a value of 196 (64+128) in Table 3.3 indicates that phases 7 and 8 are currently timing, similarly a value of 33 (1+32) indicates that phases 1 and 6 are currently timing. A sample of the phase and timing data is shown in Appendix B; the bit flags (greens, yellow, peds, ped_calls, veh_calls) are decoded using the values in Table 3.3.

As a further example, the overlays field represents overlaps that are currently displaying a green indication; a value of 1 in overlays indicates that the first overlap is displaying green; a value of 4 indicates the third overlap is displaying green. Similarly integer bit flags are used to represent phases with active vehicle and pedestrian calls as well as phases currently displaying yellow or green. For the yellow field, the value 136 (128+8) indicates that phases 8 and 4 are currently displaying yellow. These data are received automatically from the City of Portland, processed and loaded into PORTAL.

3.2.6 **Travel Times**

Readers are installed along certain segments of major corridors in the Portland metropolitan area that continuously capture MAC addresses via Bluetooth technology. MAC addresses are unique 48-bit addresses that are assigned to electronic devices by the manufacturers of these devices. Travel times, speeds and origin-destination (O-D) patterns are gathered by matching MAC addresses from between a pair of readers. The Bluetooth readers are data collection units (DCUs) that have been developed by Oregon State University. Each Bluetooth-based DCU records MAC addresses of passing vehicles and of the reader itself, date and timestamp, and strength of the signal. Data collected by these Bluetooth detectors is multimodal in nature, since the readers cannot distinguish between modes. Therefore, important considerations for this data include use of appropriate matching techniques and filtering of outliers to derive meaningful information. Currently these data are also transferred manually to PSU's Intelligent Transportation Systems Lab from the servers housed at the City of Portland. Currently, there are 11 locations within Portland where Bluetooth readers are currently deployed. Figure 3.5 shows the locations of the Bluetooth readers.

These readers are primarily present on Powell Boulevard and 82nd Avenue, both of which are major arterial corridors in Portland. The raw data file received from the city is a master file containing all Bluetooth data collected at that location. Each record in the raw data file is assigned a stationID by matching the Bluetooth MAC reader address with a file containing stationIDs and associated MAC reader address in the PORTAL database. Using a binary search process, the new data in the raw data file is identified. The new data at each location is extracted by comparing the date-timestamp of each record in the identified dataset with the latest date-timestamp in the PORTAL database. Only those records are retained whose date-timestamps are greater than the latest timestamp in the database. The obtained records are filtered to remove duplicates, by determining if each MAC ID has been already recorded within the last 20 minutes. A matching algorithm and filtering technique have been developed to obtain processed travel times from the raw data. The algorithm matches unique MAC addresses between two locations. Once an algorithm detects a unique MAC ID, it searches for the same MAC ID at the second location within a certain time range. The time range is adopted as a constraint in order to filter out trips with stops and to differentiate multiple trips occurring during a given time period. This time range is set as 20 minutes. Based on the vehicle's speed and the DCU's range of detection, a single MAC address can be read multiple times at a single location as the vehicle passes by the reader. Therefore, to get accurate travel times, matching can be performed in one of two ways: first-first match or last-last match. Our algorithm matches the MAC addresses based on first-first detection. Once a match has been found, the algorithm calculates travel time as the difference between the recorded timestamps at both locations.

The algorithm also filters the travel times for outliers. Each travel time value is checked against the previous 10 readings and is flagged as an outlier if the current value exceeds the mean of the previous 10 readings plus one standard deviation. Outliers are discarded from the final travel time's file. The final Bluetooth travel-times file contains a list of all matched travel times, from-and-to location IDs and a timestamp. In order to preserve anonymity, raw MAC addresses are not included in the output and are not stored in PORTAL. Appendix B shows a sample of matched Bluetooth travel times. The IDs of the Bluetooth readers, timestamp and matched travel times are included.
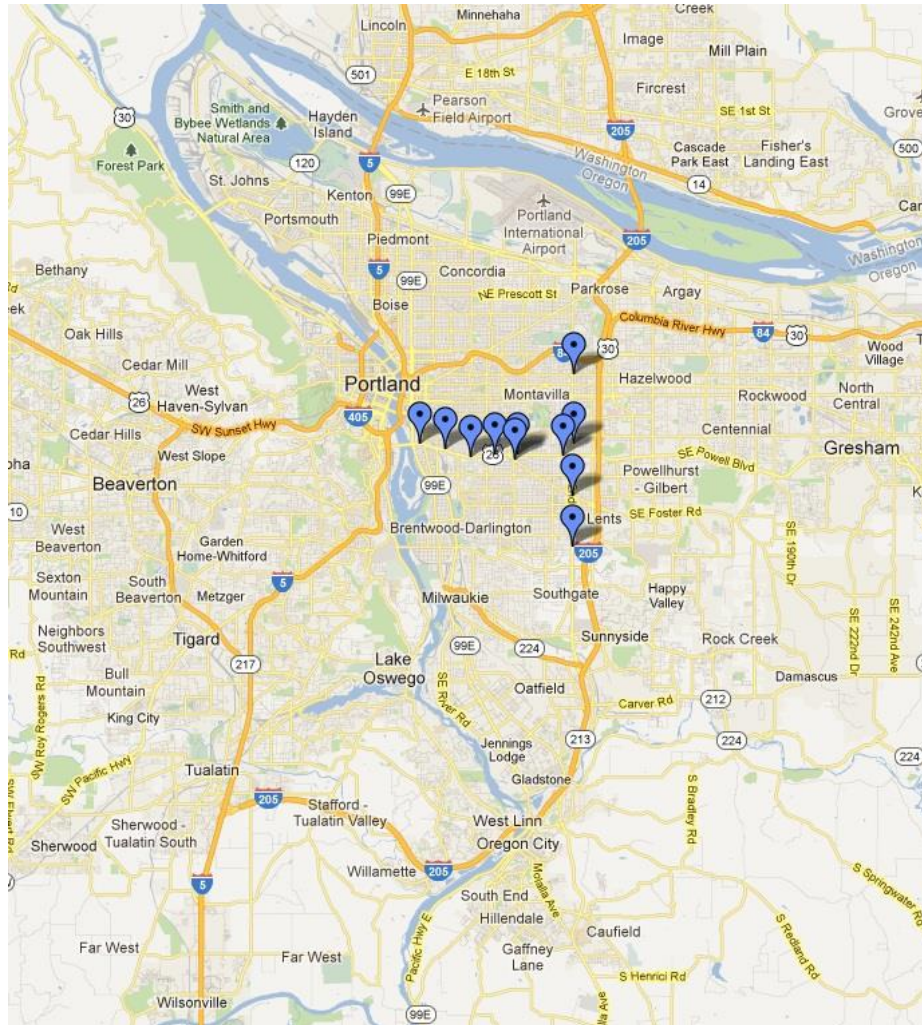
Figure 3.5 Bluetooth Reader Locations

3.3   SPATIAL AND TEMPORAL COVERAGE

Table 3.4 describes the spatial and temporal availability of the different types of arterial data. In general, temporal coverage is relatively good with many data sources available 24 hours a day, every day (24/7). Spatial availability ranges from very good – with intersection phase and timing data and transit activity available across the city to very limited availability for Bluetooth travel time data, which is installed along one corridor and bicycle and pedestrian counts, which have limited installation locations. Note that the Oregon Department of Transportation (ODOT) has expanded coverage of its Bluetooth system data in southwest Portland, but PORTAL is not currently archiving these data.

Table 3.4 Arterial Data Sources - Data Availability

Data Type Name	Spatial Availability	Temporal Availability
COP- Vehicle Counts	Available at intersections where count stations have been installed. Approximately 60 locations in City of Portland. Limited spatial availability.	Available 24x7 at 5-minute granularity.
COP-Travel Times	Bluetooth travel time available primarily along Powell Blvd with 2-3 stations along 82nd Ave. Very limited spatial availability.	Available 24x7, all Bluetooth readings are recorded and sent to PSU.
COP-Intersection Phase and Timing	Available at all intersections in City of Portland that feed into TransSuite. Good spatial availability.	Available 24x7 at cycle-level granularity.
Clark County – Vehicle Counts – Radar	Available where radar count stations installed. Approximately 10 stations in Clark County currently. Limited spatial availability, but expanding.	Available 24x7 at a 5-minute granularity.
Clark County – Vehicle Counts – Loop	Available where loop count stations installed. Limited spatial availability.	Available 24x7 at a 5-minute granularity.
TriMet – Transit Activity (AVL/APC)	Available where TriMet buses travel. Good spatial availability.	Available during times when transit service is running. Good daytime coverage.
C-TRAN – Transit Activity (AVL/APC)	Available where C-TRAN buses travel. Good spatial availability.	Available during times when transit service is running. Good daytime coverage.
COP – Bicycle Counts	Available where bicycle detectors installed in City of Portland. Approximately 15 installations. Very limited spatial availability.	Available 24x7.
COP – Pedestrian Actuation Counts	Available where pedestrian-specific detection hardware installed. Approximately 5 locations in City of Portland. Very limited spatial availability.	Available 24x7.

4.0   A FRAMEWORK FOR ARCHIVING

PORTAL stores data in a PostgreSQL-relational database management system (RDBMS). This archive implements a data warehousing strategy in that it retains large amounts of raw operational data for analysis and decision-making processes, and in that these data are stored independently of their operational sources, allowing the execution of time-consuming queries with no impact on critical operations uses. The RDBMS stores data physically on a redundant array of independent disks (RAID) with approximately 20TB of usable space; the RAID system provides both high-speed access and increased reliability through redundancy in the event of hardware failure. Offsite backups of the raw data are done regularly. A web interface that uses PHP, JavaScript, the Open Layers mapping API and the HighCharts graphing package provides data visualization tools, performance metrics and data downloads of the archived data.

The original and largest data source in the PORTAL archive is the freeway data. It is stored within a hierarchical framework consisting of detectors, stations, freeways and corridors. A detector represents a single dual-loop inductive loop detector; a station represents a set of detectors at a specific freeway location; a freeway consists of all stations along a single direction of a freeway; and a corridor consists of a set of stations along a directional freeway segment. The arterial schema was set up to mimic the existing freeway schema.

The multimodal schema conceptually consists of four components: detectors, stations, intersections and corridors, similar to the freeway network. Figure 4.1 is a visual representation of the data sensors and relationships that we intend to incorporate in the archive. A detector refers to an individual detector. A station is a set of related detectors. Stations are associated with intersections. Detectors are contained in stations and stations are contained in intersections, as shown in Figure 4.1.
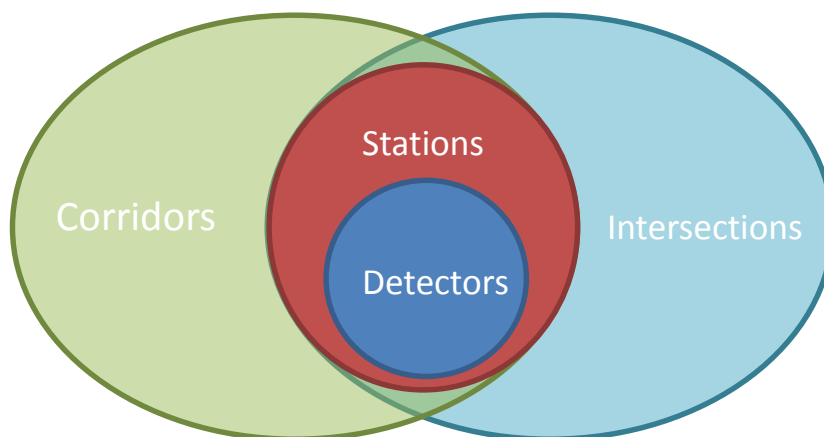
Figure 4.1 Relationship between Archive Components

A simple schematic of how this arrangement works in demonstrated by Figure 4.2.
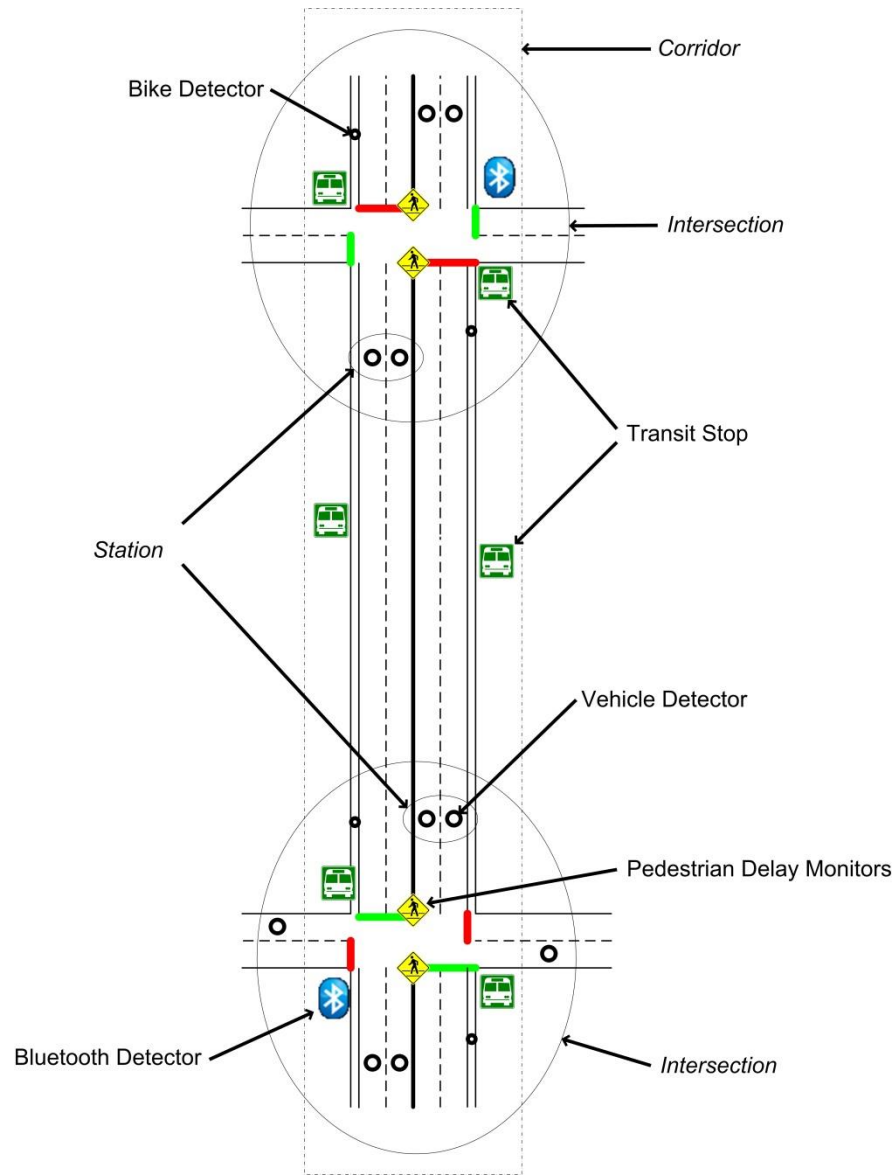
Figure 4.2 Arterial Archive Components

Vehicle detectors are represented by black circles. Presently, these are single-loop system detectors placed upstream of intersections with one detector in each lane for the purposes of counting. There can be up to one station per instrumented approach at an intersection (though at this time, most intersections are only instrumented on two approaches). Similarly, single-loop advance bike detectors (smaller open black circles) are placed upstream of the intersection in the bike lane. Pedestrian delay and actuations are being gathered at select locations using the signal controller and software. Details are presented elsewhere *(Kothuri, 2012)*. In Figure 4.2, the pedestrian delay monitors are represented by the pedestrian crossing warning sign. The transit data are organized around stops and routes. The stops are shown represented by the bus signs.

Finally, Bluetooth monitoring stations gather MAC addresses from passing vehicles, bicyclists and pedestrians. These are represented in the figure with the Bluetooth logo.

All of these elements can be associated with one or more higher-level groups. In correspondence with the concept of a freeway station, we have defined an arterial station to be a grouping of vehicle detectors by approach. Stations are contained in intersections and corridors. An intersection is a grouping of all detectors within the bounds of the functional area of the intersection. The grouping includes stations (vehicle detectors), pedestrian delay monitors, transit stops, bicycle counters, and Bluetooth detectors. A corridor is a grouping of detectors along the roadway. Logically an intersection may belong to two (or more) corridors; for example, a corridor for the main street and a corridor for the cross street. Only elements of the intersection associated with the corridor direction are associated with the corridor. Finally, sensors placed midblock are only associated with a corridor, not an intersection. The database contains tables associated with each of the four concepts: detectors, stations, intersections and corridors, in addition to data tables containing system detector and intersection signal data. A description of the database schema is provided in the next section.

4.1   DATA CLASSIFICATION

In the sections above, we described the high-level archive framework. In this section, we provide a classification of these data by spatial extent and collection location, and list applicable metrics for each data type. The classification of these data affects which arterial performance metrics these data can be used for.

For the spatial extent classification, we classify arterial data into two spatial types or extents: point-based data and segment-based data. Point-based data are collected by a detector at a specified location – for example, a volume detector at a fixed location. Segment-based data are collected over a segment of arterial roadway – for example, collection of travel times based on timestamps of vehicles entering and exiting the segment. The interrupted-flow nature of arterials amplifies the differences in usage of these two types of data. On non-interrupted flow roadways such as freeways, point-based data and segment-based data are somewhat interchangeable – point-based data can be used to approximate segment-based data. On interrupted-flow arterials, these two data types are not interchangeable and point-based data cannot be used to approximate segment-based data. For point-based data, we divide into point-based data collected at midblock. Two primary collection locations on arterials are intersection and midblock. As with spatial extent, detection location affects what the data can be used to calculate.

Primary performance metrics on arterials are volume and travel time. Volume is a point-based metric while travel time is a segment-based metric. Volume requires point-based data, and travel time requires segment-based data. Volume data are commonly collected. Table 4.1 provides a location and metric-based classification for arterial data.

Table 4.1 Arterial Data Classification – Location and Metrics

Data Type Name	Segment vs. Point (Intersection, Midblock)	Potential metrics this data are applicable to:
COP- Count Station	Point – midblock	Arterial volume, volume to capacity
COP-Travel Time	Segment	Arterial travel times
COP-Signal Cycle Data	Point – intersection	Traffic signal performance
Clark County – Count Station – Radar	Point – midblock	Arterial volumes, volume to capacity. Speeds could be used for rough travel time estimates.
Clark County – Count Station – Loop	Point – midblock	Arterial volume, volume to capacity
TriMet – Transit AVL/APC Data	Point – bus stop	Transit arterial travel time, and transit-specific metrics such as loads and reliability.
C-TRAN – Transit AVL/APC Data	Init System	Transit arterial travel time, and transit-specific metrics such as loads and reliability.
COP – Bicycle Data	Point – intersection	Bicycle volumes
COP – Pedestrian Actuations	Point – intersection	Pedestrian delay

4.2 PORTAL DATABASE STORAGE

Table 4.2 shows a listing of the files that are created in the database to store the arterial data and their descriptions. These data sources are described in Chapter 3. The data dictionaries for all the files listed in Table 4.2 are included in Appendix B. In the database organization, system detectors are classified in a hierarchical organization consisting of detectors, stations and intersections. A detector refers to an individual loop detector. A station is a grouping of loop detectors by approach. Stations are contained in intersections. An intersection is a grouping of all detectors within the bounds of its functional area. Phase and timing data are classified primarily by intersectionID. The Bluetooth readers on arterials are classified in a hierarchical manner consisting of detectors, stations and intersections. A detector refers to an individual Bluetooth reader. A station is a grouping of Bluetooth readers at a particular location. Stations are contained in intersections. An intersection is a grouping of all detectors within the bounds of its functional area. This classification is similar to the classification of system detector data described previously.

Table 4.2 List of Arterial System Tables

Table Name	Primary Key	Description
arterial_detectors	Detectorid	Metadata for each detector in these dataset.
arterial_intersections	Intersectionid	Metadata for each intersection in these dataset.
arterial_stations	stationid	Metadata for each station in these dataset.
raw_detector_archive	detectorid	Primary loop detector data table. Contains one-minute volume, occupancy and data quality flags.
Phase_and_timing_data	Intersectionid	Primary signal timing and phase data table. Contains bit flags for phases that are currently displaying green, yellow, pedestrian and vehicle call indications.
Bluetooth_stations	Bluetoothid	Metadata for Bluetooth stations
Bluetooth_traveltimes	From_bluetooth_id,to_bluetooth_id Timestamp	Primary travel time table. Contains matched travel times between pairs of Bluetooth detectors.
Ped_actuations	Timestamp, intersectionid	Primary pedestrian data table. Contains push button actuations and binned delay for select intersections.
Bike_counts	Timestamp, intersectionid	Primary bicycle data table. Contains bicycle counts at select locations.

5.0   ARCHIVING PROCESS

A variety of arterial data are collected and archived in the PORTAL transportation database. We describe the data collection and archival process for the arterial data types discussed in the previous chapters.

5.1   OVERVIEW

Table 5.1 describes in detail the data retrieval process, coverage and status for these data types described in the previous sections. We note that most data collection and archival processes are semi-automated – involving some human work and some computer work. In some cases, the semi-automated process is an effective near-term solution while the long-term solution of a fully automated feed is acceptable. In other cases, the overhead of the semi-automated process is too high. Information about the current retrieval process, the interval at which current data are retrieved, an evaluation of the level of overhead of the semi-automated process as well as status and limitations are all discussed. These tables provide an excellent summary of the current data retrieval processes for the PORTAL archive. Following the tables, sections on lessons learned for these data types are included.

The majority of the arterial data are currently retrieved by semi-automated download processes. In general, for arterial data, the data are transferred from the sensors to a central system (such as a central signal system). Given the current state of most central signal systems, fully automated data download is not currently possible. Work is in process with the central system vendors to create data feeds that can support fully automated download. Until those feeds are created, we use semi-automated processes in which a human user is involved in doing part of the download work. For a semi-automated download to be feasible, the amount of human involvement needs to be below a certain threshold; a level of one-two hours per month generally seems acceptable. Above this threshold, we consider the download infeasible.

Table 5.1 Arterial Data Archival – Data Retrieval Process, Interval and Status

Data Type Name	Data Retrieval Process	Data Retrieval Interval & Level of Effort	Data Retrieval Status & Limitations
COP- Count Station	TranSuite data are stored in files on the TransSuite server. Graduate student loads files from TransSuite onto a flash drive and brings it to PSU. PORTAL staff run scripts to load these data into the PORTAL database.	Periodically. Depends on availability of student to download data from city. Load process into PORTAL is straightforward. Overhead of semi-automated process is acceptable in the short term.	Data loaded through early December. TransSuite stores data for two months; transfer must be done at least every two months to avoid losing data.
COP-Travel Time	Data are stored on the Bluetooth devices. Scripts must be run on a City of Portland server to download data from collection devices in the fields. Files are loaded onto a flash drive and brought to PSU by a graduate student. PORTAL staff run scripts to load these data into the PORTAL database	Periodically. Depends on availability of student to go to the city, run the scripts and bring these data to PSU. Load process into PORTAL is currently functional, but scripts are brittle and may need Overhead of semi-automated process is higher than desired for the short term.	Loaded through early December. Bluetooth devices store up two weeks of data. Transfer must be done every two weeks to avoid data loss.
COP-Signal Cycle Data	The Green Driver company produces a feed of the City of Portland's signal cycle data. This feed is	Signal cycle changes. Data retrieval is fully automated	Data retrieval is ongoing. Data will be stored in the PORTAL database once the new storage is fully in place.
Clark County – Count Station – Radar	Clark County staff generates data files using the Wavetronix report generation system. These files are manually transferred to PSU. PORTAL staff uses scripts to load these data into the PORTAL database.	An interval for loading this data has not been established. Overhead of the semi-automated process is acceptable in the short term.	Sample data loaded. PSU is prepared to establish a periodic semi-automated transfer process for this data.
Clark County – Count Station – Loop	Clark County staff generates data files using the ATMS. Now reporting system. These files are manually transferred to PSU. PORTAL staff uses scripts to load these data into the PORTAL database.	An interval for loading this data has not been established. Overhead of the semi-automated process is not acceptable. The overhead of	Sample data loaded. Data feed must be established before more data can be loaded.
TriMet – Transit AVL/APC Data	TriMet staff creates a dump of data tables in their archive and provide those files to PSU. PORTAL staff loads the files into the PORTAL archive.	Data are transferred to PSU as needed. Overhead of the semi-automated process is acceptable in the short term.	Several samples of quarterly data loaded. Loading of transit data was limited by disk space limitations. These limitations have been removed and expect to load transit data on a regular basis in the near future.

Data Type Name	Data Retrieval Process	Data Retrieval Interval & Level of Effort	Data Retrieval Status & Limitations
C-TRAN – Transit AVL/APC Data	C-TRAN staff generates data files using the Init reporting system. These files are manually transferred to PSU. PORTAL staff uses scripts to load these data into the PORTAL database.	Data are transferred to PSU as needed. The overhead of the semi-automated process is acceptable in the short term.	Sample data loaded. PSU is prepared to establish a periodic semi-automated transfer process for this data.
COP – Bicycle Data	Bicycle data are stored in the TransSuite system. Graduate student retrieves data from TransSuite, generates files and manually transfers them to PSU. Graduate student runs data through a macro to filter out extraneous data. PORTAL staff uses scripts to load these data into the PORTAL database.	Periodically. Depends on availability of student to go to the city, run the scripts and bring these data to PSU as well as run analysis. Load process into PORTAL is straightforward. Overhead of semi-automated process is higher than desired for the short term.	Loaded through early December.
COP – Pedestrian Actuations	Pedestrian actuation data are stored in the TransSuite system. Graduate student retrieves data from TransSuite, generates files and manually transfers them to PSU. PORTAL staff uses scripts to load these data into the PORTAL database.	Periodically. Depends on availability of student to go to the city, run the scripts and bring these data to PSU as well as run analysis. Load process into PORTAL is straightforward. Overhead of semi-automated process is higher than desired for the short term.	Loaded through early December.

5.2 OBSERVED CHALLENGES

5.2.1 Arterial Count Stations

The PORTAL team has worked with three different data systems for accessing counts and other data – TransSuite, Wavetronix and ATMS.Now (Naztec). The complexity of getting data from the three systems varies significantly. TransSuite stores its data in files on disk. These files are retrieved from disk and loaded into the PORTAL database. This type of functionality works reasonably well for a semi-automated process, as most data that we desire to archive exists in text file format. In contrast, Wavetronix stores its data in a database. The Wavetronix system has reasonable reporting mechanisms so that a report can be created to output enough data (data for a long enough period for all detectors) so that outputting that report isn't overly human-intensive and makes for a reasonable semi-automated process. ATMS now stores its data in a database, but has limited reporting capability. Regarding the extraction of data from the ATMS, interface is too human-intensive and is not a feasible alternative for moving data to the PORTAL archive.

Discussions are occurring with both TransSuite and Naztec vendor to create data feeds that will support fully automated downloads.

5.2.2 Transit Data

The PORTAL team has worked with two transit providers – TriMet and C-TRAN. C-TRAN uses the Init system to collect AVL and APC data. TriMet has recently transitioned from Orbital to Init. While TriMet and C-TRAN use the same data collection and storage system, they have distinctly different data processing methodologies. TriMet has staff assigned to data analysis; this staff member retrieves, processes and cleans the TriMet data to produce numerous internal reports. To transfer data to PSU, TriMet staff dumps tables from their database and send the resulting files to PSU where they are loaded into PORTAL. This process is working effectively for the quarterly data; a different process will be needed when the move is made to use daily TriMet data. The motivation for using quarterly data, but not daily data, is the numerous transit schedule changes. By using quarterly data, we can minimize the effects of those changes.

Data from C-TRAN is retrieved through output of reports from the Init system. These data are processed in a similar way as the data retrieved from TriMet. The C-TRAN data are cleaned less before it reaches PORTAL than the TriMet data. This is an issue that will need to be addressed as more C-TRAN data are loaded into PORTAL.

5.2.3 Bluetooth Data

The City of Portland has placed Bluetooth readers at locations along Powell Boulevard and 82nd Avenue. The Bluetooth readers store data local to the devices. To retrieve that data, the devices are periodically polled using batch scripts run on a server at the City of Portland. Those files are then transferred to PSU, where the Bluetooth readings are matched and cleaned and the resulting travel times are loaded into PORTAL. The PORTAL interface supports display of Bluetooth travel times. Work has recently been completed to automate the Bluetooth process. The process is now nearly fully automated.

5.3 DEVELOPING A PORTAL DATA FEED SPECIFICATION

Data for transportation archives is available from a number of sources ranging from data feeds designed for the archive to tables dumped from operations databases to CSV (or other file format) reports exported from operational systems. Incorporating data from such diverse sources can be challenging and, in the worst cases, requires archive implementers to do custom work for each data source. Specifying common data formats can be specified for each type of data to be archived allows data provided in these common formats to be incorporated into the archive with limited work from the archive implementers. In Appendix A, we provide draft specifications for PORTAL arterial count and Bluetooth data feeds. For transit data, we recommend using Google Transit Feed Specification (GTFS) for stop and schedule information, and GTFS-real-time for real-time arrival information.

The draft PORTAL feed specifications are based on the GTFS, which is rapidly gaining acceptance for transit schedule and real-time arrival information. The intention of these specifications is that data feeds that meet the formats in the specifications can be directly incorporated into PORTAL.Work is in progress in both Portland and Vancouver to develop

arterial counts and Bluetooth data feed, and the feed specifications provide guidance for the development of these feeds. The feed specifications also serve the dual purpose of specifying the metadata required for each of these detection types.

6.0   EXPLORING DATA QUALITY

As with any automated sensor system, data quality and completeness is an important issue. Sensor or communication failures can leave data gaps; sensor errors can produce incorrect data. The challenge of automating this data quality effort is large—this chapter explores methodology for reviewing the data quality of the arterial vehicle, bicycle and pedestrian count.

6.1   OVERVIEW

Table 6.1 provides a summary of the status of the data quality and validations efforts to date.

Table 6.1 Arterial Data Sources – Software Systems and Data Quality

Data Type Name	Software System	Data Quality and Validation
COP- Count Station	TransSuite	Data have not been validated by PSU. Some efforts underway to develop automated quality metrics
COP - Travel Time (Bluetooth)	Custom software running on City of Portland server.	Data have not been validated by PSU.
COP-Signal Cycle Data	TransSuite. Data fed to PSU through Green Driver company.	Data quality has not been verified, but quality is likely high as data are cycle changes reported by the signal controllers and does not involve sensors.
Clark County – Count Station – Radar	Wavetronix	Data quality has not been verified by PSU.
Clark County – Count Station – Loop	ATMS.Now (Naztec)	Data quality has not been verified by PSU.
TriMet – Transit AVL/APC Data	TriMet in-house database based on Orbital System. TriMet moving to Init system.	Data quality verified by TriMet. Data are of high quality.
C-TRAN – Transit AVL/APC Data	Init System	Data quality has not been verified by PSU. Data are coming directly from C-Tran Init system.
COP – Bicycle Data	TransSuite	Data quality has been verified by PSU. See section 6.3,
COP – Pedestrian Actuations	TransSuite	Data quality has been verified by PSU. See section 6.4,

6.2   VEHICLE COUNTS

It is important to understand and document the quality of the data that is being archived prior to any analysis. Documenting detector health can assist agencies in targeting resources towards malfunctioning detectors. This is especially important when detectors are informing real-time operational strategies. Similar to the detector quality monitoring that we have implemented for freeway detectors, detector health can be communicated by visual inspection or logical flags and upper reasonable and lower bounds for count, speed, and occupancy values from the count stations. An example of a malfunctioning detector is shown in Figure 6.1.

The Sunnyside at 117th eastbound station is comprised of three detectors located in the left, center, and right through lanes for traffic heading eastbound on Sunnyside Road. The plot indicates that detector #7217 has flow readings much higher than that of the other two detectors. Early-morning volumes are unreasonable and peak flows for a single lane of urban arterial are also suspicious. This higher volume could indicate a need for detector #7217 to be calibrated or replaced.
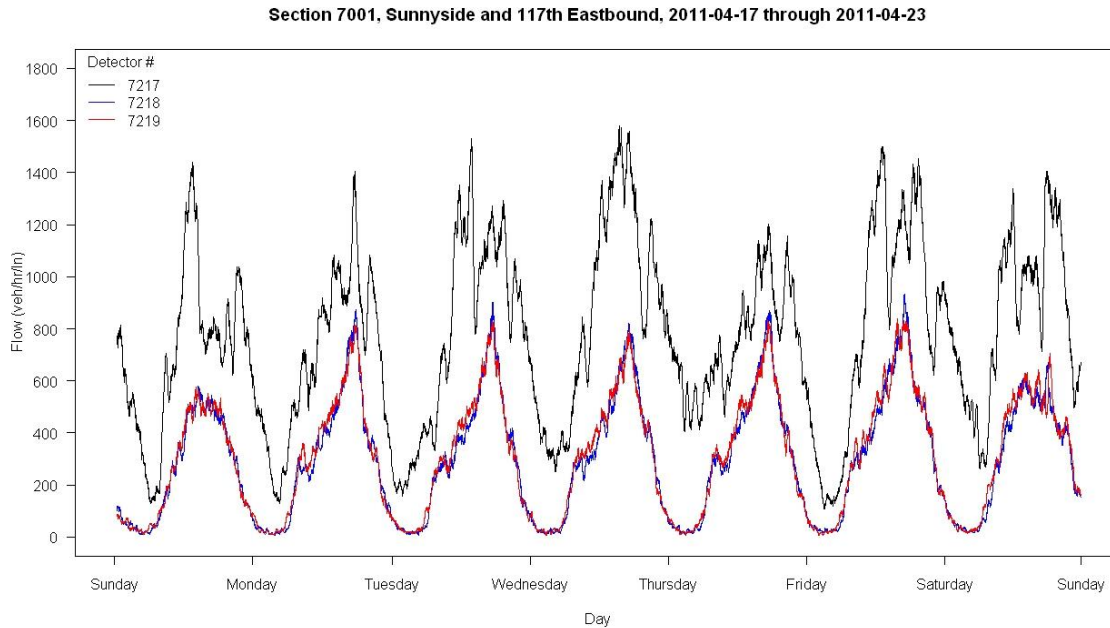
Figure 6.1 Time-series of Vehicle Counts at Eastbound Detectors at Sunnyside and SE 117th

Data quality flags are assigned for volume and occupancy values based on maximum and minimum theoretical limits. These tests are listed in Table 6.2. The tests apply to data aggregated by the sample period. The flags are Boolean and can either be true or false. Each test includes a condition; data samples that satisfy that condition are considered improbable and may indicate a malfunction and the flag is set to "true." For example, a report of a one-minute count greater than 30, which corresponds to a flow of over 1,800 vehicles/hour/lane, is considered improbable.

These data quality flags appear as a Boolean field for each data record. The flags are either true or false. These data quality flags for volume are set to true if the volume is negative or greater than the theoretical maximum for the sample period (calculated based on saturation flow). Similarly, these data quality flags for occupancy are set to true if the occupancy is negative or greater than 1,000 (occupancy is recorded in tenths of a percent). Finally, the visual inspection flag is set to true if the volumes or occupancies look improbable (for example, consistent high volumes, but lower than the max volume, at night).

Table 6.2 Vehicle Count Post-Processed Data Quality Flags

Condition	Condition Name	Flag
Volume < 0	dq_minvol	T
Volume > (s*sat flow)/3600	dq_maxvol	T
Occupancy < 0	dq_minocc	T
Occupancy > 1000	dq_maxocc	T
Visual Inspection	dq_visual	T

6.2.1 **Automated Non-functioning Detectors Report**

A non-functioning detectors report was created to identify detectors that do not produce data. This report is a first step in improving the quality of arterial data. These detectors need to be investigated by City of Portland operations and maintenance staff to determine why they are not producing data. The non-functioning detectors report for November 5, 2012, is provided in Appendix C.

6.3 **BICYCLE COUNTS**

Prior to using the bicycle counts obtained from loops for analysis, verification of the counts for accuracy is essential. Work is currently underway to verify the accuracy of counts obtained from all existing bicycle count locations. The objective in the future is to develop adjustment factors at each location to minimize error and enable the measured counts to be accurate. A limited duration test for accuracy was conducted at one intersection – North Wheeler and Williams. Video data was recorded using existing traffic surveillance cameras deployed by the City of Portland. This data was collected for a 1.5-hour peak period in the afternoon (5:30-7 p.m.) on May 10 and 11, 2011. Bicycle counts measured by the advance loops at this intersection were compared to those obtained from video data manually. Figure 6.2 shows an aerial shot of the location.

The test location was a three-legged intersection, with advance loops present on North Wheeler Avenue in the north and south bound bicycle lanes, respectively. Fifteen-minute bicycle counts obtained from the loops were compared to the 15-minute counts obtained from video data. A time sync was performed to ensure that counts were compared during the same time period.
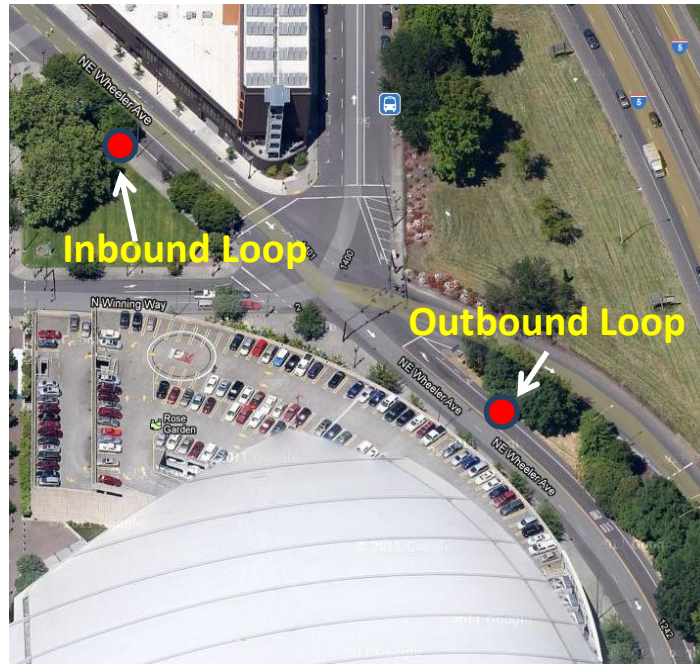
Figure 6.2 Bicycle Count Verification Location

Figure 6.3 shows the northbound (outbound) counts on May 10 and 11, 2011, during the PM peak. The plots show expected trends; bicycle counts are lower in the inbound direction and higher in the outbound direction during the PM peak period for the comparisons between the loop and video counts. The plots also show that counts produced from the loop counts are underestimated when compared to video (manual) counts, confirming a similar finding by prior research (Berkow, 2009). The mean absolute percent error (MAPE) was calculated as ((video count – loop count)/ video count) *100. Taking the absolute values prevents the positive and negative values from canceling each other out. MAPE was estimated as 14.15% and 14.75 % for the inbound and outbound directions, respectively. There are two possible reasons for underestimation: lower sensitivity setting of the advance loops and bicycles not riding over the detector in the bicycle lane. The sensitivity parameter setting allows a loop to distinguish between bicycles and vehicles. Adjusting the sensitivity to a higher value could possibly enable the detection of more bicycles. The other source of error could be attributed to the position of a cyclist in the bicycle lane. The advance detector can only detect bicycles that go over the detector and, therefore, will not be able to count any bicycles that do not travel in the lane and/or are not positioned correctly over the detector. A site visit to this intersection confirmed the hypothesis that some bicyclists do not ride in the bicycle lane and do not get counted.

Understanding the accuracy of the counts at each location is important if these data are intended to be used for the development of performance measures. Calibration of the loop, adjustment of sensitivity parameter, site visit to the location to observe bicyclist riding patterns, and development of adjustment factors are recommended to ensure accuracy of the counts from inductive loops.

Inbound Bicycle Counts (5/10/11)

Inbound Bicycle Counts (5/11/11)

Outbound Bicycle Counts (5/10/11

Outbound Bicycle Counts (5/11/11)
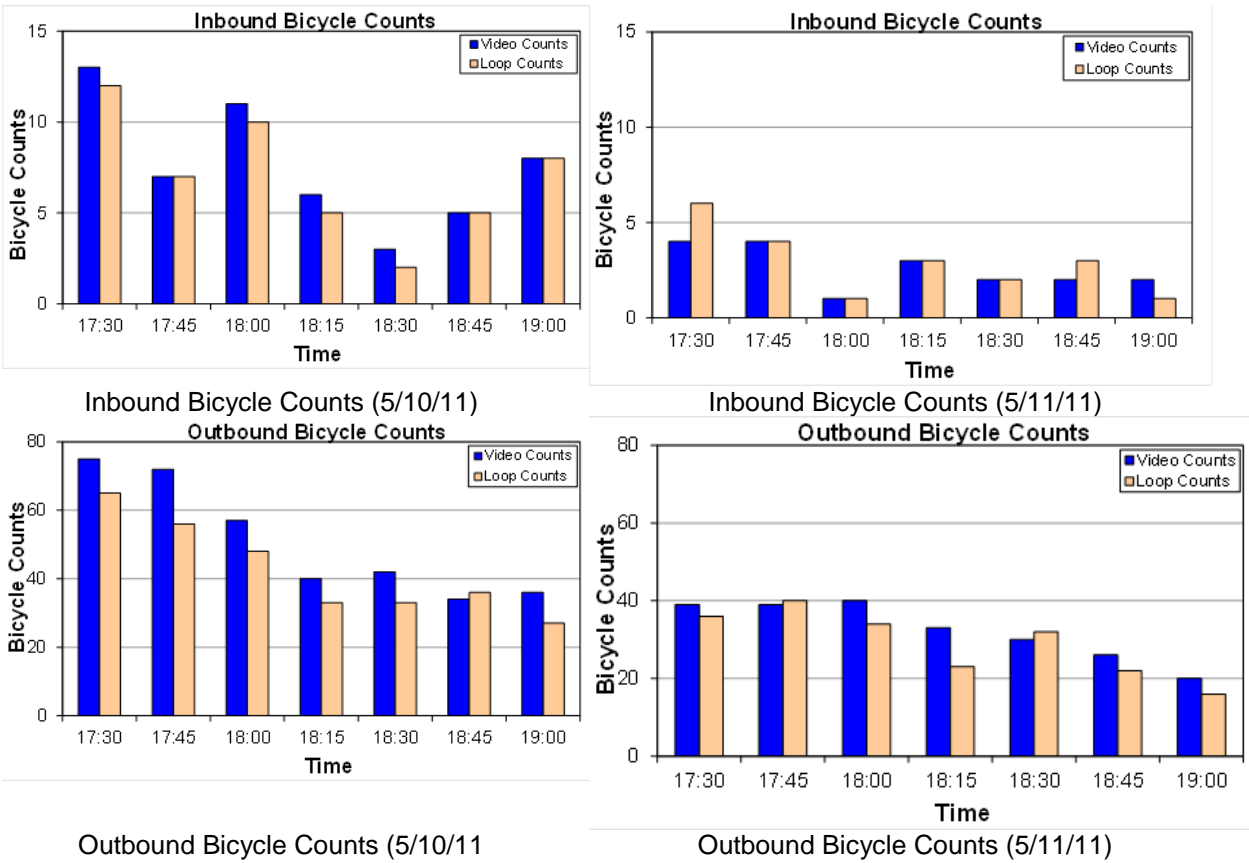
Figure 6.3 Count Plots of Inbound and Outbound Bicyclists at Validation Sites

6.4 PEDESTRIAN COUNTS

In order to develop pedestrian performance measures, custom internal logic commands were developed within Voyage, the signal controller software used by the City of Portland. These commands allowed the ability to record actuations and delay for actuated pedestrian phases. Verification of these data obtained automatically from the controller was performed using manually collected data at the same location. This verification was performed at Northeast Sandy Boulevard and 50th Avenue, a half-signal location in Portland. At this intersection, the approaches on Northeast Sandy Boulevard were controlled by a signal, whereas 50th Avenue was stop-controlled for vehicles. In order to cross Sandy, pedestrians have to activate push buttons present at each of the four corners of this location. Figure 6.4 shows the location where the verification was performed.

Figure 6.4 Pedestrian Actuation Verification Location

The verification was performed on Nov. 8, 2012, between 3:15-5 p.m. Observers were positioned at both crosswalks and recorded the time of the first push-button actuation while the pedestrian signal indication displayed a solid, "Don't walk" (solid red hand). The time when the walk indication was displayed was also recorded, and the difference between the two recorded times is a delay. Table 6.3 shows the comparison between the automated and manually collected actuation and delay data. This data shows that the actuations matched up perfectly between the two data sets. In the majority of the cases, the delay values also matched up, and minor discrepancies could be attributed to differences with syncing clocks. There were two instances where discrepancies in estimated delay were greater than 10 seconds between the automated and manual data.

Table 6.3 Pedestrian Actuations and Delay Verification

	Automated data from controller			Manually collected data		
Date	**Push-button activation**	**Walk**	**Delay**	**Push-button activation**	**Walk**	**Delay**
11/8/2012	15:15:11	15:15:55	44	15:15:11	15:15:55	44
11/8/2012	15:17:23	15:18:15	52	15:17:24	15:18:15	51
11/8/2012	15:20:58	15:21:45	47	15:20:58	15:21:46	48
11/8/2012	15:23:23	15:24:05	42	15:24:29	15:25:07	38
11/8/2012	15:24:34	15:25:15	41	15:24:38	15:25:15	37
11/8/2012	15:26:36	15:27:35	59	15:26:36	15:27:36	60
11/8/2012	15:28:34	15:29:55	81	15:28:37	15:29:55	78
11/8/2012	15:37:55	15:39:13	78	15:37:28	15:39:16	108
11/8/2012	15:46:39	15:47:33	54	15:46:40	15:47:33	53
11/8/2012	15:48:52	15:49:13	21	15:48:53	15:49:14	21
11/8/2012	15:50:13	15:50:53	40	15:50:13	15:50:54	41
11/8/2012	15:52:45	15:54:13	88	15:52:46	15:54:14	88
11/8/2012	15:55:28	15:55:53	25	15:55:28	15:55:54	26
11/8/2012	15:56:45	15:57:33	48	15:56:45	15:57:45	60
11/8/2012	16:01:26	16:02:33	67	16:01:26	16:02:33	67
11/8/2012	16:03:04	16:04:13	69	16:03:04	16:04:16	72
11/8/2012	16:05:05	16:05:53	48	16:05:05	16:05:53	48
11/8/2012	16:07:59	16:09:13	74	16:08:00	16:09:14	74
11/8/2012	16:14:59	16:15:53	54	16:15:00	16:15:54	54
11/8/2012	16:19:37	16:20:53	76	16:19:38	16:20:54	76
11/8/2012	16:27:08	16:27:32	24	16:27:08	16:27:32	24
11/8/2012	16:28:41	16:29:12	31	16:28:42	16:29:14	32
11/8/2012	16:31:00	16:32:33	93	16:30:55	16:32:30	95
11/8/2012	16:33:36	16:34:12	36	16:33:55	16:34:08	13
11/8/2012	16:34:42	16:35:52	70	16:34:30	16:35:48	78
11/8/2012	16:38:48	16:39:13	25	16:38:43	16:39:08	25
11/8/2012	16:41:17	16:42:32	75	16:41:12	16:42:28	76
11/8/2012	16:42:58	16:44:13	75	16:42:54	16:44:08	74
11/8/2012	16:45:06	16:45:52	46	16:45:01	16:45:49	48
11/8/2012	16:47:03	16:47:33	30	16:46:59	16:47:28	29
11/8/2012	16:48:15	16:49:13	58	16:48:11	16:49:08	57
11/8/2012	16:50:27	16:50:53	26	16:50:22	16:50:48	26
11/8/2012	16:52:29	16:54:12	103	16:52:25	16:54:13	108
11/8/2012	16:58:27	16:59:13	46	16:58:22	16:59:08	46

7.0   DATA VISUALIZATIONS

Data can be made more useful though the use of improved visualization techniques. As described in Chapter 2, there are five data sources associated with arterial networks – vehicle counts from system detectors, signal phase and timing data, travel times from Bluetooth readers, and bicycle and pedestrian counts. Visualizations and performance measures for each of these data sources have been developed in PORTAL and are described below.

7.1   VEHICLE COUNTS

Visualizations for system detector vehicle count data can be found under the Arterial tab on PORTAL's demonstration site (see demo.portal.its.pdx.edu). A map showing the location of the count stations is provided on the front page. The map helps the users to spatially identify the locations of current and functioning vehicle count data.

Figure 7.1shows the main map for the PORTAL Arterial Count Stations Page. The count stations in the City of Portland are displayed on this map. Each station is color-coded by direction of travel. Clicking on a station provides information about the name and location of that station and a link that can be clicked on to view data for that station.

Figure 7.2 shows the charts on the Arterial Count Station Detail Page. The top chart shows volume over the selected two-day period. The presence of an AM peak on both days can be seen. The bottom chart in Figure 7.2 shows a comparison of volume for the two selected days and it shows that the volume is similar for the two selected days. The top of the detail page contains a set of selectors to allow the user to select these data to be plotted. Date and time ranges, as well as resolution and days of week, can be selected.

For example, the selectors on the detail page allow users to create plots of volume in varying resolutions (five-minute, 15-minute, one-hour, one-day, one-month and one-year) are provided. Figure 7.3, Figure 7.4 and 7.5 show the volume plots for one day at Southwest Naito Parkway and Jefferson Street at five-minute, 15-minute and one-hour resolution, respectively.

Trends in counts can also be viewed on a weekly or monthly basis. Figure 7.6 and Figure 7.7 show the volumes for one week and one month, respectively. These plots are useful in visualizing daily and weekly trends. Figure 7.7 indicates that weekday volumes are fairly constant and a drop in volume is recorded during the weekend.

Figure 7.1 PORTAL Arterial Count Station - Main Map
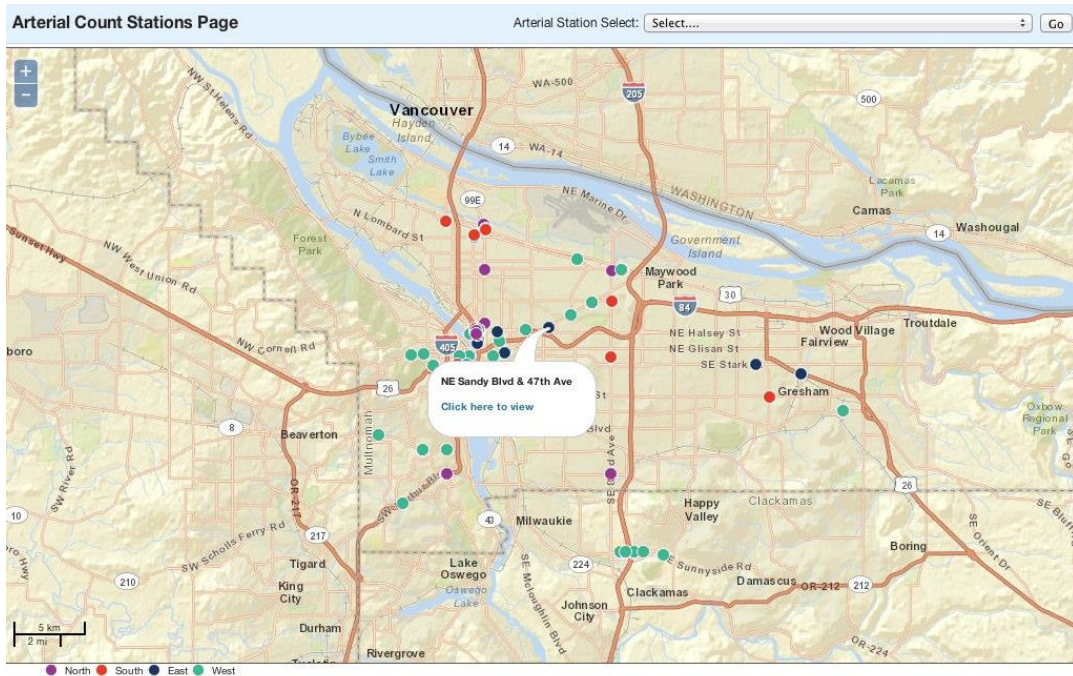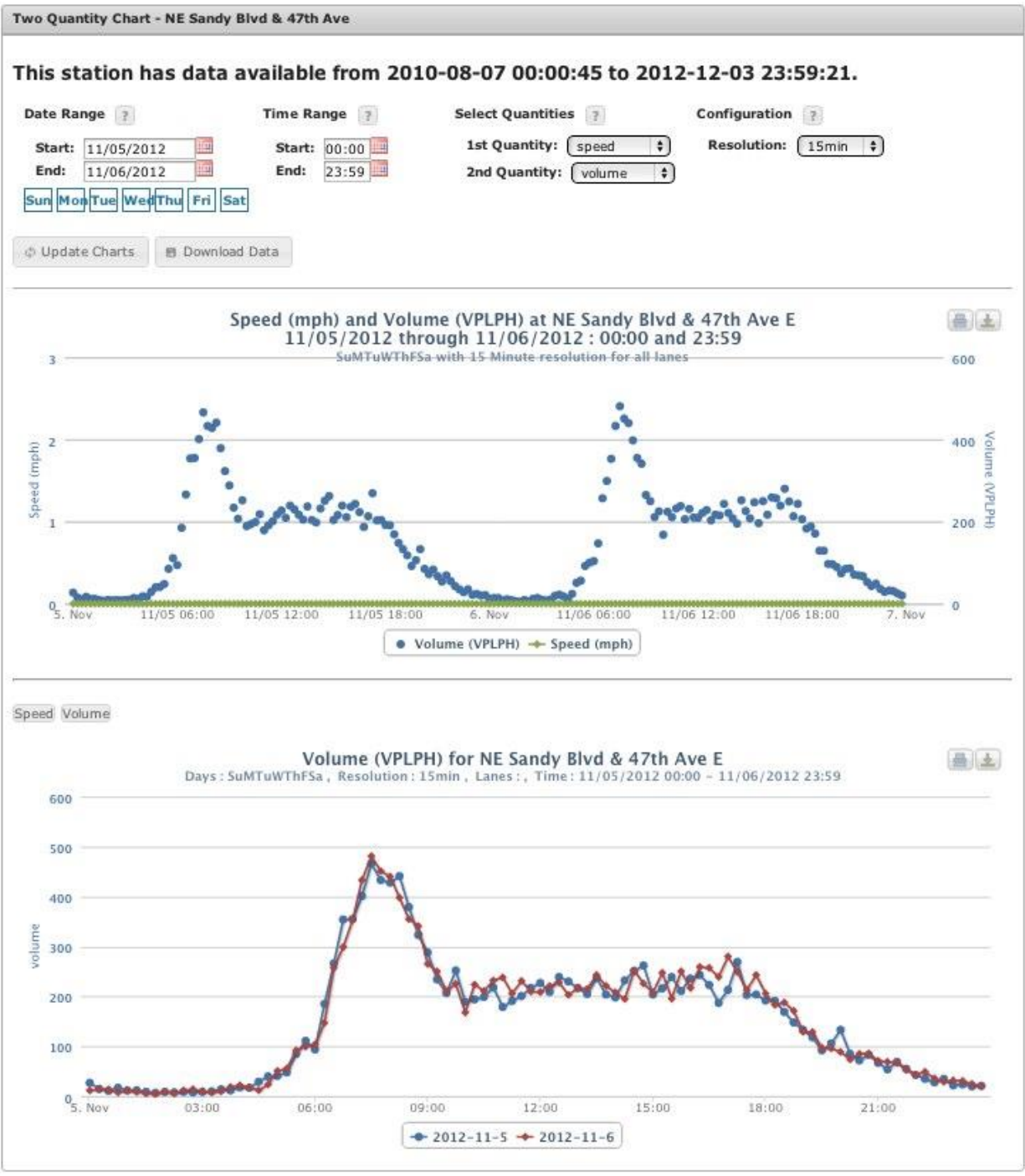
Figure 7.2 PORTAL Arterial Count Station Detail Page
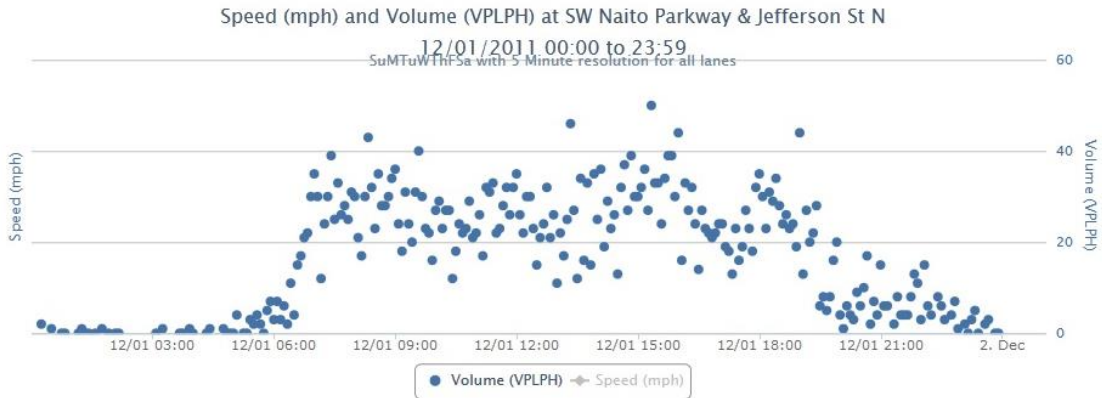
Figure 7.3 Volume Plot (5-minute Resolution)
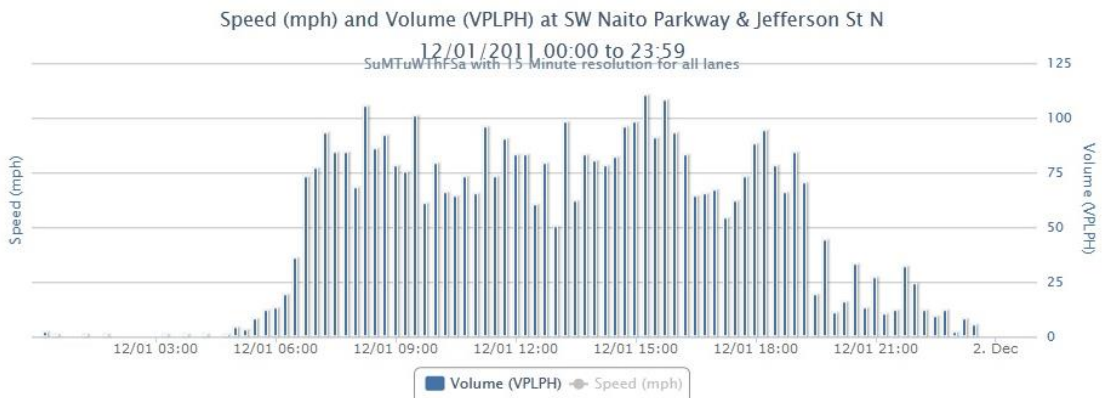
Figure 7.4 Volume Plot (15-minute Resolution)

Figure 7.5 Volume Plot (One-hour Resolution)

Figure 7.6 Volume Plot (One week)
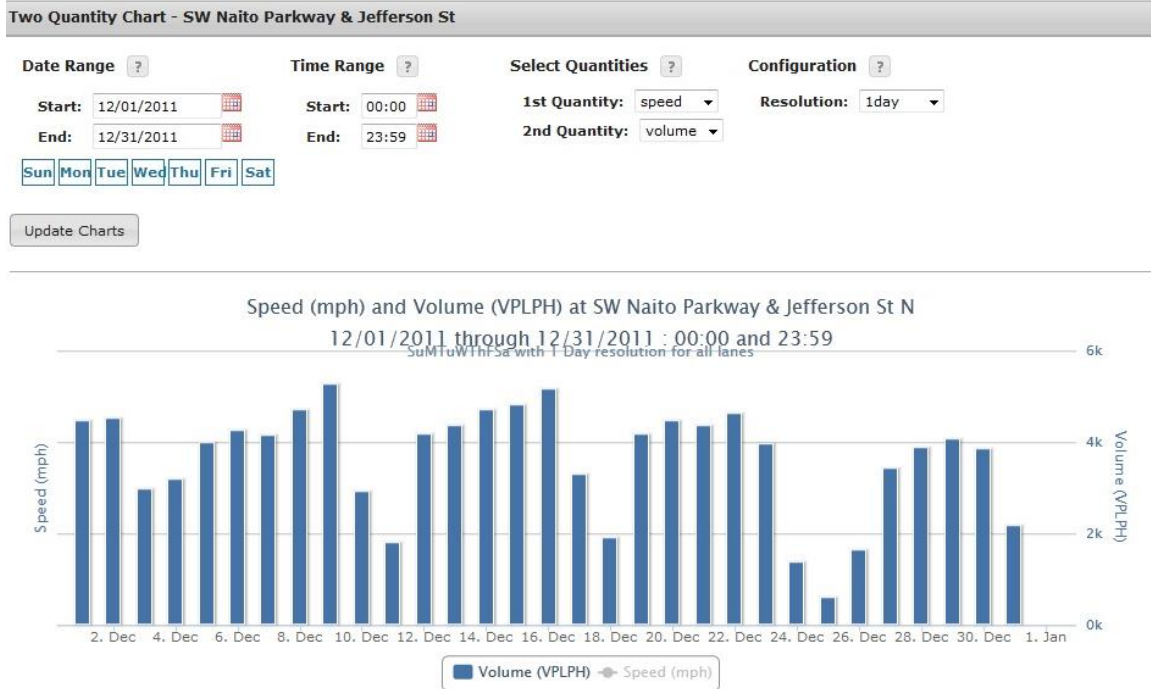
Figure 7.7 Volume Plot (One month)

Figure 7.8 presents an example of how vehicle count data and the intersection timing data can be viewed together for a simple operational snapshot evaluating operational strategies such as time-of-day coordination plans. Vehicle counts at one-minute resolution for all weekdays in February 2011 for station 135 (the northbound approach to the intersection of Southeast 82nd Avenue and Burnside Avenue) were queried from the archive. The count data has been converted to flow—at this location there are two detectors so the *y*-axis units are in total vehicles per hour. To aid in visualization, we apply a simple moving average over the previous 15 observations to smooth the count line. Each 24-hour time series is then plotted in light grey. Where flow patterns are consistent across the month, the time series are tightly grouped. As expected, variations in the flow patterns can be seen throughout the day, particularly in the morning peak around 8 a.m. and the afternoon around 5 p.m.. One day where the flow pattern substantially deviates from all other days is plotted in dark grey (which is discussed later).

Time-of-day coordination plans are activated based on fixed times for a particular intersection, and are typically based on flow patterns through the day. Since flow fluctuates throughout the day, different coordination plans are created to deal with AM and PM peak and off-peak periods. Time-of-day coordination plan boundaries are represented in Figure 7.10 with solid vertical lines. For example, from 10 p.m. to 6:30 a.m. timing plan 4 is active, which then switches to timing plan 2. For clarity, we have not shown transition plans in the figure. Identical plots to Figure 7.8 were created for every month from August 2010 to May 2011. Flow patterns are remarkably similar, though the usual seasonal volume trends are apparent. As more data are archived, signal engineers could make quick visual checks about changing demand trends that may warrant signal retiming efforts.

44
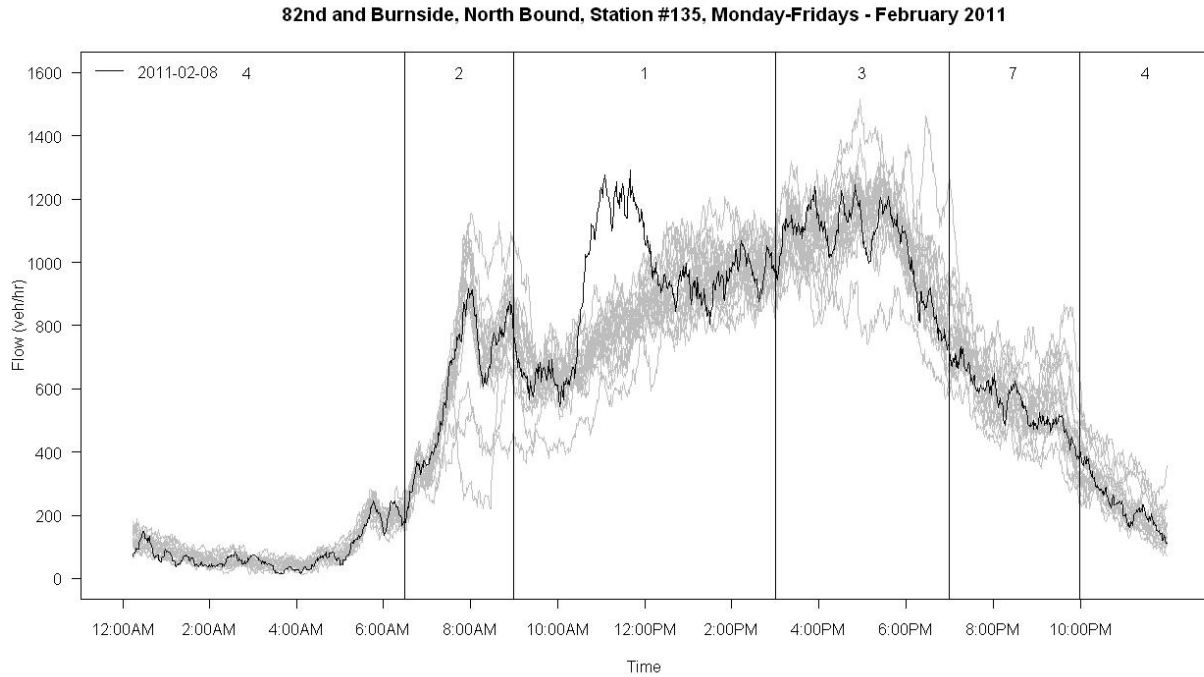
Figure 7.8 Volume and Signal Timing

In Figure 7.8, the dark grey time-series is for Feb. 8, 2011. The counts shortly after 10 a.m. deviate from all other days, as there was an increase in flow that lasted until after 12 p.m. (Southeast 82nd Avenue parallels the I-205 freeway). The PORTAL archive was queried to return speed-flow data from the nearest freeway station at I-205 northbound at Division (located to the southeast of the 82nd Avenue and Burnside intersection). Figure 7.9 shows a graph of speed and volume for the morning of Feb 8, 2011, for this station. As shown in the figure, traffic speed and volumes on I-205 clearly drop precipitously just after 10 a.m. on Feb. 8. Freeway speeds begin recovery at 11:15 a.m. and completely recover at 11:50 a.m.

Figure 7.9 Speed and Volume Plot for I-205 North at Division NB

45

Indeed, the incident database confirms that a one-vehicle crash resulted in debris blocking three lanes. The incident started at 10:11 a.m. and ended at 11:16 a.m. As a result, the volumes on Southeast 82nd Avenue at Burnside increased. This decrease in volume on I-205 and increase of volume on 82nd Avenue suggests that vehicles travelling north on I-205 diverted to 82nd Avenue as an alternate route. Operationally, an inspection of the timing plans reveal that signal timing was not changed from the normal time-of-day operation though the increased volumes resemble afternoon flow patterns. Through a post-hoc analysis, it is clear that monitoring algorithms could be applied to these data streams in real time to provide traffic diversions alerts.

With the inclusion of the other data sets, a more complete picture of an incident's total cost could be assessed. In addition to traditional freeway delays, impacts of this incident on transit users, pedestrians and arterial motor vehicles could easily be assessed and quantified.

To highlight an operational visualization, we relate these data to a corridor running traffic responsive timing strategy in Clackamas County. These plans are activated based on certain thresholds of flow as opposed to fixed time-of-day plans, as presented above. Based on input from the operating engineers, we have created a set of visualizations that can provide useful feedback in tracking the onset of these plans along a corridor based on flow progression. Figure 7.10 shows a plot of flow over time, with the actual timing plan onsets shown as the vertical lines and numbering on the top of the figure for a section of Sunnyside Road between the intersections of I-205 and 105th Avenue for two different days (April 13, 2011, and May 5, 2011). To replicate the threshold criteria used, the y-axis is flows at stations 7003 and 7004. Two different sets of detectors are plotted for each direction on this section of Sunnyside Road and overlaid with the timing plans that were present for that day. The figures confirm that different plans are being implemented based on flows. Volume thresholds can also be quickly checked. In the figure, the variability in timing plan start/stop times between days, along with the differences in flow both east and west, is easily seen. A traffic signal operator can look at the above plots and see if his flow responsive timing plan is functioning as intended. For example, in the lower plot for May 3, 2011, a timing plan of 255 occurred at 8 p.m. for a short time. Timing plan 255 in the above plot represents when the signal was "free" or fully actuated, operating without a background cycle length. The volume time-series shows after an increase in vehicle flows, the system switched back to plan 3.
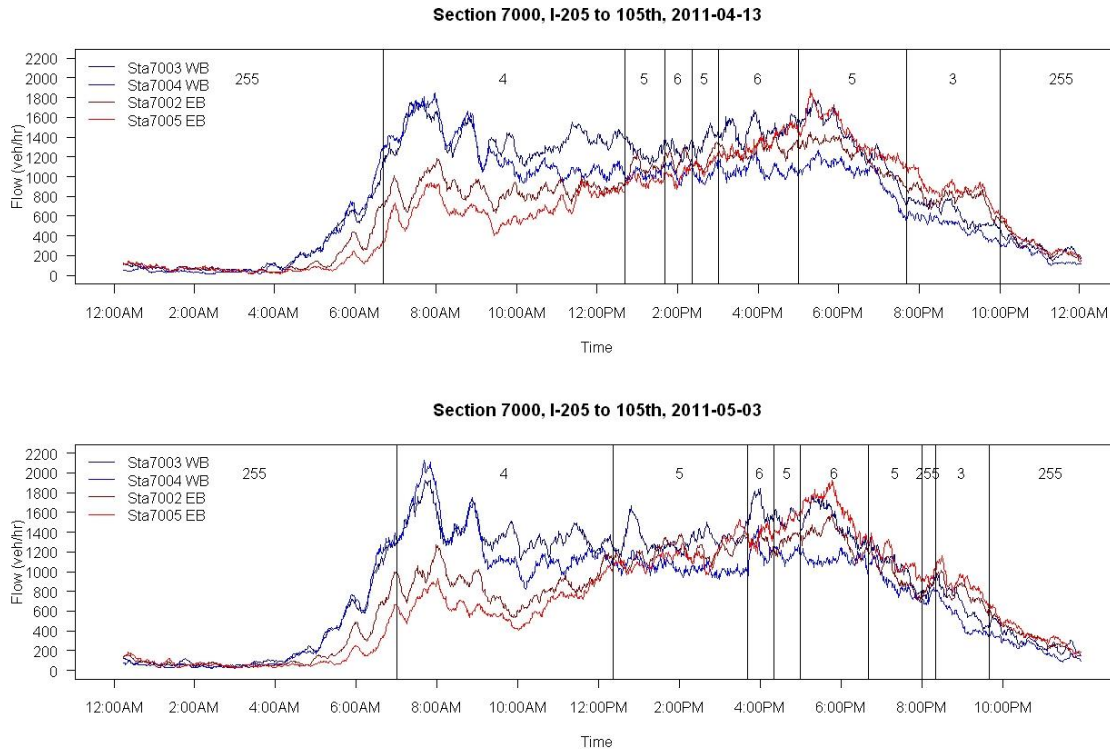
Figure 7.10 Traffic Responsive Signal Timing

7.2   BICYCLE COUNTS

Visualizations based on bicycle counts have also been created. Similar to the arterial data, the user has to select the type of data, station, and date and time range. Daily, weekly and monthly plots of counts can be generated by PORTAL. Figure 7.11 shows the plot of bicycle counts for one day at one location. The map on the top shows the location of the detector. Higher counts are seen in the AM peak as expected based on the location of the detector (along the direction of peak traffic).

Figure 7.12 shows the time series plot of bicycle counts. The gaps represent missing data. Seasonal trends, if they exist, can be determined from a time series plot. Figure 7.12 shows higher counts during the warmer months (August – October) and lower counts during the colder months (November – April), as expected. Figure 7.13 shows the bicycle counts for one week period showing weekday and weekend trends. Figure 7.14 shows the bicycle counts for one month.

47

Figure 7.11 Daily Bicycle Counts
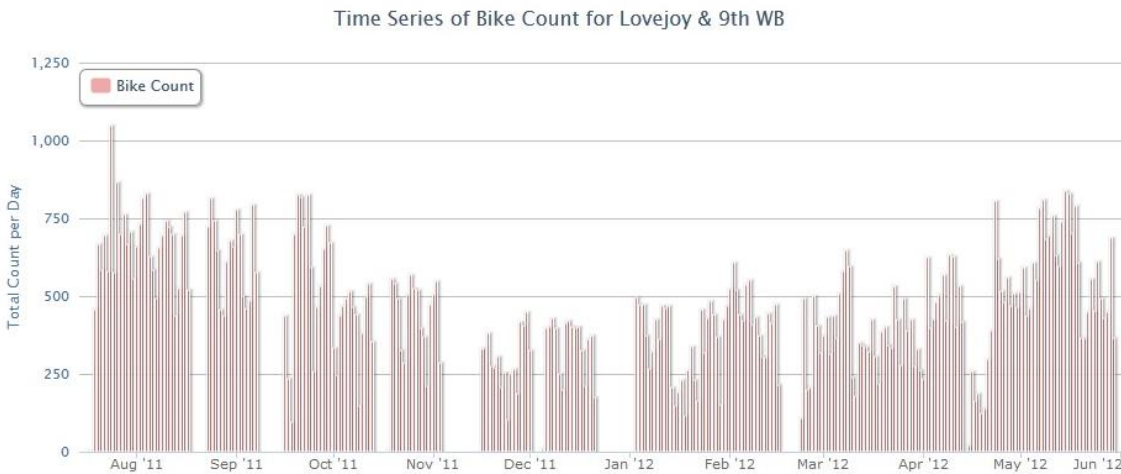
Figure 7.12 Time-series Plot of Bicycle Counts
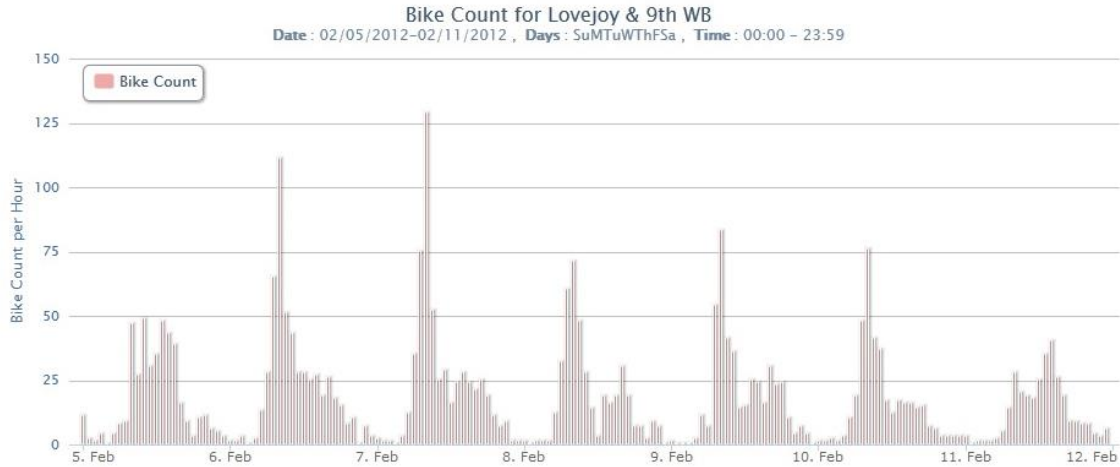
Figure 7.13 Weekly Bicycle Counts
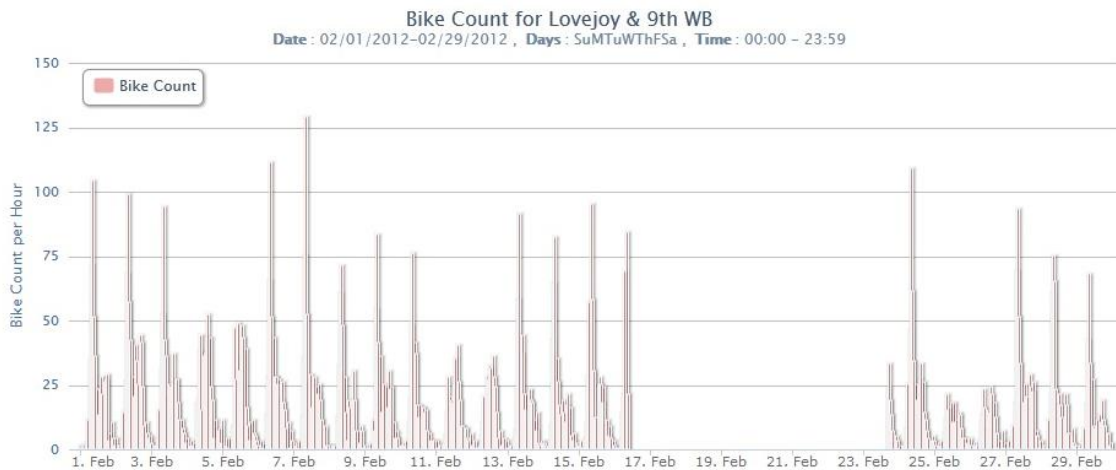
Figure 7.14 Monthly Bicycle Counts

7.3   PEDESTRIAN COUNTS

Visualizations of pedestrian data can be found under the Ped/Bike tab on PORTAL's demo site. This page allows the user to select either pedestrian or bicycle data type and, depending on the selected data type, appropriate options for statistics and plots are displayed. For pedestrian data, users can select a station and date and time ranges. Based on the selection, pedestrian push-button actuations for that particular station are displayed. In addition, binned delay data are also displayed for the chosen time period. These delay bins have been previously described in Chapter 2. Figure 7.15 shows the statistics displayed by PORTAL for pedestrian actuations and delay.
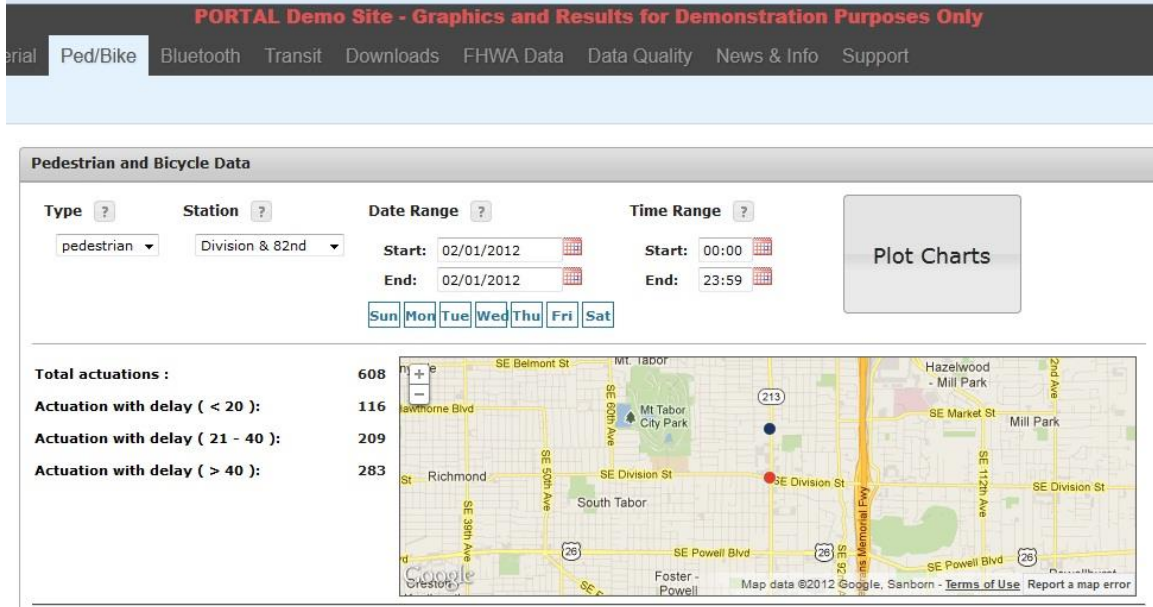
Figure 7.15 Pedestrian Actuation Statistics

The statistics shown here allow practitioners to get an idea of the total pedestrian calls for service at a particular intersection. Figure 7.16 shows the trend in pedestrian actuations for one day and the distribution of delays. The plot reveals important information, such as the levels of pedestrian activity are highest during the mid-day hours and low levels of activity are encountered during late evening and early morning hours. The plot also allows the user to identify the hour with the highest pedestrian actuation activity.

The bar chart on the right shows the distribution of delay. At this location, the majority of the pedestrians experienced delays greater than 40 seconds. These statistics provide a means to evaluate current pedestrian activity and delays and make changes if necessary.

Figure 7.16 Daily Pedestrian Actuations

50

PORTAL also displays a plot of all the actuations that are being archived in the database. Figure 7.17 shows the plot of actuations for a period of seven months. Gaps in these data are easily seen in the plot. These gaps are either due to errors in these data or loss of communication to the signal controller, which results in failure to retrieve the logs. As with the Bluetooth data, weekly and monthly data can also be displayed to visually inspect trends. Figure 7.18 shows the weekly trend; the plot reveals that actuations are fairly constant Monday-Thursday, are lower on Friday and are lowest on the weekend
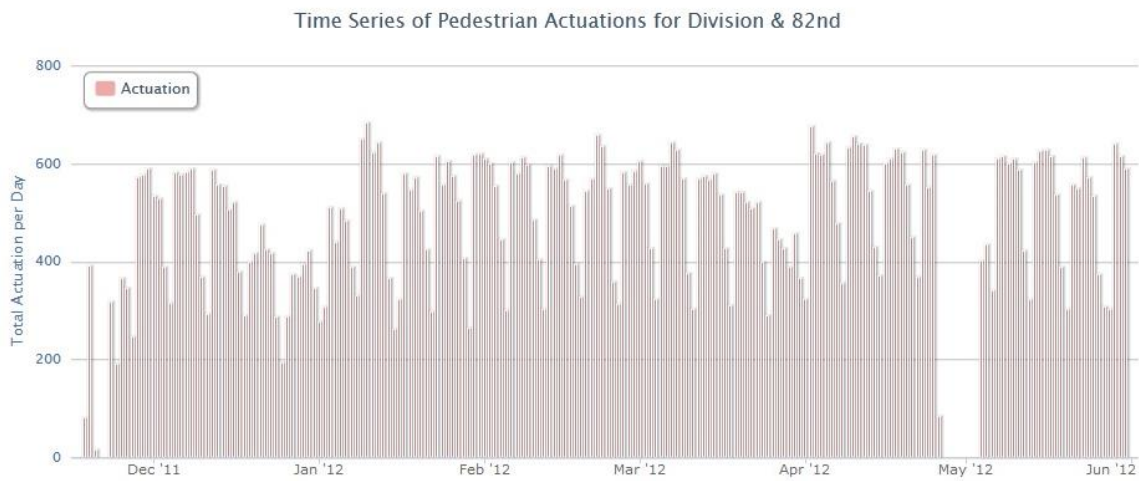
Figure 7.17 Time-series Plot of Pedestrian Actuations

Consider adding Figure 7.18 here since text below refers to Fig. 7-19

Figure 7.19 shows the monthly trend in pedestrian actuations and the associated statistics. The statistics reveal that at this location, there were 15,000 push-button actuations within one month, and the actuations with delays greater than 40 seconds are almost twice as many as the actuations with delay less than 20 seconds.
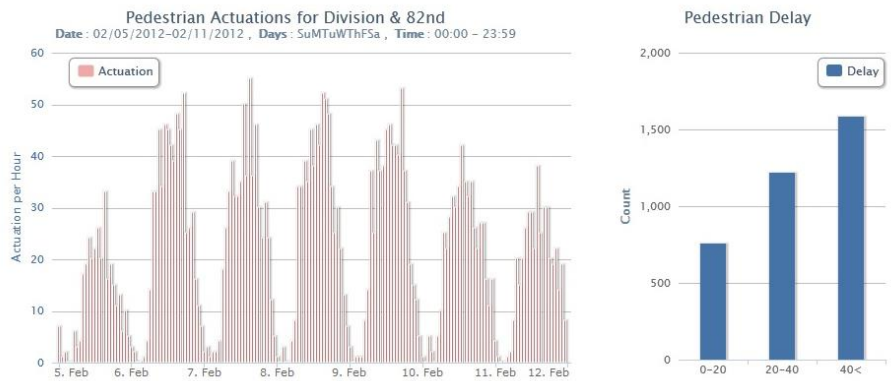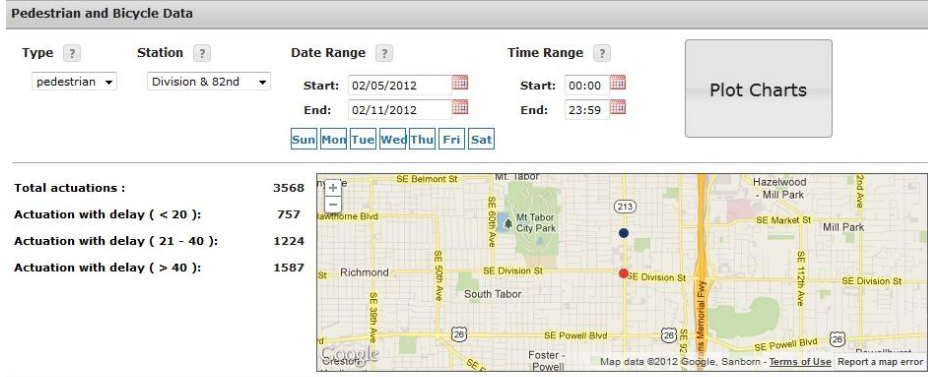
Figure 7.18 Weekly Pedestrian Actuations
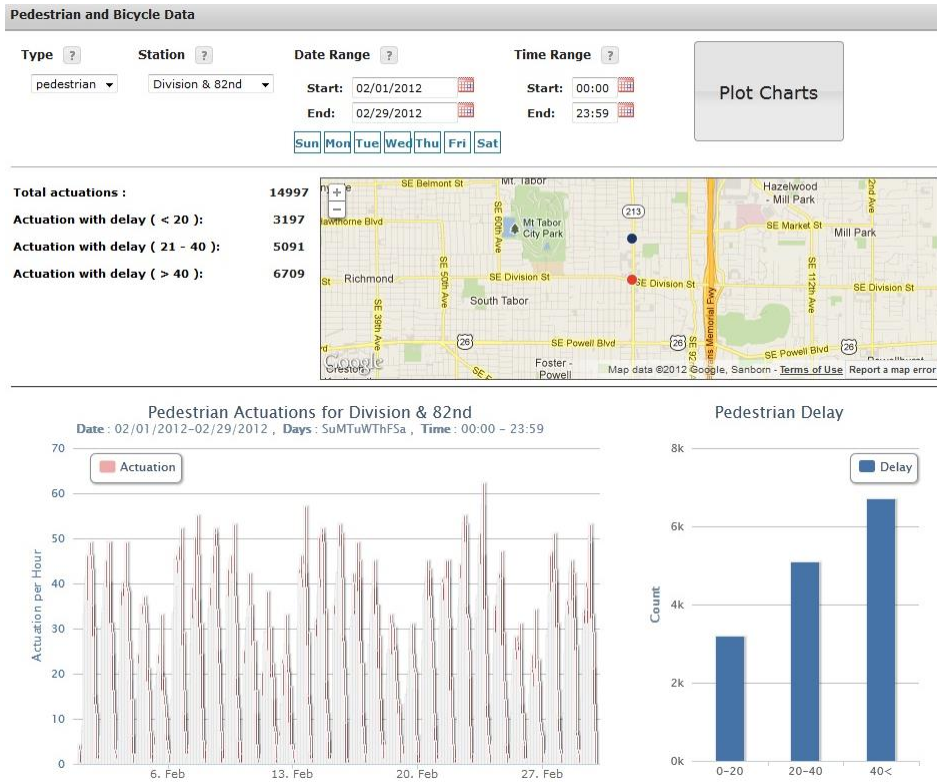
Figure 7.19 Monthly Pedestrian Actuations Trend

7.4 TRAVEL TIMES

Visualizations based on the Bluetooth data can be found under the Bluetooth tab on PORTAL's demo site. Only certain logical groupings of segments (start and end station combinations) are allowed. Once the user makes the selection of the start and end stations, date and time range, PORTAL processes the appropriate data and generates statistics and plots based on the data. Figure 7.20 shows the statistics for a given segment and time period.
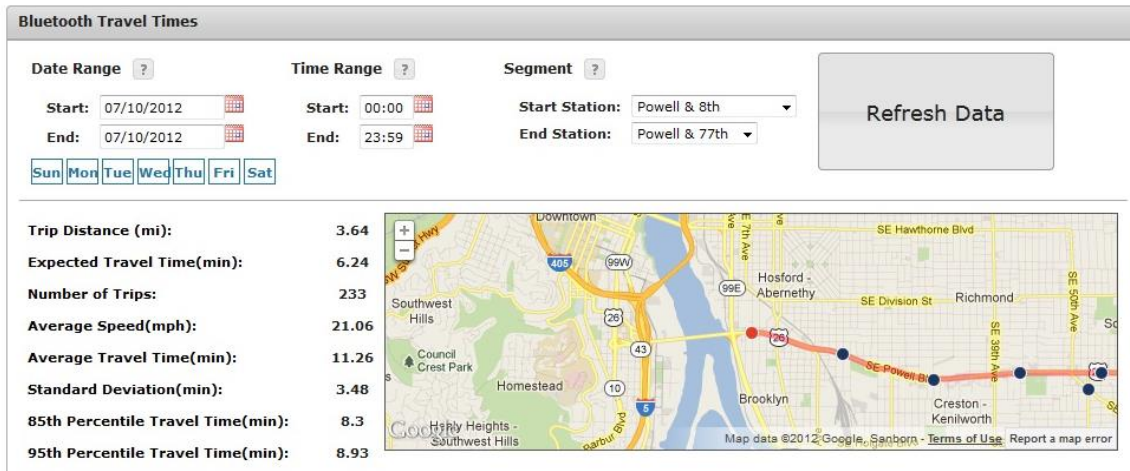
Figure 7.20 Bluetooth Travel Time Statistics

Various statistics such as the distance between the chosen start and end stations, expected travel time, average speed, average travel time, standard deviation and 85th and 95th percentile travel times are displayed. The number of trips refers to the resulting number of matched and filtered travel times obtained after processing the raw Bluetooth data for the given time period. In addition to the statistics, PORTAL also plots travel times and speeds for the chosen time period. Figure 7.21 shows the travel times for the chosen pair of stations. The points represented in the graph are the matched travel time readings. The blue line represents rolling average travel time. A download button is provided for the user to gather the travel times being displayed in the plots. Figure 7.22 shows the travel speeds. Similar to the functionality provided with the system detector plots, the user can retrieve data for a week or a month. In addition, the mean travel time can be turned off, allowing the user to view the individual travel time readings. Figure 7.23 and Figure 7.24 show the Bluetooth travel times and speeds for one week, respectively.
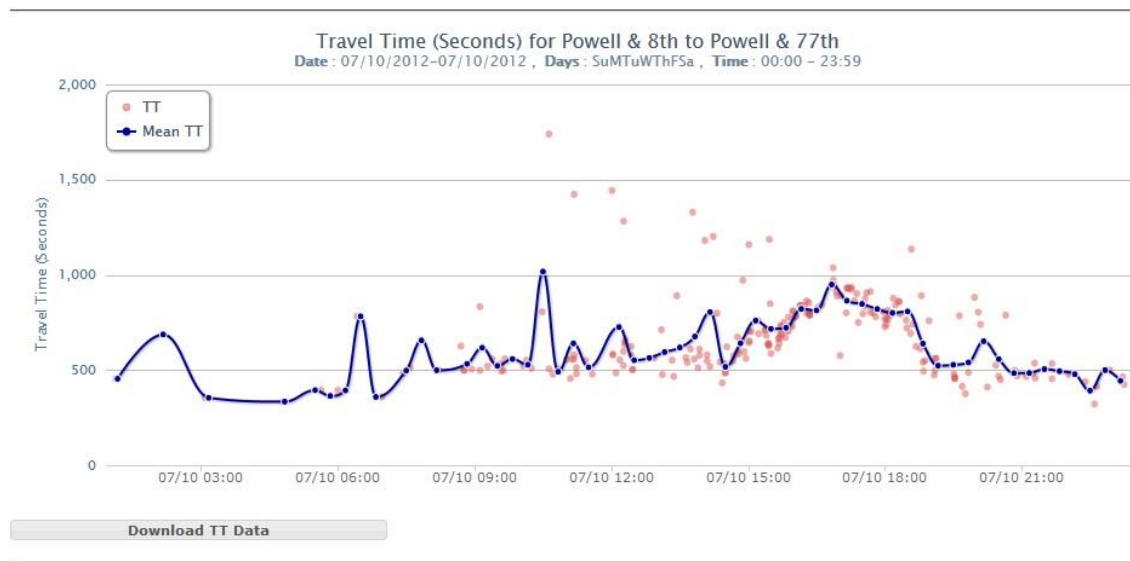
Figure 7.21 Bluetooth Travel Times

54
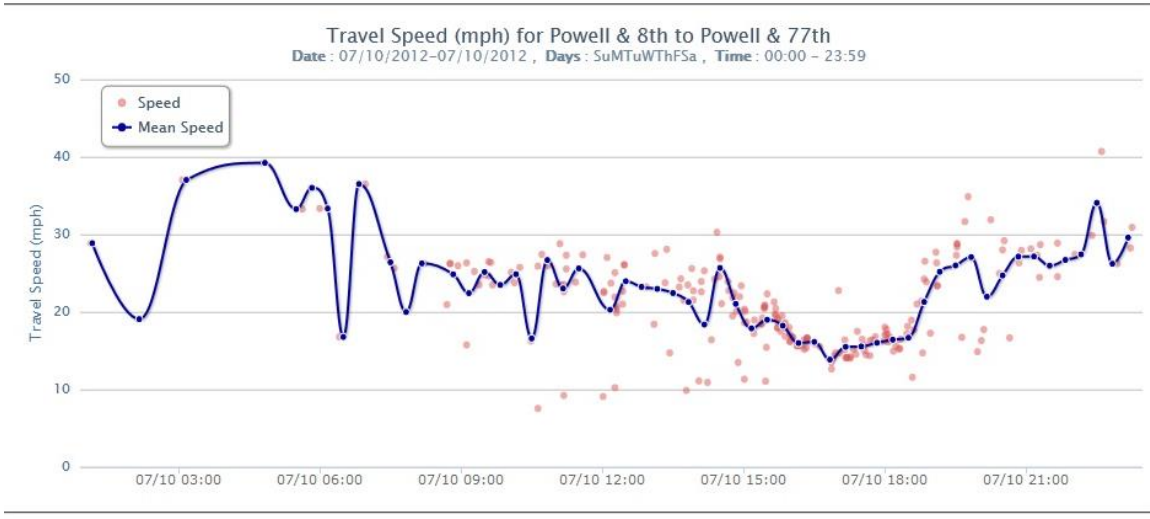
Figure 7.22 Bluetooth Speeds

Figure 7.23 Bluetooth Travel Times (One week)
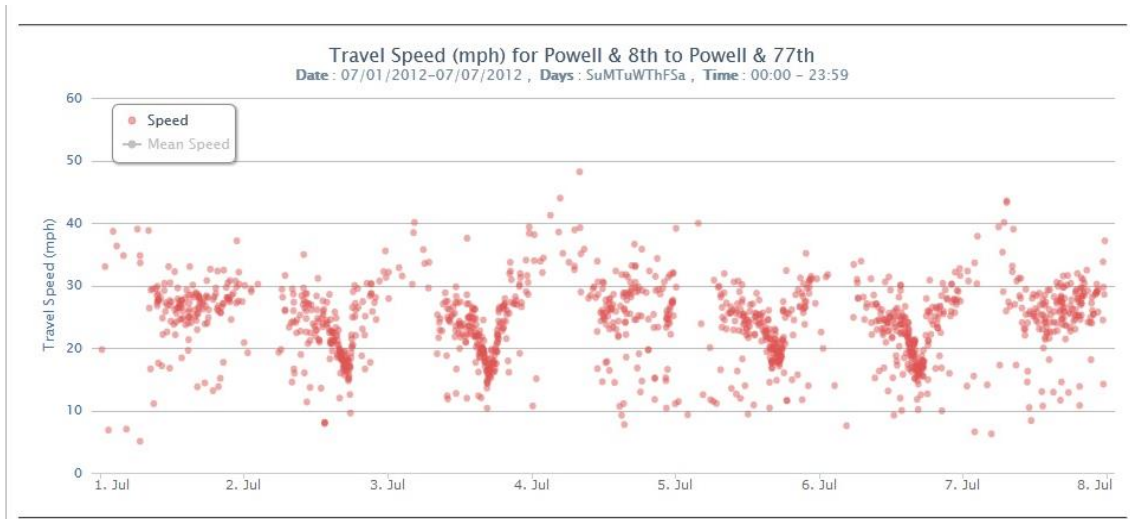
Figure 7.24 Bluetooth Speeds (One week)

Longer travel times and slower speeds during the peak periods on weekdays are easily visible in the plots. These plots provide a great visual representation of these data. and the accompanying statistics are helpful in understanding the performance of the chosen segment. Figure 7.25 and Figure 7.26 show the travel times and speeds along the chosen segment for one month.
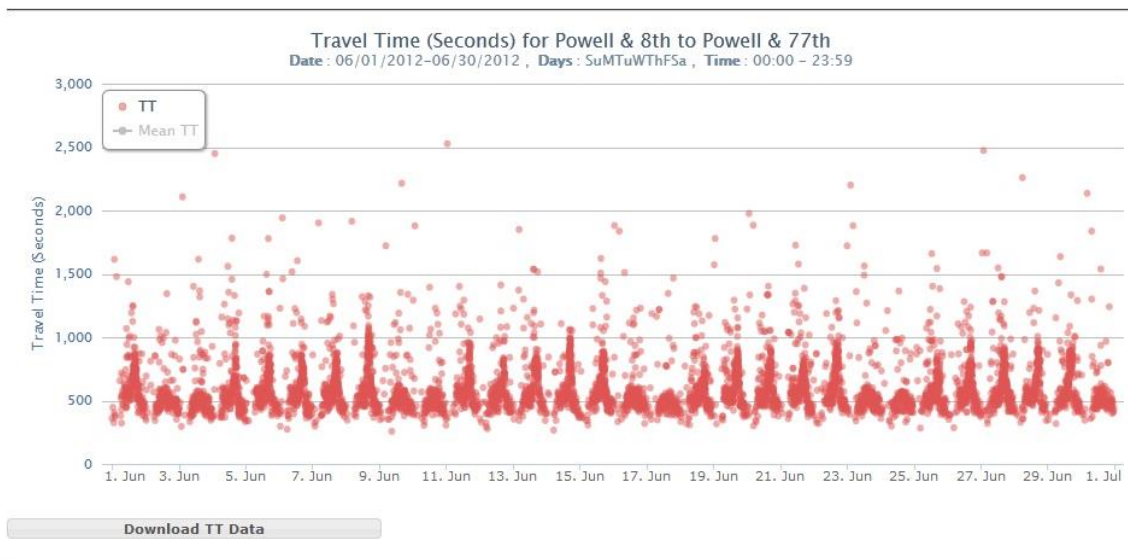
Figure 7.25 Bluetooth Travel Times (One month)
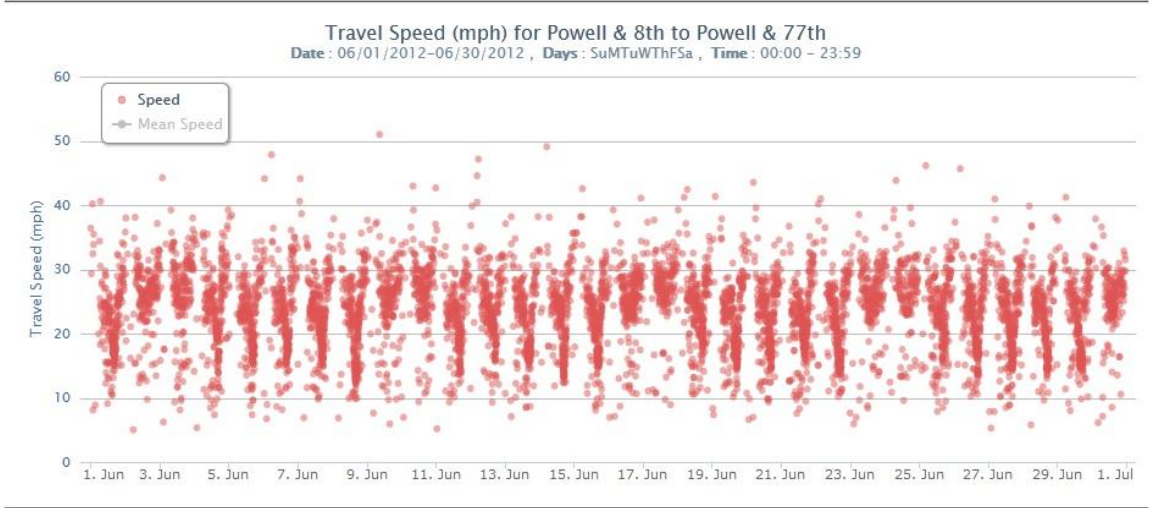
Figure 7.26 Bluetooth Travel Speeds (One month)

8.0   SAMPLE ANALYSIS

There are many possible uses for the data being archived on the arterial level. Clearly, performance measures can easily be calculated for the various modes on each corridor where complete data are available. As the archive matures, it is easy to envision a multimodal corridor performance assessment. In this chapter, we highlight two efforts to utilize multiple sources of the data for some sample analysis. In this chapter, we highlight the combination of the transit and travel time data and the use of the signal timing and vehicle count data to generate some sample congestion measures.

8.1   TRANSIT AVL AND TRAVEL TIME DATA

In this section, we investigate the fusion of transit and Bluetooth MAC address data. We explore the use of AVL-equipped buses for generating arterial travel times. While MAC-address matching alone provides accurate travel time estimates, MAC-address readers are not available along all corridors; thus, AVL data has the potential to provide additional travel time coverage. This study builds on prior research work examining the feasibility of generating arterial travel time estimates using bus probes by comparing bus travel times to MAC travel times (Bertini, 2004), but uses a significantly larger ground truth sample.

This study compares 1,720 bus travel runs to 9,060 MAC data matches for the week of May10 - 16, 2010, on Southeast Powell Boulevard in Portland. The findings from the study indicate that bus travel times are consistently biased, and the average travel time errors are dependent on the length of the segment and the number of bus stops along a route.

8.1.1   Study Area and Data Sources

The study area location is along a 2.97-mile section of Powell Boulevard in Portland. The study area corridor begins at Southeast 21st Avenue and continues to 77th Avenue, as illustrated in Figure 8.1. This corridor is a four-lane undivided arterial with two lanes in each direction, with left-turn bays present at intersections. There are 10 signalized intersections along the study corridor (SE 21st, SE 26th, Se 33rd, SE 39th, SE 50th, SE 52nd, SE 65th, SE 69th, SE 71st and SE 72nd). This corridor is heavily traveled and serves approximately 37,751 vehicles per day. Since the corridor connects the central business district to the eastside suburbs, the westbound direction experiences heavy traffic in the morning peak period and, similarly, the eastbound direction has traffic in the afternoon peak period. The study period was the week of May 10-17, 2010.
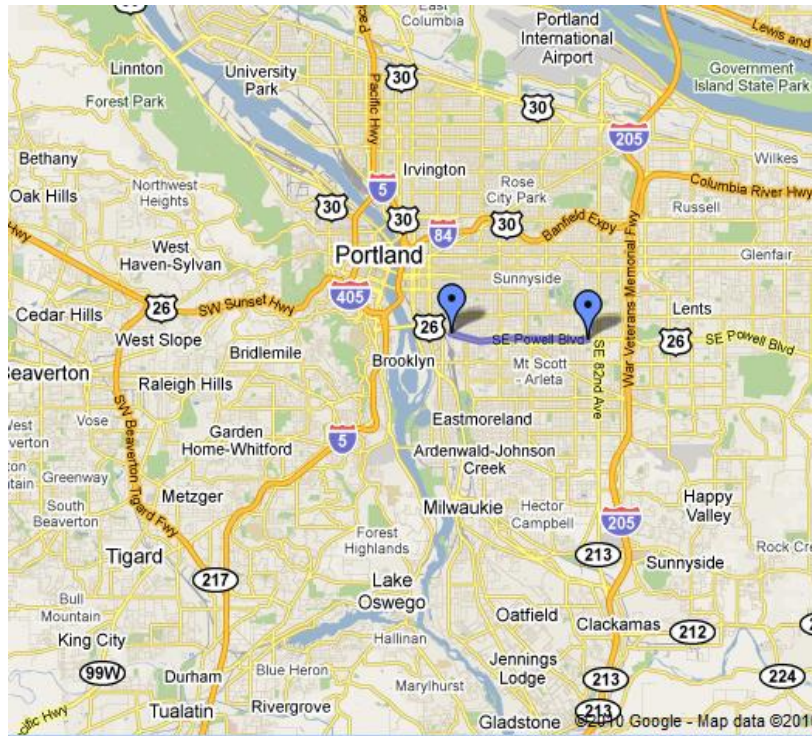
Figure 8.1 Map Showing Study Area

The data for this study came from two sources: MAC and Transit AVL. The MAC data comes from Bluetooth readers at four intersections along Powell Boulevard (SE 7th, SE 21st, SE 33rd and SE 77th). Based on 24-hour tube counts on the corridor selected for the study, an approximate capture rate of 5-11% was achieved. The MAC addresses recorded by readers at different intersections are matched using software-processing methodology developed by Quale et al. (2010). This software processing matches MAC addresses recorded at two readers and removes outliers. MAC readers cannot distinguish between continuous and discontinuous trips and multimodal trips, so removal of outliers is critical. A moving standard deviation method is incorporated into the software and is used to remove outliers (Quayle, 2010). Figure 8.2 (a) shows the observed travel times from Southeast 33rd Avenue to 77th Avenue in the eastbound direction on Powell Boulevard from May 10-17, 2010, obtained from the MAC matching process. Figure 8.2 (b) shows the westbound travel times for the same time period along the same segment.

The eastbound section of Powell Boulevard connects downtown Portland to the eastside suburbs and, therefore, higher travel times are expected in the PM peak period as seen in Figure 8.2 (a). Conversely, Figure 8.2 (b) shows higher travel times during the AM peak period. In addition to the travel times, hourly average travel time is also represented in Figure 8.2 (a) and (b).

(a) Observed Travel Times on EB Powell Blvd. (SE 33rd to SE 77th)

(b) Observed Travel Times on WB Powell Blvd. (SE 77th to SE 33rd)

Figure 8.2 MAC-generated Travel Times, Powell Blvd., May 10-17, 2010

TriMet, the transit agency for the Portland metro area, has equipped all of its buses with an AVL system based on Global Positioning System (GPS) (Bertini, 2004). AVL data for Route 9, which traverses the study corridor, was obtained for May 10-16. On each weekday, there were between 65-80 round trips in each direction; some trips were removed from the analysis due to poor data quality. On the weekend, there were 40-50 trips in each direction undertaken by Route 9. Figure 8.3 shows the trajectories for May 11, 2010. for Route 9 traveling in the eastbound direction between Southeast 33rd Avenue and 77th Avenue on Powell Boulevard between 3-6 p.m. Each trajectory in Figure 8.3 represents one trip made by the bus from origin to destination. The gray circles on each trajectory represent stops along the route. It is apparent from the figure that the bus stops are closely spaced and if passenger demand exists, the bus makes stops frequently to serve the demand. As a result, bus travel times will always be higher than automobile travel times during any time period.
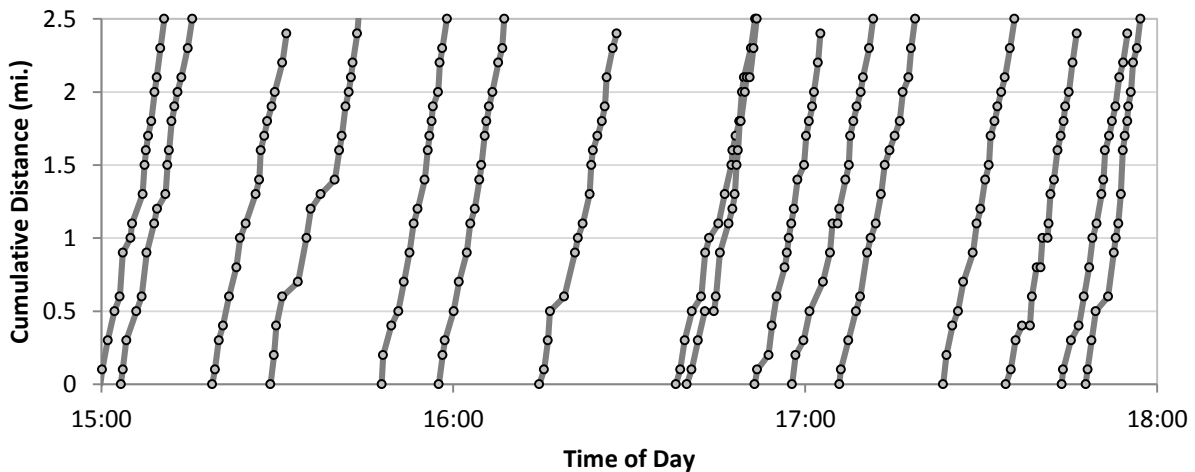
Figure 8.3 Eastbound Bus Trajectories for May 11, 2010, 3-6PM

8.1.2   Data Analysis

The objective of this study is to compare the travel times obtained from MAC readers and AVL equipped buses, and to determine a method for using travel times obtained from buses in locations where MAC readers may not be present. Travel times from both data sources were computed and compared. Table 8.1 shows the summary statistics for travel times as obtained from MAC readers after the matching process was implemented and outliers were eliminated.

The analysis for this study was done in four segments: segments 1 and 2 represented the eastbound direction, and segments 3 and 4 were defined in the westbound direction. The segments were defined in this manner because MAC readers were deployed at each of the three intersections (SE 21st, SE 33rd and SE 77th) and the goal was to obtain maximum possible matches. Figure 8.4 shows the locations of segments 1-4 and MAC readers. Segments 1and 4 are shown in purple, and segments 2 and 3 are shown in orange on the map.

61

Figure 8.4 Segments and MAC Reader Locations

Table 8.1 indicates that the most matched readings were obtained for the shorter segments (1 and 4). The average speed was fairly consistent over all segments, and the average travel time varied between 1.5 to 2 minutes for the shorter segments and between 5.9 and 6.4 minutes for the longer segments.

Table 8.1 Summary Statistics from MAC Data

	Segment 1 (EB 21st - 33rd)	Segment 2 (EB 33rd - 77th)	Segment 3 (WB 77th - 33rd)	Segment 4 (WB 33rd - 21st)
ALL (PEAK + OFF PEAK)				
Distance	0.7	2.3	2.3	0.7
Count (Number of matches)	3352	1204	1319	3185
Average Speed (mph)	23.6	24.0	23.0	27.4
Average Travel Time (s)	111.02	357.11	380.79	98.11
Standard Deviation (s)	34.65	66.06	89.24	41.99
AM PEAK (7AM – 9AM)				
Count (Number of matches)	147	40	135	301
Average Speed (mph)	24.5	26.1	20.5	19.7
Average Travel Time (s)	104.14	322.18	426.96	161.85
Standard Deviation (s)	26.16	31.98	95.36	92.35
PM PEAK (4PM - 6PM)				
Count (Number of matches)	586	269	183	388
Average Speed (mph)	19.2	19.7	22.9	24.6
Average Travel Time (s)	139.27	432.24	372.53	101.06
Standard Deviation (s)	47.05	67.52	69.06	16.46

The BDS obtained from TriMet consisted of a total of 1,867 trips on all four segments for the one week period (May 10-16). This data was subjected to a basic data cleaning process and records that failed to meet the set criteria were discarded from the analysis. For some bus trips, the arrival time at the downstream stop was earlier than the leave time at the previous stop, thereby leading to negative travel times. In such cases, the entire trip was eliminated from the analysis. In addition, trips containing zero values for maximum speed between stops were also removed. After the data cleaning process, a total of 147 trips were eliminated, leaving a total of 1,720 trips for further analysis. The data was divided into the four segments as defined earlier and travel times were computed for each trip. Travel times were computed in two different ways:

$$TT1 = a_j - a_i - \Sigma d_{ij} \ ;$$

$$a_j = arrive\ time\ at\ last\ stop\ j$$
$$a_i = arrive\ time\ at\ first\ stop\ i$$
$$\Sigma d_{ij} = Sum\ of\ dwell\ times\ between\ stops\ i\ and\ j$$

$$TT2 = \Sigma x_{ij};$$

$$x_{ij} = travel\ time\ between\ stops\ i\ and\ j = \frac{d_{ij}}{v_{ij}}$$
$$d_{ij} = distance\ between\ stops\ i\ and\ j$$
$$v_{ij} = \max speed\ between\ stops\ i\ and\ j$$

Dwell time is recorded by the BDS system and is defined as the number of seconds the door is open. Subtracting the total dwell time from the difference of the arrival times at the last and first stop will yield travel time for the bus, which can be compared to the MAC address travel times. The second method of computation of travel times utilizes the maximum speed achieved by the bus between stops. Travel time between a pair of stops is computed as the ratio of the distance to the maximum speed. The travel time for the trip is computed as the sum of travel times between pairs of stops. Previous research showed that the maximum speed achieved between a pair of stops was found to explain the movement of automobile traffic reasonably well (Chakroborty, 2004). Table 8.2 shows summary statistics for Route 9 from May 1-16.

Table 8.2 Summary Statistics from TriMet Data

	Segment 1 (EB 21st - 33rd)	Segment 2 (EB 33rd - 77th)	Segment 3 (WB 77th - 33rd)	Segment 4 (WB 33rd - 21st)
ALL (PEAK + OFF PEAK)				
Distance	0.7	2.3	2.3	0.7
Count (Number of Runs)	456	410	414	440
Average TT1 (s)	142.24	501.74	533.15	130.16
Stdev TT1 (s)	53.69	89.22	81.72	29.88
Average TT2 (s)	84.36	309.38	316.85	82.0
Stdev TT2 (s)	12.71	70.77	31.77	12.51
AM PEAK (7AM – 9AM)				
Count (Number of Runs)	37	35	53	59
Average TT1 (s)	134.97	450.51	621.55	158.51
Stdev TT1 (s)	25.11	39.19	74.62	51.99
Average TT2 (s)	77.28	293.01	340.52	88.15
Stdev TT2 (s)	10.77	23.19	29.20	17.63
PM PEAK (4PM - 6PM)				
Count (Number of Runs)	67	51	36	42
Average TT1 (s)	164.37	618.49	549.22	133.35
Stdev TT1 (s)	37.29	72.81	66.43	17.43
Average TT2 (s)	93.66	364.22	323.28	84.39
Stdev TT2 (s)	15.13	171.82	24.64	10.73

Comparing the MAC and bus travel times (TT1), it is apparent that the bus travel times are much higher (20-40%) than the MAC travel times across all four segments, even though the dwell times have been removed. This is not surprising for a number of reasons. First of all, buses are large vehicles and have different operating characteristics than automobiles. They accelerate and decelerate over longer distances, and the delay caused by these two phases has not been accounted for in the travel time computations. In addition, buses tend to travel slower and use the shoulder lane primarily.

Alternatively, bus travel times (TT2) using maximum speed for travel time computations are consistently lower than MAC travel times (15-25%). While maximum speeds achieved by the bus between stops may reflect the automobile travel speeds reasonably well during off-peak hours, during peak periods – with the onset and dissipation of congestion – this assumption is not likely to be valid. It is unreasonable to expect the bus to travel at maximum speed for the total distance between stops during congested periods. Therefore, the underestimation of bus travel times (TT2) as compared to MAC travel times is not unexpected. Figure 8.5 shows a time series plot of segment 2 (EB 33rd – 77th) travel times for one week (May 10 -16).
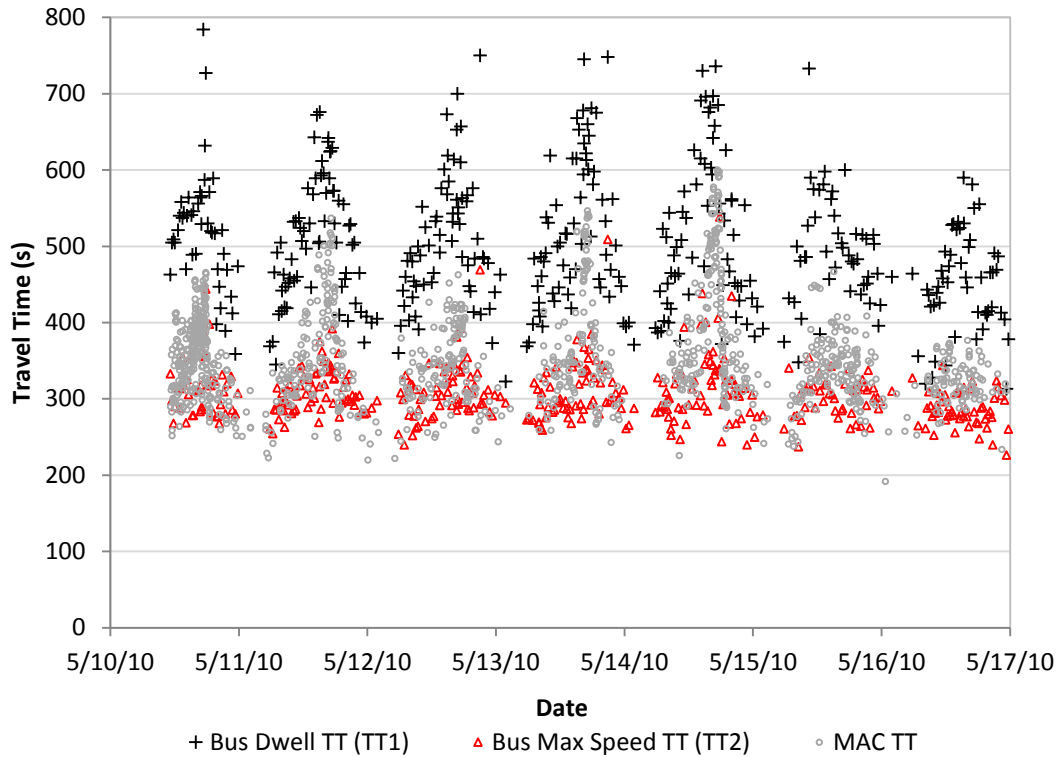
Figure 8.5 Time-series Plot of Travel Times for Segment 2 (EB 33rd – 77th)

Bus travel times (TT1 and TT2) are represented by black crosses and red triangles, respectively. MAC travel times are represented as gray circles. From Figure 8.5, it is apparent that TT1 overestimates and TT2 underestimates travel times as compared to MAC estimates. However, it is interesting to note that the patterns in the scatter plot shown in the above figure for TT1 and MAC are quite similar, with both patterns showing corresponding increases in travel times depending on the time of day. On the other hand, the distribution obtained using maximum speeds (TT2) appears to be flat and does not show a similar trend. Visually, the travel time distributions (TT1 and MAC) appear to be correlated with TT1 estimates exhibiting a positive bias, as compared to MAC estimates.

To further enhance the graphical exploration of the time-series scatter plot in Figure 8.5, a Lowess (Locally Weighted Scatterplot Smoothing) line was fit to each day's time-series. In Figure 8.5, Lowess curves are fit for both the bus and MAC travel time estimates for all four segments. The fit lines give additional weight to previous observation; that bus predicted travel times are positively and consistently biased.

65
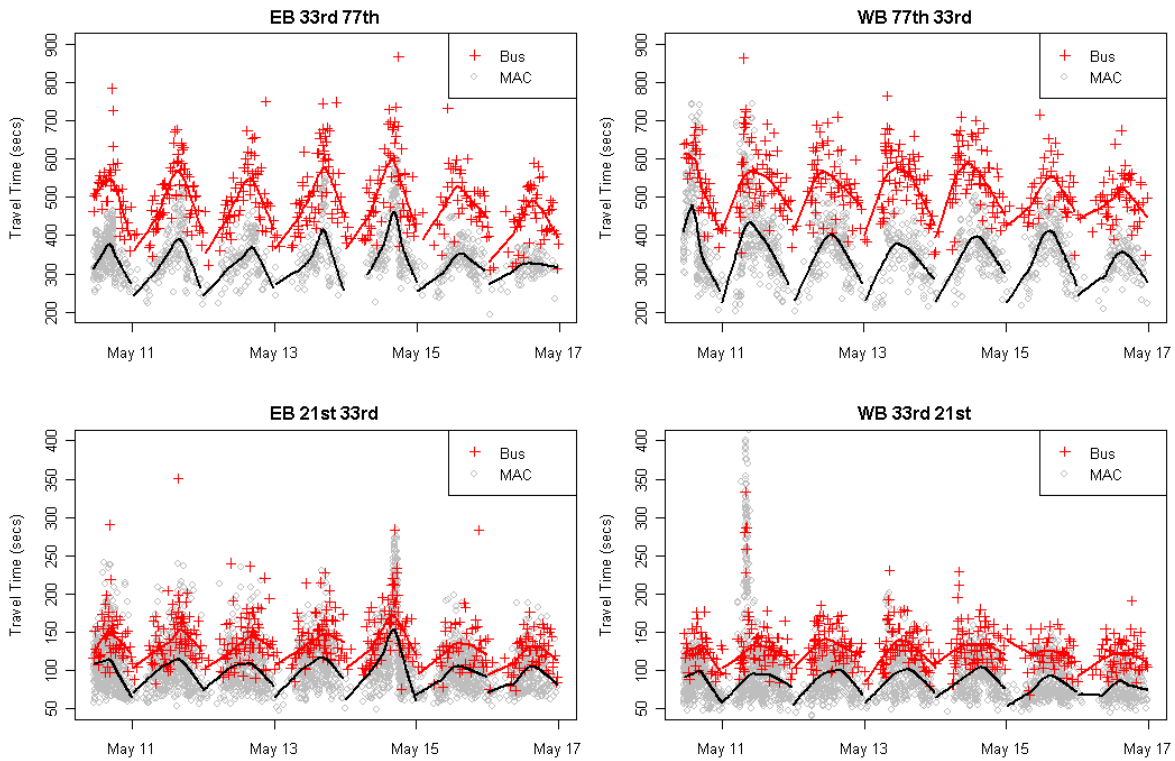
Figure 8.6 Lowess Curves for Bus and MAC Travel Times

Figure 8.5 and Figure 8.6 indicate the bus travel times (TT1) exhibit a bias when compared to the MAC travel times. It is necessary to quantify the observed bias in order to derive the automobile travel times using transit data. Table 8.3 shows the summary of error statistics as obtained from the comparison of the bus and MAC travel time estimates. Since the start times of the bus and MAC data did not correspond exactly, each data set was aggregated to five-minute time bins and travel times within each time bin were compared.

Table 8.3 Travel Time Error Statistics

	ME (s)	MAPE (%)	MPE (%)	SDPE
Segment 1	-35.02	42.44	-38.59	63.21
Segment 2	-165.03	48.75	-48.52	25.01
Segment 3	-172.48	52.19	-51.05	30.71
Segment 4	-34.19	45.04	-43.20	39.11

Error is computed as the difference between the MAC travel time estimate and the bus travel time estimate (TT1). Positive errors indicate underestimation and negative values imply overestimation of travel times. Mean Error (ME) is the average of all errors. Percent Error is calculated as the ratio of error to the MAC travel time estimate and is expressed as a percentage. Mean Absolute Percent Error (MAPE) represents the average of the absolute percent errors.

66

Taking the absolute value prevents the positive and negative values from cancelling out. Mean Percent Error (MPE) and Standard Deviation of Percent Error (SDPE) represent the average and standard deviation of percent error, respectively.

As the bus travel times are larger than the MAC travel times in most instances as explained earlier, the negative values for ME and MPE as seen in the table are expected. The values for ME, as seen in Table 8.3 for various segments, indicate that the error increases with an increase in the length of the segment. Therefore, ME for segments 1 and 4 are lower than MEs for segments 2 and 3. The MAPE values for the four segments range between 40-55%, indicating that the bus travel times need to be adjusted significantly in order to predict automobile travel times. Figure 8.7 (a – d) show time-series plots of the difference in travel time estimates, as a function of time of day for each of the segments, respectively. Figure 8.7 (a, b) show errors for the eastbound segments, whereas Figure 8.7 8 (c, d) represent errors for the westbound segments. Each gray circle in the plots represents a travel time error along that segment. Figure 8.7 (a-d) indicate that the travel time errors are clustered and lie within a range for a value, barring some outliers. This trend is replicated over all days. For example, the range of travel time errors for segments 2 and 3 is between 0 and 300 seconds, and for segments 1 and 4 the errors range between 0 and 100 seconds.

(a)

(b)

(c)

(d)
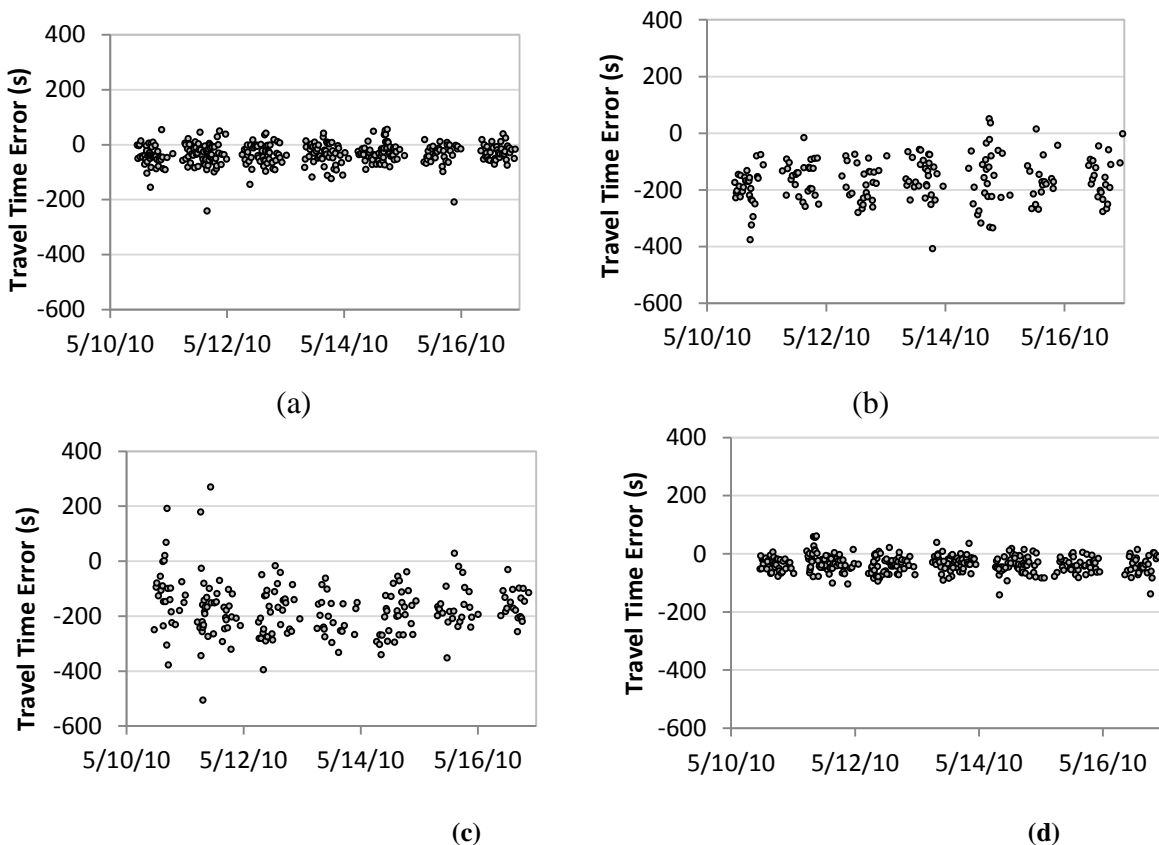
Figure 8.7 Travel Time Error Plots for Segments 1-4

67

The magnitude of errors are higher for segments 2 and 3 and lower for segments 1 and 4, as indicated in Figure 8.7. There may be two main contributing factors: a) length of segment and b) number of stops. These factors are consistent with findings by other researchers (Chakroborty, 2004). Segments 2 and 3 are longer (2.3 miles) and have more stops (Wunnava and Wasson, 2008) than segments 1 and 4 (0.7 miles, 4-5 stops). The effect of an increase in the length of the segment implies that the bus has to travel for longer distances at lower speeds than that of an automobile, thereby causing travel time errors to accrue. The effect of increases in the number of bus stops indicates that the buses have to stop and serve passenger demand more often. The resulting acceleration and deceleration phases due to an increase in the number of stops are not experienced by automobiles and result in longer travel times for buses, thereby contributing to larger magnitudes of error. The availability of limited data makes it impossible to quantify the effects of length of segment and number of stops on average error. Other factors that could contribute to higher bus travel times include age, operating characteristics of the bus and driver behavior, which are hard to quantify.

This section studied the application of transit probe data to estimate arterial travel times. Travel time comparisons made between a large ground truth data set derived from MAC readers and a large bus data set obtained from TriMet indicate that the bus travel times are consistently and positively biased. The results obtained based on the comparison of estimates indicated that errors are larger for longer segments with more stops. Other sources of error include age of bus, operating characteristics of bus and driver behavior. The findings from this research can be incorporated in the development of a method to estimate corridor travel times using bus probe data.

8.2   SIGNAL TIMING AND VEHICLE COUNTS

Visualizations pertaining to the intersection phase and timing are not yet available in PORTAL; however, they will be implemented in the future. Challenges with this data source include very large amounts of data, which requires enormous data storage, handling and processing capabilities. Sample visualizations have been generated in Excel and R and are shown below. Figure 8.8 shows a plot of phases and splits for one cycle at an intersection.
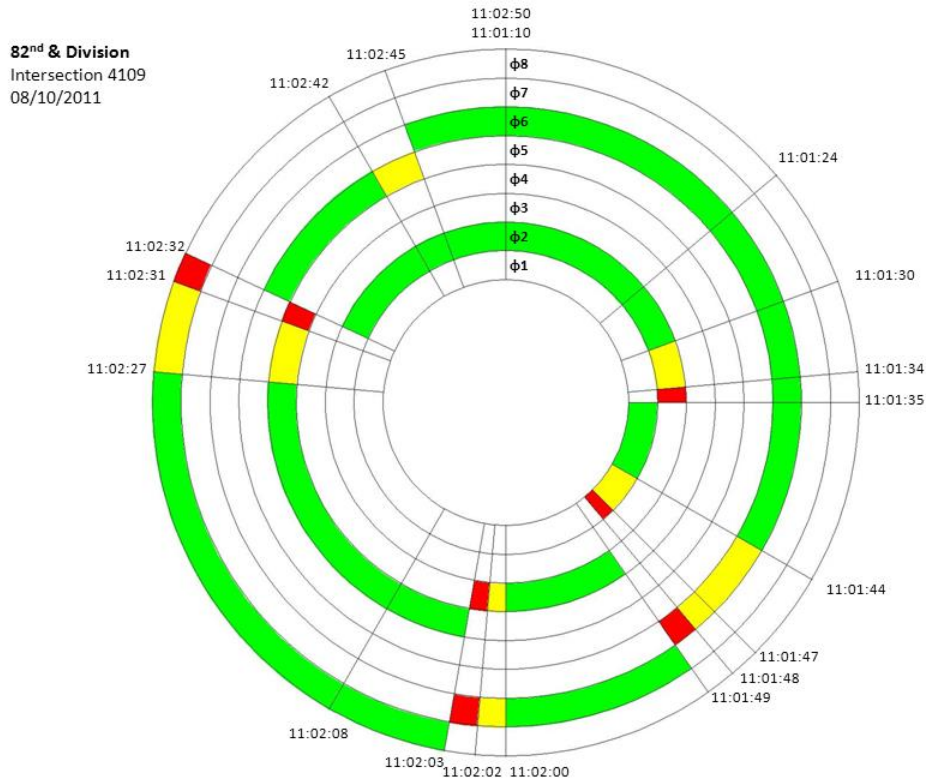
Figure 8.8 SE 82nd Ave and Division, Intersection Phase Timing Pinwheel from 11:01:50 to 11:02:50

The plot shows actual green, yellow and red times for all eight phases, as timed by the signal controller for a cycle at an intersection. From the plot, it is easy to see that phases 2 and 6 were served the highest amount of green, followed by phases 4 and 8. Such plots allow the user to visualize the intersection operation in a graphical manner.

To compare signal-cycle data, we wish to compare the percent of signal green time for the signal phase (or phases) associated with an approach with the volume for that approach. We will compare these data at a 15-minute resolution. To make this comparison, the percent of green for each phase for each 15-minute interval in the study time period must be calculated. That green percentage will be compared with 15-minute volumes, which are calculated by aggregating volumes to a 15-minute resolution. We note that volumes are represented in vehicles per lane per hour (VPLPH).

8.2.1   Traffic Signal Phase and Timing Data

The signal-cycle data are received by PSU as a feed from the Green Driver company. Recall that for the signal-cycle data, we need to calculate the percent of signal green time for each phase.

As discussed in Chapter 2, the signal phase and timing data includes phasing information for each signal represented as a series of bit flag fields are used to indicate phase information. Bit flag fields use the bits in an integer value to represent a series of yes/no flags, as described in Chapter 2. For example a value of 136 (128+8) in the Green field indicates that phases 8 and 4 are currently displaying green. Table 8.4 shows the values associated with each phase.

Table 8.4 Phases And Values

Phase Number	Value
1	1
2	2
3	4
4	8
5	16
6	32
7	64
8	128

To calculate the green times for each phase from the signal-cycle data, we use the matrix of possibilities presented in Chapter 2 and repeated, for convenience, in Table 8.5. The matrix contains eight columns – one for each phase. In each column, we list all possible values of a 16-bit integer for which the phase associated with the column is active. For example, in column 1 (Phase 1), the numbers 1, 3, 17, 33, 19, 35, 49 and 51 are listed. This means that if the Green field in the signal-cycle data is equal to one of these values, then Phase 1 is active. So, given a value for the Green field from the signal-cycle data, to determine which phases are active locate all occurrences of the value in the table – the phases active are the column headings. Note that the same value can appear in multiple columns, which indicates that those corresponding phases are active together. For example, a value of 196 indicates that phases 3, 7 and 8 are active.

Table 8.5 Phase Table

Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7	Phase 8
1	2	4	8	16	32	64	128
3	3	12	12	17	33	68	132
17	18	68	72	18	34	72	136
33	34	132	136	48	48	192	192
19	19	76	76	19	35	76	140
35	35	140	140	49	49	196	196
49	50	196	200	50	50	200	200
51	51	204	204	51	51	204	204

To calculate the time for which a phase was active, we must determine when the phase became active and active and then when the phase is no longer active. As shown in the example in

Table 8.6, the signal-cycle data are a series of Green values – 34, 34, 0 in this example. We wish to calculate the time for which Phase 2 was green. From Table 8.5, we observe that Phase 2 is green when the Green value is 34, thus Phase 2 becomes green at 34.24 and is no longer green at 43.23. Thus the length that Phase 2 was green is (43.23-34.24) or approximately nine seconds.

70

Table 8.6 Active Phase Calculation Example

Green	Timestamp	Phase 2 Green Flag	Phase 2 Green Time
34	12/1/2011 0:00:34.24	1	
34	12/1/2011 0:00:36.24	1	
0	12/1/2011 0:00:43.23	0	8.98

Once the green times for each phase are calculated, the green times are aggregated to 15-minute intervals. The percent of green for each phase for that 15-minute interval is calculated using the following formula:

Phase X % Green = 15-Min. Aggregated Green Time * 100 / (15 * 60)

This calculation is done for all phases for all 15-minute intervals in the study time period and the results are recorded.

8.2.2 Vehicle Count Data

Detector volume data are received by PSU from the City of Portland at regular intervals. An example of the detector volume data are provided in Table 8.7. As shown, volumes are received on a minute-by-minute level and, in addition, there are often multiple detectors on a single approach. To calculate the 15-minute volume for an approach, we first sum the one-minute volumes for all detectors at an approach to create a total one-minute granularity approach volume. Then, the one-minute volumes are aggregated into 15-minute volumes. The total 15-minute aggregation for detectors 411 and 412 for the first 15 minutes of December 1 is shown in Table 8.7.

Table 8.7 Arterial Volume Data Aggregation Example

Date and Time	Detector ID	Volume	Detectorid	Volume	SB Vol	SB Agg
12/1/2011 0:00	411	0	412	0	0	
12/1/2011 0:01	411	2	412	2	4	
12/1/2011 0:02	411	2	412	1	3	
12/1/2011 0:03	411	1	412	0	1	
12/1/2011 0:04	411	1	412	0	1	
12/1/2011 0:05	411	1	412	1	2	
12/1/2011 0:06	411	0	412	2	2	
12/1/2011 0:07	411	0	412	0	0	
12/1/2011 0:08	411	3	412	0	3	
12/1/2011 0:09	411	1	412	1	2	
12/1/2011 0:10	411	1	412	0	1	
12/1/2011 0:11	411	0	412	1	1	
12/1/2011 0:12	411	1	412	0	1	
12/1/2011 0:13	411	2	412	0	2	
12/1/2011 0:14	411	0	412	0	0	23

After the aggregation procedure described above and shown in Table 8.7, the result is a total volume for a 15-minute period. We represent volumes in units of vehicles per lane per hour (VPLPH). We convert each 15-minute aggregation from a total volume to VPLPH by

multiplying the volume by 4 (there are 4 15-minute periods in an hour) and then dividing by the number of detectors at the approach (two in this case).

Finally, we calculate the volume per hour of green by taking the VPLPH number and multiplying by 100 and dividing by the phase percent green for that approach. This gives us the volume number used in the comparisons in the next section.

8.2.3 Comparison of 82nd Avenue Signal Performance

The signal-cycle data and volume data were compared for two locations along 82nd Avenue. The two locations are 82nd Avenue and Burnside and 82nd Avenue and Flavel. The locations are shown in Figure 8.9

Figure 8.9 Study Locations

Figure 8.10 and Figure 8.11 show the volume and percent green for the study locations on 82nd Avenue for a selected day. Volumes are plotted as lines on the left y-axis; green percent is plotted on the right *y*-axis. The *x*-axis is time. Northbound volume and percent green are shown in blue in both plots; southbound volume and percent green are shown in red. There is a very limited relationship between percent green and volume in these data sets. In Figure 8.10, there is a small drop in percent green associated with the increased AM peak volume. We conclude that there is a limited correlation between volume and percent green in this data set, and as a result do not believe the percent green can be used as a surrogate for a volume count using the current detection.
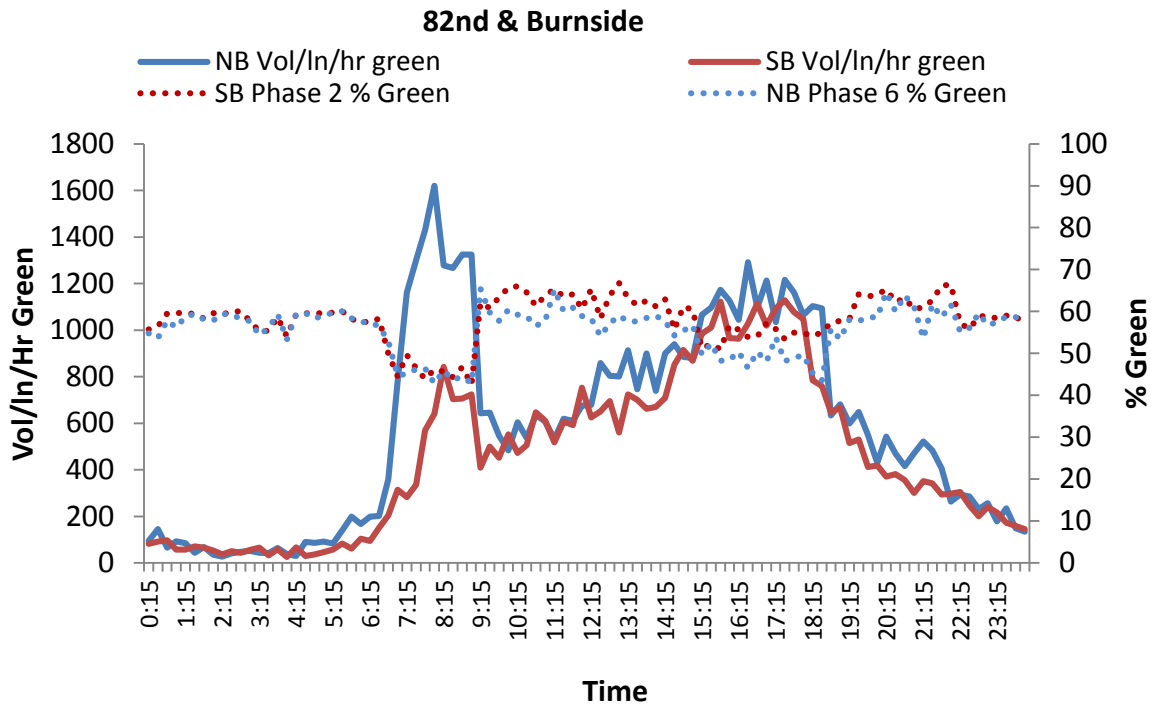
Figure 8.10 Phase - Volume Comparison - 82nd & Burnside
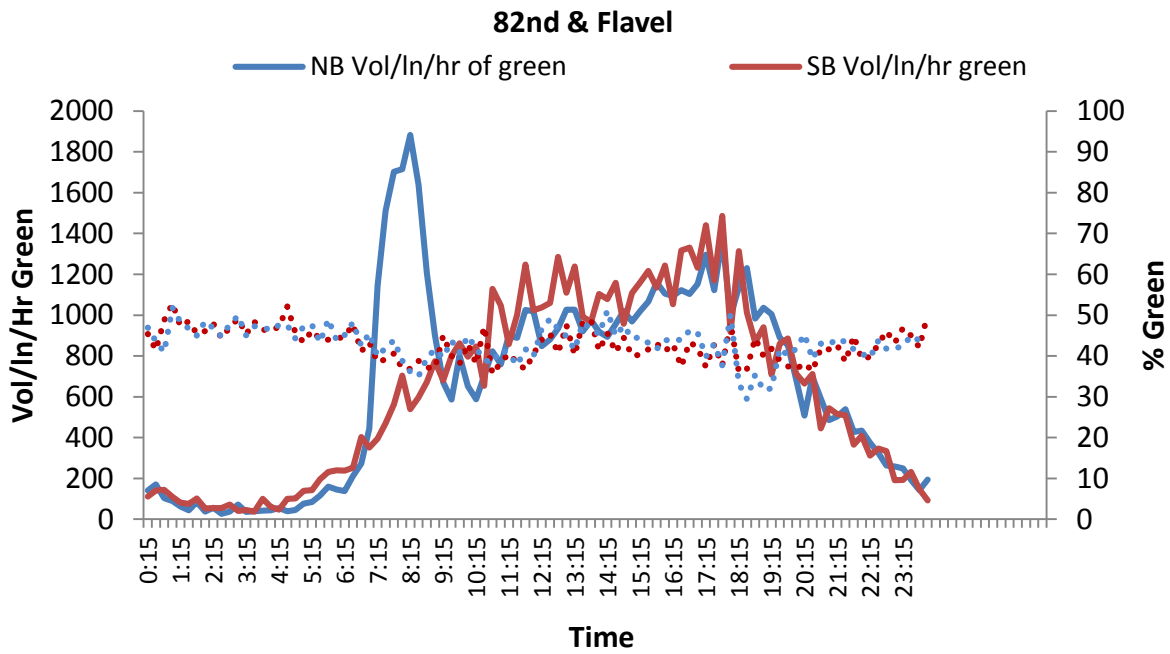
Figure 8.11 Phase-Volume Comparison 82nd & Flavel

9.0   CONCLUSION

The creation of an arterial data archive is fundamental towards understanding transportation system performance. The inclusion of multimodal data in transportation archives is necessary to quantify demand, and aids in the establishment of a balanced transportation system that serves all users.

Multimodal data are necessary if agencies are to implement the policies that are described in transportation planning documents. The Portland metropolitan region has set targets for encouraging pedestrian and bicycle travel, and for making the existing roadway infrastructure operate as effectively and efficiently as possible. For arterial control, understanding the number of pedestrians per hour can inform whether coordinated operation is unduly delaying the person trying to cross the street. Similarly, it may inform the operating engineer when operating a traffic signal in traffic responsive control would benefit the system's users. It is possible to extend this thinking to the competitiveness of transit on the arterial, identifying control decisions that would enable buses to closer match the performance of single occupant vehicles on the network. This data collected as a result of these efforts will lead to decisions about how the Portland region is implementing a transportation system that meets the goals of encouraging travel decisions that promote active living while moving freight reliably.

The data archival process allows for the provision of real-time traveler information, long-term performance monitoring, operational improvements and development of long-range plans. A framework for establishing a multimodal data archive is outlined in this paper. Detailed descriptions of these data sources and sample visualizations that aid in understanding system performance have been developed. A web interface of this data archive is in development. The inclusion of arterial, transit, Bluetooth, bicycle and pedestrian data sets complement the existing freeway data sets in PORTAL and offer tremendous potential for operations and research.

Currently, an effort is underway by regional transportation agencies in Portland to identify performance measures that can be used to assess an arterial system's efficiency. As part of the Multimodal Arterial Performance Management Regional Concept of Operations, a comprehensive list of performance measures has been developed for the region and these measures have been classified by mode and ranked by priority. The purpose of the list was to develop a regional concept for collecting and archiving multimodal travel data on Portland's arterial network.

9.1   NEXT STEPS

Over the course of these research projects, we have identified a number of next steps that are required to improve the PORTAL archive for a regional program. These next steps are summarized in Table 9.1

Table 9.1 Arterial Data Archival – Data Retrieval Process – Next Steps and Coverage Percentage

Data Type Name	Data Retrieval Process - Next Steps	Coverage – Amount of available data in PORTAL
COP- Vehicle Count	Long term, the Portland Metro region is working with TransSuite to develop a data feed from the TransSuite system that can be used to transfer the arterial count data to PSU and to ODOT. Near-term improvements to the retrieval process are desired to automate the transfer of the files from TransSuite to PSU to avoid a student needing to transfer the files manually.	The majority of City of Portland count stations are in the PORTAL archive.
COP-Travel Time	Long term, PSU should work with the City of Portland to automate the Bluetooth data fetching process.	Bluetooth stations are located strategically along Powell Blvd and 82nd Ave. All Bluetooth stations in existence are available in PORTAL.
COP- Traffic Signal Phase and Timing	The signal-cycle feed is fully automated. The only remaining steps are to load these data into the PORTAL archive. The loading process was on hold due to disk space constraints.	The coverage of the signal-cycle data depends on the Green Driver feed. Coverage is believed to be currently complete.
Clark County – Vehicle Counts – Radar	Near term, the semi-automated process for archiving data is acceptable. Long term, a fully automated process would be ideal but can be prioritized lower due to the efficiency of the current feed.	Approximately five of the Clark County radar stations are in the PORTAL archive. Approximately 5-10 more stations are available and need to be loaded into PORTAL.
Clark County – Vehicle Counts – Loop	Near term, the semi-automated process for archiving data from ATMS is not acceptable and the overhead is too high. Discussions are underway with Naztec, the ATMS.Now vendor, to create a data feed that would support automated data download.	Coverage is minimal. Due to the overhead of extracting data from the ATMS.Now system, only one Clark County loop station is in the PORTAL archive.
TriMet – Transit AVL/APC Data	Currently, PSU is using TriMet quarterly data. The current transfer process is acceptable for quarterly data. When the move is made to use TriMet daily data, an automated transfer process should be investigated.	Coverage is dependent on TriMet's collection ability. Coverage is relatively complete and coverage percentage will increase with TriMet's transition to the Init system.
C-TRAN – Transit AVL/APC Data	The semi-automated transfer process is acceptable in terms of overhead. Next steps would be to evolve the semi-automated process into a regular, periodic transfer.	A small sample of C-TRAN data is in PORTAL. The spatial coverage is good, but the time coverage is not.

Data Type Name	Data Retrieval Process - Next Steps	Coverage – Amount of available data in PORTAL
COP – Bicycle Data	The overhead of the semi-automated transfer process is higher than desired. Work should be done to reduce the time involved for the graduate student doing the transfers. Cost: 5-10 minutes per intersection to pull bike or ped data. Ten minutes to process these data per intersection. Twenty minutes per intersection. Ten hours of work (30 intersections) to pull one set of data.	Bicycle count data are pulled from 11 intersections, not all are in PORTAL. In addition, there are more potential intersections from which we could get counts.
COP – Pedestrian Actuations	The overhead of the semi-automated transfer process is higher than desired. Work should be done to reduce the time involved for the graduate student doing the transfers.	Pedestrian actuation data are pulled from 17 intersections, but only seven of the intersections are in PORTAL.

10.0 REFERENCES

Berkow, M., C.M. Monsere, P. Koonce, R.L. Bertini, and M. Wolfe. Prototype for Data Fusion Using Stationary and Mobile Data Sources for Improved Arterial Performance Measurement. *Transportation Research Record: Journal of the Transportation Research Board, No. 2099*, TRB, National Research Council, Washington, D.C., 2009, pp. 102-112.

Berkow, M., Wolfe, M., Monsere, C., Bertini, R. "Using Signal System Data and Buses as Probe Vehicles to Define the Congested Regime on Arterials." 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.

Bertini R.L. and S. Tantiyanugulchai. Transit buses as traffic probes – use of geolocation data for empirical evaluation. *Transportation Research Record* 1870, TRB, National Research Council, Washington D.C., pp 35-45, 2004.

Bertini, R.L., Hansen, S., Bryd, A. and T. Yin. Experience Implementing a User Service for Archived Intelligent Transportation Systems Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1917*, TRB, National Research Council, Washington, D.C., 2005, pp. 90-99.

Bertini, R.L., M. Lasky, and C. M. Monsere. Validating Predicted Rural Corridor Travel Times from an Automated License Plate Recognition System: Oregon's Frontier Project. *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, 2005.

Bhaskar, A., E. Chung and A.G. Dumont. "Fusing loop detector and probe vehicle data to estimate travel time statistics on signalized urban networks. Computer-Aided Civil and Infrastructure Engineering, 26, pp. 433-450. 2011.

Cathey, F.W. and D. J. Dailey. Estimating Corridor Travel Time by Using Transit Vehicles as Probes. *Transportation Research Record* 1855, TRB, National Research Council, Washington D.C., pp.60-65, 2003.

Cathey, F.W. and D. J. Dailey. Transit Vehicles as Traffic Probe Sensors. *Transportation Research Record No.* 1804, TRB, National Research Council, Washington D.C., pp 23-30, 2002.

Chakroborty, P. and S. Kikuchi. Using bus travel time data to estimate travel times on urban corridors. *Transportation Research Record* 1870, TRB, National Research Council, Washington D.C., pp 18-25, 2004.

Chen C., Petty, A., Skabardonis, A., Varaiya, P. and Z. Jia. Freeway Performance Measurement System: Mining Loop Detector Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1748*, TRB, National Research Council, Washington, D.C., 2001, pp. 96-102.

Choi, K. and Y. Chung. A Data Fusion Algorithm for Estimating Link Travel Time. *Journal of Intelligent Transportation Systems Vol. 7,* pp 235-260, 2002.

Coifman, B. Vehicle Reidentification and Travel Time Measurement in Real-Time on Freeways using the Existing Loop Detector Infrastructure. *Transportation Research Record No. 1643*, TRB, National Research Council, Washington D.C., 1998, pp. 181–191.

Dion, F. and Rakha, H. Estimating Dynamic Roadway Travel Times Using Automatic Vehicle Identification Data for Low Sampling Rates. In *Transportation Research Part B 40*, pp 745–766, 2006.

Elango, C. and D. Dailey. "Irregularly Sampled Transit Vehicles used as a Probe

Haghani, A., M. Hamedi, K. Sadabadi, S. Young, and P. Tarnoff. Freeway Travel Time Ground Truth Data Collection Using Bluetooth Sensors. *Proceedings of 89th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2010.

Hall, R. and N. Vyas. Buses as a Traffic Probe: Demonstration Project In *Transportation Research Record No. 1731*, TRB, National Research Council, Washington D.C. pp 96–103, 2000.

Hallenbeck, M.E., J.M. Ishimaru, K.D. Davis and J.M. Kang. Arterial Performance Measurement Using Stop Bar Data. *Proceedings of the 87th Annual Meeting of the Transportation Research Board*, TRB, National Research Council, Washington, DC, 2008.

Kothuri, S., Reynolds, T., Monsere, C. and P. Koonce. Preliminary Development of Methods to Automatically Collect Pedestrian Counts and Bicycle Delay at Signalized Intersections. *Proceedings of the 91st Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2012.

Kothuri, S., Reynolds, T., Monsere, C. and P. Koonce. Testing Strategies to Reduce Pedestrian Delay at Signalized Intersections: A Pilot Study in Portland, Oregon. *Proceedings of the 92nd Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2013.

Kwon, J. and K. Petty. An Arterial Performance Measurement System. Presentation at the *Annual Meeting Workshop of the TRB Signal Systems Committee*, Washington DC, 2006.

Kwong, K., R. Kavaler, R. Rajagopal, and P. Varaiya. A Practical Scheme for Arterial Travel Time Estimation Based on Vehicle Re-identification Using Wireless Sensors. *Proceedings of 87th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2008.

Li, H., Hainen, A., Day, C., Grimmer, G., Sturdevant, J. "Longitudinal Performance Measures for Assessing Agency-Wide Signal Management Objectives." 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013

Liu, H. and P. Michalopoulos. Development of a Real-Time Arterial Performance Monitoring System Using Traffic Data Available from Existing Signal Systems. Research Report, University of Minnesota.

Liu, H., and W. Ma. Time-Dependent Travel Time Estimation Model for Signalized Arterial Network. *Proceedings of the 86th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2007.

Liu, H., and W. Ma. Time-Dependent Travel Time Estimation Model for Signalized Arterial Network. *Proceedings of the 86th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2007.

Liu, H., Ma, W. "Performance Measurement System for Arterial Traffic Signals." 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.

Lund, A. and M.L. Pack. Dynamic Wide-Area Congestion and Incident Monitoring Using Probe Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2174*, TRB, National Research Council, Washington, D.C., 2010, pp. 1-9.

Malinovskiy, Y., Y. Wu, Y. Wang, and U. Lee. Field Experiments on Bluetooth-based Travel Time Data Collection. *Proceedings of 89th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2010.

Ndoye, M., V. Totten, B. Carter, D.M. Bullock, and J.V. Krogmeier. Vehicle detector Signature Processing and Vehicle Reidentification for Travel Time Estimation. *Proceedings of 88th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2008.

Northwest Signal Supply Inc. Voyage Software Operating Manual.

Oh, C. and S.G. Ritchie. Real-time Inductive-Signature-Based Level of Service for Signalized Intersections. In *Transportation Research Record No. 1802*, TRB, National Research Council, Washington D.C. pp 97–104, 2002.

Olson, C., Kothuri, S., Monsere, C., Koonce, P., Tufte, K. "A Framework for Multimodal Arterial Data Archiving." 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.

PeMS Iteris http://www.iteris.com/iperform.aspx

PORTAL. PORTAL transportation archive at Portland State University. http://demo.portal.its.pdx.edu

Pu, W., J. Lin, and L. Long. Real-Time Estimation of Urban Street Segment Travel Time Using Buses as Speed Probes. *Proceedings of 88th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2009.

Quayle, S.M., P. Koonce, D. DePencier, and D. Bullock. Arterial Performance Measures Using MAC Readers: Portland Pilot Study. *Proceedings of 89th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2010.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

RCTO TAC. Metro Multimodal Arterial Performance Measure Regional Concept of Operations. February 2013.

Remias, S, Hainen, A., Day, C., Li, H., Rivera-Hernandez, E., Sturdevant, J., Young, S., Bullock, D. "Characterizing Performance of Arterial Traffic Flow Using Probe Vehicle Data"

RITIS. RITIS at the CATT Lab at University of Maryland. http://www.cattlab.umd.edu/?portfolio=ritis

Sadabadi, K., M. Hamedi, and A. Haghani. Evaluating Moving Average Techniques in Short-Term Travel Time Prediction Using an AVI Dataset. *Proceedings of 89th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2010.

Skabardonis, A., and N. Geroliminis. Real-Time Estimation of Travel Times on Signalized Arterials. *Proc., 16th International Symposium on Transportation and Traffic Theory,* University of Maryland, College Park, July 2005.

Slavin, C., Feng, W., Figliozzi, M., Koonce, P. "A Statistical Study of the Impacts of SCATS Adaptive Traffic Signal Control on Traffic and Transit Performance." 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013

Sun, C. S.G. Ritchie, K. Tsai, and R. Jayakrishnan. Use of Vehicle Signature Analysis and Lexicographic Optimization for Vehicle Reidentication on Freeways. *Transportation Research, C*, 7:167–185, 1999.

Sun, C. S.G. Ritchie, K. Tsai, and R. Jayakrishnan. Use of Vehicle Signature Analysis and Lexicographic Optimization for Vehicle Reidentication on Freeways. *Transportation Research, C*, 7:167–185, 1999.

Sun, C.C., G.S. Arr, R.P. Ramachandram, and S.G. Ritchie. Vehicle Re-identification Using Multidetector Fusion. *IEEE Transactions on Intelligent Transportation Systems*, Volume 5, No.3, September 2004, pp 155-164.

Sun, C.C., G.S. Arr, R.P. Ramachandram, and S.G. Ritchie. Vehicle Re-identification Using Multidetector Fusion. *IEEE Transactions on Intelligent Transportation Systems*, Volume 5, No.3, September 2004, pp 155-164.

Tarnoff, P.J., J. S. Wasson, S.E. Young, M. Ganig, D.M. Bullock, and J.R. Sturdevant. The Continuing Evolution of Travel Time Data Information Collection and Processing. *Proceedings of the 88th Annual Meeting of the Transportation Research Board*, TRB, National Research Council, Washington, DC, 2009.

Texas Transportation Institute. Guidelines for Developing ITS Data Archiving Systems. Report 2127-3, 2001.

U.S. Department of Transportation, Research and Innovative Technology Administration. ITS Strategic Research Plan 2010-2014. Accessed from
http://www.its.dot.gov/strat_plan/strategic_plan2010_2014/index.htm

Uno, N., F. Kurauchi, H. Tamura and Y. Iida. Using Bus Probe Data for Analysis of Travel Time Variability. *Journal of Intelligent Transportation Systems,* Volume 13 Issue 1, pp 2-15, 2009.

Using Media Access Control Address Matching. *ITE Journal*, 78(6), 20-3., 2008.

Vehicle Traffic Sensor." *Transportation Research Record 1719*, TRB, National Research Council, Washington D.C., pp. 33–44, 2002.

Washington State Department of Transportation. The Gray Notebook. http://www.wsdot.wa.gov/Accountability/GrayNotebook/navigateGNB.htm

Wasson, J. S., Sturdevant, J. R., and Bullock, D. M. "Real-Time Travel Time Estimates Using Media Access Control Address Matching." ITE Journal, 78(6), 20-23, 2008.

Wolfe, M., Monsere, C., Bertini, R. "Exploring the Effects of Sensor Data Aggregation on Measuring Arterial Performance." 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013

Wunnava, S.V., K. Yen, T. Babij, R. Zavaleta, R. Romero and C. Archilla. Travel Time Estimation Using Cell Phones for Highways and Roadways. Report prepared for the Florida Department of Transportation.

Zhang, H. M. Link–Journey–Speed Model for Arterial Traffic. *Transportation Research Record: Journal of the Transportation Research Board, No. 1676,* TRB, National Research Council, Washington, D.C., 1999, pp. 109–115.

APPENDIX A

PORTAL DATA FEED SPECIFICATION

In this Appendix we provide a PORTAL data feed specification. The intent is to specify common formats to allow data to be incorporated directly into the archive. Common formats allow the addition of more data sources to the archive at minimal cost.

The Appendix provides draft specifications for PORTAL arterial count and Bluetooth data feeds. We call these specifications: PORTAL Arterial Feed Specification (PAFS) and PORTAL Bluetooth Feed Specification (PBFS). For transit data, Google Transit Feed Specification (GTFS) for stop and schedule information and GTFS-real-time for real-time arrival information is recommended.

The feed specifications are based on the GTFS, which is rapidly gaining acceptance for transit schedule and real-time arrival information. GTFS uses comma-delimited text files, which are simple to create, easy to load into databases and much smaller than similar XML or JSON files. Their simple structure results in more efficient data transfer and loading. Taking the lead from GTFS, the PORTAL Feed Specifications also use comma-delimited files. PORTAL Feed Specifications use term definition, file requirements and field names similar to GTFS wherever possible. Finally, the PORTAL Feed Specifications follow the format of GTFS attributes where possible.

PORTAL Arterial Feed Specification

The PAFS specifies a set of metadata files and a data feed file for arterial count station data to be incorporated into the PORTAL archive.

Table A-1 PORTAL Arterial Feed Specification (PAFS) Files

Filename	Required	Defines
intersections.txt	Required	Arterial intersections covered by this feed.
approaches.txt	Required	Approaches to intersections covered by this feed. Each intersection will have multiple approaches.
lanes.txt	Required	Lanes within approaches covered by this feed. Each approach will have one or more lanes.
detectors.txt	Required	????? Not sure if I need this as there may be one detector per lane???
arterial_data.txt	Required	Arterial count station data feed.

Table A-2 intersections.txt

Field Name	Required	Details
intersection_id	Required	The **intersection_id** field is a unique integer that identifies an intersection. The **intersection_id** is dataset unique.
intersection_name	Required	The **intersection_name** field contains the name of the intersection.
intersection_lat	Required	The **intersection_lat** field contains the latitude of the intersection. The field value must be a valid WGS 84 latitude.
intersection_lon	Required	The **intersction_lon** field contains the longitude of the intersection. The field value must be a valid WGS 84 longitude value from -180 to 180.

Table A-3 approaches.txt

Field Name	Required	Details
approach_id	Required	The **approach_id** field is a unique integer that identifies a directional intersection approach. The combination of the **intersection_id** and **approach_id** fields is dataset unique.
intersection_id	Required	The **intersection_id** field contains the ID for the intersection with which this approach is associated. The **intersection_id** is referenced from the intersections.txt file.
approach_name	Required	The **approach_name** field contains the name of the approach.
Bound	Required	The **bound** field contains a one- or two-character identifier of direction of the approach. Valid values are: N, S, E, W, NE, NW, SE, SW.

Table A-4 lanes.txt

Field Name	Required	Details
lane_id	Required	The **lane_id** field contains the number of the lane of this approach. The combination of the **lane_id** and **approach_id** fields is dataset unique.
approach_id	Required	The **approach_id** field contains the ID of the approach with which this lane is associated. The **approach_id** is referenced from the approaches.txt file.
intersection_id	Required	The **intersection_id** field contains the ID for the intersection with which this lane is associated. The **intersection_id** is referenced from the intersections.txt file.
lane_type	Required	The **lane_type** field contains the type of the lane. Valid values are: RIGHT, STRAIGHT, LEFT (NEED HELP)
detector_id	Required	The **detector_id** field is a unique integer that identifies the detector. The **detector_id** field is dataset unique.

Table A-5 arterial_data.txt

Field Name	Type	Required	Description
detector_id	Integer	Required	The **detector_id** field is the ID of the detector for this data reading. The **detector_id** field is referenced from the arterial_lanes.txt file.
Timestamp	Timestamp	Required	The **timestamp** field contains the date and time of this data reading. Timestamp is specified as YYYY-MM-DD HH:MM:SS.
Volume	Integer	Required	The **volume** field contains the volume reading. Volume is a count.
Speed	Double	Optional	The **speed** field contains the speed reading. Speed must be reported as miles per hour.
Occupancy	Double	Optional	The **occupancy** field contains the occupancy. Occupancy must be a percentage between 0 and 100.
Status	Integer	Optional	The **status** field contains the status of this reading. This field may be used to indicate data quality or detector issues.

PORTAL Bluetooth Feed Specification

The PBFS specifies a set of metadata files and a data feed file for arterial count station data to be incorporated into the PORTAL archive.

Table A-6 PORTAL Bluetooth Feed Specification (PBFS) Files

Filename	Required	Defines

bluetooth_detectors .txt	Required	Bluetooth detectors covered by this feed.
bluetooth_data.txt	Required	Bluetooth data feed.

Table A-7 bluetooth_detectors.txt

Field Name	Required	Details
bt_detector_id	Required	The **bt_detector_id** field is a unique integer that identifies a Bluetooth detector. The **bt_detector_id** is dataset unique.
bt_detector_name	Required	The **bt_detector_name** field contains the name of the Bluetooth detector.
bt_detector_lat	Required	The **bt_detector_lat** field contains the latitude of the Bluetooth detector. The field value must be a valid WGS 84 latitude.
bt_detector_lon	Required	The **bt_detector_lon** field contains the longitude of the Bluetooth detector. The field value must be a valid WGS 84 longitude value from -180 to 180.

Table A-8 bluetooth_data.txt

Field Name	Type	Required	Description
bt_detector_id	Integer	Required	The **bt_detector_id** field is the ID of the detector for this data reading. The **bt_detector_id** field is referenced from the bluetooth_detectorss.txt file.
Timestamp	Timestamp	Required	The **timestamp** field contains the date and time of this data reading. Timestamp is specified as YYYY-MM-DD HH:MM:SS.
Macid	Text	Required	The **macid** field contains the anonomized MAC address of this data reading.
signal_strength	Double	Optional	The **signal_strength** field contains the signal strength of this data reading.

Term Definitions and File Requirements

The Term Definitions and File Requirements are adopted from GTFS with slight modifications. We credit the Google team for these definitions and requirements as well as the general feed structure.

Term Definitions

- Field required: The column must be included in your feed and a value must be provided for each record.
- Field optional: The column may be omitted from your feed. If you include an optional column, each record in your feed must have a value for that column.
- Dataset unique: The field contains a value that maps to a single distinct entity within the column. For example, in the highways.txt file highway_id is dataset unique. In other words, each highway must have a unique identifier which is the highway ID.

File Requirements

The following are requirements for the format and content of the feed files:

- All PORTAL Feed Specification files must be saved as comma-delimited text.

- The first line of each file must contain field names. The tables above list the field names that may be used in a file.

- Field names are case sensitive.

- Field values may not contain tabs, carriage returns or new lines.

- Field values that contain quotation marks or commas must be enclosed within quotation marks. In addition, each quotation mark in the field value must be preceded with a quotation mark. This is consistent with the manner in which Microsoft Excel outputs comma-delimited (CSV) files. For more information on the CSV file format see http://tools.ietf.org/html/rfc4180. The following example demonstrates how a field value would appear in a comma-delimited file:

  - **Original field value:** Contains "quotes," commas and text.

  - **Field value in CSV file:** Contains "quotes," commas and text.

- Field values must not contain HTML tags, comments or escape sequences.

- Remove any extra spaces between fields or field names. Many parsers consider the spaces to be part of the value, which may cause errors.

- Each line must end with a CRLF or LF linebreak character.

- Files should be encoded in UTF-8 to support all Unicode characters. Files that include the Unicode byte-order mark (BOM) character are acceptable. Please see the (*Unicode FAQ*) for more information on the BOM character and UTF-8.

- Zip the files in your feed.

APPENDIX B

DATA DICTIONARIES

Table B-1 Data Dictionary for Arterial Detector File (arterial_detectors.csv)

Attribute Name	Attribute Type	Description
detectorid	Integer	ID of the detector (key)
lane	Character (30)	Lane in which the detector is located
stationid	Integer	ID of the station

Sample

Attribute Name	Example 1	Example 2
detectorid	253	254
lane	RIGHT	LEFT
stationid	156	156

Table B-2 Data Dictionary for Arterial Intersection File (arterial_intersections.csv)

Attribute Name	Attribute Type	Description
intersectionid	Integer	ID of the intersection (Key)
location	Character (100)	Location of the Intersection
lat	Character (30)	Latitude of the Intersection
lon	Character (30)	Longitude of the Intersection

Sample

Attribute Name	Example 1	Example 2
intersectionid	2117	4110
location	NE Fremont St & 82nd Ave	SE Powell Blvd & 82nd Ave
lat	45.548142	45.497411
lon	-122.578737	-122.578702

Table B-3 Data Dictionary for Arterial Station File (arterial_stations.csv)

Attribute Name	Attribute Type	Description
stationid	Integer	ID of the station (Key)
bound	Character (1)	Direction of travel based on station location (N, S, E or W)
intersectionid	Integer	ID of the associated intersection
lat	Character (30)	Latitude of the station
lon	Character (30)	Longitude of the station

Sample

Attribute Name	Example1	Example 2
stationid	135	136
bound	N	S
intersectionid	6012	6012
latlon	45.5226	45.522595
lon	-122.578983	-122.579149

Table B-4 Data Dictionary for Raw Detector Archive File (raw_detector_archive.csv)

Attribute Name	Attribute Type	Description
detectorid	integer	ID of the detector (Key)
timestamp	timestampwithtimezone	Upload date and time

status	text	Displays the upload status (Good, Timeout, Bad Response)
sampleperiod	integer	Duration of time (seconds) during which data was collected
sequencenumber	integer	Number that is incremented by the controller when these data in its detector upload buffer has changed; used to determine whether the uploaded data are new, old or whether an upload is missed; ranges between 0-255
volume	integer	Volume in vehicles per sample period
occupancy	integer	Occupancy in units of tenth of a percent
dq_minvol	boolean	Data quality flag check for min volume; T if vol < 0
dq_maxvol	boolean	Data quality flag check for max volume; T if vol > (sampleperiod *sat flow)/3600
dq_minocc	boolean	Data quality flag check for min occupancy; T if occ < 0
dq_maxocc	boolean	Data quality flag check for max occupancy; T if occ > 1000
dq_visual	boolean	Data quality flag check for visual inspection of volume; T if volume data are suspect

Table B-5 Data Dictionary for Raw Detector Archive File (raw_detector_archive.csv)

Attribute Name	Attribute Type	Description
detectorid	integer	ID of the detector (Key)
timestamp	timestampwithtimezone	Upload date and time
status	text	Displays the upload status (Good, Timeout, Bad Response)
sampleperiod	integer	Duration of time (seconds) during which data was collected
sequencenumber	integer	Number that is incremented by the controller when these data in its detector upload buffer has changed; used to determine whether the uploaded data are new, old or whether an upload is missed; ranges between 0-255
volume	integer	Volume in vehicles per sample period
occupancy	integer	Occupancy in units of tenth of a percent
dq_minvol	boolean	Data quality flag check for min volume; T if vol < 0
dq_maxvol	boolean	Data quality flag check for max volume; T if vol > (sampleperiod *sat flow)/3600
dq_minocc	boolean	Data quality flag check for min occupancy; T if occ < 0
dq_maxocc	boolean	Data quality flag check for max occupancy; T if occ > 1000
dq_visual	boolean	Data quality flag check for visual inspection of volume; T if volume data are suspect

Table B-6 Data Dictionary for Phase and Timing Data File (phase_and_timing_data.csv)

Attribute Name	Attribute Type	Description
fromtopofcycle	integer	The value of the local timer (i.e., the number of seconds elapsed after local zero)
greens	integer	Contains bit flags representing the phases that are currently displaying a green indication

yellow	integer	Contains bit flags representing the phases that are currently displaying a yellow indication
peds	integer	Contains bit flags representing the phases that are currently displaying a Pedestrian Walk indication
ped_calls	integer	Contains bit flags representing the phases that have active ped calls
veh_calls	integer	Contains bit flags representing the phases that have active vehicle calls
status	smallint	Indicates signal status: 0 – Normal, 1 - Preempt, 2- Transition, 3 – Flash, 4 – Free, 6 - Stop
online	smallint	whether or not the light controller is 'online'
intersectionid	integer	ID of the intersection
overlays	integer	Contains bit flags representing the overlaps that are currently displaying a green indication. The least significant bit represents Overlap 1, the next least significant bit represents Overlap 2 and so on. If a bit is set, the corresponding Overlap is currently displaying green
plan_num	integer	The current timing plan number in effect
timestamp	timestamp	The time at which this data was recorded

Table B-7 Data Dictionary for Bluetooth Stations File (Bluetooth_stations.csv)

Attribute Name	Attribute Type	Description
bluetoothid	integer	ID of the detector (Key)
intersectionid	integer	ID of the associated intersection
lat	Character (30)	Latitude of the bluetooth detector location
lon	Character (30)	Longitude of the bluetooth detector location

Sample

bluetoothid	10001	10002
intersectionid	4117	4113
lat	45.5013	45.482784
lon	-122.5790	-122.5790

Table B-8 Data Dictionary for Bluetooth Travel Times File (bluetooth_traveltimes.csv)

from_bluetooth_id	integer	ID of the detector at the start of the segment (Key)
to_bluetooth_id	integer	ID of the detector at the end of the segment (Key)
timestamp	timestamp with date	Date and time of the observed travel time (Key)
traveltime	integer	Matched travel time between a pair of Bluetooth detectors
dq_tt	boolean	Data quality flag check for travel times; T if data are bad, F if data are good

Table B-9 Sample of Raw Detector Data

Attribute Name	Example 1	Example 2
detectorid	253	255
timestamp	9/15/2011 1:20:59	9/20/2011 6:15:34
status	GOOD	GOOD
sampleperiod	15	15
sequencenumber	129	155

	1	0
volume	1	0
occupancy	0	0
dq_minvol	f	f
dq_maxvol	f	f
dq_minocc	f	f
dq_maxocc	f	f
dq_visual	f	t

Table B-10 Sample of Bicycle Counts

StationID	Date	Time	Detector	Volume
200090	Saturday, September 1,2012	0:30	24	15
200090	Saturday, September 1,2012	1:30	24	3
200090	Saturday, September 1,2012	2:30	24	5
200090	Saturday, September 1,2012	3:30	24	2
200090	Saturday, September 1,2012	4:30	24	2
200090	Saturday, September 1,2012	5:30	24	7
200090	Saturday, September 1,2012	6:30	24	3
200090	Saturday, September 1,2012	7:30	24	17
200090	Saturday, September 1,2012	8:30	24	18
200090	Saturday, September 1,2012	9:30	24	20
200090	Saturday, September 1,2012	10:30	24	21
200090	Saturday, September 1,2012	11:30	24	31
200090	Saturday, September 1,2012	12:30	24	39
200090	Saturday, September 1,2012	13:30	24	31
200090	Saturday, September 1,2012	14:30	24	25
200090	Saturday, September 1,2012	15:30	24	36
200090	Saturday, September 1,2012	16:30	24	21
200090	Saturday, September 1,2012	17:30	24	24
200090	Saturday, September 1,2012	18:30	24	20
200090	Saturday, September 1,2012	19:30	24	13
200090	Saturday, September 1,2012	20:30	24	11
200090	Saturday, September 1,2012	21:30	24	10
200090	Saturday, September 1,2012	22:30	24	8
200090	Saturday, September 1,2012	23:30	24	10

Table B-11 Sample of Pedestrian Actuations and Delay Data

Date	Time	Detector	Count
Wednesday, August 1,2012	0:30	33	1
Wednesday, August 1,2012	0:30	34	1
Wednesday, August 1,2012	0:30	35	0
Wednesday, August 1,2012	0:30	36	1
Wednesday, August 1,2012	0:30	37	0
Wednesday, August 1,2012	0:30	38	0
Wednesday, August 1,2012	0:30	39	0
Wednesday, August 1,2012	0:30	40	0
Wednesday, August 1,2012	1:30	33	1
Wednesday, August 1,2012	1:30	34	0
Wednesday, August 1,2012	1:30	35	2

Wednesday, August 1,2012		1:30	36	0
Wednesday, August 1,2012		1:30	37	0
Wednesday, August 1,2012		1:30	38	0
Wednesday, August 1,2012		1:30	39	0
Wednesday, August 1,2012		1:30	40	0
Wednesday, August 1,2012		2:30	33	0
Wednesday, August 1,2012		2:30	34	1
Wednesday, August 1,2012		2:30	35	1

Table B-12 Stop Event Data

Attribute Name	Example 1	Example 2
service_date	15SEP2011:00:00:00	15SEP2011:00:00:00
vehicle_number	2927	2927
leave_time	21198	21444
train	1267	1267
badge	1149	1149
route_number	4	4
direction	0	0
service_key	W	W
trip_number	1020	1020
stop_time	21181	21417
arrive_time	21188	21402
dwell	0	15
location_id	1486	13297
door	0	2
lift	0	0
ons	0	0
offs	0	1
estimated_load	3	3
maximum_speed	33	24
train_mileage	34.6	35.7
pattern_distance	26054.15129	32002.65884
location_distance	176.6057792	44.76730189
x_coordinate	7666298.96	7672220.638
y_coordinate	677239.7283	676713.8646
data_source	2	2
schedule_status	2	2

Table B-13 Sample of Phase Timing Data

Attribute Name	Example 1	Example 2
fromtopofcycle	1	62
greens	32	34
yellow	2	0
peds	32	34
ped_calls	42	34
veh_calls	3	132
status	2	2
online	1	1
intersectionid	4107	4109
overlays	0	0
plan_num	4	4
timestamp	9/15/2011 12:00:00 AM	9/15/2011 12:00:00 AM

Table B-14 Sample of Stored Matched Travel Times

Attribute Name	Example 1	Example 2
from_ bluetooth_id	10001	10002
to_bluetooth_id	10002	10001
timestamp	10/20/11 13:25	11/06/2011 20:19:03
traveltime	427	133

APPENDIX C

NON-FUNCTIONING DETECTORS

Table C-1 Detectors Producing Zero Volume on November 5, 2012

Detectorid	Location	Bound	Lane
31	SW 2nd Ave & Alder St	EB	CENTER
32	SW 2nd Ave & Alder St	EB	RIGHT
36	SW 1st Ave & Main St	W	RIGHT
37	SW 1st Ave & Main St	W	LEFT
45	SW Naito Parkway & Harrison St	N	RIGHT
47	SW Naito Parkway & Harrison St	S	LEFT
49	SW Naito Parkway & Morrison St	E	CENTER
50	SW Naito Parkway & Morrison St	E	CENTER
61	SW Broadway Ave & Jackson St	S	ON RAMP
62	SW Broadway Ave & Jackson St	S	LEFT
63	SW Broadway Ave & Jackson St	S	CENTER
64	SW Broadway Ave & Jackson St	S	RIGHT
202	NE Martin Luther King Jr Blvd & Columbia Blvd	S	LEFT
220	NE Broadway St & 30th Ave	E	LEFT
229	NE Sandy Blvd & 16th Ave	E	LEFT
230	NE Sandy Blvd & 16th Ave	E	RIGHT
232	NE Sandy Blvd & 18th Ave	W	RIGHT
237	NE Martin Luther King Jr Blvd & Columbia Blvd	S	RIGHT
238	NE Martin Luther King Jr Blvd & Columbia Blvd	S	LEFT
240	NE Martin Luther King Jr Blvd & Columbia Blvd	S	RIGHT
241	NE Sandy Blvd & 70th Ave	E	RIGHT
242	NE Sandy Blvd & 70th Ave	E	LEFT
243	NE Sandy Blvd & 70th Ave	W	LEFT
244	NE Sandy Blvd & 70th Ave	W	RIGHT
245	NE Portland Hwy & 60th Ave	W	RIGHT
246	NE Portland Hwy & 60th Ave	W	LEFT
300	SW Beaverton Hillsdale Hwy & 25th Ave	E	BIKE
301	SW Beaverton Hillsdale Hwy & 25th Ave	E	RIGHT
302	SW Beaverton Hillsdale Hwy & 25th Ave	E	LEFT
303	SW Beaverton Hillsdale Hwy & 25th Ave	W	LEFT
304	SW Beaverton Hillsdale Hwy & 25th Ave	W	RIGHT
305	SW Beaverton Hillsdale Hwy & 25th Ave	W	BIKE
413	NE 82nd Ave & Fremont St	N	RIGHT
414	NE 82nd Ave & Fremont St	N	LEFT
415	NE 82nd Ave & Fremont St	S	RIGHT
416	NE 82nd Ave & Fremont St	S	LEFT
506	W Burnside St & MacLeay Blvd	W	RIGHT
507	W Burnside St & MacLeay Blvd	W	LEFT
602	N Interstate Ave & Argyle St	S	LEFT
603	N Interstate Ave & Argyle St	S	RIGHT
611	NE Weidler St & Victoria Ave	E	LEFT
612	NE Weidler St & Victoria Ave	E	CENTER 1
613	NE Weidler St & Victoria Ave	E	CENTER 2
614	NE Weidler St & Victoria Ave	E	RIGHT
615	NE Weidler St & Victoria Ave	N	RIGHT

616	NE Weidler St & Victoria Ave	N	LEFT
617	NE Weidler St & Victoria Ave	N	CENTER
661	N Broadway St & Larrabee Ave	E	RIGHT
662	N Broadway St & Larrabee Ave	E	LEFT
663	N Broadway St & Larrabee Ave	W	LEFT
664	N Broadway St & Larrabee Ave	W	RIGHT
665	N Broadway St & Larrabee Ave	E	RIGHT
666	N Broadway St & Larrabee Ave	E	CENTER
667	N Broadway St & Larrabee Ave	E	LEFT
668	N Weidler St & Flint Ave	S	RIGHT
669	N Weidler St & Flint Ave	S	CENTER
670	N Weidler St & Flint Ave	S	LEFT
685	N Weidler St & Williams Ave	N	RIGHT
686	N Weidler St & Williams Ave	N	CENTER
687	N Weidler St & Williams Ave	N	LEFT
693	SW Barbur Blvd & Bertha Blvd	NE	RIGHT
709	SW Barbur Blvd & Hooker St	N	RIGHT
710	SW Barbur Blvd & Hooker St	N	LEFT
711	SW Barbur Blvd & Hooker St	S	RIGHT
722	NE Killingsworth at Columbia Pkwy NB	E	LEFT
725	NE Killingsworth at Columbia Pkwy NB	W	RIGHT
726	NE Killingsworth at Columbia Pkwy NB	W	CENTER
727	NE Killingsworth at Columbia Pkwy NB	W	LEFT
1013	SE Stark & 174th	E	RIGHT
1022	NW Burnside & 201st	E	LEFT

OTREC
**OREGON TRANSPORTATION RESEARCH
AND EDUCATION CONSORTIUM**

P.O. Box 751
Portland, OR 97207

OTREC is dedicated to
stimulating and conducting
collaborative multi-disciplinary
research on multi-modal surface
transportation issues, educating
a diverse array of current
practitioners and future leaders
in the transportation field, and
encouraging implementation of
relevant research results.