Materials for Review and Progress Update
November 4, 2014 Technical Advisory Committee (TAC) Meeting
Online Non-motorized Traffic Count Archive Project

Mt McLoughlin conference room, ODOT Mill Creek Building (555 13th St NE, Salem, Oregon)
November 4, 2014, 10am-12pm.

Dear TAC Members: This document and the included set of attached documents are submitted for your review and feedback prior to the November 4th meeting of the Archive TAC. Please review and comment where you like. Please submit comments by EOB October 31st so that we can summarize comments in advance of the meeting.

Packet Contents

ITEM	*DESCRIPTION*
1. Materials for Review and Progress Update	*This document* explaining TAC packet and outlining Quarterly progress
2. Revised Schema	Updated Schema Diagram based on feedback and testing
3. Database Tables and PostgreSQL script	Documents demonstrating the database architecture and script based on revised schema
4. Metadata User Forms	Screenshots of the current metadata user forms
5. QA/QC Plan Memo	Memo describing the QA/QC plan

Quarterly Progress Update:

In the current quarter, we have focused our efforts on further refinement of the archive database schema, development of the metadata (e.g. segment, detector and count location information) user interface, count upload interface, work toward developing a count upload preferred format, the development of a Quality Assurance and Quality Control (QA/QC) plan, continued development of the database backend pursuant to the above tasks, and development of a refined software development timeline. In further detail:

- **Refinement of the archive database schema:** Based on feedback from the TAC and testing of the initial schema with test count data, we have made improvements to the schema. The updated schema has undergone scenario testing, and we believe it to be versatile and robust. Updates made to the schema design include:
    - To accommodate counters that may be moved from one location to another, the count detector equipment information is no longer directly connected to the segment area and descriptor, but is instead connected to the traffic flow and count data via a separate association table called the "Flow-Detectors" table.
    - The table formerly known as the "Descriptor" has been split into two tables: "facility" and "flow." The "facility" is linked to the "segment area" and the traffic "flow" type. The traffic "flow" is linked to the "detector" which counts it through the association table, "Flow-Detectors," which also includes the latitude and longitude of the count location. The count records are connected to the Flow-Detectors table. This allows multiple detectors to count the same flow or multiple flows to be counted by one detector. This approach greatly increase the schema's flexibility and allows it to better accommodate mobile counters and cases where manual counts are used to check automated counts.
    - On the "Data" table, count Duration has been added to allow users to input either the start and end times or the start time and the duration

  Related attachment:
    - *Diagram of the revised schema*

- **Database backend development:** Much of the programming work this quarter has been focused on the development of the database schema. An initial schema was created and presented at the last TAC meeting. Based on feedback from the TAC (we particularly thank Jeremy Raw), the schema was revised. The new schema separates out the concepts of Path (Facility), Flow and Detector. The new schema supports permanent detectors, mobile detectors, replacement of permanent detectors and allows changes to meta-data (i.e. a change in surrounding land use) without losing the identity of the detector. Related attachments:
    - Tables, output (ERD)
    - PostgreSQL script

- **Development of the metadata user interface:** Metadata upload user forms include segment area, facility, flow, detector and count descriptor forms, all the upload of the data describing a particular count (but not the actual count itself). In coordination with the database schema, metadata forms are being developed and mapped to the schema. Related attachments:
    - *Screenshots of draft metadata user forms*

- **Count upload interface:** Count upload forms enable the user to upload count data, which includes the count, start time, and duration only. The forms are currently in development.

- **Work toward developing a count upload preferred format:** In order to facilitate the upload of count data, we have identified a preferred count data format that will most easily map data to the archive database. The use of this simplified data format (consisting of the count, start time, and duration) will be encouraged, though effort will be made to accommodate other standardized data formats. We have been in contact with major counter manufacturers, TRAFx and Ecocounter, to identify strategies to import their data outputs into the archive as directly as possible. EcoCounter has agreed to talk to us directly about the format and work with us to ensure that their output format is compatible with our format. TRAFx has not been willing to make any changes, but offers a simple data export format.

- **Outreach and engagement Engagement:** Communication with Jeremy Raw, (with Steven Jessberger cc'ed), and Mark Hallenbeck concerning the Traffic Monitoring Guide format has clarified key points. We have verified that our new schema can be translated into TMG format with minor changes to be made in the winter. In addition, we've been discussing potential future data collaborations with parties interested and involved in bicycle and pedestrian count efforts (including UCLA, which runs the Los Angeles Bike Count Data Clearinghouse, and the Minnesota Department of Transportation). These outreach efforts will continue and should lead to fruitful collaborations beyond Phase 1 of the project.

- **Development of a Quality Assurance and Quality Control (QA/QC) plan:** One focus of this quarter has been on developing a quality assurance and quality control plan that will guide our QA/QC approach as the archive is developed and launched. The plan has three levels of quality checks:
    - Database Constraints
    - Automated Checks
    - User Observation

  Data that fails to pass the database constraints is automatically rejected with comments to the user. The automated checks will flag data that is unusually high, low, or overly consistent. Current thresholds are set based on test data, but can be updated in future phases. For each flag, the user can either indicate that the data should be accepted as valid, should be accepted but flagged, or is invalid and should be accepted but hidden from public use. In addition, the user will be shown a graph of the data and can add comments on data. At this phase of work, we do not recommend a quality rating system, except that data with no flags or comments marked as "invalid" is "gold star" data. Related attachment:
    - *Memo outlining draft QA/QC plan.*

Overall Progress to Date:

As reflected in the quarterly update and packet contents, we are moving from archive planning to implementation, particularly in the software development. An updated estimate of completion by task is included in Table 1 below.

Table 1: Progress on Online Non-motorized Traffic Count Data Archive

Tasks	% Complete
1. Review the State of the Practice	100%
2. Establish Non-motorized Data Collection Methods and Formats	80%
3. Develop an Online Tool	
Schema Development	95%
Upload Forms	40%
QA/QC plan and implementation	25% (plan)
Output forms	0%
4. Stakeholder Engagement	25%
5. Research Deployment	0%

Goals Update:

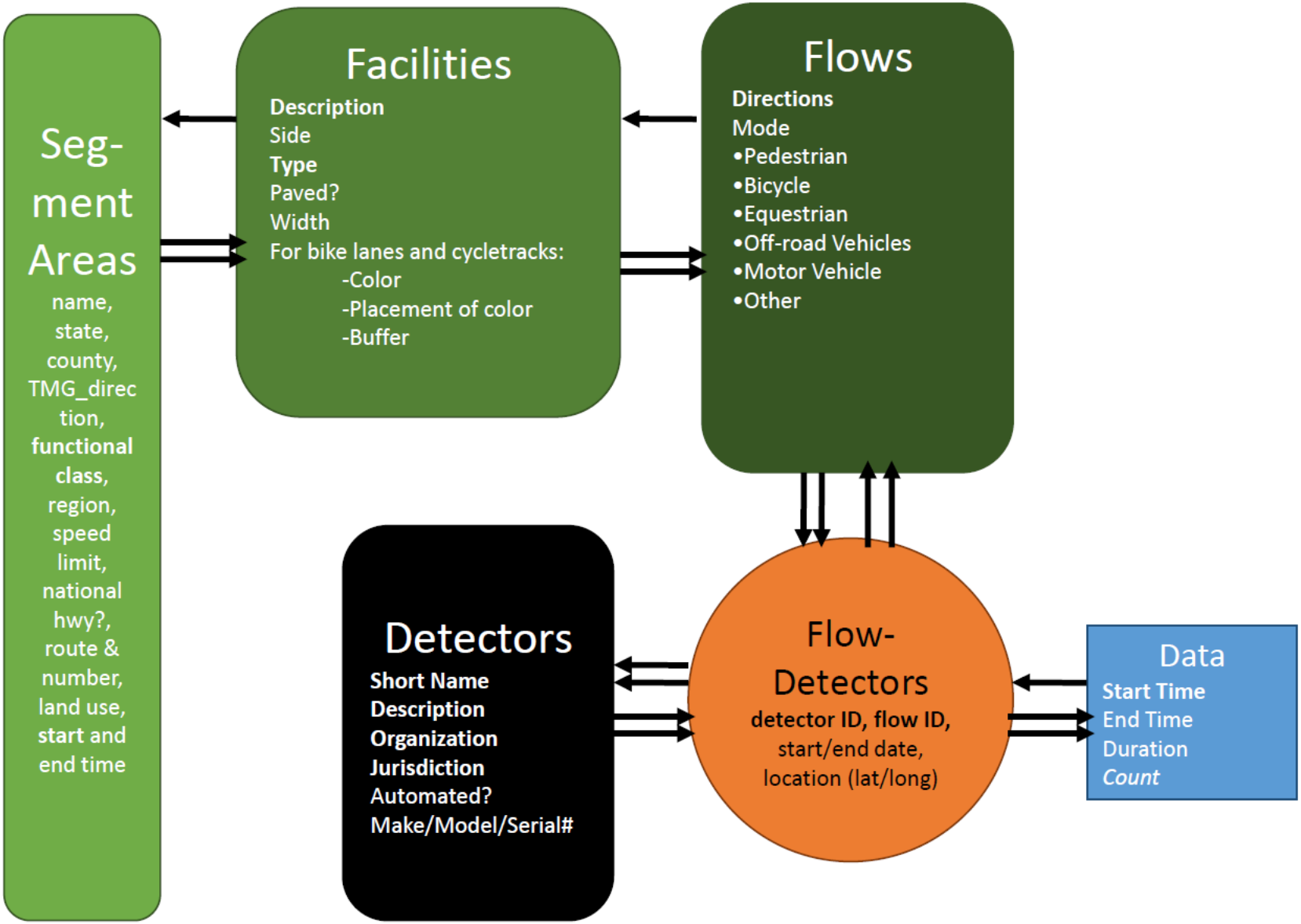
In order to focus our efforts on developing a database and user interface with a solid foundation for future data upload, storage, and access ("Phase 1 Priority" items from our Functional Requirements), we have been working on narrowing our focus to accomplish these goals, while exploring further ideas and implementation options for a project "Phase 2".

- **Phase 1 Priority:** Phase 1 Priority items remain as outlined in the functional requirements distributed to TAC members following the July TAC meeting (in an email dated August 7th, 2014). The focus is on creating a solid archive infrastructure that can import permanent counter data, perform basic QA/QC checks, and allow for data to be extracted by users. Manual (non-intersection) counts can be accommodated in Phase 1, but will need to be formatted to match the archive data format. Accommodating continuous counter data will be the priority in Phase 1. Weather data will not be incorporated in outputs in Phase 1.
- **Phase 2 Planning:** Due to our extremely limited budget, development of the items previously classified as "Possibly Phase 1, Likely Phase 2" remains unlikely for Phase 1. Items in this list (including uploading manual count data, complex data visualization (the pro version of the output tools), and more robust QA/QC checks) remain important goals that we will turn to once the base archive and database is in solid working order.

- **Project Expansion Wish List:** As the project progresses, we continue to identify more longer term archive needs, exciting opportunities, and continue to chart a path for the expansion of the archive beyond the initial Phase 1 development period.
    - Detailed project documentation to include a complete data dictionary describing all meta-data fields providing name, data type and sample data. In addition, database tables and constraints would be described and explained - including the motivation for the constraints. Finally all data cleaning and processing scripts would be made publicly available and documented.
    - Development of a 10 year vision for the Archive.
    - Exploration of approaches and funding sources to expand the reach and availability of archive data.
    - Exploration of methods of enhancing the data output and visualization. Building off of the Archive API that will be built into the backend of the data base, we will be encouraging users and partners to develop exciting data access and visualization tools. Among the ongoing considerations are the creation of a staff position to serve this role, crowdsourcing and/or competitions, etc.

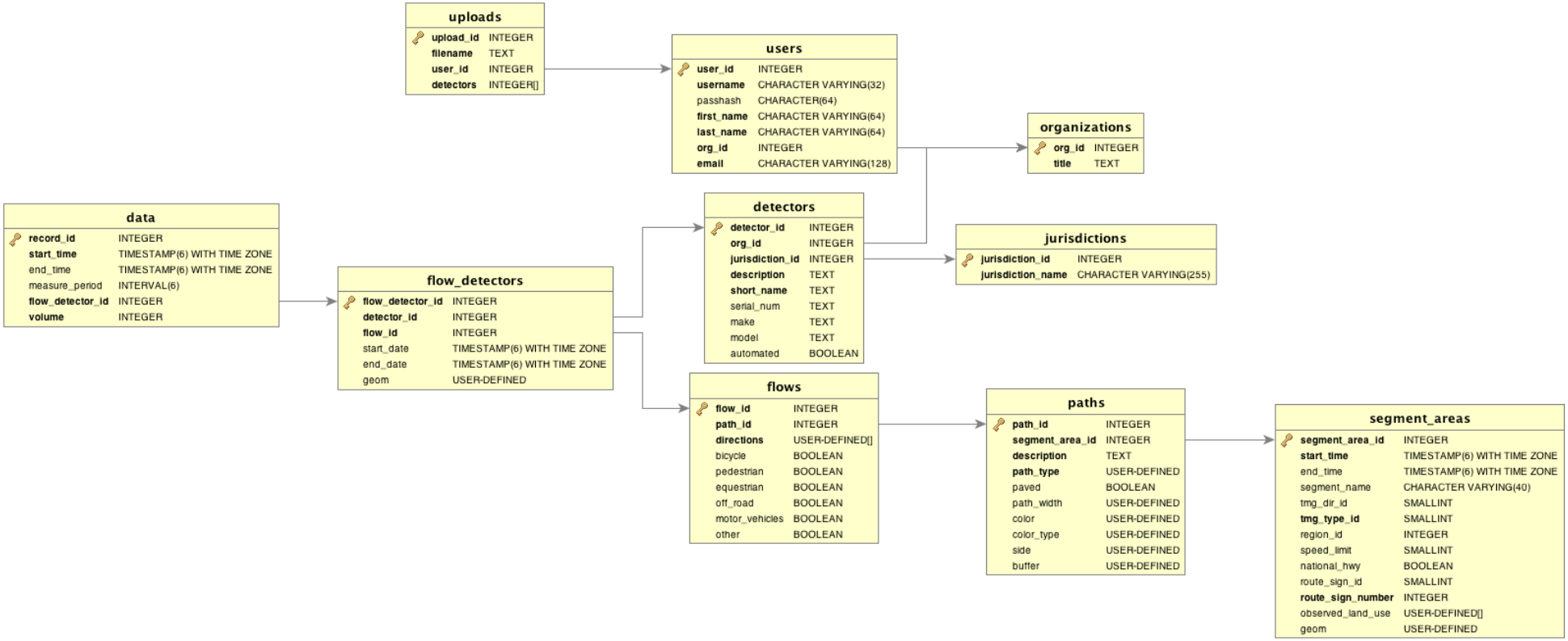
2. Revised Schema

Schema


Seg-ment Areas
name, state, county, TMG_direction, functional class, region, speed limit, national hwy?, route & number, land use, start and end time
Facilities
Description
Side
Type
Paved?
Width
For bike lanes and cycletracks:
-Color
-Placement of color
-Buffer
Flows
Directions
Mode
•Pedestrian
•Bicycle
•Equestrian
•Off-road Vehicles
•Motor Vehicle
•Other
Detectors
Short Name
Description
Organization
Jurisdiction
Automated?
Make/Model/Serial#
Flow-Detectors
detector ID, flow ID, start/end date, location (lat/long)
Data
Start Time
End Time
Duration
Count

3. Database Tables and PostgreSQL script

Database Tables/ERD



PGSQL for Bike/Ped Portal Schema - October 2014

```
CREATE ROLE bike_ped WITH PASSWORD 'b1kesnp3dz';
CREATE ROLE bike_ped_ro WITH PASSWORD 'Thisiz@password';

CREATE SCHEMA bike_ped;

ALTER SCHEMA bike_ped OWNER TO bike_ped;

GRANT USAGE ON SCHEMA bike_ped TO bike_ped_ro;

ALTER DEFAULT PRIVILEGES IN SCHEMA bike_ped GRANT ALL ON TABLES TO bike_ped;
ALTER DEFAULT PRIVILEGES IN SCHEMA bike_ped GRANT EXECUTE ON FUNCTIONS TO bike_ped;
ALTER DEFAULT PRIVILEGES IN SCHEMA bike_ped GRANT ALL ON SEQUENCES TO bike_ped;

ALTER DEFAULT PRIVILEGES IN SCHEMA bike_ped GRANT SELECT ON TABLES TO bike_ped_ro;
ALTER DEFAULT PRIVILEGES IN SCHEMA bike_ped GRANT EXECUTE ON FUNCTIONS TO bike_ped_ro;
ALTER DEFAULT PRIVILEGES IN SCHEMA bike_ped GRANT USAGE, SELECT ON SEQUENCES TO
bike_ped_ro;

SET SCHEMA 'bike_ped';

CREATE TABLE fips_state (
  statefp smallint PRIMARY KEY NOT NULL,
  state   char(2) UNIQUE NOT NULL,
  name    character varying ( 128 ) UNIQUE NOT NULL,
  gnisid  bigint UNIQUE NOT NULL,
  geom    public.geometry(Polygon,4326)
);

--SELECT public.AddGeometryColumn('fips_state','geom',4326,'POLYGON',2);

COMMENT ON TABLE fips_state         IS 'American National Standards Institute (ANSI) Codes for States,
the District of Columbia, Puerto Rico, and the Insular Areas of the United States';
COMMENT ON COLUMN fips_state.statefp IS 'FIPS State Code';
COMMENT ON COLUMN fips_state.state   IS 'Official United States Postal Service (USPS) Code';
COMMENT ON COLUMN fips_state.name    IS 'Name';
COMMENT ON COLUMN fips_state.gnisid  IS 'Geographic Names Information System Identifier
(GNISID)';
COMMENT ON COLUMN fips_state.geom    IS 'State boundaries';

CREATE TABLE fips_county (
  state      char(2) NOT NULL REFERENCES fips_state (state),
  statefp    smallint NOT NULL REFERENCES fips_state,
  countyfp   smallint NOT NULL,
```

```
  countyname character varying ( 128 ) NOT NULL,
  classfp    char(2) CHECK ( classfp IN ( 'H1', 'H4', 'H5', 'H6', 'C7' ) ) NOT NULL,
  geom       public.geometry(Polygon,4326),
  PRIMARY KEY (statefp, countyfp),
  UNIQUE (statefp, countyname)
);

--SELECT AddGeometryColumn('bike_ped'::text,'fips_county'::text,'geom'::text,4326,'POLYGON',2);

COMMENT ON TABLE fips_county            IS 'FIPS Codes for Counties and County Equivalent Entities';
COMMENT ON COLUMN fips_county.state      IS 'State Postal Code';
COMMENT ON COLUMN fips_county.statefp    IS 'State FIPS Code';
COMMENT ON COLUMN fips_county.countyfp   IS 'County FIPS Code';
COMMENT ON COLUMN fips_county.countyname IS 'County Name and Legal/Statistical Area
Description';
COMMENT ON COLUMN fips_county.classfp    IS 'FIPS Class Code';
COMMENT ON COLUMN fips_county.geom       IS 'County boundaries';

CREATE TABLE regions (
  region_id serial PRIMARY KEY NOT NULL,
  description character varying (255) UNIQUE NOT NULL,
  geom public.geometry(Polygon,4326)
);

--SELECT AddGeometryColumn('bike_ped','regions','geom',4326,'POLYGON',2);

CREATE TABLE tmg_dir (
  tmg_dir_id  smallint PRIMARY KEY NOT NULL,
  description character varying (255) UNIQUE NOT NULL
);

CREATE TABLE tmg_types (
  tmg_type_id smallint PRIMARY KEY NOT NULL,
  description character varying (255) UNIQUE NOT NULL
);

CREATE TABLE route_signings (
  route_sign_id smallint PRIMARY KEY NOT NULL,
  description character varying (255) UNIQUE NOT NULL
);

CREATE TYPE bp_land_use AS ENUM (
  'Residential','Office','Retail','School',
  'University','Industrial','Agricultural',
  'Park','Mixed use'
);
```

```
CREATE TABLE segment_areas (
  segment_area_id      serial PRIMARY KEY NOT NULL,
  start_time          timestamp with time zone NOT NULL DEFAULT now(),
  end_time            timestamp with time zone DEFAULT NULL,
  segment_name         character varying ( 40 ) DEFAULT NULL,
  tmg_dir_id          smallint REFERENCES tmg_dir DEFAULT NULL,
  tmg_type_id          smallint REFERENCES tmg_types NOT NULL,
--   statefp            smallint NOT NULL,  -- Spatial query
--   countyfp             smallint NOT NULL,  -- Spatial query
  region_id           integer REFERENCES regions DEFAULT NULL,
  speed_limit          smallint DEFAULT NULL CHECK (speed_limit BETWEEN 0 AND 100 ),
  national_hwy         boolean DEFAULT NULL,
  route_sign_id        smallint REFERENCES route_signings DEFAULT NULL,
  route_sign_number    integer NOT NULL DEFAULT 0,
  observed_land_use    bp_land_use[] DEFAULT NULL,
  geom                public.geometry(Polygon,4326),
  CONSTRAINT enforce_sign_number CHECK ( ( route_sign_id IS NOT NULL AND route_sign_number > 0 )
OR ( route_sign_id IS NULL AND route_sign_number = 0 ) ),
  CONSTRAINT enforce_dates CHECK ( (end_time IS NULL) OR (end_time > start_time) )
);

COMMENT ON TABLE segment_areas IS 'Areas containing a group of devices performing various types of
non-motorized vehicular traffic measurements.';
COMMENT ON COLUMN segment_areas.segment_area_id IS 'A unique ID for each segment area.';
COMMENT ON COLUMN segment_areas.segment_name      IS 'A short name describing this segment,
e.g. Hawthorne Bridge';
COMMENT ON COLUMN segment_areas.tmg_dir_id       IS 'The FHWA segment direction, if known.';
COMMENT ON COLUMN segment_areas.tmg_type_id      IS 'The FHWA functional class.';
COMMENT ON COLUMN segment_areas.region_id        IS 'The region to which the segment belongs.';
COMMENT ON COLUMN segment_areas.speed_limit      IS 'A posted speed limit, if available.';
COMMENT ON COLUMN segment_areas.national_hwy     IS 'Whether the defined segment is part of the
National Highway System.';
COMMENT ON COLUMN segment_areas.route_sign_id    IS 'FHWA posted route signing.';
COMMENT ON COLUMN segment_areas.route_sign_number IS 'US Bike Route number';

CREATE TYPE bp_direction AS ENUM (
  'N', 'S', 'E', 'W',
  'NW', 'SW', 'NE', 'SE'
);

CREATE TYPE bp_path_type AS ENUM (
  'roadway',
  'path/trail',
  'sidewalk',
  'cycle track',
```

```
  'bike lane',
  'general activity count'
);

CREATE TYPE bp_path_width AS ENUM (
  'miniscule',
  'tiny',
  'small',
  'moderate',
  'wide',
  'wider',
  'widest',
  'huge'
);

CREATE TYPE bp_color AS ENUM (
  'red',
  'blue',
  'green',
  'other',
  'none'
);

CREATE TYPE bp_color_type AS ENUM (
  'throughout',
  'solid at intersections',
  'dashed at intersections',
  'other'
);

CREATE TYPE bp_side AS ENUM (
  'left', 'right', 'center'
);

CREATE TYPE bp_buffer AS ENUM (
  'none','colored','painted stripe',
  'flexible posts','planters','parking',
  'concrete barrier','other physical barrier'
);

CREATE TABLE count_types (
  type_id char(1) PRIMARY KEY NOT NULL,
  description text NOT NULL
);

CREATE TABLE jurisdictions (
```

```
  jurisdiction_id   serial PRIMARY KEY NOT NULL,
  jurisdiction_name character varying(255) NOT NULL
);

CREATE TABLE paths (
  path_id        serial PRIMARY KEY NOT NULL,
  segment_area_id integer NOT NULL REFERENCES segment_areas,
  description     text NOT NULL,
  path_type       bp_path_type NOT NULL,
  paved          boolean DEFAULT NULL,
  path_width      bp_path_width DEFAULT NULL,
  color          bp_color DEFAULT NULL,
  color_type      bp_color_type DEFAULT NULL,
  side           bp_direction DEFAULT NULL,
  buffer         bp_buffer DEFAULT NULL,
  CONSTRAINT enforce_bike_meta CHECK ( ( path_type NOT IN ('bike lane', 'cycle track') AND
path_width IS NULL AND color IS NULL AND color_type IS NULL AND side IS NULL AND buffer IS NULL )
OR ( path_type = 'sidewalk' AND path_width IS NOT NULL ) OR ( path_type IN ('bike lane', 'cycle track') )
)
);

COMMENT ON TABLE paths IS 'Path components of a segment area.';
COMMENT ON COLUMN paths.segment_area_id IS 'A reference to the segment area in which this path
exists.';
COMMENT ON COLUMN paths.description IS 'A human-readable description of the path.';
COMMENT ON COLUMN paths.path_type IS 'The type of path.';
COMMENT ON COLUMN paths.paved IS 'A boolean indicating wether the path has a paved surface.';
COMMENT ON COLUMN paths.path_width IS 'The estimated width of the path.';
COMMENT ON COLUMN paths.color IS 'The color of the path surface.';
COMMENT ON COLUMN paths.color_type IS 'The pattern of the path surface color.';
COMMENT ON COLUMN paths.side IS 'The directional side of the segment on which the path occurs.';
COMMENT ON COLUMN paths.buffer IS 'The physical buffer on the path.';

CREATE TABLE flows (
  flow_id        serial PRIMARY KEY NOT NULL,
  path_id        integer NOT NULL REFERENCES paths,
  directions      bp_direction[] NOT NULL,
  bicycle        boolean DEFAULT NULL,
  pedestrian      boolean DEFAULT NULL,
  equestrian      boolean DEFAULT NULL,
  off_road       boolean DEFAULT NULL,
  motor_vehicles  boolean DEFAULT NULL,
  other          boolean DEFAULT NULL,
  UNIQUE ( path_id, directions, bicycle, pedestrian, equestrian, off_road, motor_vehicles, other )
);
```

```
COMMENT ON TABLE flows IS 'Describes a specific type of measured traffic. For example, eastbound pedestrian traffic.';
COMMENT ON COLUMN flows.flow_id IS 'A sequential numeric ID for each flow.';
COMMENT ON COLUMN flows.directions IS 'Direction of travel for the flow.';
COMMENT ON COLUMN flows.bicycle IS 'Boolean indicating whether the measured flow includes bicycles.';
COMMENT ON COLUMN flows.pedestrian IS 'Boolean indicating whether the measured flow includes pedestrians.';
COMMENT ON COLUMN flows.equestrian IS 'Boolean indicating whether the measured flow includes horses.';
COMMENT ON COLUMN flows.off_road IS 'Boolean indicating whether the flow includes motorized off-road traffic.';
COMMENT ON COLUMN flows.motor_vehicles IS 'Boolean indicating whether normal motorized traffic is included in the flow.';
COMMENT ON COLUMN flows.other IS 'Boolean indicating whether the measured flow includes other types of traffic.';


CREATE TABLE organizations (
  org_id serial PRIMARY KEY NOT NULL,
  title  text NOT NULL
);

CREATE TABLE detectors (
  detector_id     serial PRIMARY KEY NOT NULL,
  org_id         integer NOT NULL REFERENCES organizations,
  jurisdiction_id integer NOT NULL REFERENCES jurisdictions,
  description     text NOT NULL,
  short_name      text NOT NULL,
  serial_num      text DEFAULT NULL,
  make           text DEFAULT NULL, -- Create a lookup table of device
  model          text DEFAULT NULL, -- manufacturers and models?
  automated       boolean DEFAULT true,
  UNIQUE ( serial_num, make, model )
);

COMMENT ON TABLE detectors IS 'Metadata pertaining to specific physical detectors.';
COMMENT ON COLUMN detectors.detector_id IS 'A sequential numeric ID for each detector.';
COMMENT ON COLUMN detectors.description IS 'A meaningful description of the detector.';
COMMENT ON COLUMN detectors.short_name IS 'A short description of the detector as assigned by the owning organization.';
COMMENT ON COLUMN detectors.serial_num IS 'The serial number as assigned by the manufacturer.';
COMMENT ON COLUMN detectors.make IS 'The device manufacturer''s name.';
COMMENT ON COLUMN detectors.model IS 'The device model.';
COMMENT ON COLUMN detectors.automated IS 'Whether the device produces an automated count.';
```

```
--CREATE EXTENSION btree_gist;

CREATE TABLE flow_detectors (
  flow_detector_id serial PRIMARY KEY NOT NULL,
  detector_id      integer NOT NULL REFERENCES detectors,
  flow_id          integer NOT NULL REFERENCES flows,
  start_date       timestamp with time zone DEFAULT now(),
  end_date         timestamp with time zone DEFAULT NULL,
  geom             public.geometry(Point,4326),
  CONSTRAINT no_overlapping_flow_detectors EXCLUDE USING GIST (
    detector_id WITH =,
    flow_id WITH =,
    tstzrange(start_date, COALESCE(end_date, 'infinity'::timestamptz)) WITH &&
  )
);

COMMENT ON TABLE flow_detectors IS 'Describes where a specific detector was placed and when and
what it was set to measure.';

CREATE TABLE users (
  user_id    serial PRIMARY KEY NOT NULL,
  username   character varying (32) UNIQUE NOT NULL,
  passhash   char(64) DEFAULT NULL,
  first_name character varying (64) NOT NULL,
  last_name  character varying (64) NOT NULL,
  org_id     integer NOT NULL REFERENCES organizations,
  email      character varying (128) UNIQUE NOT NULL
);

CREATE TABLE uploads (
  upload_id serial PRIMARY KEY NOT NULL,
  filename  text NOT NULL,
  user_id   integer NOT NULL REFERENCES users,
  detectors integer[] NOT NULL,
  CONSTRAINT no_detectorless_data CHECK ( array_length(detectors, 1) > 0 )
);

CREATE TABLE data (
  record_id       serial PRIMARY KEY NOT NULL,
  start_time      timestamp with time zone NOT NULL,
  end_time        timestamp with time zone DEFAULT NULL,
  measure_period   interval DEFAULT NULL,
  flow_detector_id integer NOT NULL REFERENCES flow_detectors,
  volume          integer NOT NULL,
  CONSTRAINT no_negative_volumes CHECK ( volume >= 0 ),
  CONSTRAINT validate_periods CHECK (
```

```
    (end_time IS NULL AND measure_period IS NOT NULL) OR
    (measure_period IS NULL AND end_time IS NOT NULL) OR
    (start_time + measure_period = end_time)
  )
);

INSERT INTO bike_ped.tmg_types (tmg_type_id, description) VALUES
  (1, 'Interstate'),
  (2, 'Principal Arterial - Other Freeways and Expressways'),
  (3, 'Principal Arterial - Other'),
  (4, 'Minor Arterial'),
  (5, 'Major Collector'),
  (6, 'Minor Collector'),
  (7, 'Local'),
  (8, 'Trail or Shared Use Path'),
  (9, 'General Activity Count');

INSERT INTO bike_ped.tmg_dir (tmg_dir_id, description) VALUES
  (0, 'East-West or Southeast-Northwest combined (volume stations only)'),
  (1, 'North'),
  (2, 'Northeast'),
  (3, 'East'),
  (4, 'Southeast'),
  (5, 'South'),
  (6, 'Southwest'),
  (7, 'West'),
  (8, 'Northwest'),
  (9, 'North-South or Northeast-Southwest combined (volume stations only)');

INSERT INTO bike_ped.route_signings (route_sign_id, description) VALUES
  (1, 'Not signed'),
  (2, 'Interstate'),
  (3, 'U.S.'),
  (4, 'State'),
  (5, 'Off-Interstate Business Marker'),
  (6, 'County'),
  (7, 'Township'),
  (8, 'Municipal'),
  (9, 'Parkway Marker or Forest Route Marker'),
  (10, 'None of the above');
```

4. Metadata User Forms

Segment Area form:

General Information about the Count Site

Note: The first time you enter a new count site, you'll be asked some detailed information about the site. This information will be saved so that in the future you will only need to select and confirm the count site before entering the count data.

Please answer the following questions as completely and accurately as possible to help us properly use the data. You can return later to add additional information.

Segment Name

Hawthorne Bridge

This is a relatively short name that you use to refer to the area.

Segment Type

Local

For the purposes of adding this data to the FHWA Travel Monitoring Analysis System, we would like to know the type of roadway. If you do not know, please leave it blank.

Region

Portland

The municipal or administrative region containing the segment.

Speed Limit

35

The posted speed limit along the segment.

Part of National Highway System

Posted Route Sign

The type of signage along the route.

Route Sign Number

0

If count is taken on a facility with a posted route number indicate that number. Otherwise leave blank.

Observed Land Use

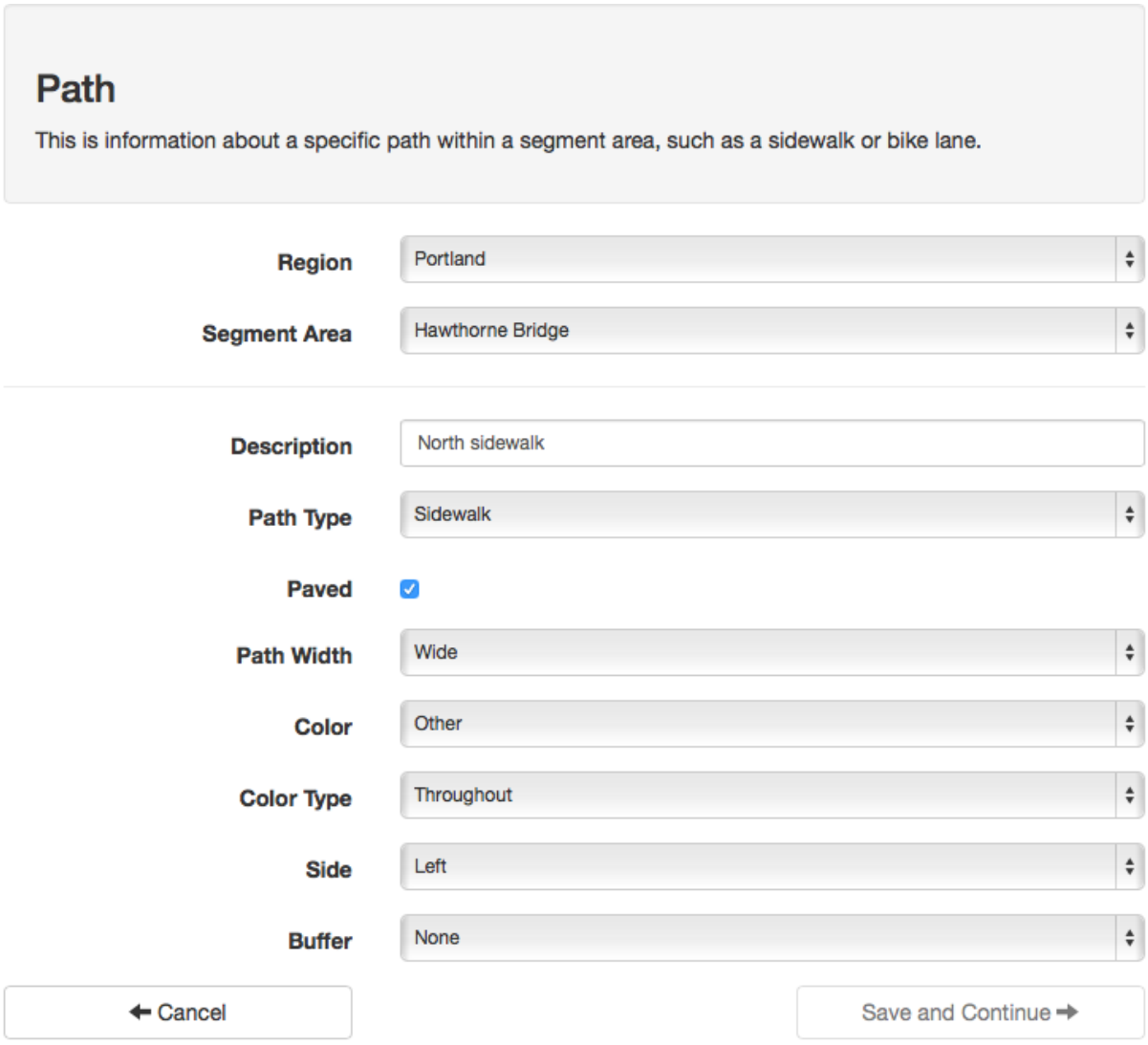
- [x] **Residential**
- [x] **Office**
- [ ] **Retail**
- [ ] **School**
- [x] **University**
- [ ] **Industrial**
- [ ] **Agricultural**
- [x] **Park**
- [ ] **Mixed Use**

Pick all that apply and are immediately adjacent to the count site. Do not include land uses that are nearby, but not immediately adjacent (e.g. a block away).

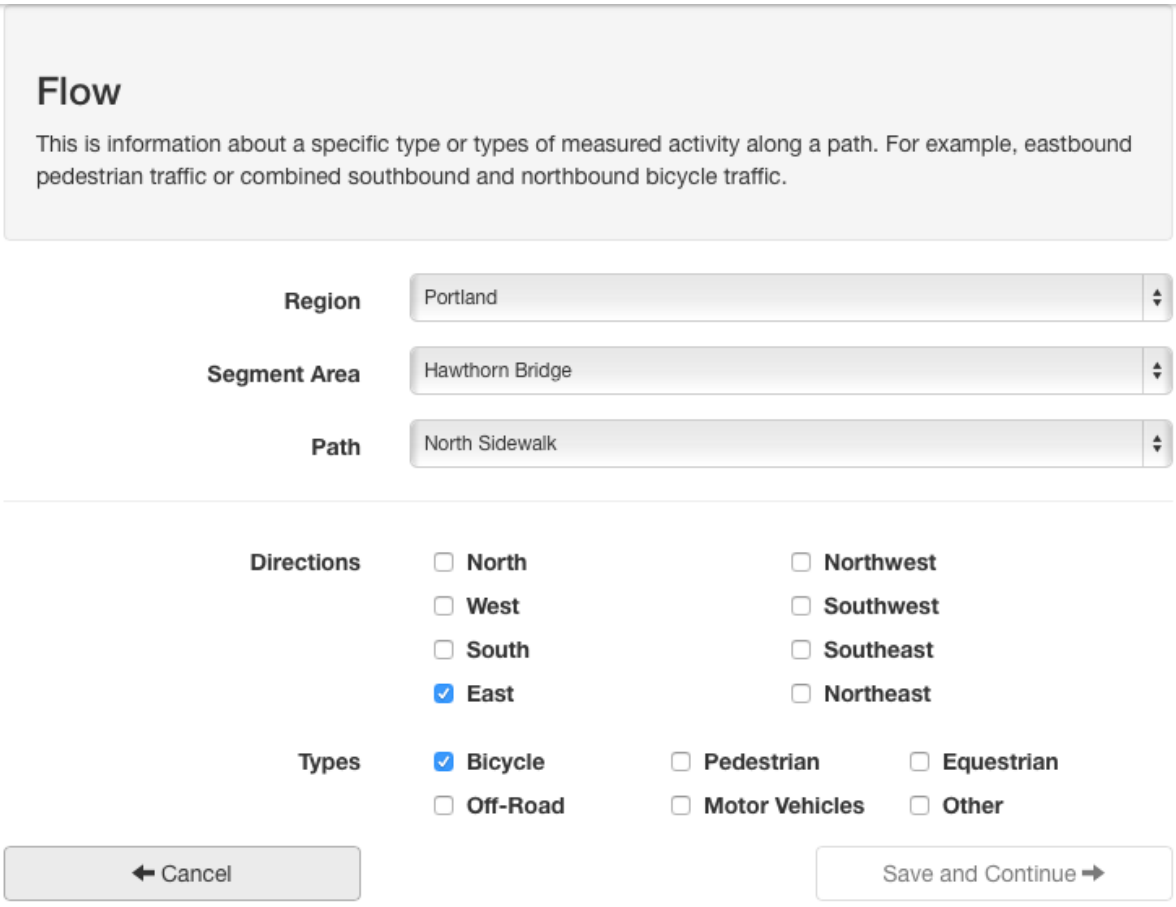
← Cancel

Save and Continue →

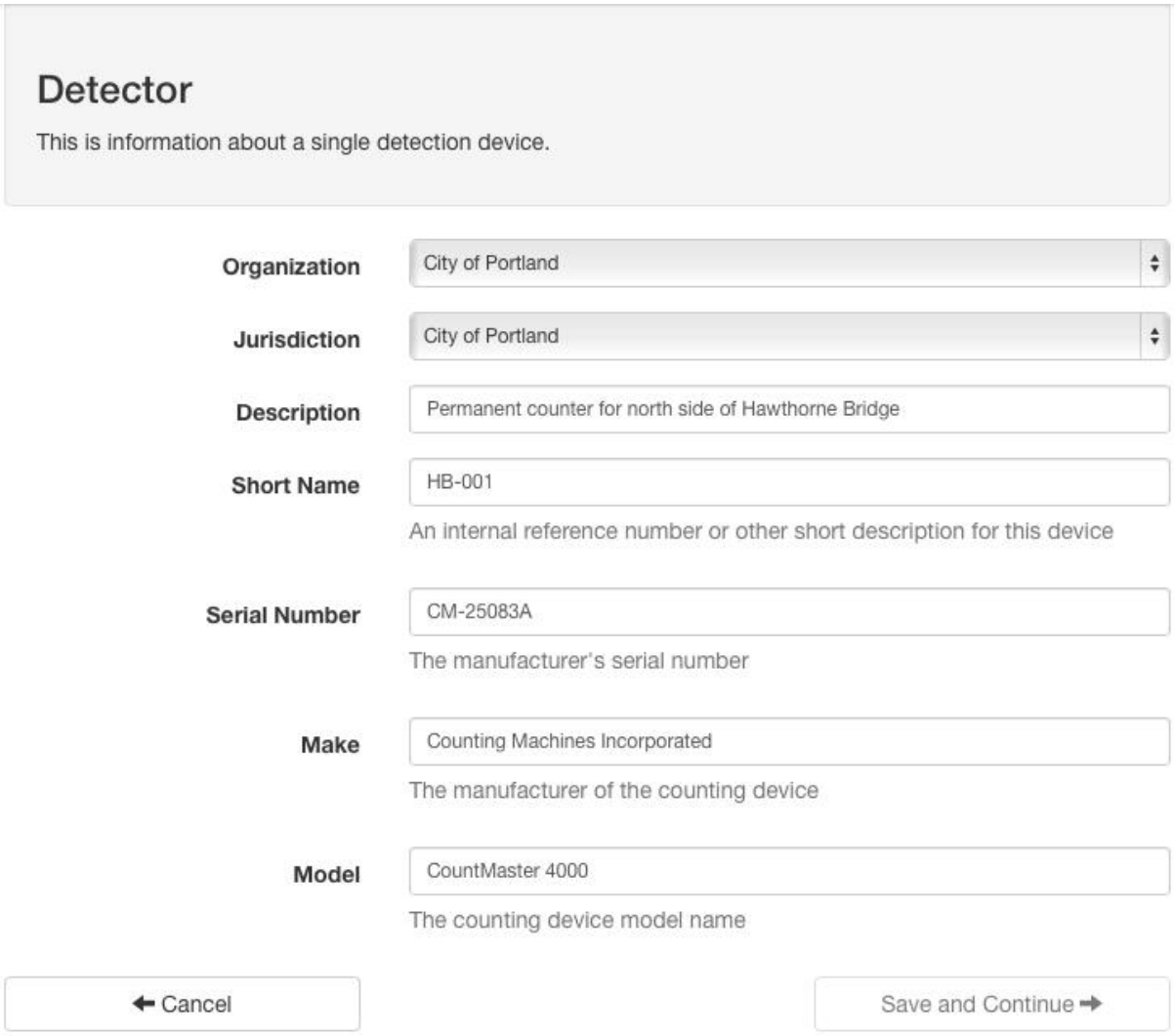
Facility/Path form:



Flow form:


Flow
This is information about a specific type or types of measured activity along a path. For example, eastbound pedestrian traffic or combined southbound and northbound bicycle traffic.
Region
Portland
Segment Area
Hawthorn Bridge
Path
North Sidewalk
Directions
North
Northwest
West
Southwest
South
Southeast
East
Northeast
Types
Bicycle
Pedestrian
Equestrian
Off-Road
Motor Vehicles
Other
Cancel
Save and Continue

Detector Form:



5. QA/QC Plan Memo

Quality Assurance / Quality Control (QA/QC) Plan

Online Non-motorized Traffic Count Archive ("Bike/Ped Portal")

Prepared for November 4th, 2014 Technical Advisory Committee Meeting

The Bike/Ped Portal will have several layers of quality assurance and quality control that aim to improve the quality of data. The QA/QC measures will either act to a) automatically reject certain data (while providing feedback as to why the data was rejected so that the uploader may correct the problem); b) alert the uploader about an issue observed in the data and require affirmation before allowing the upload; or c) trigger an alert on the back end while allowing the data to be uploaded, though not yet accepted until an archive representative approves the data. QA/QC measures will be implemented to various degrees during metadata upload (e.g. inputting information about the count detectors, location, etc.) and during the count data upload process (i.e. uploading the actual count numbers).

Three levels of QA/QC will be discussed in this document:

1. Database Constraints
2. Automated Checks
3. User Observation

Each level of QA/QC is detailed below. The Appendix reviews the background material on this subject. In this memo the term "user" refers to the person who is inputting the data. These three steps apply to data directly uploaded to the website and data that are automatically uploaded. For automatically uploaded data, users will need to log in to the site to QA/QC the data and will be emailed if more than a month of data has been left unchecked. Unchecked data stays in a holding area and is not accepted into the database.

1. Database constraints

The first layer of control on data quality begins with data validation according to the field constraints imposed on both the metadata and count data. The field constraints will limit the acceptable data to appropriate formats, lengths, etc.

1. Database Constraints

- Required fields must be non-null.
- If there is a route type indicated (interstate, US hwy, state hwy, county road, …), there must also be a route number associated with it. If there is no route type indicated, there also can't be a route number.
- End time must be after the start time, or the end time can be null.
- Speed limit is blank by default, must be between 0 and 100 if specified.
- The pavement color, prevalence of color through intersections and driveways, and buffer are specific to cycle tracks and bike lanes, and must be blank for other types of facilities.
- Overlapping counts (counts for the same detector of the same flow at the same time) will not be allowed. Feedback will be provided so that the uploader may edit the data appropriately. To

replace data already in the database, users will need to contact Portal staff to first delete the old data before they attempt to add new counts for that detector/flow/time period.

- If no detector is specified, no data can be uploaded.
- Each count must include either the start and end date/time or the start date/time and the duration. If start and end date/time and duration are all included the start date/time plus the duration is required to equal the end time.
- Counts must be integers greater than or equal to zero, not NULL with default zero.

Some constraints will be added in future phases of work, as additional funding becomes available.

Constraints for future phases:

- Ability to identify data for which daylight savings time is in the wrong place.
- Missing data. (gaps in the time stamps) One and two hour gaps could be associated with false time stamps or incorrect or correct daylight savings times.

2. Automated Checks

Once count data have passed initial constraint validation (including having the appropriate fields, having a valid timestamp, etc.), a set of automated checks will examine if the dataset includes unusually high or low counts, and other checks as listed below.

Flag	**Accept as Valid**	**Accept but keep Flag**	**Invalid data, accept but hide**
On the hourly level			
Flag if >15 consecutive zeros			
Flag data with >6 identical non-zero values			
Flag hours > 1,500			
On the daily level			
Flag days > 10,000			
Flag days with zero counts (if hourly counts not provided)			

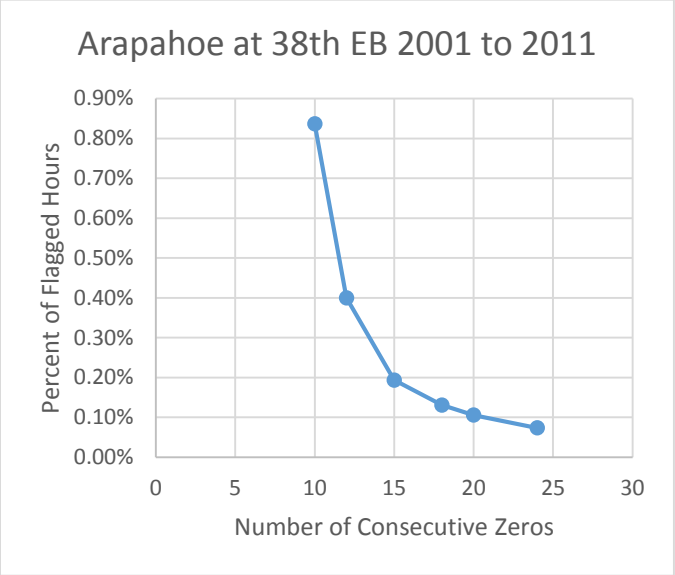
For each flag, the user will be asked if the data should be accepted as valid; if the data should be accepted but kept flagged; or if the data are invalid. Invalid data can still be included in the archive, but will be marked as invalid and will be hidden from the general public.

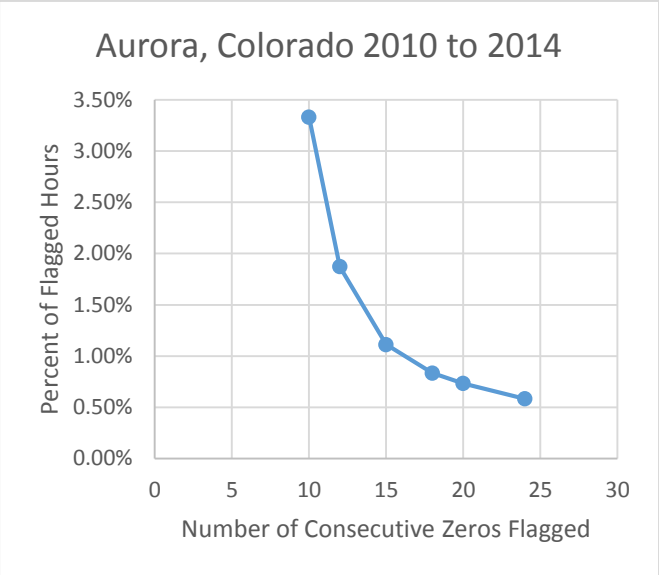
The numbers supplied in the flags above can be easily changed in the software, but for now, we are not allowing the user to change them and we are not using the data in the database to determine them, because there is not enough test data yet available. However, we have used the test data we have to verify that these are reasonable for the first Phase, and expect that future experience will allow us to modify them in future phases of the work.

Each flag was evaluated based on test data. For flags of lower bound thresholds two datasets were examined:

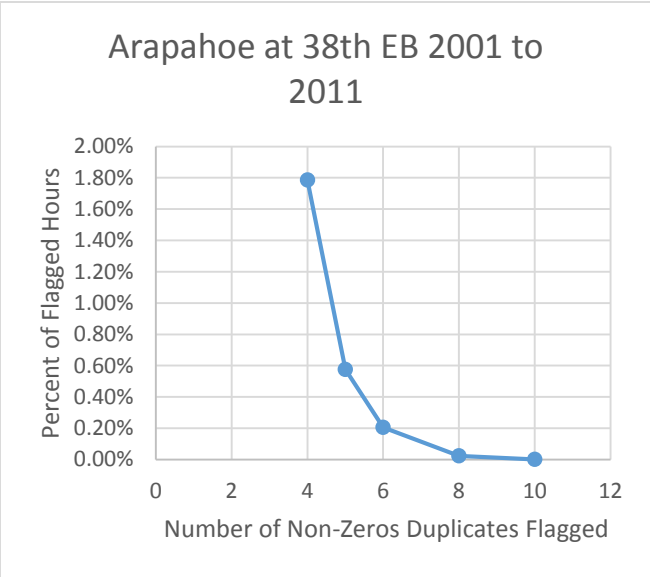
- A set of 10 years of test data from eastbound bicycle counts from an inductive loop detector operated by the city of Boulder at a low volume location with Annual Average Daily Bicyclists (AADB) of 164 in suburban Boulder, Colorado, on a wide sidewalk shared-use path on the north side of a six-lane highway (Arapahoe Blvd.). The test data have already been validated and checked, so they represent a correct dataset, although the location is known to consistently under count cyclists. Of the 10 years of data only 73% of it is present in the test data. The rest was either not collected or rejected.
- A set of bicycle counts from inductive loops in Aurora, Colorado, collected by the Colorado Department of Transportation on a residential street near a school in a suburban area for a four year period Oct. 2010 to Oct. 2014. While these data have not been validated or cleaned, they do represent a site with very low counts (AADB=27 in 2011). For this reason, these data are a good test case for low volume thresholds.

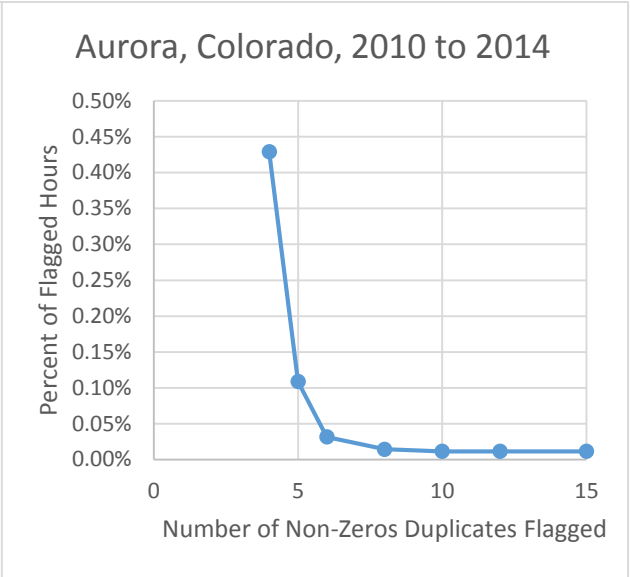
When more than 15 consecutive hours with zero count are present, these data will be flagged. The value, 15, was chosen because for less than 15 consecutive zeros the percent of flagged hours in the test data increased sharply. This threshold of 15 consecutive zeros represents 0.2% of the Arapahoe data and 1.1% of the Aurora data, which seems reasonable in both cases.



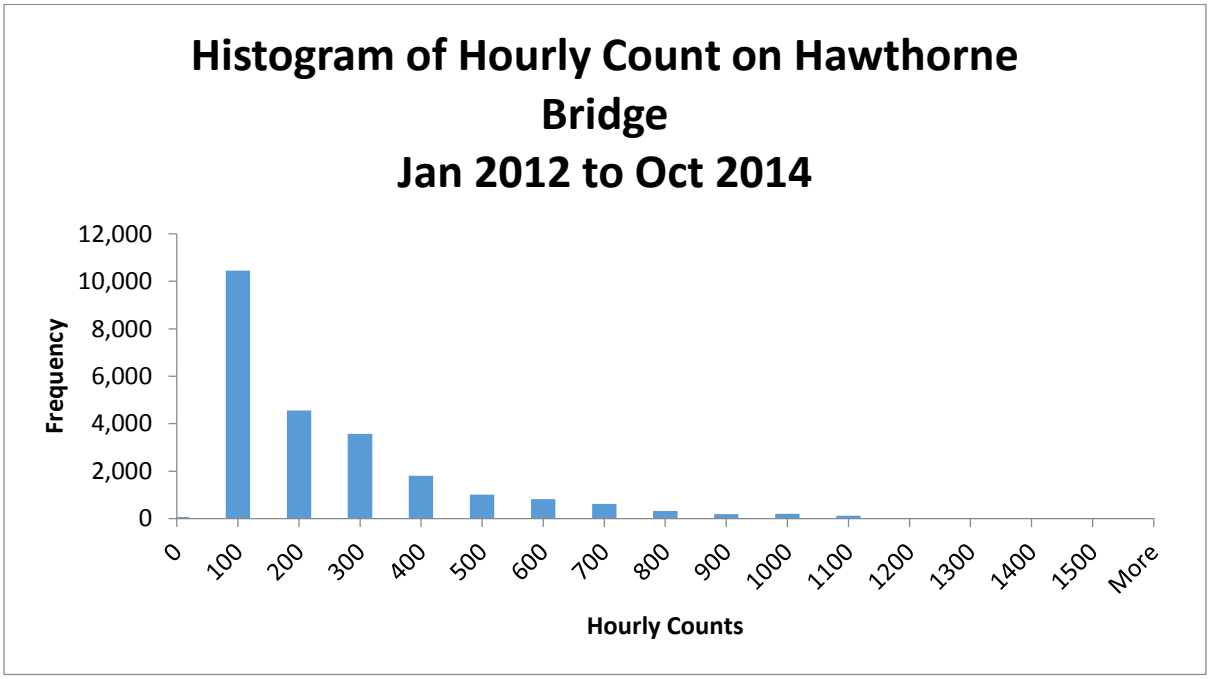


The value, 6, consecutive non-zero values was chosen based on the percent of data flagged. This threshold was chosen, because for more fewer than six consecutive non-zeros the percent of flagged hours rose sharply. At the threshold of 6 consecutive non-zero values 0.2% of the Arapahoe test data and 0.3% of the Aurora test data were flagged. This seems reasonable in both cases.



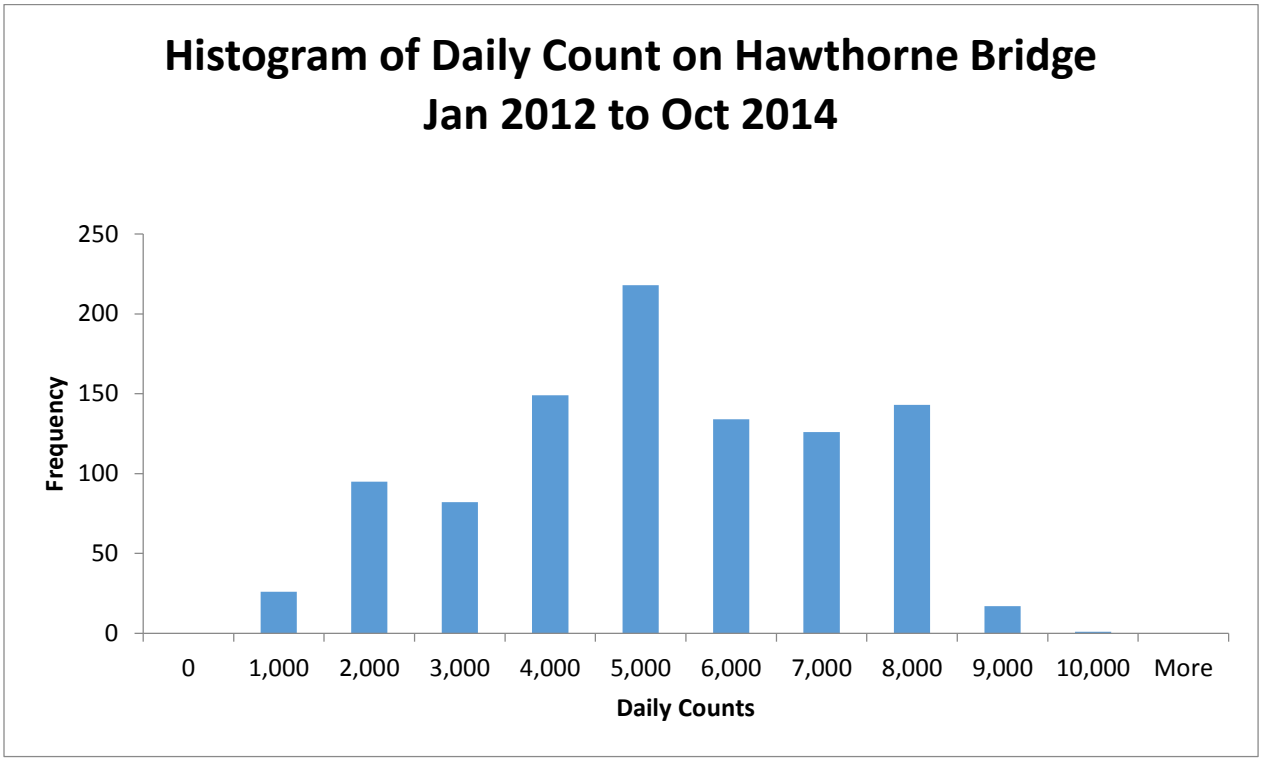


The maximum hourly value, 1,500, was chosen based on data from Portland's Hawthorne Bridge, one of the locations in the US with the highest bicycle volumes, for the period of January 2012 to October 2014. The highest count recorded on the bridge during that time period was 1,697 (both directions) bicycles on Saturday, June 8 at 9pm, the date of the World Naked Bike Ride. The next highest hours were 1,535 on Sunday, August 11, 2013 and 1,502 on Sunday, August 12, 2012 both at 8am corresponding to Portland's annual Bridge Pedal. The choice of an hourly volume that would result in flagging these exceptional hours seemed appropriate, as it would allow users to identify hours with particularly high volumes and note them appropriately in the comments.



The maximum daily value, 10,000 count per day, was again chosen based on data from Portland's Hawthorne Bridge, which recorded the highest daily count of 9,834 on Saturday, June 8, 2013 (*1*), the date of the World Naked Bike Ride. Here we chose a value above the maximum to prevent flagging too

much. The hourly counts should already flag most events, and this additional check serves to flag events that are either exceptional or that represent technical problems.



Flagging days with zero counts is redundant with flagging over 15 consecutive zeros, but is appropriate when the data is only provided at the daily level and no hourly values are available.

These checks do not include all the checks desired. Some potentially useful checks could be included in future phases of work as additional funds become available.

Potential automated checks to be implemented in future phases:

- Direction Distribution Check: For data with separated two directional traffic flows, flag data with too much of a directional bias.
- Comparison of Counts: Where counts are collected for the same flow, at the same time, but by different detectors, check if the percent difference is above a given threshold.
- Repeating Counts: Check for count patterns that repeat in order to identify if data has been copied from one time period and pasted to a different time period.
- Inconsistent Counts: When counts jump or decline precipitously, there may be a problem. As more test data become available we can better study this to understand how best to identify such jumps and if a generalized threshold can or should be established.
- Global, State, Regional, Equipment-specific, User-specified, or Location-specific Thresholds: Future phases of work could examine how thresholds should be set and allow them to be set by jurisdiction, by individual location, by users, or based on historic data at the site. Future research is needed to identify how to set thresholds.

3. User Observations

In addition to the automated checks the user (in this case, the person who inputs the data) can add comments to the data for user specified time periods. The user will also be shown a graph (below) of the total counts over time and asked to identify problem data as either questionable or incorrect. The user can indicate if the comment refers to counts which the archive will "accept" or to "invalid data" which would be hidden from the public web interface.

Data Rating System:

At this time it is not feasible to develop a comprehensive data rating system. Such a system can be developed in the next phase of work when more data are available to be evaluated. For now, we suggest the following simplistic system.

If the user has identified any of the data as "invalid" either through the automated checking or visual inspection processes, the data will not be available to anyone except official users of the system; in other words, it will be hidden from the general public.

All the data quality information, both automated flags that have not been indicated as "accept as valid" by the user and user specified comments will be supplied when data are accessed.

The data will obtain a gold star rating if it has no automated flags, except those for which the user has indicated "accept as valid", and no user specified "invalid data." For example, gold star data may contain the user comments, such as "heavy snow caused path to close" or "Cycling event passed by counter."

At this time, we do not propose any further rating of the data. Such a rating system will be easier to develop in the next phase of work when sufficient data have been uploaded to the system to test such ratings.

Ongoing maintenance:

In future phases of the project, maintenance checks can be added for the metadata as well. For example, as a means of confirming that metadata is current for each detector, the user who input the metadata could be asking to confirm the station data annually.

References

1. EcoCounter. Bike Barometer: Portland Hawthorne Bridge.

QA/QC Memo Appendix

Quality Assurance/Quality Control Supporting Documentation

Prior to preparing the recommended QA/QC for the Technical Advisory Committee, the project team investigated similar QA/QC checks used by others for both non-motorized and motorized traffic. This appendix documents some of the tests used by others.

Sprinkle

Elizabeth Stolz with Sprinkle Consulting provided a list of potential checks based on her work with motor vehicle monitoring counts QA/QC. The table is provided in the following pages. Some of these checks were included in our QA/QC plan as indicated by the designation "M" for minimum requirement and some are considered for future inclusion as indicated by the designation "P" for potential requirement.

FHWA

Jeremy Raw from the Federal Highway Administration (FHWA) provided a draft document he is working on for Travel Monitoring Analysis System (TMAS) which included quality checks for fatal errors, critical errors, and warnings. Since this document is still in draft form, it is not provided in this appendix. The fatal errors seemed similar to our constraints, but are specific to TMAS format. Consistency checks are covered, which were also specific to the Traffic Monitoring Guide (TMG) format used by TMAS. We may add additional consistency constraints in future phases of the work. The document also contained tests for reasonableness, some of which we chose to include and others

Turner & Lasley

Another useful document was a paper by Shawn Turner and Philip Lasley of the Texas Transportation Institute which examined data quality from a (Turner & Lasley, 2013). They suggest six aspects of data quality: accuracy, validity, completeness, timeliness, coverage, and accessibility. Our QA/QC plan focuses on validity, although the purpose of the bike/ped Portal is to increase data accessibility. The paper lists three types of automated validity criteria:

- Quality Control Checks
- Validity Checks
    - Univariate and multivariate range criteria: such as maximum traffic counts per time period.
    - Spatial and temporal consistency critera such as directional ratios should be less than 80% of traffic in one direction unless it's oneway and percent deviation from previous hour, and maximum ratio of peak hour to daily volume.
    - Detailed diagnostics which are technology specific
- Business Rules

The authors examine an example data set and use counts in one direction to check and adjust counts that were unusually high in the opposite direction.

Traffic Monitoring Guide

The Federal Highway Administration's TMG also includes guidance on data quality for motor vehicle monitoring data, which is not reproduced here, but includes case studies (in TMG's Appendix E) from Virginia, Vermont, Pennsylvania, Washington State, and New York State (Federal Highway Administration, 2013). For example, Vermont includes monthly manual inspection of graphs of traffic over a 24 hour period from each day of the week for a given month to identify problems. Automated checks identify monthly volumes that are 10% different from the previous year.

Bike/Ped Travel Monitoring Data - Quality Assurance and Quality Control Checks*

Valid Check	#		Check Type	Description	Comments
Formatting Checks				**Description**	
✓	1		Formatting	Data file contains numbers where numbers are expected	
✓	2		Formatting	Data file contains spaces where spaces are expected	Gaps Check
Location and Classification Checks				**Description**	
✓	3		Location	Data associated with correct site location	
✓	4		Location	Data is associated with correct road characteristics	example: roadway only, multi-use path, etc.
✓	5		Classification	Data associated with correct classification scheme	bike only, bike/ped combo, ped only, etc.
✓	6		Classification	Data are associated to the correct collection method	Automated, Manual, IR, Video, etc.
✓	7		Location	Data are associated with correct submitting Agency	ODOT, City of Eugene, etc.
Data Validation Checks				**Description**	
X	8		Data Validation	Directional Distribution	Data is too variable to set a threshold until a year of data is collected
X	9		Data Validation	Hourly Directional Distribution	Data is too variable to set a threshold until a year of data is collected
✓	10	P	Data Validation	Interquartile (IQ) Range Check	Total Daily Volume upper limit
X	11		Data Validation	Ratio Check – Lane 1 to Lane 2	No Sites with this type of Configuration / Instrumentation
✓	12	P	Data Validation	Ratio Check	Noon Volume to Midnight Volume Check
✓	13	M	Data Validation	Number of Consecutive Zero's	Need to set threshold….motorized is 7 but non-motorized might be 10 (KLN to use 24)
✓	14	P	Data Validation	Number of Allowable Zero Intervals from 7am and 7pm	Need to set threshold
✓	15	M	Data Validation	Number of Consecutive Identical Values Check	
X	16		Data Validation	Volume Interval	Data too variable from hour to hour to check volume interval changes
X	17		Data Validation	Continuous AADT Percent Difference check	Not enough data to run year to year AADT checks
X	18		Data Validation	Continuous MADT Percent Difference Check	Not enough Data to run year to year MADT checks
X	19		Data Validation	Short-term AADT Percent Difference Check	Need a full year of continuous count data to factor short duration count data
X	20		Data Validation	Classification Percent Threshold Checks	Not enough knowledge to set threshold for bike versus ped
X	21		Data Validation	Ratio of Bike/Ped Check	Need to establish threshold
X	22		Data Validation	Total Vehicle Length Check	Need to set threshold

X	23		Data Validation	Vehicle Speed Check	Can speed data be acquired from collection equipment?
✓	24		Data Validation	All Bike Volumes Associated with Bike Volume data, etc.	
X	25		Data Validation	Minimum Number of Hours for a Short-term Count	Should be 24 hours
✓	26		Data Validation	Hourly Volume Math Check	Direction one + direction two = Total Volume
✓	27		Data Validation	Count not Complete	Must have 24 hours of consecutive hourly data
✓	28	M	Data Validation	Total Daily Volume Zero Check	Check daily volume equal to zero, throw this out for AADT creation purposes

M = Minimum QA/QC check

P = Potential for in future

*Table provided by Elizabeth Stolz with Sprinkle Consulting

References

Federal Highway Administration. (2013). Traffic Monitoring Guide. Washington, DC: U.S. Department of Transportation.

Turner, S., & Lasley, P. (2013). *Quality Counts for Pedestrians and Bicyclists: Quality Assurance Procedures for Non-Motorized Traffic Count Data*. Paper presented at the 92nd Annual Meeting of the Transportation Research Board, Washington, D.C.