U.S. Department of Transportation

ADVANCING TRANSPORTATION DATA UTILITY WHILE MITIGATING PRIVACY RISK

Ariel Gold, Data Program Manager ITS JPO
Jason M. Carter, Oak Ridge National Laboratory

JANUARY 17, 2018

PROJECT VISION AND OBJECTIVES

To develop tools and practices that manage privacy risk while maintaining the utility of ITS data for research and development of innovative applications that improve safety, mobility, and environmental protection.

OBJECTIVES

- **Improve Public Understanding of Geolocation Privacy**
- **Identify Privacy Weaknesses and Develop Solutions**
- **Collaborate with ITS Deployers to Build Capabilities**

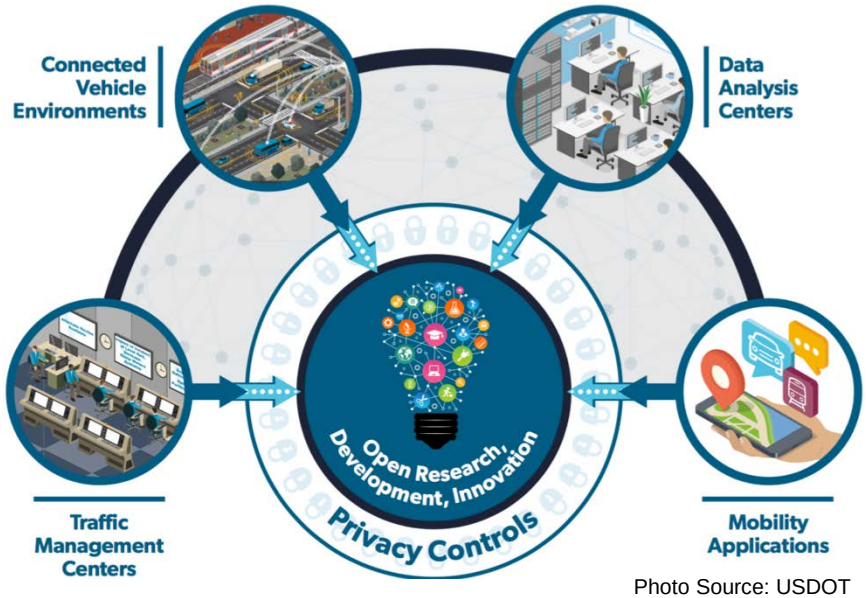
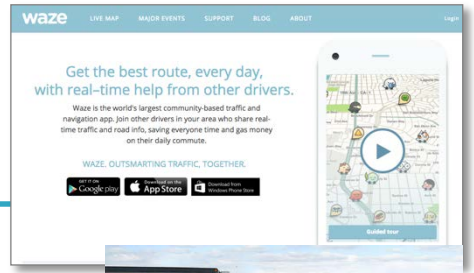
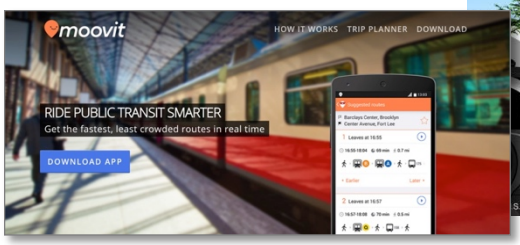


Photo Source: USDOT

TRANSPORTATION DATA HAS VALUE

- Mobility applications and location-based services
- Driving safety and advanced driving systems
- Product improvement
- Situational awareness: traffic, navigation, routing
- Traffic management





Sharing data allows more people to perform research, innovate, and meet open data requirements

Google Maps

GEOLOCATION, MOBILITY, LINKED SOURCES, AND PRIVACY

- People consider presence and absence at certain locations sensitive. Some sensitive locations also have value to analysts.

- Linking datasets may introduce problems that did not exist in the separate datasets.

Source: www.washingtoncountywisdems.org

- Limiting collection to areas that exclude sensitiv locations can help. Providing useful summary information may also mitigate problems.

Source: johnrodandco.com

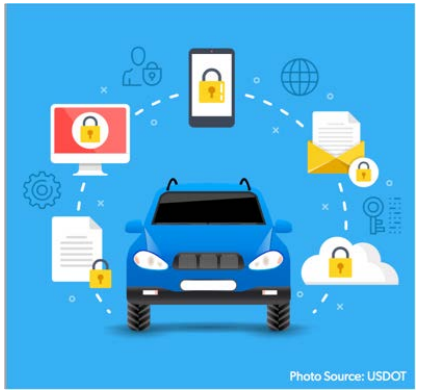
UNDERSTANDING PRIVACY AND OUR DATA IS ESSENTIAL TO MANAGING RISKS

Information privacy is defined in many ways…

Control over information about ourselves [1]
Conditions that deprive other's access to information about you [2]

Privacy language and risk mitigation is evolving... [3]

Identifying Information,
Personal Information,
Anonymization,
De-Identification,
Re-Identification

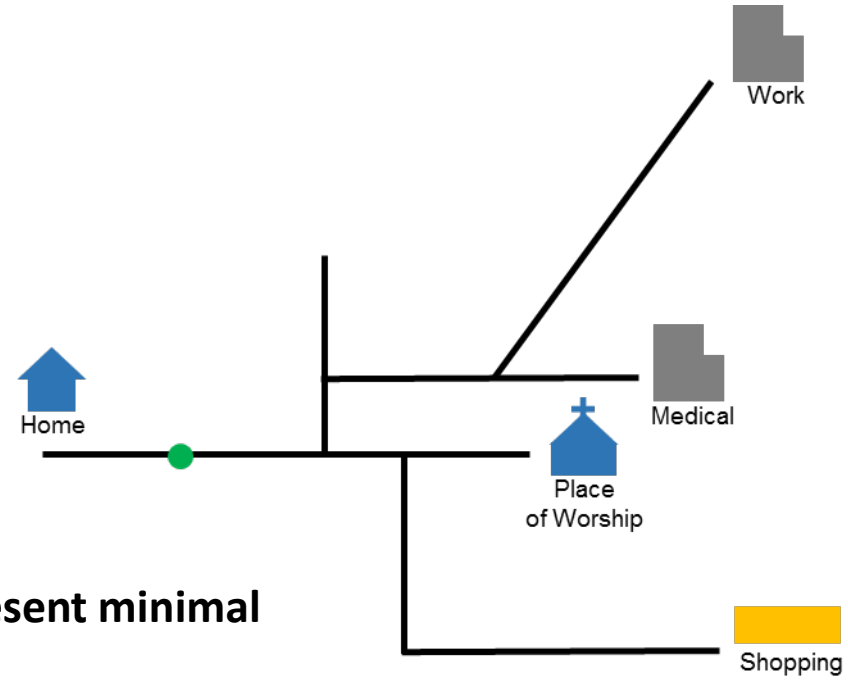


To mitigate privacy risk we must understand the characteristics that make our data personal and identifying

[1] Solove, Daniel J.; Conceptualizing Privacy, California Law Review, 2002
[2] Reiman, Jeffrey H.; Privacy, Intimacy, and Personhood; Philosophy and Public Affairs, 1976
[3] Garfinkel, Simson L.; De-Identification of Personal Information (NISTIR 8053); 2015

PRIVACY PROTECTION IS "DESIGNED INTO" ITS DATA SPECIFICATIONS

But, ordering and collecting data may present unanticipated challenges.

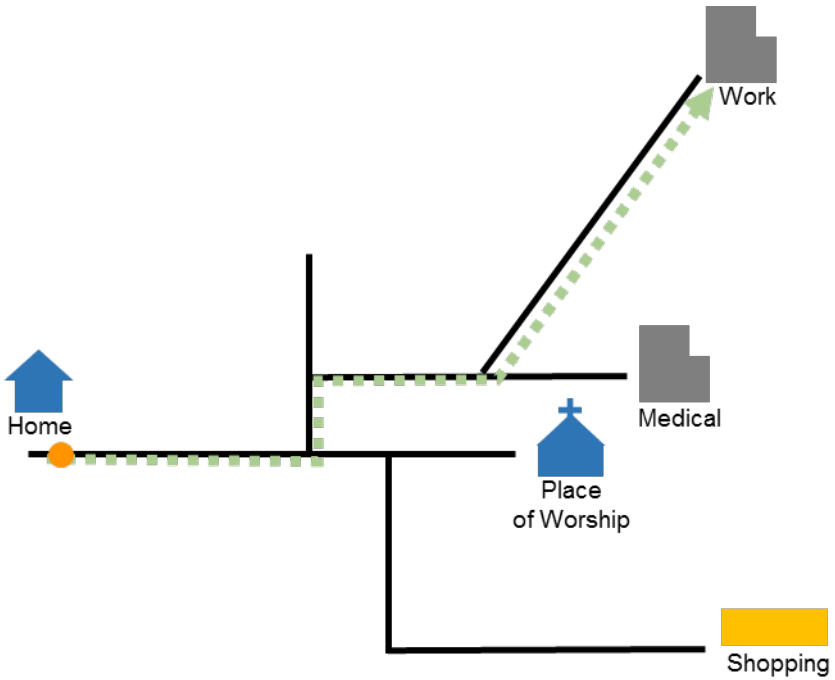


Individual locations present minimal privacy risk

PRIVACY PROTECTION IS "DESIGNED INTO" ITS DATA SPECIFICATIONS

But, ordering and collecting data may present unanticipated challenges.

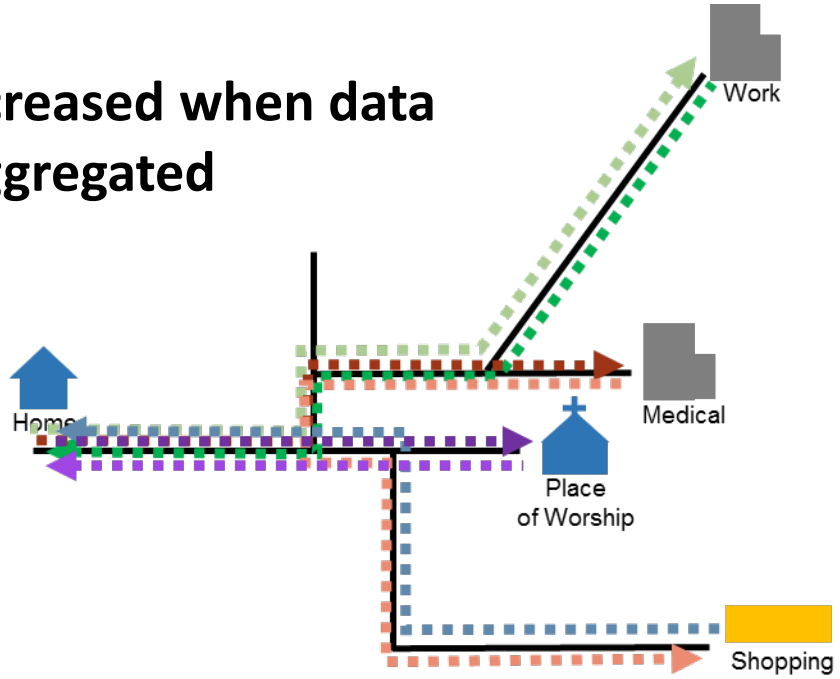
Risk is increased when data can be ordered



PRIVACY PROTECTION IS "DESIGNED INTO" ITS DATA SPECIFICATIONS (cont.)

But, ordering and collecting data may present unanticipated challenges.

Risk is further increased when data is ordered and aggregated

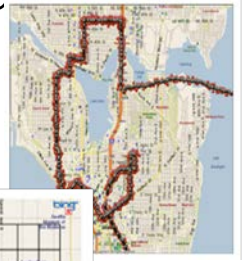


Examples:

- Learning the route a person takes to get home
- Learning how many stop signs a person "rolls through"
- Learning when a person's home is likely vacant throughout the week
- Learning when a business's operational tempo changes

CHALLENGES, CONSIDERATIONS, AND CURRENT APPROACHES

- Perspectives on what makes a location private are subjective
- Open source information aids re-identification efforts
- Once data is released, it is hard to recall
- How data is used dictates location and time fidelity requirements
  - Safety and traffic application development require precise data
- Location privacy algorithms modify or eliminate original data
  - Summarize, reposition, reduce fidelity, remove records



Risk, Protection, and Utility Requirements should be Balanced

WE MANAGE LOCATION PRIVACY RISK BY IDENTIFYING SENSITIVE LOCATIONS AND HIDING THEM

1. **Identify** privacy-sensitive locations/behavior using trip and map features
   - Loitering behavior (where we stop)
2. **Remove** enough data to hide sensitive locations
   - Improve results using relevant external data
   - Incorporate data privacy strategies proven in other areas
     k-Anonymity [8,9], Information Theory [10]

Two tools have been developed that apply this approach to different data contexts:

- **Privacy Protection Algorithm (PPA)**
- **Privacy Protection Module (PPM)**

[8] k-anonymity: A Model for Protecting Privacy, Latanya Sweeney; 2002
[9] k-secure path: Hiding Sequential and Spatiotemporal Patterns; Abul, Bonchi, Giannotti; 2010
[10] A Mathematical Theory of Communication; Claude E. Shannon; 1948

PRIVACY PROTECTION ALGORITHM:

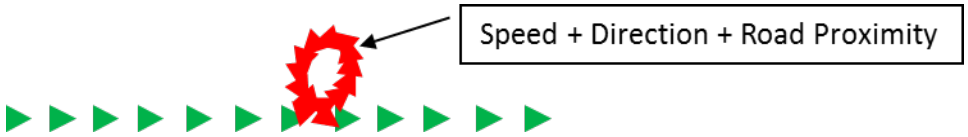
A Tool for Trip Databases

DRIVING BEHAVIOR AND MAP INFORMATION ARE USED TO IDENTIFY SENSITIVE LOCATIONS

- We classify **driving behavior (many points)** in conjunction with **where it occurs**
  - Rule 1: Trip begin and end points are assumed sensitive

  - Rule 2: Stops at certain locations

  - Rule 3: Turnaround or drop-off behavior



- Road proximity, road type, and area type help make decisions
  - Example: Stops on interstates are not sensitive
  - Example: Stops at a traffic light are not sensitive

Signals and signs

Schools, churches, medical facilities

Source: lifeofdad.com

Traffic

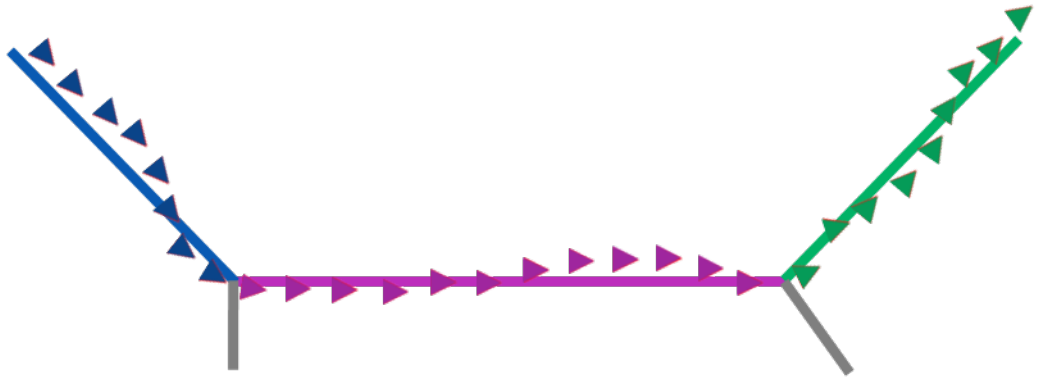
Source: zmescience.com

Home driveways

Source: jonnylives.com

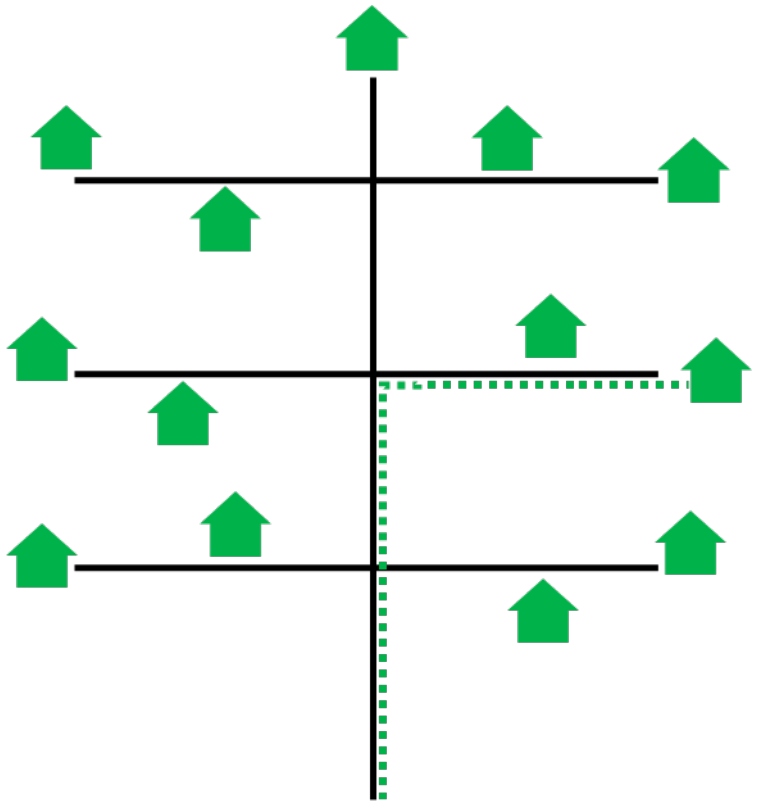
ASSOCIATING GEOLOCATIONS WITH ROADS PROVIDES INFORMATION USEFUL FOR PRIVACY PROTECTION

- Time-ordered locations with heading and speed only provide part of the picture



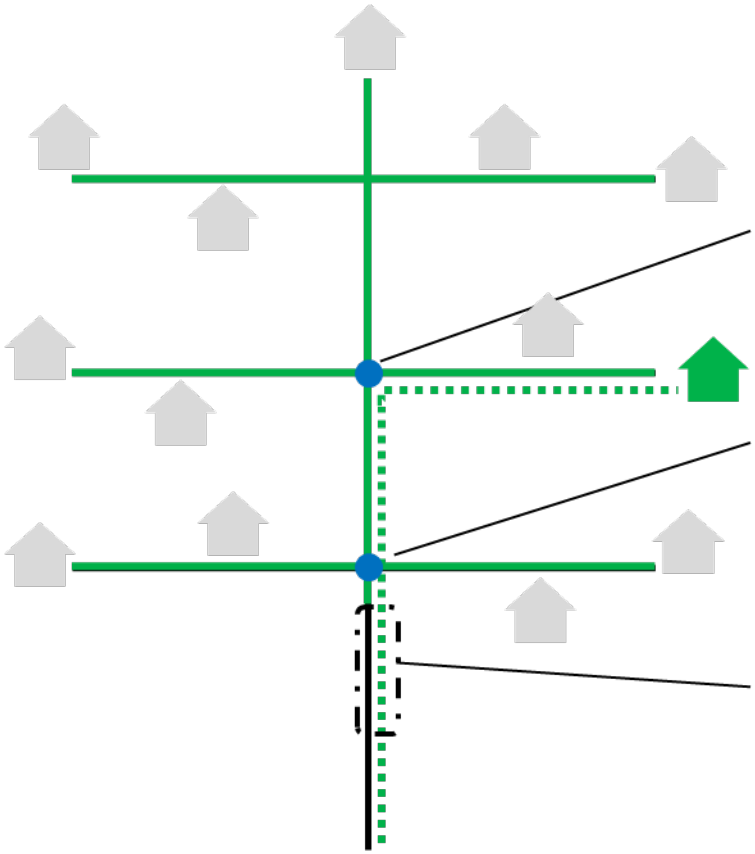
- To use this information, locations must be “matched” to road segments
  - Each trip point can be labeled with data from its “matched” road
  - Location measurements can be corrected
  - Off-road driving can be detected
  - Intersection pass-through can be determined.

LEARNING SOMEONE'S DESTINATION BY REDUCING POSSIBILITIES



- As we approach a destination, the number of possible alternative destinations decreases.
- We reverse this process to **hide sensitive locations**.

HOW WE HIDE SENSITIVE LOCATIONS USING THE ROAD NETWORK



The first 4-way intersection hides the actual destination within 3 road branches

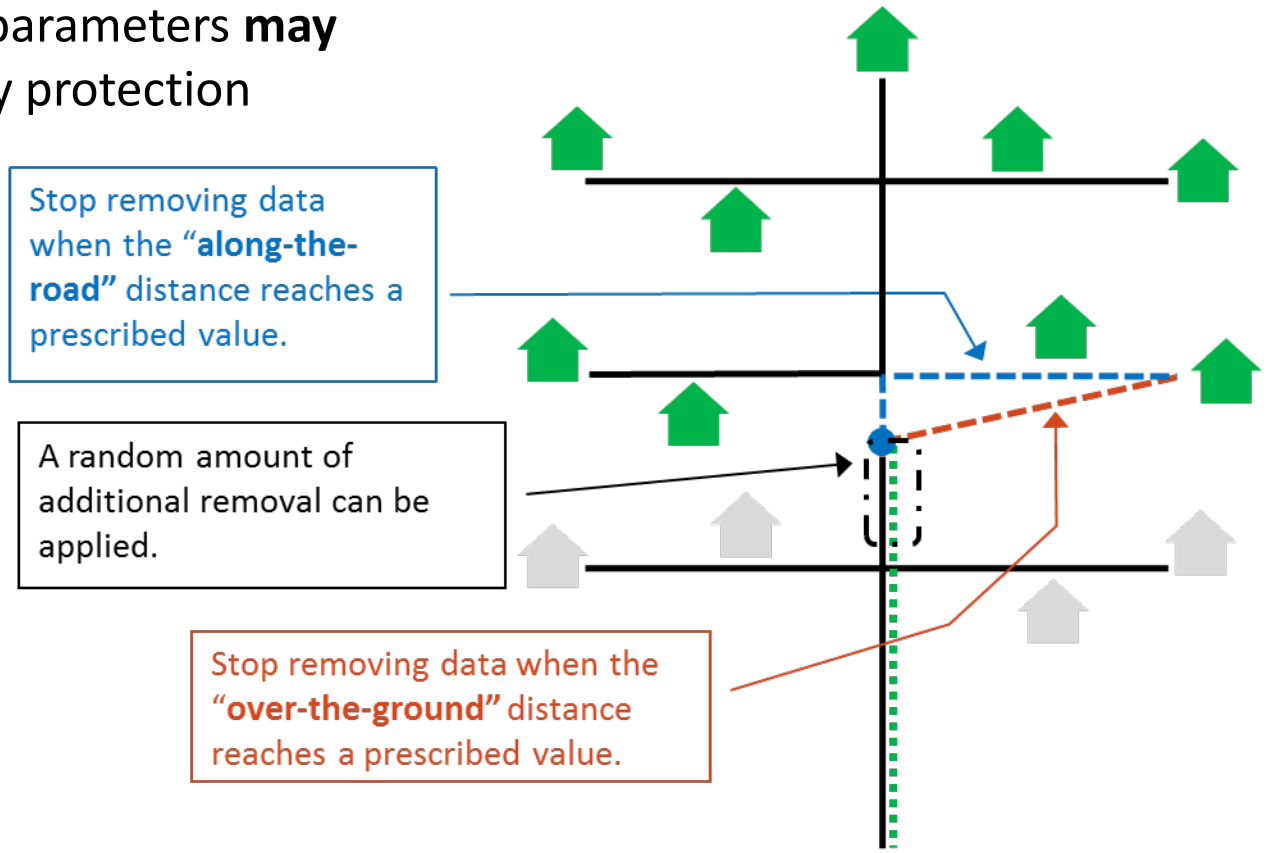
The second 4-way intersection hides the actual destination within 2 additional road branches

A random amount of additional removal can be applied.

… A MEASURABLE WAY TO HIDE A SENSITIVE LOCATION INDEPENDENT OF DISTANCE!

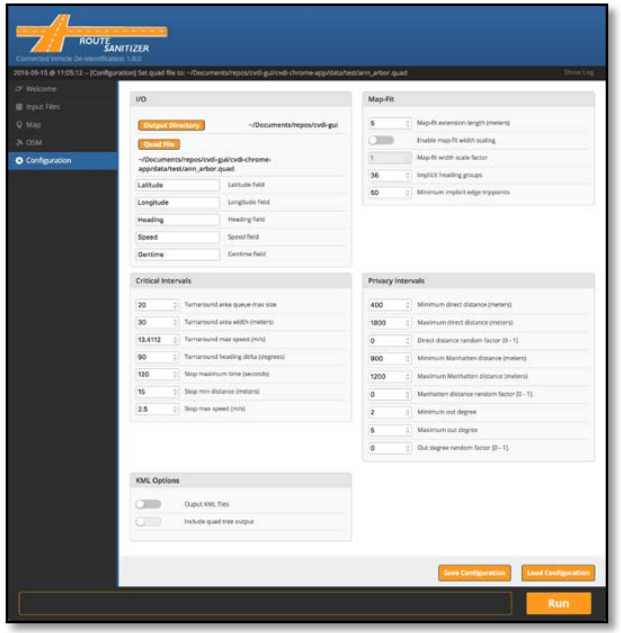
ALTERNATIVES FOR PRIORITIZING DATA RETENTION

- Distances from sensitive locations can be used to **maximize data retention,** if desired
- Data retention parameters **may preempt** privacy protection parameters



PRIVACY PROTECTION ALGORITHM (PPA): A TOOL FOR LARGE MOVING OBJECT DATABASES

- Used to open source Ann Arbor Safety Pilot data
  - ~460,000 trips
- User-friendly interface
- Well documented, highly configurable
- Uses OpenStreetMap data (free)
- Optional annotated KML output for visual inspection of results
- Input and output data is comma-separated values (CSV)
  - Required fields: Latitude, Longitude, Time, Heading, Speed
  - Additional fields are retained in output



PRIVACY PROTECTION MODULE:

A Tool for Streaming Data

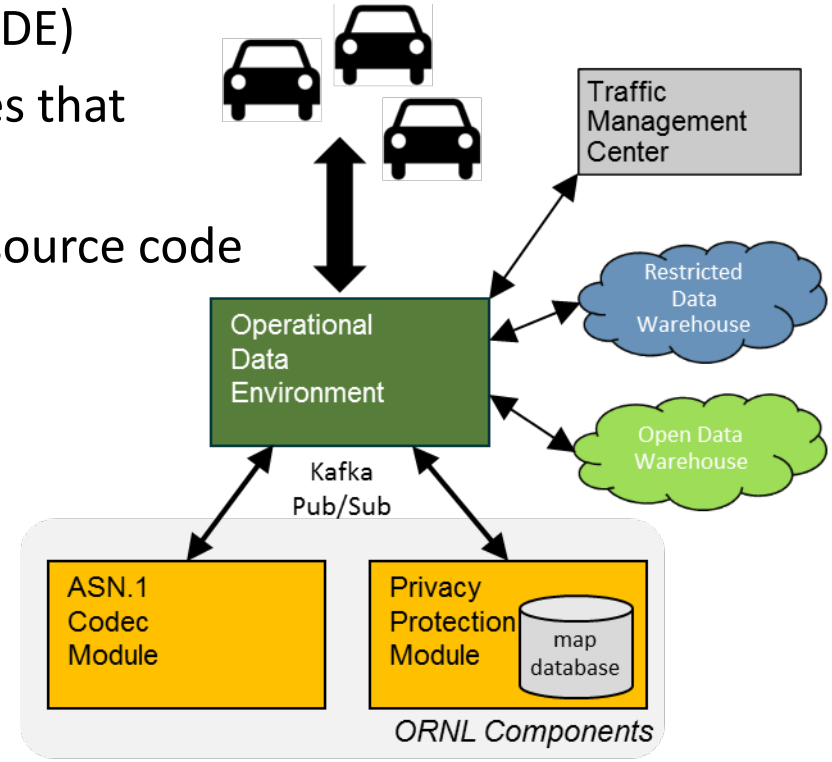
PROTECTING A HIGH-SPEED, LOW-DENSITY INTERSTATE

- Streaming data applications present unique privacy protection challenges

- Geofencing: Identify and isolate a known, irregular area.
- Locations are **classified as outside or inside the fence** in real time.
- Additional filters can be applied: low and high speeds, vehicle dimensions, identifiers

PRIVACY PROTECTION MODULE (PPM): A TOOL FOR STREAMING LOCATION DATA

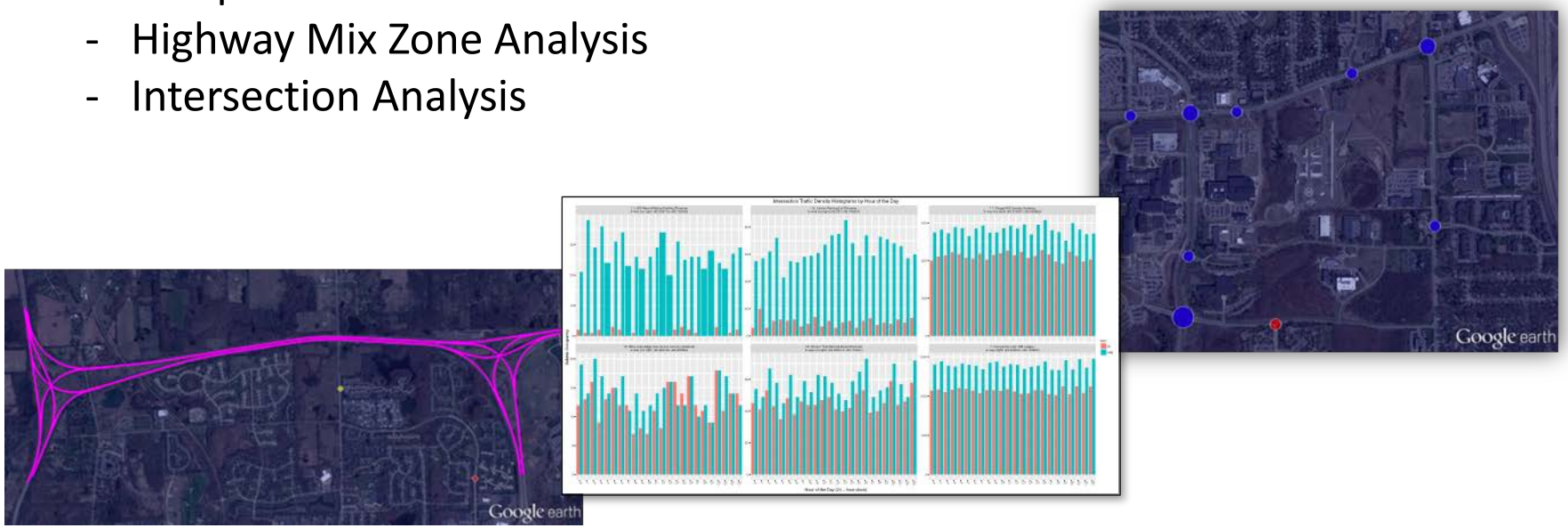
- Developed using Agile practices for the WYDOT Safety Pilot
- Example of **application-specific** privacy protection
- A standalone capability that integrates with the Operational Data Environment (ODE)
- Handles multiple J2735 message types that contain location information
- All code and documentation is open source code
  - *https://github.com/usdot-jpo-ode*



EVALUATING THE METHOD

HOW WELL ARE WE MANAGING PRIVACY RISK? AND, DOES THE DATA REMAIN USEFUL?

- Privacy Vetting by Independent Teams
  - No successful re-identifications
  - Incorrect claims of driver identity remain a possibility
- Processed data remains useful in many scenarios
  - Independent evaluation
  - Highway Mix Zone Analysis
  - Intersection Analysis



ADDITIONAL WORK

WORK IN PROGRESS

- Continued advancement of geolocation privacy tools
- Application-specific approaches to geolocation privacy
- Measuring data utility in relation to privacy protection
- Quantitative evaluation of privacy designs

JOIN THIS EFFORT

- We want to understand your ITS data privacy concerns and **help you address them**!
- Contact us for help and questions:

Ariel Gold
Data Program Manager
(202) 366-4374
ariel.gold@dot.gov

Jason M. Carter
Principal Investigator, ORNL
(865) 574-1480
carterjm@ornl.gov