CWGP: CATT Works Georeferencing Protocol

Version 1.1 – 2021 March 01

Versioning Notes:
The version number is formatted as [Major]. [Minor], with minor increments (e.g., 1.0 to 1.1) reflecting minor changes and major increments (e.g., 2.3 to 3.0) reflecting significant changes.

Version 1.1 includes general proofreading and minor edits to clarify the motivation and provide more precise explanations for certain data items (e.g., offset in *CWPoint*).

1. Introduction and Background

In an effort to streamline the Eastern Transportation Coalition (formerly I-95 Corridor Coalition) Vehicle Probe Project (VPP) probe data validation program and to build toward more interoperable geo-spatial data systems, this document defines a simple georeferencing protocol to communicate roadway locations in a basemap-agnostic (or even map-free) manner. The validation process uses field measurements to regularly evaluate probe-based VPP travel time data from multiple vendors. The process is currently hampered by location referencing inefficiencies, the most notable being that the validation road segments rarely align with the segmentation that probe vendors use for reporting traffic data, which are typically pre-coded Traffic Message Channel (TMC) segments. Even if attempts are made to place validation sensors at TMC segment endpoints (a potentially difficult task given constraints imposed by sensor mounting considerations) differences in TMC basemap implementations across vendors lead to varying interpretations of segment codes. Additionally, most data vendors have introduced proprietary, high-granularity segmentation schemes, introducing even greater variance in the way roadway locations are communicated. When coupled with licensing issues (e.g., data vendors are often unable to share the details of their basemap due to licensing restrictions from a map provider) and periodic map updates, the process of spatially aligning each vendor's traffic data with the validation segment as collected in the field has become particularly arduous, time inefficient, and consumes resources that could be better spent evaluating vendor data quality.

Rather than describe road segments based on a map representation with statically pre-coded nodes or segments on a versioned basemap, the approach herein is based on geodetic coordinates (a standardized spatial coordinate system, not a map). This approach is not tied to any specific basemap or map version. Instead, it describes locations, called *CWPoint* and *CWSegment* locations, primarily via information about the endpoint geodetic coordinates and headings. Furthermore, it allows the sender to add additional information (e.g., additional coordinate geometries, road name, functional class) to help disambiguate complex locations when necessary. By providing a language to describe locations in a fluid and dynamic manner without using pre-defined identifiers that are proprietary to a static basemap, the data validation team can request traffic data from vendors along a roadway with precision, without having to map sensor locations and custom segment geometries to each vendor's customized, and many times proprietary, electronic basemap.

The Eastern Transportation Coalition (ETC) is adopting this georeferencing standard in the next cycle of the VPP procurement, now renamed the Traffic Data Marketplace (TDM), requiring data vendors to support this language, and thus simplify, and ease the burden of transmitting data across disparate basemap systems which will allow the validation program to regain efficiency. This protocol is a useful langue for exchanging location information, and is open for any organization to adopt beyond just the TDM. The purpose is not to replace static map encoding, but to ensure that data from any such static encoding basemap system can easily and seamlessly be communicated to other mapping systems independent of any map representation.

The CWGP protocol shares similarities with the OpenLR protocol (https://www.openlr-association.com/openlr-association.html) which is an open source georeferencing protocol supported by industry, particularly in European markets, for dynamic location referencing. The CWGP differs is that it is lighter to implement, constrained in use cases, simpler, and in human-readable format.

2.0 CATT Works Georeferencing Protocol (CWGP) Concepts

At a high level, CWGP seeks to provide a basic mechanism for communicating roadway locations, specifically segments and points, in a simple and basemap-agnostic (or even basemap free) manner. Through this protocol, researchers, transportation agencies, and industry can talk and exchange information about the same roadway locations without the overhead of managing a mutually versioned basemap. The key goals are as follows:

- Describe locations primarily via geodetic coordinates and headings with minimal attributes
- Provide optional fields to assist when location coordinates alone are insufficient to disambiguate complex locations or roadway geometries. Additional information is encouraged, but not required.
- Describe locations with sufficient detail and in human readable, compact format (compression and low bandwidth considerations are not the primary motivation)
- There are no requirements on either the sender or receiver's basemaps (or that they even possess a static basemap) other than the requirement that any such maps are accurately geo-referenced to the earth. Basemaps may be routable or non-routable.
    - *Sender*: should be able to use known geodetic coordinate locations or basic GIS tools to encode such locations to request or communicate data, along with (optional) attribute fields
    - *Receiver*: free to associate the locations with its own internal basemap however desired (e.g., GIS conflation, map matching algorithms, etc.) In the high majority of cases simply latitude/longitude/heading information will be sufficient to uniquely identify roadway segments and points.

The purpose of this protocol is not to communicate roadway geometry to a level of fidelity to transmit/recreate a complete geometry of a roadway network. Rather, the perspective for use is that of a roadway/highway engineer or planner that requires a concise and compact method to clearly communication location of road segments and roadway point locations without having the overhead associated with map representations. The protocol is detailed enough to allow road crews to precisely locate points and segments in the field with simple GPS equipment. It is also detailed enough that any office personnel could identify the location or segments precisely on an electronic basemap with minimal effort, and without ambiguity.

3.0 CWGP Data Elements

This section defines two basic data elements at the conceptual level, *CWPoint* and *CWSegment*. It is important to keep in mind that this CWGP protocol was developed with the initial and primary use case of supporting validation exercises currently in process with the Eastern Transportation Coalition (ETC), and thus the authors' goal was to create a protocol with as much pliability as possible to serve that particular role based on experience of running a multi-vendor validation program over the past few years. It is not intended as a stand-alone map representation, but rather as a shorthand notation to precisely communicate segment and point locations between and between two entities operating with different electronic basemaps.

The two data items defined are the *CWPoint* and *CWSegment*. *CWSegment* is defined first, as this is the used most often within the ETC Vehicle Probe Project (VPP) validation process. *CWPoint* is also defined in anticipation of validation exercises for volume data in the next iteration of the VPP to move forward in 2021. Note that the "CW" prefix refers to "Catt Works" in Catt Works Georeferencing Protocol, and this naming convention – typically presented in italics - is used to differentiate data elements from the use of "point" and "segment" in the common vernacular.

Lastly, this generalized attribute table is presented conceptually to convey the information intended for *CWPoint* and *CWsegment*. These conceptual definitions reflect the data order and technical-structure of the implementation method. The following sections will convey and give examples for this protocol implemented with such data file protocols such as Comma Separated Variables (CSV) (and extensible markup language (XML) in the future). The freedom and flexibility to define the attributes may vary with implementation type.

3.1 CWSegment

CWSegment is used to describe linear sections along a roadway and communicate segment-based traffic data (e.g., travel time, space-mean speed). Minimally, a *CWSegment* must include a beginning and ending point and heading vectors to communicate location and directionality at both start and end locations (see Figure 1 below). When road systems are more complex, such as in more highly urbanized areas, additional attributes may be needed when the spatial location precision and heading at the endpoints are insufficient alone to uniquely identify the intended segment.

Figure 1: *CWSegment* geometry defined between points A and B.

Table 1: *CWSegment* attribute descriptions.

ATTRIBUTE	DESCRIPTION
Required	
Unique Identifier	A unique textual or numeric identifier for the segment.
Geo Coordinate A, B	A geo coordinate value (comprised of latitude and longitude) that defines the beginning position (A), which is the start point (upstream starting point) of the roadway segment, and the ending position (B), which is the end point (downstream ending point) of the roadway segment. Each geo coordinate uniquely identifies a point on the centerline of the roadway (or centerline of a group of lanes in the case of a divided road). Longitude and latitude are measured in decimal degrees.
Heading A, B	The heading at point A communicates the direction a vehicle is pointing (or traveling) when it departs from Position A – the start point of the segment. Likewise, the heading at point B communicates the direction a vehicle is pointing (or traveling) when it arrives at Position B – the end point of the segment. Heading is measured in degrees ranging from 0-360, where 0 degrees is North, 90 degrees is East, 180 degrees is South, and 270 degrees is West. For example, if the heading is 45 degrees it means that a vehicle is pointing (or traveling) North-East direction.
Optional	
Geo Coordinate(s) [Interim]	If geo coordinate endpoints A and B are not adequate to uniquely describe the path from A to B, one or more interim geo coordinate points can be provided. Each geo coordinate value must contain a set of geo-coordinate (longitude, latitude) values
Heading(s) [Interim]	Each interim geo coordinate may have corresponding heading value to communicate the instantaneous 0-360 degree travel direction at the coordinate location.
Road Name	Road name or route number used to describe the roadway location. This field can be used to provide either a road name (e.g., Main Street) or route number (e.g. US-23). In cases where the provided latitude and longitude values are ambiguous, perhaps between a highway and parallel frontage road, this attribute helps clarify the road to which the point belongs.
Road Classification	Classification scheme to communicate the relative importance of the roadway. Road classes are ranked in order of importance, 1. N (e.g., 1 is more important than 2). See discussion in appendix. There are many road classifications systems, all of which provide a general ranking starting at either 0 or 1 for the highest class roadways (such as interstates) and preceding to 7 to 9 for local roads. Unless otherwise defined, road class is defined as normalized 1 to 7 CWGP ranking, which is consistent with HPMS and OpenStreetMap as shown in the appendix.

Road Form	Classification scheme to communicate the road's physical form. The main purpose of this attribute is to be able to distinguish mainline facilities from ramps. 1 = Divided mainline 2 = Undivided mainline 3 = Ramp 4 = Non-mainline (other than ramp) 5 = Other
Segment Length	Length of the segment in feet. Note that this length accounts for the true road geometry, and is not the straight-line distance between endpoints A and B.
Direction	Cardinal (N, S, E, W) or Ordinal (NE, SE, SW, NW) direction associated with the *CWSegment* location. Ex. Most even number Interstate routes run E-W so a section of I-70 may have cardinal direction of E or W depending on the direction of travel.
Previous, Next segment(s)	This field provides a way to list the *CWSegment* that immediately precedes and/or follows the current segment in the same direction of travel. Having knowledge of nearby segments can be useful from multiple perspectives. First, it can help the receiver accurately decode the segment location; if the current location is ambiguous (e.g., multiple segments have similar endpoints and headings), knowing the previous segment can help provide clarity. Additionally, it makes it easier to link nearby segments to facilitate corridor-level analyses.
User Defined Attribute and Value	Recognizing that additional fields may be required in a way that cannot be anticipated at this time, users will be able to add an attribute and value as needed.

3.2 CWPoint

Anticipating the need to validate volume estimates for the upcoming re-issue of the VPP and subsequent validation program, a single location along a roadway will be required to be communicated. A *CWPoint* data location uniquely defines a roadway point as defined below. This could be the location of a sign, a sensor, an entry point (break in access or driveway). The interpretation of a *CWPoint* will be application specific. The CWGP definition for required and optional attributes are given below to explicitly define and disambiguate the location reference relative to a roadway as clearly as possible. The *CWPoint* is defined as a point along a roadway, and not a general point in space. Many of the fields parallel that of *CWSegment.*

Figure 2: *CWPoint* geometry

Table 2: *CWPoint* attribute definitions.

ATTRIBUTE	DESCRIPTION
Required	
Unique Identifier	
Geo Coordinate A	A (longitude, latitude) value that uniquely identifies a point on the centerline of the roadway (or centerline of a group of lanes in the case of a divided road). Longitude and latitude are measured in decimal degrees.
Heading A	See description under *CWSegment*. Because the *CWPoint* data structure was designed for applications that have an implied heading (e.g. traffic volumes at a point along the roadway) this attribute is mandatory.
Optional	
Road Name	See description under *CWSegment*
Road Classification	See description under *CWSegment*
Road Form	See description under *CWSegment*
Direction	See description under *CWSegment.* For example, if the point location represents a traffic sensor, the direction attribute can be used to indicate the direction(s) for which volumes are collected.
Offset	Lateral offset from the road centerline for precise location in feet. This measurement, which may be used to locate an asset (e.g., sign, sensor, signal) is made relative to the roadway centerline, with positive offsets to the right and negative offsets to the left when using the Heading value to specify the “forward” direction of travel.
User Defined Attribute and Value	See description under *CWSegment*

4.0 CWGP Specification

The initial version of the CWGP data specification consists of a Comma-separated Value (CSV) data format. The CSV specification is useful for transferring simple tabular attribute data in lightweight human-readable text files. Note, since CSV is a ‘flat file’ format, it can be challenging to express certain logic. For example, an arbitrary number of intermediate coordinates along a segment cannot be encoded easily in a matrix format. Also note that CSV is ill-suited for API applications.

It is envisioned that future versions of CWGP will incorporate additional data formats – likely Extensible Markup Language (XML) and/or JavaScript Object Notation (JSON). These formats – while more complicated – will provide a robust way to express *CWPoint* and *CWSegment* locations, enforce logic rules that define data elements, and are well-suited for exchanging location information via an API.

CSV Specification

Table 3 lists the CSV column names associated with each previously-described *CWSegment* attribute, while Table 4 does the same thing for *CWPoint* attributes. Additional comments are provided to clarify implementation details when necessary.

Table 3: *CWSegment* attributes for CSV specification

COLUMN NAME	ATTRIBUTE	COMMENTS
Required		
Id	Unique Identifier	
LonA	Geo Coordinate A (longitude)	
LatA	Geo Coordinate A (latitude)	
HeadA	Heading A	
LonB	Geo Coordinate A (longitude)	
LatB	Geo Coordinate A (latitude)	
HeadB	Heading A	
Optional		
LonInterim	Interim Geo Coordinates (longitude)	Space-separated list of intermediate longitude values. Note that commas are not used to separate values because CSV files are already comma-delimited. E.g., lon1 lon2 … lonN
LatInterim	Interim Geo Coordinates (latitude)	Space-separated list of intermediate latitude values E.g., lat1 lat2 … lonN
HeadInterim	Interim Headings	Space-separated list of intermediate heading values E.g., head1 head2 … headN
RoadName	Road Name	
RoadClass	Road Classification	
RoadForm	Road Form	
LengthFeet	SegmentLength	
Direction	Direction	
PrevSegId	Previous Segment Identifier	
NextSegId	Next Segment Identifier	
[UserDefined]	User Defined	User may define the name of the column to communicate additional attributes.

Table 4: *CWPoint* attributes for CSV specification

COLUMN NAME	ATTRIBUTE	COMMENTS
Required		
Id	Unique Identifier	
LonA	Geo Coordinate A (longitude)	
LatA	Geo Coordinate A (latitude)	
HeadA	Heading A	
Optional		
RoadName	Road Name	
RoadClass	Road Classification	
RoadForm	Road Form	
Direction	Direction	
Offset	Offset	
[UserDefined]	User Defined	User may define the name of the column to communicate additional attributes.

5.0 Encoding and Decoding CWGP locations

Encoding CWGP locations from the sender's map

Encoding CWGP locations (i.e., transferring location information from the sender's basemap into a CWGP location reference in CSV format) is intended to be as straightforward as possible. It intentionally does not require the sender to have access to a routable basemap to be able to generate the location references, recognizing that many potential users, such as state transportation agencies, use GIS basemaps that are not suitable for routing purposes.

Instead, CWGP simply requires the sender to prepare a CSV file using the basic coordinate or attribute information previously described. Any modern GIS software (e.g., QGIS, ArcGIS) and many geospatial software libraries (e.g., GeoPandas, Shapely) are capable of extracting geocoordinates from GIS basemap layers, which can be used to obtain the require coordinate and heading information. Additionally, it would be conceptually straightforward to develop an open-source software tool to facilitate the conversion process in an automated (or semi-automated) manner.

Decoding CWGP locations on the receiver's map

Decoding CWGP locations (i.e., associating the CWGP location references with locations on the receiver's basemap) is not prescribed as part of the specification. One reason for this is that the most efficient approach for each basemap is dependent on the structure of each basemap, supporting software tools, and how much information is provided by the sender. Reasonable approaches may include various combinations of the following strategies, beginning with simple approaches and moving to more complex ones when it is warranted for disambiguation:

- *Simple spatial search:* Finding nearest Point or Segment location based on spatial proximity and headings

- *Utilizing optional textual data:* Information about road names / route numbers can be integrated to make simple spatial searches more robust.

- *Utilizing optional topological data:* Information about previous and next CWSegments (optional attributes) may be helpful for disambiguating complex locations. If the previous or next segments were successfully located, it may give insight into the current location.

- *Map Matching / Routing:* When only endpoints (or endpoints plus a handful of intermediate "via points") are provided, map matching algorithms -- or existing software that implements them -- can be employed to find the most plausible set of intermediate coordinates, and thus form a more granular LineString geometry that can be used for spatial searches or in conflation algorithms. Examples of these of map matching and routing services include Open Source Routing Machine (OSRM) and pgRouting.

- *Conflation techniques*: Advanced conflation algorithms typically require full segment geometries, and thus require a detailed polyline representation of each *CWSegment* – a task that likely first requires map matching or routing. Afterwards, the segments can be assigned to a target basemap via various network conflation approaches – e.g., curve similarity matching, etc.

6.0 Examples

The following two examples illustrate common uses cases for CWGP. The first uses *CWSegments* to define custom road segments for travel time validation purposes, while the second uses *CWPoints* to describe individual point locations along a roadway for validating traffic volumes.

CWSegment for VPP travel time validation

VPP probe speed and travel time data are regularly validated by spot-checking data quality at select roadway locations. Ground truth travel time data is collected along custom road segments, which are defined such that the endpoints align with the locations of traffic sensors in the field. *CWSegments* provide a simple way to communicate these validation locations to vendors, who must subsequently provide their probe-based speed/travel time data for the same locations.

For example, Figure 3 shows the locations of wireless re-identification traffic monitoring (WRTM) sensors placed along two corridors near Boston, MA during a recent VPP travel time validation effort. Custom roadway segments are defined between adjacent WRTM sensors in the direction of traffic flow, and used as the spatial basis for comparing VPP speed/travel time data.

Figure 4 shows three such sample segments, which are described via *CWSegment* specification in CSV format in Table 5. For each segment, "A" represents the segment start point, "B" represents the endpoint, and the arrows communicate the instantaneous travel direction at each endpoint. Note that all mandatory attributes (i.e., Id, LonA, LatA, HeadA, LonB, LatB, HeadB) are provided, and a handful of others are included for clarity. In these cases, there is not much additional information needed beyond the endpoint locations and headings, but the road name/class/form can help disambiguate in more complex scenarios.

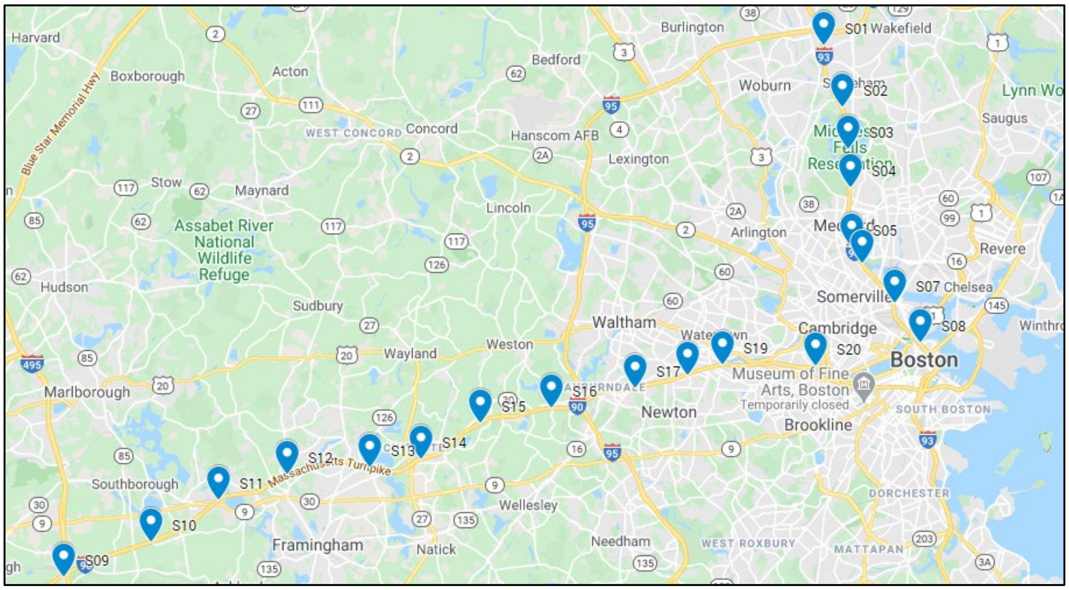

Figure 3. Traffic sensor locations from a recent VPP travel time validation effort.

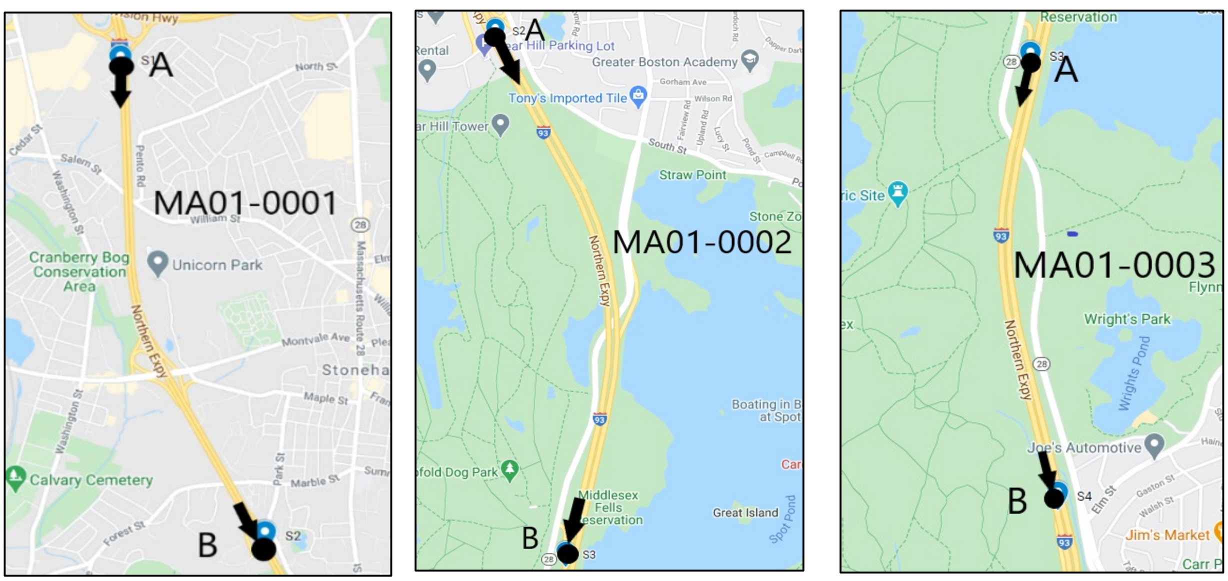

Figure 4. Validation segments are defined between adjacent sensors, and can be communicated via *CWSegments.* A is the start point, B is the endpoint, and the arrows indicate the travel heading.

Table 5. Sample validation segments from Figure 4 in *CWSegment* CSV format

Id	LonA	LatA	HeadA	LonB	LatB	HeadB	RoadName	RoadClass	RoadForm	LengthFeet
MA01-0001	-71.1183	42.4962	180	-71.1079	42.4696	159	I-93	1	1	10301

MA01-0002	-71.1079	42.4696	156	-71.1042	42.4512	193	I-93	1	1	7102
MA01-0003	-71.1042	42.4512	193	-71.1028	42.4348	167	I-93	1	1	6083

CWPoint for VPP volume validation

In anticipation of traffic volumes being included in the next iteration of VPP, the validation team envisions using CWPoint data elements to reference the roadway location (and specific traffic direction) associated with volume data. In this context, the validation team may obtain ground-truth traffic counts (e.g., via field measurements or video-based processing) at a known location, and request that VPP vendors provide a corresponding volume estimate for the same time period at the location described via a *CWPoint.*

As an example, Figure 5 shows a sample location where ground truth count data was collected along I-93 in the Northbound direction. Table 6 provides an example of a CSV-based *CWPoint* that communicates this location. Note that in additional to the mandatory attributes (Id, LonA, LatA, HeadA), additional attributes are included (RoadName, RoadClass, RoadForm, Direction) to help disambiguate the roadway from a nearby frontage road (Mystic Valley Pkwy) that carries traffic the same direction. However, by specifying RoadName = I-93, RoadClass = 1 (i.e., highest functional class), RoadForm = 1 (Divided mainline), it's clear that the location being referenced is indeed the freeway.

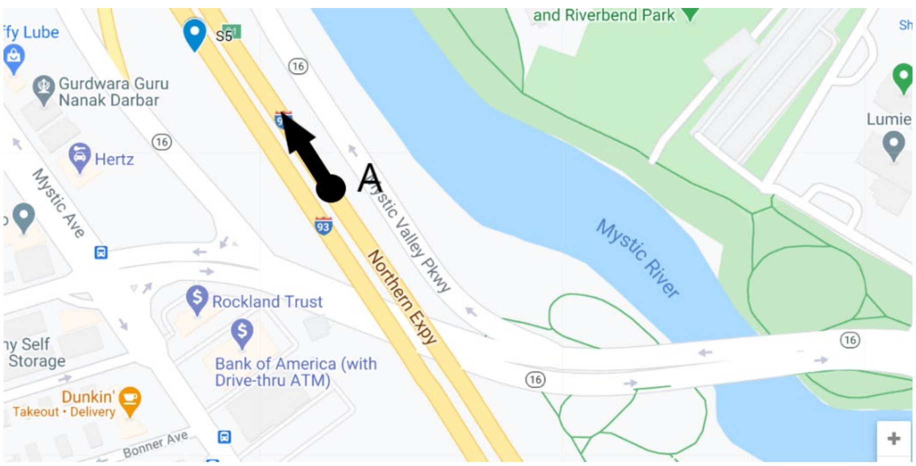

Figure 5. Sample traffic volume validation location.

Table 6: Sample traffic volume validation location from Figure 5 in *CWPoint* CSV format

Id	LonA	LatA	HeadA	RoadName	RoadClass	RoadForm	Direction
MA01-VOL-1	-71.09991	42.40679	327	I-93	1	1	N

Appendix A: Road Classification

The CWGP road classification system uses a simple ranking of functional road classes from most important to least important. Many map providers have different ways of classifying roadways (e.g., TMC maps use functional road class (FRC), HPMS uses Functional System (F_SYSTEM), OpenStreetMap uses a set of allowable 'highway' tag values). To make this classification attribute compatible with any desired scheme, it simply uses integer values to specify relative importance. Below are examples of common mappings between the CWGP integer classification values and other functional classification systems.

Example – HPMS Functional System

CWGP Road Classification	HPMS Functional System (F_SYSTEM)
1	Interstate
2	Principal Arterial – Other Freeways and Expressways
3	Principal Arterial – Other
4	Minor Arterial
5	Major Collector
6	Minor Collector
7	Local

Example – OpenStreetMap (link)

CWGP Road Classification	OpenStreetMap Highway Values
1	Motorway
2	Trunk
3	Primary
4	Secondary
5	Tertiary
6	Unclassified
7	Residential

Example – TomTom Functional Road Class (link)

CWGP Road Classification	TomTom FRC
1	FRC 0: Motorways; Freeways; Major Roads
2	FRC 1: Major Roads less important than Motorways

3	FRC 2: Other Major Roads
4	FRC 3: Secondary Roads
5	FRC 4: Local connecting roads
6	FRC 5: Local roads of high importance
7	FRC 6: local roads
8	FRC 7: Local roads of minor importance
9	FRC 8: Local roads