Connected Vehicle Pilot Deployment Program Phase 2

Data Privacy Plan – New York City

www.its.dot.gov/index.htm
Final Report — December 27, 2016
Publication Number: FHWA-JPO-17-453

U.S. Department of Transportation

Produced by Connected Vehicle Pilot Deployment Program Phase 2
New York City Department of Transportation
U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology
Intelligent Transportation Systems Joint Program Office

Notice

Technical Report Documentation Page

1. Report No. **FHWA-JPO-17-453**	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Connected Vehicle Pilot Deployment Program Phase 2, Data Privacy Plan- New York City		**5. Report Date** December 27, 2016 **6. Performing Organization Code**
7. Author(s) Drew Van Duren, Scott Cadzow, Jonathan Petit, William Whyte, Security Innovation Robert Rausch, TransCore		**8. Performing Organization Report No.**
9. Performing Organization Name And Address New York City Department of Transportation (NYCDOT) Traffic Operation, ITS Management division 34-02 Queens Boulevard Long Island City, NY 11101		**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.** DTFH6116RA00007
12. Sponsoring Agency Name and Address ITS-Joint Program Office 1200 New Jersey Avenue, S.E., Washington, DC 20590		**13. Type of Report and Period Covered** Final Report **14. Sponsoring Agency Code** HOIT-1

15. Supplementary Notes
Work performed for:
Program Manager: Kate Hartman
Contracting Officer's Representative (AOR): Jonathan Walker

16. Abstract

This document represents a data privacy plan for ensuring the data privacy and security of those participating in the New York City connected vehicle pilot. Personally Identifiable Information (PII) and Sensitive PII (SPII) will be collected from various participants during the course of the pilot, protected, anonymized, obfuscated, and studied. Privacy information may pertain to the individual registrants along with the V2I, V2V and Pedestrian-related safety and mobility application data that will be output from Aftermarket Safety Devices (ASD), Roadside equipment, in-vehicle telematics systems, and hand-held devices. Privacy protection planning also includes the necessary protective controls and procedures for aggregating large quantities of data that, in isolation, are not private, but in the aggregate, may disclose PII and individual location histories.
The data privacy plan includes an identification of the New York City connected vehicle pilot privacy-related data, its security treatment and the necessary filtering, anonymization and obfuscation requirements needed for distributing the data for Independent Evaluator (IE), USDOT, and research data exchange (RDE) use.

17. Key Words Data, Connected Vehicle, Privacy, Management, New York City	**18. Distribution Statement**		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** **43**	**22. Price**

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Acknowledgements

The New York City Department of Transportation (NYCDOT) and its pilot deployment team thanks the many fleet owners dedicated to bringing connected vehicle technology to New York City. These stakeholder organizations demonstrate their commitment towards attaining Vision Zero's goals through their participation. The various NYCDOT fleets, taxi owners, UPS, MTA/NYCT, the NYC Sanitation Department, NY State Motor Truck Association, and Pedestrians for Accessible and Safe Streets (PASS) have expended considerable resources participating in the development of this overall Concept of Operations including this plan.

Finally, the team wants to thank the USDOT for sponsoring this project and laying the foundation for future connected vehicle deployments.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Scope and purpose of document

The scope and purpose of this document, drawn from the Statement of Work for the Pilot Deployment Phase 2, is as follows.

Purpose: to detail data privacy controls sufficient to mitigate the risk of harm to individuals that would result in the improper handing or disclosure of the Personally Identifiable Information (PII) and Sensitive Personally Identifiable Information (SPII) collected from individuals in connection with the New York City (NYC) Connected Vehicle Pilot Deployment (CVPD).

This document's use of the terms 'PII' and 'SPII' are formally provided in the List of Definitions and Acronyms.

In general, PII are any information that can be used to identify an individual rather than a device, vehicle, or transmitter. Basic information such as name, address, telephone number, and equipment identification can be considered PII.

SPII are any information if lost, compromised or disclosed without authorization, could result in substantial harm, embarrassment, inconvenience, or unfairness to an individual. Sensitive PII requires stricter handling guidelines because of the increased risk to an individual if the data are compromised. SPII may contain any of the following:

- Social Security number (SSN)
- Passport number
- Driver's license number
- Vehicle Identification Number (VIN)
- Biometrics, such as finger or iris print
- Financial account number such as credit card or bank account number
- The combination of any individual identifier and date of birth, 39 or mother's maiden name, or last four of an individual's SSN

This document is the Data Privacy Plan (DPP) for NYC CVPD, developed to satisfy the requirement under Agreement No. DTFH6116H00026 that NYC CVPD devote sufficient resources, and develop and adhere to policies and procedures to ensure that privacy-risks stemming from deployment are mitigated appropriately. This DPP documents the technical, policy, standards, and physical controls that NYC CVPD will put in place (and require its sub-recipients and contractors to put in place) to protect PII and SPII in order to mitigate potential privacy harms. These controls will ensure that NYC CVPD will interact with PII only on infrastructure that is subject to appropriate security controls. Sub-recipients, contractors, and others who handle or may access PII or SPII developed by NYC CVPD shall be required to adhere to this DPP.

The intended audience of this document includes:

- Project oversight staff from USDOT
- All project staff within NYC CVPD who may have access to, manage systems that store, or form policies that affect sensitive data and its treatment
- All subcontractors whose work may impact sensitive data
- The Independent Evaluator

1.2 New York City (NYC)

1.2.1 Text Summarized from CV Pilot Program

The Connected Vehicle Pilot Deployment program seeks to spur innovation among early adopters of connected vehicle application concepts, using best available and emerging technologies. The pilot deployments are expected to integrate connected vehicle research concepts into practical and effective elements, enhancing existing operational capabilities. The intent of these pilot deployments is to encourage partnerships of multiple stakeholders (e.g., private companies, States, transit agencies, commercial vehicle operators, and freight shippers) to deploy applications utilizing data captured from multiple sources (e.g., vehicles, mobile devices, and infrastructure) across all elements of the surface transportation system (i.e., transit, freeway, arterial, parking facilities, and tollways) to support improved system performance and enhanced performance-based management. The pilot deployments are also expected to support an impact assessment and evaluation effort that will inform a broader cost-benefit assessment of connected vehicle concepts and technologies.

The primary objective of the NYC CVPD team is to improve the safety of travelers and pedestrians in New York City through the application of connected vehicle technologies. This objective of the CV Pilot NYC directly aligns with New York City's Vision Zero initiative, which seeks to reduce pedestrian fatalities and make the City's streets safer for travelers in all modes of transportation. The NYC site provides an ideal opportunity to evaluate the CV technology and applications in tightly-spaced intersections typical in a dense urban transportation system. Connected vehicle technologies and associated applications will be deployed along heavily traveled high accident rate arterials in Manhattan and Brooklyn (see Figure 1-1) to provide a comparative sample that can be used to verify benefits against those for locations that are not instrumented.

Source: NYCDOT, 2016

Figure 1-1. New York City Pilot Deployment Site Map

The NYC pilot deployment will feature the installation and utilization of vehicle to vehicle (V2V) technology in approximately 8000 city-owned and other fleet vehicles. Traffic signals in the high-priority corridors in Manhattan and Brooklyn will be upgraded with vehicle to infrastructure (V2I) communications capabilities. Table 1-1 lists the devices to be deployed.

Table 1-1. Devices to be Deployed

NYCDOT – Devices	Estimated Number
Roadside Unit (RSU) at Manhattan and Brooklyn Intersections and FDR Drive	353
Taxi Equipped with Aftermarket Safety Device (ASD)*	5,850
MTA Fleet Equipped with ASD*	1,250
UPS Truck Equipped with ASD*	400
NYCDOT Fleet Equipped with ASD*	250
DSNY Fleet Equipped with ASD*	250
Vulnerable Road User (Pedestrians/Bicyclists) Device	100
PED Detection System	10 + 1 spare
Total Equipped Vehicles	**8,000**

*In addition, 600 spare ASDs will be purchased.

Applications to be deployed include Red Light Violation Warning, Pedestrian in Signalized Crosswalk Warning, Vehicle Turning Right in Front of Bus, Mobile Accessible Pedestrian Signal System (PED-SIG), six (6) Vehicle-to-Vehicle applications, and several speed compliance applications to help control speeds and enhance intersection and pedestrian safety.

Table 1-2 lists the applications to be deployed.

The New York City Department of Transportation leads this deployment effort.

Table 1-2. List of NYC CVPD Applications

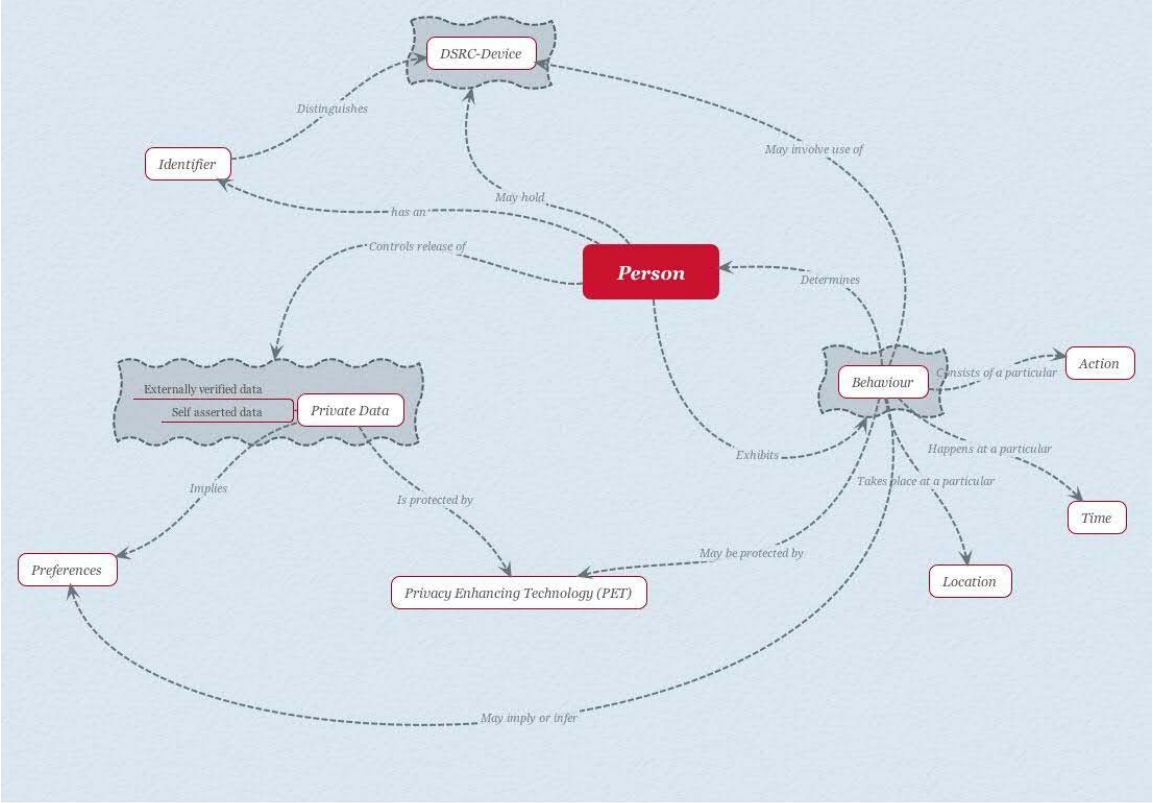
ID	Category	NYCDOT – CV Application
1	V2I/I2V Safety	Speed Compliance
2		Curve Speed Compliance
3		Speed Compliance/Work Zone
4		Red Light Violation Warning
5		Oversize Vehicle Compliance
6		Emergency Communications and Evacuation Information
7	V2V Safety	Forward Crash Warning (FCW)
8		Emergency Electronics Brake Lights (EEBL)
9		Blind Spot Warning (BSW)
10		Lane Change Warning/Assist (LCA)

ID	Category	NYCDOT – CV Application
11		Intersection Movement Assist (IMA)
12		Vehicle Turning Right in Front of Bus Warning
13	V2I/I2V Pedestrian	Pedestrian in Signalized Crosswalk
14		Mobile Accessible Pedestrian Signal System (PED-SIG)
15	Mobility	Intelligent Traffic Signal System (I-SIGCVDATA)

1.2.2 Simplified PII and Participant Analysis of NYCDOT Planning

1.2.2.1 Overview: General Model

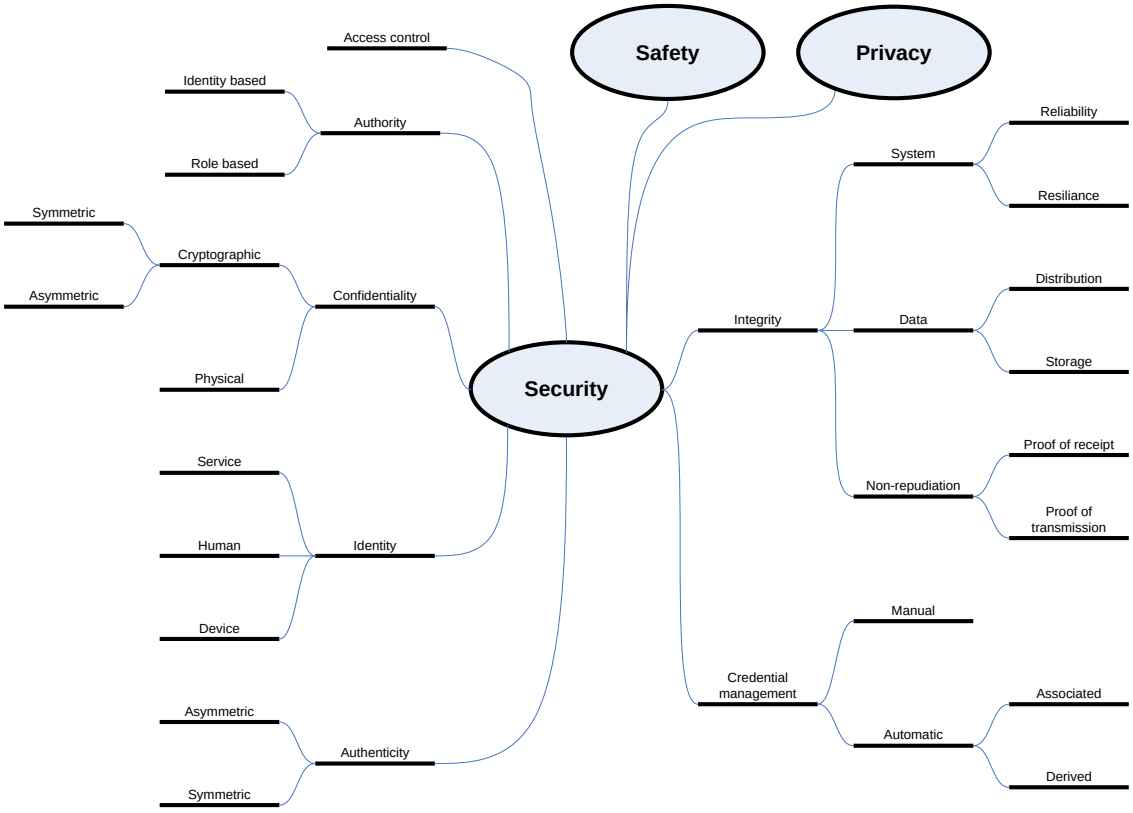
The generalized model of how data relates to privacy is examined in Figure 1-2 where PETs (Privacy Enhancement Technologies) may be used to assist in the protection of private data.



Source: NYCDOT, 2016

Figure 1-2. General Model for Link Between a Person, Their Behavior, and the Role of PETs

Figure 1-3, Security Relationships and Functions, shows privacy and safety as having relationships to security but not as part of security. Thus privacy can be protected by security functions if the data is visible to them. Note that data that is published is by definition public and not protected from exploit. The data in CV/ITS is not published in that sense but is broadcast for the "greater good" as it is intended to aid situational awareness.



Source: NYCDOT, 2016

Figure 1-3. Security Relationships and Functions

1.2.2.2 Participation and Consent of Users within Pilot

First, the term “users” needs to be clarified for the NYC pilot project.

User group "A" consists of the drivers of fleet vehicles for fleets (DOT, NYC, UPS, Taxi companies) that are participating in the CVPD. The project will be dealing with the fleet owners (vehicle owners) which are typically not the drivers. The fleet owner hires the drivers and the NYC project has no relationship with the drivers (except in the special case of the taxi owner/operator, which is not the group we have targeted for the taxi fleet). The project has no identifying ***PII or SPII*** related to these drivers.

The DSRC message set does not of itself contain data that is regularly classified as PII, and the DSRC enabled device will be bound with a registered vehicle rather than a user. The NYC CVPD project will not record the names, or any other identifying information, of any individual drivers. Given the open-ended remit of the pilot to give an insight to the use of CV/ITS technology in order to reduce

the impact of travel on users there may be significant post event data analysis to identify trends amongst vehicles. For maintenance purposes, equipment identification will be used, but equipment identities used for maintenance will not be combined with driver identities and, as stated above, NYC CVPD will not record driver identities.

Only fleet owners/operators can associate drivers to vehicles via schedules and their organizational-specific methods of identifying vehicles (e.g., VIN, license plate, etc.). While fleet owners can visually observe OBU identifying information (e.g., serial number) on installed equipment, they will not have access to any information regarding securely installed certificates.

The NYC CVPD program will maintain an inventory of what OBU is installed in which vehicle (by VIN or license plate), but will have no access to the drivers or driver schedules pertinent to the vehicles (these shall be maintained solely by their operating organization). Like the fleet owners, NYC CVPD will have no access to installed OBU certificates and therefore will not have an associative mechanism to connect them to vehicles.

We therefore do not consider it necessary to obtain consent from driver participants (group “A”).

User group "B" consists of visually impaired pedestrians using Personal Information Devices (PID) to aid their navigation of City streets. This user group is recruited, and members consent to using the application. Where consent is made on hard copy, the hard copy will be scanned and stored electronically in a secure enclave. The retained hard copies will be stored securely during the course of the New York pilot or stored until such time as a second, off-site, secure electronic backup is made of the scanned agreements. Group “B” users will be given subsequent access to verify or correct the consent agreements, as requested. User group “B” consent agreements will include terms that only allow NYC CVPD operators and researchers to perform research-oriented evaluation and analysis of the personally identifiable data as required by the Independent Review Board (IRB) according to USDOT’s guidance on Human Use Approval (HUA)[1]. Group “B” participants may be provisioned anonymous polls or surveys by the IE to improve research quality with regard to user experience. These survey instruments (including written surveys, polls, or interviews) will be developed and administered by the IE and participating researchers, as approved by the IRB. All PII collected by the IE in the course of this data collection will be protected by the IE and NYC CVPD according to the terms of the privacy agreements and IRB approval.

1.2.2.3 Participant Notice

Notice pertains to NYCDOT’s provision to provide participants in user group “B” with a clear and understandable set of the privacy risks associated with the NYC CVPD. Consent processes and forms will clearly detail the specific privacy risks associated with collecting and transmitting event data from PIDs. Clear notice will also be given of privacy risks to data post-collection, and the controls that are put in place to mitigate those risks.

1.2.2.4 Participant Data Use and Transparency

The NYC CVPD consent forms and consent processes for participants will include provisions to be informed of data use, data treatment, anonymization, obfuscation, security, onward transfer and storage policies. Data use policies, specifically, will be specified in consent documentation provided to individual and organizational participants. The data use policies will clearly describe the collection, use, dissemination and maintenance policies regarding the data, along with and distinctions regarding

how the data is used internally (by the Independent Evaluator) vs. externally (i.e., the research data exchange).

Transparency provisions will make clear the different levels of privacy protection afforded the data based on which parties have access to that data. Specifically, the Independent Evaluator will have access to some privacy-protected data (specifically, surveys and poll data from PID users) as a party to the privacy agreement, whereas the Research Data Exchange will not.

1.2.2.5 Participant Redress

The NYC CVPD consent forms and processes for participants will include provisions for participating individuals and organizations to:

1. Correct and update information they have provided
2. Seek clarification on data use and protection
3. Subsequently opt-out of the pilot, if requested

1.2.2.6 Numbers, Protocols, Messages

User group A. In 2015 there were 2.1 million vehicles registered in NYC with a further 8.9 million vehicles registered outside of NYC in the state of New York (From https://dmv.ny.gov/statistic/2015reginforce-web.pdf), and obviously there is a certain number of out of state vehicles using the NYC road network every day. The pilot is therefore relatively very small (compared to the overall number of potential ITS equipped vehicles) and this should be taken into account in any analysis of the overall results. One concern here is that with a small population of pilot users they may be easy to spot and isolate during the pilot phase as ITS transmissions will not be constant but only occur when devices pass. This contrasts with the long term goal of the ITS industry wherein devices will be ubiquitous and isolation of one device from millions of potential hosts becomes increasingly difficult without a lot of effort. The pilot assumes use of aftermarket safety devices that it is presumed implement the DSRC message set and operate in the reserved ITS frequency band. During the initial phase of the pilot such devices will be relatively rare and may be isolated by an attacker with appropriate equipment. The nature of the DSRC basic safety message, containing detail location and speed information, eases the isolation task when only small populations exist thus there may be a higher risk of violation of location privacy for a low device population than in a higher device population scenario.

We consider the privacy risk from this to be manageable, since the messages do not contain PII and cannot be associated with a particular driver by an eavesdropper. Additionally, vehicles will be provided with 50 BSM-signing certificates per week and will change the certificates periodically (approx.. once per 5 minute period), limiting an eavesdropper's ability to link together two transmissions from the same vehicle.

User Group B. The number of vulnerable road users represented by the pilot is quite low. The DSRC message related to Vulnerable Road User, the "MSG_PersonalSafetyMessage", is noted as experimental and support has to be mandated in the course of the trials. There is no explicit PII content in the Personal Safety Message (although the message does have the potential to contain significant behavioral data). The DSRC protocol will be implemented in a manner that the identity element of the Personal Safety Message is a pseudonym, and furthermore the pseudonym will be varied over time (as allowed in DSRC).

Note that vulnerable road users (VRUs) are not fully defined but globally the majority of definitions suggest that a VRU is anything without a vehicle protecting them, thus including pedestrians, cyclists, motorcyclists, horse-riders. The rationale is that in the event of a collision a person in a vehicle has the frame of the vehicle, airbags and features such as crumple zones to absorb and deflect any hit. A VRU has none of these features. Sources for definitions include:
http://ec.europa.eu/transport/themes/its/road/action_plan/its_and_vulnerable_road_users_en,
http://www.grsproadsafety.org/our-knowledge/safer-road-users/vulnerable-road-users, and
http://safety.transportation.org/doc/Vulnerable%20Users%20White%20Paper.pdf

1.2.2.7 Safety and Risk

The pilot program does not place any users (drivers, pedestrians, cyclists) at greater risk of death or injury. The pilot program will use a control group that is monitored to assist in the assessment of any results from the pilot. Specifically, the purpose of the control group is to gather results contemporaneous to the pilot "treatment" group such that any reduction in risk of death or injury can be compared against the control as well as against historic data.

The pilot program considers the use of co-operative ITS in the context of the connected vehicle only. The risk controller in such a scenario remains the driver, pedestrian, or cyclist using the equipment. There is a risk escalation in using DSRC as it may distract the user from the very thing that it is intending to reduce (see NHTSA report on distraction https://www.nhtsa.gov/risky-driving/distracted-driving). The DSRC message set does not give guidance on how to present messages to the affected driver even though the intent is to improve overall situational awareness of the affected drivers but it is expected that audio alerts will be given to drivers (the specific format of the audio may be device specific and will be configurable as required by the device specifications).

For analysis purposes at registration (see above), Group B users may be asked to offer an assessment of their reaction time and to make a declaration of any cognitive or physical disorders that may influence the results. This data may be made available to researchers. Group “B” users will be given the opportunity to grant consent to this via consent forms.

Chapter 2. Data Lifecycle

As noted in the New York City Connected vehicle requirements document, the evaluation data will be collected in the vehicle in "real time" by capturing the vehicle location, speed, heading, acceleration, steering wheel angle, and alerts/alarms in upon the occurrence of a "triggered" event. This data will be collected for a configurable time (~10 seconds) before the trigger and will continue to be collected for a configurable time after the event. Events can be configured and will include the issuance of an audible alert, but could also be related to accelerometer values, steering wheel angle, speed, etc. – i.e. any data known to the onboard unit. For analysis purposes, this data will include the received BSMs from nearby vehicles (configurable distance), and SPaT and MAP messages received from RSUs within a configurable range. All of this data constitutes the creation of an "event record". The event record is encrypted with the public key of the TMC and stored in the vehicle. Note that the maximum data collected is 5 minutes before the trigger and up to 5 minutes after the trigger, but this is expected to be closer to 10 seconds in each case such that the evaluation team can determine what triggered the alert (or event) and the driver reaction to that alert. For the Control group, no alert will be provided, but the ASD will use the same event conditions to trigger event data recording and the same data will be collected as for the non-control group, to be used for comparison to help determine the benefits of the CV technology.

When the vehicle passes an RSU which is equipped to upload the event logged data, the ASD will transmit the event log (in the encrypted form in which it was created) to the evaluation system at the TMC; during its transport, the data will remain encrypted and signed so that it cannot be intercepted or corrupted. When it reaches the TMC, it will be "batched" and processed to normalize and remove both the exact time and location (obfuscated) so that it cannot be matched with any external data source (e.g. police records) to make it identifiable to a specific vehicle or driver. If the normalization process fails, TMC personnel will be informed for troubleshooting and re-initialization purposes. Once the data has been processed, it will be:

1. made available to the independent evaluator
2. made available to the RDE
3. purged

Raw data will only be available to the normalization and obfuscation process for no more than 24 hours. If, for whatever reason the data fails to be sanitized and normalized in this period of time, the data will be purged.

The in-vehicle systems will only transmit their log data to the TMC server when they receive a properly authenticated certificate requesting the data, and once the TMC receives the data, it will transmit a "purge" request to the vehicle system and the data will be purged from all logs.

In addition, the in-vehicle systems will contain a configurable maximum data age such that data will be purged, based on the date in the log file, if power is present or when power is restored and life of the data is older. This is intended to protect the privacy of the data in the event the vehicle is involved in a crash or is impounded or out of service for an extended period of time. Date/time information settings are to be strongly access-controlled during and between power-up events in the in-vehicle systems. Additionally, log data is to be integrity protected to prevent log time forgeries.

Chapter 3. PII Content in SAE J2735, J2945 Message Set

3.1 Overview of DSRC Message Set

The co-operative ITS message set to be used in the New York Connected Vehicle Pilot is defined in SAE J2375 and J2945/1, the PII content of which is discussed in this section as it pertains to each of the applications addressed in the NYC pilot. A high-level view of the PII content of the DSRC message set is given here along with a description of each application and the type of PII data that may be collected in each. See section 5.3 and Table 5-1 for more information on each application's data obfuscation.

3.1.1 Basic Safety Message

The core message for the safety goal of the pilot is the DSRC Basic Safety Message that always contains information relating to the vehicle position, and some of its dynamic attributes, as well as a temporary identifier.

```
BSMcoreData ::= SEQUENCE {
    msgCnt              MsgCount,
    id                  TemporaryID,
    secMark             DSecond,
    lat                 Latitude,
    long                Longitude,
    elev                Elevation,
    accuracy            PositionalAccuracy,
    transmission        TransmissionState,
    speed               Speed,
    heading             Heading,
    angle               SteeringWheelAngle,
    accelSet            AccelerationSet4Way,
    brakes              BrakeSystemStatus,
    size                VehicleSize
}
```

Vehicle BSMs are collected as raw data in a variety of the connected vehicle applications under study. Vehicle IDs ("id") are temporary identifiers provided in each BSM, which are changed from time to time.

See section 5.3.1 for more information on how vehicle identification information ("id") and vehicle size ("size") are further obfuscated in preparation for data analysis.

See section 5.3.3 for more information on how location data (lat, long) is obfuscated for subsequent data analysis.

3.2 V2I/I2V Safety

3.2.1 Speed Compliance

The Speed Compliance application involves making speed limit information available to the in-vehicle systems.

If the driver is exceeding the speed limit as stated in the DF_RegulatorySpeedLimit data field this may trigger an alert condition, leading to an alert being raised for users outside the Control Group and to event data being logged as described in overview in Chapter 2. No PII is collected; the BSMs prior to and immediately after the notification are collected and will be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.2.2 Curve Speed Compliance

The Curve Speed Compliance application involves making speed limit and curve geometric information available to the in-vehicle systems.

If the driver is exceeding the speed limit as stated in the DF_RegulatorySpeedLimit data field this may trigger an alert condition, leading to an alert being raised for users outside the Control Group and to event data being logged as described in overview in Chapter 2. No PII is collected; the BSMs prior to and immediately after the notification are collected and will be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.2.3 Speed Compliance/Work Zone

The Speed Compliance/Work Zone application involves making speed limit and work zone geometry (roadway affected) information available to the in-vehicle systems.

If the driver is exceeding the speed limit as stated in the DF_RegulatorySpeedLimit data field this may trigger an alert condition, leading to an alert being raised for users outside the Control Group and to event data being logged as described in overview in Chapter 2. No PII is collected; the BSMs prior to and immediately after the notification are collected and will be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.2.4 Red Light Violation Warning

The Red Light Violation Warning application uses a combination of the SPaT and MAP information transmitted by the intersection it is approaching to determine if the vehicle is or is likely to violate the stop line at the intersection based on the vehicle location and vehicle kinematics.
The vehicle may determine that it will not be able to stop at an intersection in time. This will count as a loggable event. In this case the vehicle will transmit the V2V/V2I message containing the element VehicleEventFlags set to indicate "eventStopLineViolation", may alert the driver if the driver is not in the Control Group, and log the event data as described in overview in Chapter 2. The BSMs, SPaT and MAP messages for the subject vehicle will be collected immediately prior to the alert message (or when it would have been sent if the system is in silent mode) and following the alert message and will

be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.2.5 Oversize Vehicle Compliance

The onboard unit is configured with the vehicle's type and height and receives a TIM, or MAP message (TBD) that provides the description of the roadway the vehicle is approaching from a local RSU.

The onboard systems then use this combined information to determine if a loggable event condition is occurring. In this case the vehicle may alert the driver if the driver is not in the Control Group, and will log the event data as described in overview in Chapter 2. The BSMs, SPaT and MAP messages for the subject vehicle will be collected immediately prior to the alert message (or when it would have been sent if the system is in silent mode) and following the alert message and will be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.2.6 Emergency Communications and Evacuation Information

This is a broadcast message from the infrastructure to the vehicles and no PII is collected.
If the vehicle is within the geographic region of relevance for the message, this counts as a loggable event. In this case the vehicle may alert the driver if the driver is not in the Control Group, and will log the event data as described in overview in Chapter 2. The BSMs, TIM, and MAP messages for the subject vehicle will be collected immediately prior to the alert message (or when it would have been sent if the system is in silent mode) and following the alert message and will be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.3 V2V Safety

NOTE 1: Current Co-operative ITS messaging schemes do not engage in a dialogue between vehicles, but provide broadcast of vehicle status that is received and may be interpreted for the driver.

NOTE 2: Responsibility for avoidance of crashes and safe change of direction remains with the driver. The C-ITS/DSRC messaging is viewed therefore as an aid for the driver only.

NOTE 3: Because of the initial testing and the USDOT requirement to support a "control group", some vehicles will not issue an alert to the driver; this does not affect the operation of the application or the logging; they both occur as though it was an active alert and will be recorded, but not presented to the driver.

NOTE 4: The nature of the data collected ("event" data) and the contents of the event record are described above.

3.3.1 Forward Crash Warning (FCW)

Upon receipt of DSRC status messages from vehicles the subject vehicle can algorithmically determine the potential of colliding with either a vehicle or infrastructure ahead. If this potential passes a particular threshold it is considered a loggable event situation: if the driver is not in the Control Group an alert is raised to enable them to take evasive action.

Loggable events are logged as described in overview in Chapter 2. No PII is collected; the BSMs prior to and immediately after the notification are collected and will be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.3.2 Emergency Electronics Brake Lights (EEBL)

When braking severely/hard (defined as greater than 0.4g of deceleration) the braking vehicle transmits the DE_VehicleEventFlags with the event eventHardBraking set (depending on circumstance additional events may be transmitted alongside this and include eventABSactivated and eventTractionControlLoss, noting that if traction is lost the braking may no longer exceed the hard braking limit but the loss of control may be significant, similarly if traction is lost and ABS is activated the actual braking force may not be sufficient to engage the eventHardBraking flag although the braking may exceed the safe limit for the surface conditions).

If this results in a receiving vehicle raising an alert, the ConOps states that data around the time of the alert is gathered and reported to the TMC. No PII is collected; the BSMs prior to and immediately after the notification are collected and will be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.3.3 Blind Spot Warning (BSW)

If this results in a receiving vehicle raising an alert, the ConOps states that data around the time of the alert is gathered and reported to the TMC. No PII is collected; the BSMs prior to and immediately after the notification are collected and will be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.3.4 Lane Change Warning/Assist (LCA)

If this results in a receiving vehicle raising an alert, the ConOps states that data around the time of the alert is gathered and reported to the TMC. No PII is collected; the BSMs prior to and immediately after the notification are collected and will be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.3.5 Intersection Movement Assist (IMA)

If this results in a receiving vehicle raising an alert, the ConOps states that data around the time of the alert is gathered and reported to the TMC. No PII is collected; the BSMs prior to and immediately after the notification are collected and will be encrypted and processed in a manner that eliminates the

possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.3.6 Vehicle Turning Right in Front of Bus Warning

If this results in a receiving vehicle raising an alert, the ConOps states that data around the time of the alert is gathered and reported to the TMC. No PII is collected; the BSMs prior to and immediately after the notification are collected and will be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location.

3.4 V2I/I2V Pedestrian

3.4.1 Pedestrian in Signalized Crosswalk

If this results in a receiving vehicle raising an alert, the ConOps states that data around the time of the alert is gathered and reported to the TMC. This application does not require the OBE to transmit data (it receives only from the RSE). No PII is collected; the transmitted RSE messages received by the OBE prior to and immediately after the notification are collected and will be encrypted and processed in a manner that eliminates the possibility that the data collected in conjunction with the driver alert can be used with other external sources of data to recreate the event in terms of time and location. Note that the alert will result from the calculations of the potential "collision" with the pedestrian at the crosswalk.

3.4.2 Mobile Accessible Pedestrian Signal System (PED-SIG)

This application collects data within the Pedestrian device that indicates that the pedestrian queried the application and was provided information regarding the crossing and the status of the pedestrian signal. Such data will be transmitted to the TMC for analysis to determine the number of general events encountered and used to help assess the utility to the sight challenged pedestrians. The sensitivity of this data and its protection/obfuscation are identified in Table 5-1. Degree of Obfuscation by CV Application Warning.

3.5 Mobility

3.5.1 Intelligent Traffic Signal System (I-SIGCVDATA)

No privacy related data is collected for this application. Periodically, the system identifies vehicles as they come within the range of the RSU which records their temporary ID and their current location and sends the data to the TMC where it can be compared with other "sightings" to determine travel times for the vehicles as long as the ID remains the same. This will provide statistical data since the vehicle will be changing its ID and certificates hence limiting its ability to be tracked; the goal of this application is to measure block to block travel times which are then used by the adaptive control algorithms (ACDSS) to select the timing plan. In this instance, the NYC CV pilot is trying to determine if the limited CV data is sufficient to drive the adaptive control system and can then be used to replace the existing RFID readers with appropriate savings and increased accuracy.

Since the system is based on privacy by design and does not use any other aspects of the RSU identity, there are no issues with the privacy of the data.

Chapter 4. Data Privacy Plan

4.1 Overview of Technical Methods Proposed for Use

4.1.1 Specific Measures Identified in SP 800-122

4.1.1.1 De-identification

When event data is passed from the TMC to other users it will be stripped of identifiers, including BSM temporary ID and any certificates. The following location information emanating from the BSMs will be obfuscated as described in Section 5.3.3:

- lat
- long
- elev

The following telematics data from the BSM will be preserved to identify driver actions and responses as a function of time:

- transmission
- speed
- heading
- angle
- accelSet
- brakes
- size

'Size' will be generalized to further obfuscate and reduce correlation potential (see section 5.3.1).

Per section 5.3.3, the following BSM data will be removed and further obfuscated in the event record to reduce attacker correlation potential:

- id (temporary ID)

No data from the DSRC message (containing the BSMs) security envelope will be included in the event record. This includes:

- Signer certificate

4.1.1.2 Anonymization

Anonymization mechanisms identified in NIST SP 800-122 include

- Generalizing the data
- Suppressing the data

- Introducing noise into the data
- Swapping the data
- Replacing data with the average value

We do not propose to use any of these approaches. Our assessment is that the de-identification approach approved above is enough to provide sufficient privacy protection.

4.1.1.3 Security Measures

NIST SP 800-122 identifies the security measures enumerated in Table 4-1.

Table 4-1. NIST SP 800-122 Security Measures

Measure	Role from SP 800-122	Integration in NYC CV Pilot
Access Enforcement (AC-3)	Organizations can control access to PII through access control policies and access enforcement mechanisms (e.g., access control lists).	The raw data is accessible only to the TMC. The TMC sanitizes the data before making it available to the RDEs.
Separation of Duties (AC-5)	Organizations can enforce separation of duties for duties involving access to PII. For example, the users of de-identified PII data would not also be in roles that permit them to access the information needed to re-identify the records	The data will not be stored for more than 24 hours within any server under City control to avoid being subject to FOIA requests or subpoena. Note that with the exception of the pedestrian applications – the CV pilot does not have any user information other than the ASD serial number and the vehicle to which the ASD was attached. Any records of the relationship between the driver and the vehicle are only known to the vehicle owner and is not included in any information stored within the CV system.
Least Privilege (AC-6)	Organizations can enforce the most restrictive set of rights/privileges or accesses needed by users (or processes acting on behalf of users) for the performance of specified tasks. Concerning PII, the organization can ensure that users who must access records containing PII only have access to the minimum amount of PII, along with only those privileges (e.g., read, write, execute) that are necessary to perform their job duties	
Remote Access (AC-17)	Organizations can choose to prohibit or strictly limit remote access to PII. If remote access is permitted, the organization should ensure that the communications are encrypted	See Access Enforcement

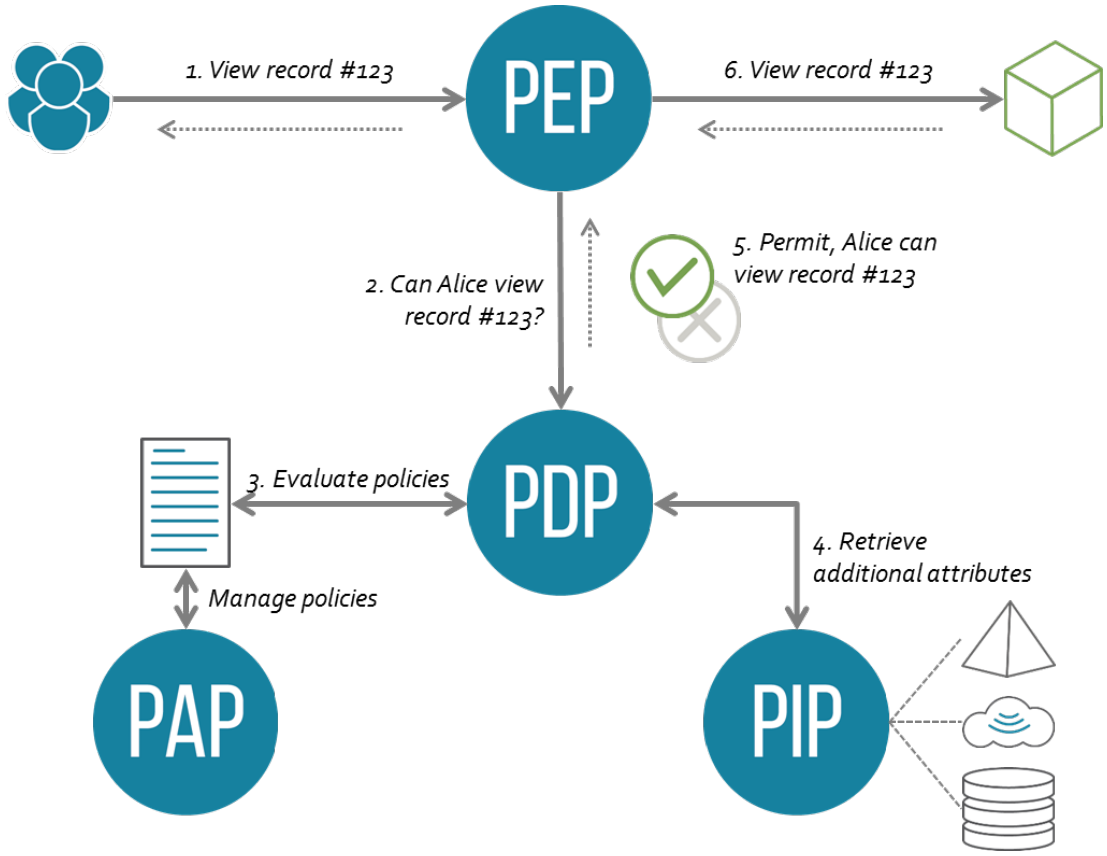
Measure	Role from SP 800-122	Integration in NYC CV Pilot
User-Based Collaboration and Information Sharing (AC-21)	Organizations can provide automated mechanisms to assist users in determining whether access authorizations match access restrictions, such as contractually-based restrictions, for PII	No user-based information sharing, access to information is configured by an administrator
Access Control for Mobile Devices (AC-19)	Organizations can choose to prohibit or strictly limit access to PII from portable and mobile devices, such as laptops, cell phones, and personal digital assistants (PDA), which are generally higher-risk than non-portable devices (e.g., desktop computers at the organization's facilities).	No mobile device access to PII
Auditable Events (AU-2)	Organizations can monitor events that affect the confidentiality of PII, such as unauthorized access to PII	The database in which event records are stored shall maintain an access log
Audit Review, Analysis, and Reporting (AU-6)	Organizations can regularly review and analyze information system audit records for indications of inappropriate or unusual activity affecting PII, investigate suspicious activity or suspected violations, report findings to appropriate officials, and take necessary actions	The access log will be available to administrators and to independent evaluators
Identification and Authentication (Organizational Users) (IA-2)	Users can be uniquely identified and authenticated before accessing PII.	Identity of users, and the authentication state of the identity, shall be used as attributes in access control policy. The system shall enforce that only certain users have access to the data on the TMC, and that only certain remote entities (the IE) have access to raw event data
Media Access (MP-2)	Organizations can restrict access to information system media containing PII, including digital media (e.g., CDs, USB flash drives, backup tapes) and non-digital media (e.g., paper, microfilm). This could also include portable and mobile devices with a storage capability	Physical access to the media shall be restricted; the PII database shall be subject to the usual backup policy but the backup shall be encrypted and physically isolated so that it can only be restored to an authorized machine and cannot be read directly.

Measure	Role from SP 800-122	Integration in NYC CV Pilot
Media Marking (MP-3)	Organizations can label information system media and output containing PII to indicate how it should be distributed and handled. The organization could exempt specific types of media or output from labeling so long as it remains within a secure environment. Examples of labeling are cover sheets on printouts and paper labels on digital media	When data is retrieved the policy enforcement engine will determine the classification based on the attribute set in the policy and mark the results accordingly. Cryptographic watermarking of digital records with the classification may be employed if suitable methods are identified and the watermark used as part of access the control mechanism (may be used to prohibit the devices used to access the data). In addition the persistent data stores (disks) used to contain PII shall be physically isolated where possible (if cloud stores are used all stores shall hold data in encrypted form only irrespective of the presence of PII or not).
Media Storage (MP-4)	Organizations can securely store PII, both in paper and digital forms, until the media are destroyed or sanitized using approved equipment, techniques, and procedures	Where PII is stored digitally (and it is proposed that only digital stores are maintained) the media shall be encrypted during its active life and sanitized in accordance with appropriate FIPS processes at the end of life.
Media Transport (MP-5)	Organizations can protect digital and non-digital media and mobile devices containing PII that is transported outside the organization‘s controlled areas	All media labelled (see above) as containing PII shall be transported in secure (locked) containers. In addition if data is transported electronically it shall be protected by transit encryption schemes (e.g. TLS).
Media Sanitization (MP-6)	Organizations can sanitize digital and non-digital media containing PII before it is disposed or released for reuse	Data sanitization shall be implemented in accordance with latest FIPS recommendations
Transmission Confidentiality (SC-9)	Organizations can protect the confidentiality of transmitted PII	As indicated above all PII transmitted shall be encrypted when transmitted
Protection of Information at Rest (SC-28)	Organizations can protect the confidentiality of PII at rest, which refers to information stored on a secondary storage device, such as a hard drive or backup tape	As indicated above all media used to hold data classified as PII shall be encrypted.

Measure	Role from SP 800-122	Integration in NYC CV Pilot
Information System Monitoring (SI-4).	Organizations can employ automated tools to monitor PII internally or at network boundaries for unusual or suspicious transfers or events	In the pilot phase all data access shall be logged in order to build up a picture of normal transfers and events. This will be used to determine the base line for identification of unusual behavior which shall be analyzed in real time, and in analysis of logs.

4.1.2 Access Control Measures

All data required for the pilot to operate shall be retained in access controlled elements. This shall apply to instances of classes as active objects, to databases and to server elements maintained in either fixed locations or distributed using cloud services as illustrated in Figure 4-1.



Source: Wikipedia, 2016 (https://en.wikipedia.org/wiki/XACML)

Figure 4-1. Example Policy Based Access Control Architecture

4.1.3 Specific Privacy Protection Measures

4.1.3.1 Pseudonymity

Pseudonymity ensures that a user may use a resource or service without disclosing its user identity, but can still be held accountable for that use. A pseudonym is an identifier allocated by an authority to a single entity or group of entities and which bears no relation to the true identity of the entity or group. In this way, it is only the authority that is able to resolve a pseudonym to a true identity. By changing pseudonyms on a regular basis the real identity can also be protected from behavioral analysis attacks.

4.1.3.2 Unlinkability

Unlinkability ensures that a user may make multiple uses of resources or services without others being able to link these uses together. This means that within the pilot from the perspective of an unauthorized party, users and their actions in the pilot are no more and no less related after an observation than they are related concerning the a-priori knowledge. Therefore, the probability of particular actions being related to particular pilot users remains the same after an observation as it was before.

The unlinkability of two (or more) messages may depend on whether their content is protected against a particular attacker. Messages may be considered to be unlinkable if the attacker is unable to acquire information on the sender or recipient of a message by analyzing the message. Nevertheless, even simple analysis of the contents of a number of messages can reveal certain characteristics which link them together.

Pseudonyms may serve as a basis for unlinkability but their use does not, on its own, guarantee that any link between users and their behavior will be hidden from an attacker.

4.2 Overview of Policy and Management Methods Proposed for Use to Support Safe Use and Management of (S)PII

4.2.1 Frameworks and Compliance

It is broadly assumed that the CVPD program shall act in accordance with the US Federal Information Security Management Act (FISMA) (http://csrc.nist.gov/drivers/documents/FISMA-final.pdf) and the accompanying FISMA Implementation Project. In addition the NYCDOT is expected to implement security management standards in close compliance to the ISO/IEC 27000 series of standards.

Note that the FISMA Implementation Project was established in January 2003 to produce several key security standards and guidelines required by Congressional legislation. These publications include FIPS 199, FIPS 200, and NIST Special Publications 800-53, 800-59, and 800-60. Additional security guidance documents are being developed in support of the project including NIST Special Publications 800-37, 800-39, and 800-53A. It should be noted that the Computer Security Division continues to produce other security standards and guidelines in support of FISMA. These publications can be located by visiting the division's Publications page at: http://csrc.nist.gov/publications/.

4.2.2 Public Access to Data

Data gathered from the CV pilot under state ownership will be made accessible to the public after appropriate sanitization. Data shall be anonymized as far as is technically possible and shall always be maintained using the key concept of pseudonymization, i.e. real identities shall be suppressed, and re-linking of the pseudonym to the real identity shall be designed in such a way that it is not readily achieved without oversight of a trusted (human) operator.

4.3 Simplified Privacy Impact Assessment (PIA)

Following the guidance given in SP 800-122, a simplified Privacy Impact Assessment has been conducted, with the results given in Table 4-2.

Table 4-2. Privacy Impact Assessment (PIA)

Id	PIA Question	Resolution
Technology		
(1)	Does the project apply new or additional information technologies that have substantial potential for privacy intrusion?	YES. The field of CV is relatively new and the DSRC message set declares the location of individuals (or their vehicles) and the nature of their transport behavior. Thus an attacker able to capture sufficient data may be able to collate it to an individual and thus PII is revealed. Measures planned to pseudonymize transmissions, to rotate pseudonyms, and to restrict the binding of transmissions to individuals will minimize the risk of PII leakage.
Identity		
(2)	Does the project involve new identifiers, re-use of existing identifiers, or intrusive identification, identity authentication or identity management processes?	YES. The CV equipment used is new, and is newly identified. There is some overhead in the management of pseudonyms to minimize the privacy exposure.
(3)	Might the project have the effect of denying anonymity and pseudonymity, or converting transactions that could previously be conducted anonymously or pseudonymously into identified transactions?	NO. Measures in place starting from the base protocols assure pseudonymity.
Multiple Organizations		
(4)	Does the project involve multiple organizations, whether they are government agencies (e.g. in 'joined-up government' initiatives) or private sector organizations (e.g. as outsourced service providers or as 'business partners')?	YES. More than one organization is involved. Data able to isolate an individual is not shared between them.

Id	PIA Question	Resolution
Data		
(5)	Does the project involve new or significantly changed handling of personal data that is of particular concern to individuals?	NO. Whilst the project gathers location and behavioral data it is not linked to an individual and is not retained in the system for more than a few seconds in order to achieve the ITS CV processing goal.
(6)	Does the project involve new or significantly changed handling of a considerable amount of personal data about each individual in the database?	NO. As above.
(7)	Does the project involve new or significantly changed handling of personal data about a large number of individuals?	NO. As above.
(8)	Does the project involve new or significantly changed consolidation, inter-linking, cross-referencing, or matching of personal data from multiple sources?	NO. As above.
Exemptions and Exceptions		
(9)	Does the project relate to data processing which is in any way exempt from legislative privacy protections?	NO.
(10)	Does the project's justification include significant contributions to public security measures?	NO.
(11)	Does the project involve systematic disclosure of personal data to, or access by, third parties that are not subject to comparable privacy regulation?	NO.

4.4 Checklist Summary

The checklist of DPP considerations from NIST SP 800-122 is presented in Table 4-3.

Table 4-3. Checklist of NIST SP 800-122 DPP Considerations

Checklist Question	DPP Consideration
Has your organization ever performed work for a Federal agency that involved handling PII?	NO, While New York City has, the NYC Department of Transportation has not dealt with PII for a federal agency.
Does your organization have any policies/procedures to protect the security and confidentiality of PII?	YES As outlined in the present document all data, not restricted to just those data elements likely to be explicit PII, are made available only on the basis of "need to know" where the need to know is determined from the role in the organization, the nature of the information requested, and additional contextual information (e.g. the location of the requesting party, the technical means of accessing the data).
Does your organization have any policies/procedures to control and limit access to PII?	YES As outlined in the present document all data, not restricted to just those data elements likely to be explicit PII, are made available only on the basis of "need to know" where the need to know is determined from the role in the organization, the nature of the information requested, and additional contextual information (e.g. the location of the requesting party, the technical means of accessing the data).
Does your organization store PII on network drives and/or in application databases with proper access controls (i.e., User IDs/passwords)?	YES All data is protected by a set of access control policies both for network drives, cloud-based systems, and databases.
Does your organization limit access to PII only to those individuals with a valid need to know?	YES. As indicated above.
Does your organization prohibit or strictly limit access to PII from portable and mobile devices, such as laptops, cell phones, and personal digital assistants (PDA), which are generally higher-risk than non-portable devices (e.g., desktop computers at the organization‘s facilities)?	YES. As indicated above the policies used to determine if an individual is permitted access to data includes assessment of the context of the request and this includes the nature of the device used and the location it is used in.
Does the information system used by your organization to store PII contain automated or easy-to-use process to ensure that only authorized users access PII – and only to the extent that each user has been authorized to do so?	YES All access control decisions are logged and processes to perform access verification are performed both in real time (the access control engine) and in behavioral analysis of the log files. The log files, as they themselves contain data that may be PII, are protected in like manner to the core data of the organization.

Checklist Question	DPP Consideration
Does your organization monitor events that may affect the confidentiality of PII, such as unauthorized access to PII?	YES. As noted above all actions are logged and the logs subject to review for verification of correct application of authorization policy.
Does your organization audit its information systems on a regular or periodic basis?	YES. Internal and external audit in line with NYC DOT IT policy
Does your organization analyze information system audit records for indications of inappropriate or unusual activity affecting PII, investigate suspicious activity or suspected violations, report findings to appropriate officials, and take necessary actions?	YES in line with NYC DOT IT policy
Does your organization restrict access to information system media containing PII, including digital media (e.g., CDs, USB flash drives, backup tapes) and non-digital media (e.g., paper, microfilm)?	YES. External ports such as USB port, on RSEs and OBEs will be sealed.
Does your organization restrict access to portable and mobile devices capable of storing PII?	YES in line with NYC DOT IT policy
Does your organization require that information system media and output (such as printed documents) containing PII be labeled to indicate appropriate distribution and handling?	YES. All information output from the information system is labeled to indicate the level of sensitivity of the data (covering commercial and public sensitivity levels).
Does your organization securely store PII, both in paper and digital forms, until the media are destroyed or sanitized using approved equipment, techniques, and procedures?	YES Where PII is retained it is subject to access control restriction and where appropriate, and where technology allows, is maintained in an encrypted form.
Does your organization sanitize digital and non-digital media containing PII before disposing of or reusing the media? **NO YES** (if **YES**, please explain or attach relevant policy)	YES. The organization sanitizes media and disposes of data in accordance with NIST Special Publication 800-88r1 (http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-88r1.pdf)

Chapter 5. Review of USDOT Data Privacy Policy

5.1 Categories of Records Collected.

The following forms of personal or potentially information about individual participants and/or their motor vehicle and motor vehicle use will be collected. We present the list first in the order provided by DOT, then in the order provided by the Concept of Operations.

5.1.1 DOT Ordering

- **Participant Background Information**
  - Individual Identifiers – only for PED application
  - Full Name (First, Middle, Last) – only for PED application
  - Demographic information, including age and gender – only for PED application
  - Individual subject research identifier created by DOT – ASD is used for surrogate for vehicle ID – owner will provide any ID to be used for matching purposes.
  - Driver's license number, issuing state, and qualifiers – None.
- **Vehicle Identifiers**
  - Vehicle Identification Number (VIN) of government issued vehicles – YES
  - Identifiers for equipment installed by DOT in personal or government issued vehicle – YES
- **Contact Information**
  - Mailing/Residential Address – Fleet owner, PED user
  - Phone number(s) – Fleet owner, PED user
  - Email address(es) – Fleet owner, PED user
  - Institutional or organizational affiliation – Fleet owner
  - Work/Business related contact information – Fleet owner
  - Occupation and work schedule – Vehicle schedule obtained from
- **Project Information**
  - Vehicle sensor information – YES
  - Video or still images, including infrared – YES
  - Audio recordings – YES
  - Dynamic information about a vehicle, including location, heading – YES
  - Proximity to and interaction with other vehicles and infrastructure – YES

- Dynamic information about a driver's interaction with the vehicle, including steering wheel, turn signal, and accelerator and brake pedal positions – YES
- Data collected from drivers by means of surveys, focus groups, or interviews – collected anonymously for vehicle drivers, targeted for PED.

5.1.2 NYC Ordering

- **ASD**
  - CAN Bus:
    - o directional signals
    - o hard braking
    - o steering wheel angle
    - o trajectory
    - o speed
  - The ASDs will use UTC (Coordinated Universal Time) time, and will include accelerometers (X, Y, Z) that can be used to detect vehicle actions and movements
- **ASD actions logs will record details of vehicle movements and the surrounding conditions surrounding a CV app warning event**
  - Details regarding the CV app which generated the warning issued
  - BSM transmitted message content of the subject vehicle (10 BSMs per second),
  - BSM content received from other CV-equipped vehicles within a configurable range of the subject vehicle (10 BSMs per second, per surrounding vehicle),
  - SPaT & MAP messages received from RSUs within a configurable range of the subject vehicle, and
  - Vehicle data available from the vehicle CAN bus (e.g., directional signals, hard braking, steering wheel angle, trajectory, speed, etc.).
  - The ASD shall log the SAE J2735 BSMs (each transmitted at 10 per second per vehicle) before and after an event.
  - The ASD shall begin logging the information 10-20 seconds before and complete logging 20-50 seconds after the event.
  - The ASD shall collect less detailed CV probe data for mobility data collection.
- **The less detailed CV breadcrumb data will be logged continuously and will be used for mobility data collection**
  - the BSM data from each CV vehicle will be logged in a less detailed manner and will be used to calculate travel times over the study roadways. This BSM 'breadcrumb' data will be logged in a pre-determined time interval (likely between 1 and 5 seconds) or possible using a distance based measure (e.g. every 100 feet) and continuously for each CV equipped vehicle. This data would not include the detailed BSMs from surrounding vehicles (since they themselves would have their own ASD breadcrumb recording).
- **After uploading to the CVPD servers (details follow), all locally stored action logs on the ASDs will be deleted.**
- **PED ASD:**
  - Ideally details similar to the action log data trajectories will be collected from the pedestrian ASD units, but details are still being developed and will be documented when

more certainty exists. The same data privacy, data obfuscation, and IRB issues that exist with the in-vehicle ASDs also exist with pedestrian units, and similar solutions to ensure the safety and privacy of the pedestrians is maintained. Additionally, however, information pertaining to the pedestrian ASDs will also include participants' personal contact information in accordance with informed consent forms. This is not the case with vehicular ASDs.

- **RSU:**
  - RF measurement evaluation data
  - The RSU shall record the BSM data from the Host Vehicle (HV).
- **Non CV data:**
  - Travel Time monitoring via bus GPS datasets: through the MTA Bus Time system, the route, position, and status of all MTA buses in service are available at 30-second intervals. While these data are made available in real-time via an online feed for app developers, NYCDOT also receives a daily flat file of all records each night. As with the taxi breadcrumbs, successive breadcrumb data points can be used for point-to-point travel time for a 30-second interval.
  - Taxi GPS trip records and breadcrumb data are supplied to NYCDOT from TLC on a monthly basis. Average speed is only available from the trip data for the overall trip (trip distance / trip time).
  - Travel time monitoring from Midtown in Motion (MIM) and other regional tag readers: the overall coverage of MIM's real-time travel time monitoring system extends from 1st Avenue to the east, 11th Avenue to the west, 23rd Street to the south and 57th Street to the north, which includes monitoring of segment travel times in real-time. Segment travel times are observed through deployment of RFID tag readers at fixed locations at major cross streets: 57th, 49th, 42nd, 34th, and 23rd Streets. Other regional agencies have also deployed tag readers on fixed highway segments and facilities for travel time monitoring. These data are compiled regionally by TRANSCOM (xcm.org) for query by its coalition agencies, including NYCDOT.
  - Volume monitoring from permanent count stations through the study area, toll collection counts, etc.: traffic counts are performed at varying frequencies, types, and durations within the CV Pilot area. Volume counts are regularly performed at fixed toll facilities (bridges, tunnels) and New York State DOT continuous count station, which can be used to assess system-wide – or zonal – changes in vehicular demand. Localized counts are typically one-off, performed for various planning projects. Both can be used for volume measures during the pilot but may not have the spatial and temporal completeness needed for the entire CV Pilot study area. In addition, the exact location, type and duration of localized counts on the study corridors for other projects are yet unknown.
- **Other non-CV data that are not available in real-time would be stored separately**. The following data will be collected and stored in real-time, and assigned based on 15-minute intervals to corresponding event logs at as data is uploaded from the ASDs to the TMC servers (anticipated to be at least daily, depending on the ASD usage and location relative to RSUs providing such uploading capabilities).
  - Traffic count data (if available) at 15-minute intervals
  - Midtown in Motion segment travel time data (if available) at 15-minute intervals
  - NWS weather data at 1-hour intervals
  - DSNY plow data (when applicable) at 15-minute intervals

  - TRANSCOM traffic incident data on study corridors at 15-minute intervals

- **The following data cannot be gathered in real-time, and instead will be stored external to the ASD Action Log data:**
  - Taxi activity and MTA Bus Time data
  - Crash data
  - Summary log/calendar of special events and related street closures
  - Summary log/calendar of short- and long-term work zone presence
  - External project related changes (Vision Zero, transit, reconstruction and signal retimings)
- **Driver:** the aggregate change in driver behavior will not only be assessed, but the different ranges of response (including no response or no change) will be tracked to arrive at a distribution of the driver responses in reaction to the deployed CV technology.

5.2 Required Privacy Controls

We will apply the following privacy controls throughout the data lifecycle:

- **Collection of PII**
  - Collect only PII that the researcher has been authorized to collect by USDOT.
  - Collect the minimum PII required for the research and not more.
- **Notice to Human Subjects**
  - Provide appropriate advanced notice, if at all possible at the point of collection, to the individuals from whom the PII is being collected. -- only for PED
  - Obtain advanced approval for the notice from the USDOT Contracting Officer.
- **Use and Sharing of PII**
  - Ensure that Recipient personnel acknowledge PII responsibilities to ensure that PII is used only as authorized.
  - Ensure that notice is given to participants, prior to consenting, concerning which organizations may have access to the PII.
  - Ensure that any changes to the list of authorized PII-accessing organizations requires notification to participants such that they may re-authorize consent.
  - Ensure that data destruction policies are followed on any PII belonging to a participant who has removed consent.
  - Retain only the PII that is necessary for the purposes of the NYC CVPD as authorized by USDOT
  - Not use PII for purposes other than those authorized by USDOT.
  - Ensure that access to PII is on a “need to know” basis for authorized purposes only.
  - Not exceed authorized access to PII, or disclose PII to unauthorized persons.
- **Security**
  - Protect all PII, electric or hardcopy, in their custody from unauthorized disclosure, modification, or destruction so that the confidentiality, integrity, and availability of the information are preserved.

- Store PII only on IT infrastructure employing security controls commensurate with the risk to the individual that would result from unauthorized access, disclosure, or use of the information.
- Encrypt all PII in transit or at rest.
- Encrypt all PII transmitted or downloaded to mobile computers/devices.
- Ensure that all individuals having access to PII have received training in the policies and procedures that protect PII.

- **Maintenance and Disposal**
  - Maintain PII in accordance with the applicable National Archives and Records Administration (NARA) records schedule (available from the USDOT Contracting Officer).
  - After conclusion of the research project, maintain PII only as permitted by the NARA schedule and, in the case of contractor-conducted research, relevant data rights classes in the applicable contract. Retention of PII that may be necessary for continued routine operations may be permitted (e.g., registration and account information).

- **Privacy Documentation**
  - Document compliance with the provisions of the Recipient's Data.
  - Privacy Plan and the Data Privacy and Security provisions in the Grant Agreement.
  - Upon request, provide to the USDOT Contracting Officer sufficient documentation to demonstrate compliance with the Recipient's Data Privacy Plan and the Data Privacy and Security provisions in the Grant Agreement.

- **Privacy Reporting and Notification**
  - Report within one business day to the USDOT Contacting Officer any suspected loss of control or any unauthorized disclosure of PII by the Recipient, its sub-recipients or contractors.
  - Report within one business day to the USDOT Contacting Officer all suspected or actual unauthorized collection, use, maintenance, dissemination, or deletion of PII by the Recipient, its sub-recipients or contractors.
  - Report within one business day (once a determination has been made) any PII or SPII breaches to the known or suspected participants affected by the breach.

5.3 ASD Action Log Data Obfuscation Procedure

Detailed trajectory data or action log data will be scrubbed of potential PII relatable data prior to storage on NYC DOT servers. The is required to lessen the likelihood of using precision time and place specifics to marry CV Pilot performance evaluation data to other existing data sources and databases which do contain PII data. Detailed data that could be construed as PII will then be deleted and not archived anywhere. Liability issues/concerns are significant when any archived data from NYC CV Pilot may be tied through location and time details to other databases/accident records.

5.3.1 Vehicle ID

Vehicle IDs will be anonymized based on the ASD device. Due to privacy and security concerns, this anonymization will be updated regularly to prohibit tracking a single vehicle throughout the study, even with an anonymized ID number. As a result, it will not be possible to measure or assess longer term

individual driver learning and behavior changes resulting from the CV deployment. No driver ID will ever be recorded in the ASD unit or recorded data.

Despite the lack of any vehicle ID in the stored data, a parameter on the ASD that identifies the vehicle size and type will be included. The vehicle type parameter will be recorded with the action log data, but will be general enough so as to not identify the particular participant fleet which the vehicles belong.

5.3.2 Time and Date Data

Time and date information will be obfuscated and recorded to have exact date and time scrubbed from records and instead will be registered in different categories or bins of date and time. For example, an action log data set recorded on Tuesday January 5, 2016 starting at 8:35 a.m. may be recorded only as 'January Tuesday 8-9 a.m.', or 'January Weekday 8-9 a.m.'. Special event days (parades with road closures, holidays, etc.) would be exist in separate bins from normal operation days.

5.3.3 Location Data

Detailed latitude and longitude data recorded by the ASD will be scrubbed and converted to an undefined Cartesian coordinate system for storage and later evaluation. Each event data set will still be recorded with full precision for all trajectory points relative to each other in the same event data set, but will not be tied to an exact real-world coordinate. This removes the detailed location data that could be tied to accident records but still preserves all relative details about vehicle movements and driver actions taken in response to the CV app warning.

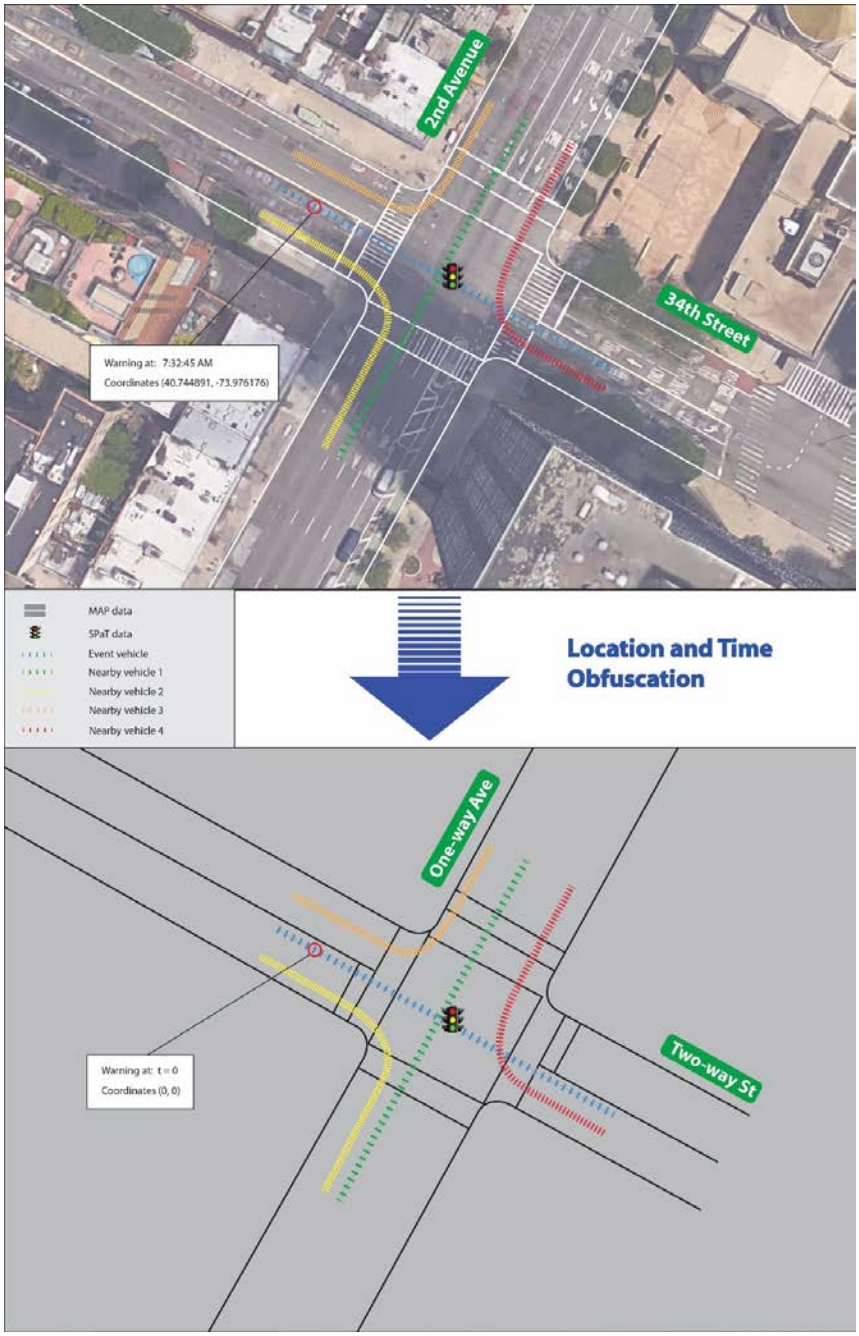
This obfuscation of location will be done independently for each event action log data set recorded. For example, if the first recorded point in an event dataset from an ASD was recorded at (Lat_0, $Long_0$) degrees, it would be scrubbed of location details and stored as point (0,0) (X_0,Y_0 in feet) and serve as the 'anchor point' and for all remaining points in the same ASD event dataset and. All successive points (X_1,Y_1…X_n,Y_n) are recorded from that first point as deltaX and deltaY in the X,Y plane, as calculated from the precise GPS data and preserving as many significant digits as needed. All actual Latitude and Longitude and delta values will be discarded, and only the (X_n,Y_n) values will be archived.

To preserve some information regarding where the event occurred, the location will be recorded as belonging to an aggregated location or bin. Possible examples of such bins could include 'Freeway' vs. 'Manhattan Arterial' vs. 'Brooklyn Arterial' (at the most aggregated level) or 'First Avenue between 23rd&34th St' vs 'First Avenue between 34th&42nd St' (at the least aggregated level).

An additional categorization will also be recorded that provides some details about the general configuration of the intersection or roadway which the event was registered (e.g. 'Intersection of One-way Avenue & One-way Street', or 'Intersection of One-way Avenue & Two-way Street', or 'Midblock on 4-lane one-way Avenue').

For illustrative purposes, a schematic of the obfuscation process is presented in Figure 5-1. In the figure, the top half is illustrative of the details that will be recorded on the ASD in the time before and after an ASD warning was issued. This includes include precise knowledge of where in time and space the vehicle was when the ASD warning was generated (e.g. traveling eastbound on 34th Street approaching 2nd Avenue at 7:32 in the morning), and the precise time and location of all nearby CV equipped vehicles through received BSMs, and the MAP and SPaT data heard from the RSE. The

bottom half of the figure represents that same event as stored in the TMC servers and to be used for evaluation after the obfuscation process scrubs the precise time and location details. Here, instead of know the exact location, we know the warning event happened approaching a one-way avenue on a two-way street at some point in time between 7 a.m. and 8 a.m. It should be noted that while the precise time and location details are scrubbed from the ASD Action Log data, the relative detail and precision of the all vehicle BSMs, MAP, and SPaT are retained and available for subsequent analysis.



(Map Source: Google Maps, 2016)

Figure 5-1. Schematic Example of ASD Time and Location Obfuscation Process

The exact level of scrubbing or obfuscation to both the time and date and the location data has not yet been finalized and cannot be fully be finalized until the deployment is underway. This is due to the fact that the obfuscation of the time/date and location data must be aggregated enough to ensure that there are enough event action log samples in each combination of time/date and location bin to maintain anonymity of the event data to any particular time and location to prevent a later correlation of a single trajectory action log record to other accident, taxi logs, or other existing databases containing PII.

Full details of the degree of obfuscation categories will be developed during the Phase 2 testing period as the CV applications are fine-tuned for the operation parameters and the number of warnings generated per application are better known. The design of the obfuscation though time and location binning will attempt to preserve as much detailed information as possible while still maintaining the anonymity of matching any one obfuscated ASD action log data to a record from another data collection method (i.e. a crash report) which does include PII data.

As different warning types will occur at different rates, the degree of aggregation may differ depending on the type of warning issued. All aggregations will be completed between upload of the stored data from ASDs and storage in NYC DOT databases. For example, the Speed Compliance application is expected to trigger many more warnings that the Vehicle Turning Right in Front of Transit Warning (VTRW) application. Therefore, a much lower degree of obfuscation of action logs for the speed compliance app will be needed as compared to the VTRW app.

Additionally, there is recognition is that the analysis of some action log data for some CV applications will require specifics to be know about the location where the action log data was recorded. As such, the following details for action log data is expected to be retained under certain application warnings.

The degree of obfuscation for each of the CV applications are shown in Table 5-1. For each of the applications, options exist as to whether or not the detailed action log data traces are retained, and if exact location, date/time, vehicle type, and weather and/or special events are preserved or obfuscated to ensure participant anonymity while still providing as much data as possible for the CVPD performance evaluation purposes.

Table 5-1. Degree of Obfuscation by CV Application Warning

User Need	Category	NYCDOT Needs	CV Application	Database storage of detailed Action Logs?	Retain Action Log (Trajectory) Details?			
					LOCATION Collected only at the instrumented locations?	TIME Keep Time of Day, Day of Week?	VEHICLE Keep Vehicle Type?	Include Weather and Special Events?
Manage Speeds	Safety, Mobility	Discourage Spot Speeding	Speed Compliance	Yes (When warning is generated)	No (Obfuscate Details)	No (Obfuscate Details)	Yes (Keep Vehicle Type)	No (Obfuscate Details)
	Safety	Improve Truck safety	Curve Speed Compliance	Yes (When warning is generated)	Yes (Keep Details)	No (Obfuscate Details)	Yes (Keep Vehicle Type)	No (Obfuscate Details)
	Safety	Improve Work Zone Safety	Speed Compliance / Work Zone	Yes (When warning is generated)	Yes (Keep Details)	No (Obfuscate Details)	Yes (Keep Details)	No (Obfuscate Details)
Reduce Vehicle to Vehicle Crashes	Safety	V2V Safety Applications	FCW EEBL BSW LCW IMA	Yes (When warning is generated)	No (Obfuscate Details)	No (Obfuscate Details)	Yes (Keep Vehicle Type)	No (Obfuscate Details)

User Need	Category	NYCDOT Needs	CV Application	Database storage of detailed Action Logs?	Retain Action Log (Trajectory) Details?			
					LOCATION Collected only at the instrumented locations?	TIME Keep Time of Day, Day of Week?	VEHICLE Keep Vehicle Type?	Include Weather and Special Events?
	Safety	Reduce Accidents at High Incident Intersections	Red Light Violation Warning	Yes (When warning is generated)	No (Obfuscate Details)	No (Obfuscate Details)	Yes (Keep Vehicle Type)	No (Obfuscate Details)
	Safety	Reduce Incidents, Improve Safety	Vehicle Turning Right in Front of Bus Warning	Yes (When warning is generated)	No (Obfuscate Details)	No (Obfuscate Details)	Yes (Keep Vehicle Type)	No (Obfuscate Details)
Reduce Vehicle to Pedestrian Crashes	Safety	Improve Pedestrian Safety on Heavily Traveled Bus Routes	Pedestrian in Signalized Crosswalk Warning	Yes (When warning is generated)	No (Obfuscate Details)	No (Obfuscate Details)	Yes (Keep Vehicle Type)	No (Obfuscate Details)
	Safety	Improve Safety of Visually and Audibly-impaired pedestrians	Mobile Accessible Pedestrian Signal System (PED-SIG)	Yes (When warning is generated)	No (Obfuscate Details)	No (Obfuscate Details)	Yes (Keep Vehicle Type)	No (Obfuscate Details)
Reduce Vehicle to Infrastructure Crashes	Mobility	Address Bridge Low Clearance Issues/Enforce Truck Route Restriction	Oversized Vehicle Compliance	Yes (When warning is generated)	Yes (Keep Details)	No (Obfuscate Details)	Yes (Keep Vehicle Type)	No (Obfuscate Details)
Inform Drivers of Serious Incidents	Mobility	Inform Drivers	Emergency Communications and Evacuation Information	No (Discard action log data, retain ASD 'ping' only)	Yes (Keep Details)	Yes (Keep Details)	Yes (Keep Vehicle Type)	Yes (Keep Details)
Provide Mobility Information	Mobility	Replace Legacy Measurements	Intelligent Traffic Signal System Connected Vehicle Data (I-SIGCVDATA)	No (ASD breadcrumbs processed and stored, detailed action log discarded)	Yes (Keep Details)	Yes (Keep Details)	Yes (Keep Vehicle Type)	Yes (Keep Details)
Manage System Operations	System Operations	Ensure Operations of the CV Deployment	NA	No (Discard action log data, retain ASD 'ping' only)	Yes (Keep Details)	Yes (Keep Details)	Yes (Keep Vehicle Type)	Yes (Keep Details)

5.4 Data Sharing Framework

A data sharing framework will be developed and delivered as part of the Performance Measurement of the NYC CVPD. The recipients of the data shared via this framework includes the USDOT and its independent evaluators working on the connected vehicle initiative, and the larger connected vehicle research community via the USDOT's Research Data Exchange (RDE). This section describes the plans for sharing performance and evaluation data from the NYC CVPD. The detailed requirements related to data sharing are included in the System Requirements Specification (SyRS). The data sharing framework is illustrated in Figure 5-2.

5.4.1 The Data

As part of the NYC CV Pilot, the data being collected can be grouped into two categories based on the data source – intrinsic and extrinsic. The former is data collected by the CV components (the

ASDs and the RSUs) and the latter is data from external sources (such as traffic control system data, incidents, crash statistics, weather data, traffic counts, travel times and speeds, etc.).

The raw data collected from the ASDs and RSUs will be processed at the NYCDOT TMC and scrubbed to remove any PII (personally identifiable information) using the data obfuscation techniques outlined above. Following the obfuscation process, the raw data will be destroyed, leaving only the obfuscated data. The obfuscated data will then be post-processed to develop performance metrics as described earlier in this document.

Both the disaggregated obfuscated data and aggregated processed performance metrics data will be shared with the USDOT and the independent evaluator for their evaluation of the CVPD. The disaggregated data will be uploaded at a shorter weekly or bi-weekly frequency, and the aggregated data with the various measures will be uploaded on a bi-monthly or quarterly basis.

The same data will also be shared with the larger research community via the RDE, but due to the sheer size of the anticipated data, a representative selected sub-set of the data may instead be selected to be uploaded and shared.

Additionally, while the ASD action log data will be obfuscated to prevent the matching of details to a specific time and location, a final review of the obfuscated data logs to ensure that privacy is maintained before posting the data to the RDE. If some groupings of the obfuscated data and the confounding data have very limited numbers of samples, that data may be withheld from being shared on the RDE.

5.4.2 Data and Privacy

Both safety and non-safety data being collected by the NYC CV Pilot is owned by the individual stakeholders. The use and sharing of the data will be governed by the agreements (MOUs) between NYCDOT and the stakeholders, which are currently being drafted and reviewed by all stakeholders. The guiding principles of these agreements are that the data will be obfuscated to remove any PII, and that the data is being collected to evaluate the system performance related to safety, consistent with NYCDOT's Vision Zero initiatives. The use of the data by the USDOT and the research community via the RDE should also respect the principles of privacy contained within those signed agreements.

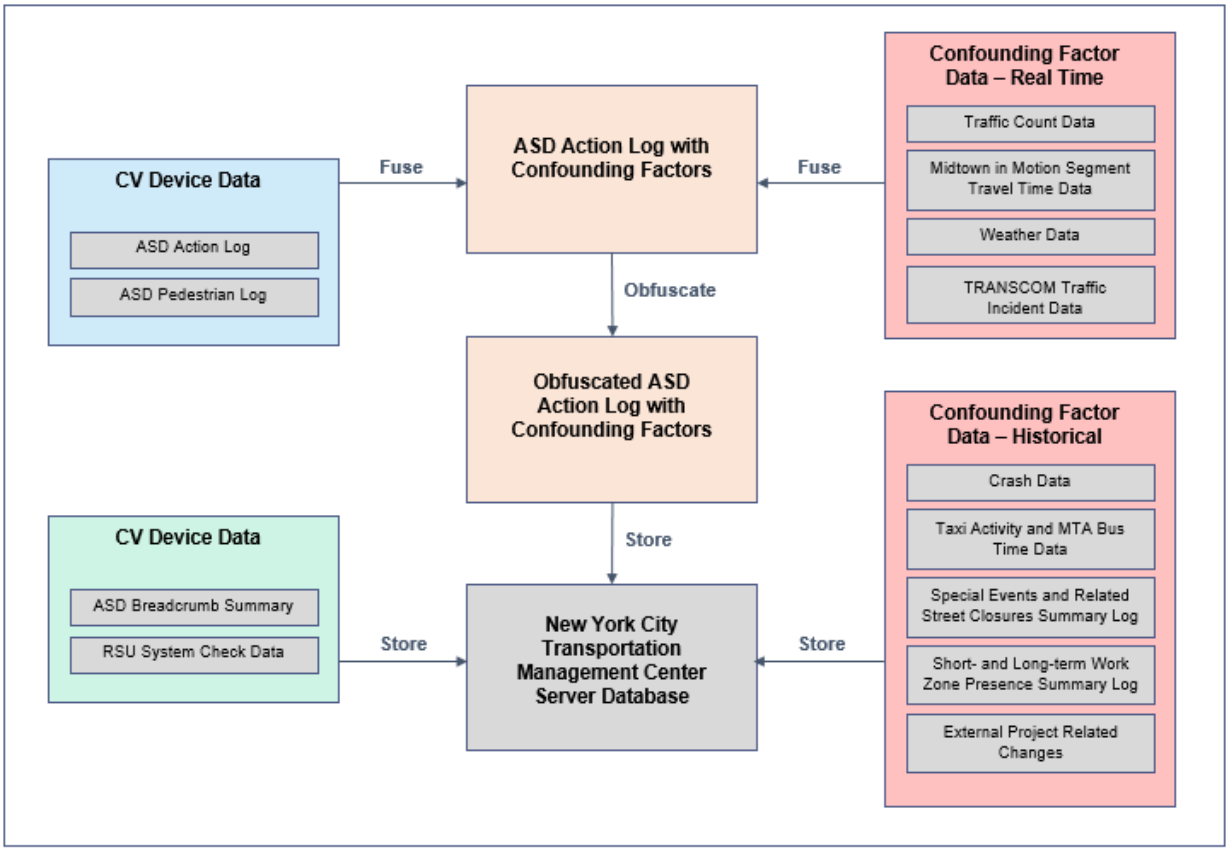
Data being collected is distinguished between vehicular and non-vehicular (pedestrian) participants. Pedestrian applications will include two types of data to protect: 1) raw trajectory data from the PID and 2) participant PII such as name and contact information that is to be used for conducting targeted interviews (see section 5.1.1). Vehicular participant data will only consist of vehicle trajectory data.

5.4.3 Data Transmission

The obfuscated data will be hosted on servers in the NYCDOT TMC. While details are yet to be finalized, the data exchange to USDOT is expected to be completed using individual flat files (e.g. CSV) or rotated tables in databases so that the data package transmitted is manageable. It is anticipated that there will be two levels of upload frequencies for the data – more frequent (daily to weekly, as possible), and less frequent (bi-monthly).

The disaggregated cleansed data from both the CV and non-CV based sources, will be uploaded in 24 to 48 hour time windows, as the data processing, cleaning, fusion to confounding factor data, and

obfuscation processing (as needed) is completed. The processed aggregated data will be uploaded on a bi-monthly or quarterly basis.



Source: NYCDOT, 2016

Figure 5-2. Overall Data Flow and Processing Methods for Performance Evaluation Data

References

#	Document
1	“USDOT Guidance Summary for Connected Vehicle Pilot Site Deployers: Human Use Approval”, Sept. 2015
2	NIST Special Publication 800-53, Appendix J (Security and Privacy Controls for Federal Information Systems and Organizations)
3	NIST Big Data Interoperability Framework: Volume 1 Definitions
4	NIST Big Data Interoperability Framework: Volume 4, Security and Privacy
5	NIST SP 800-88r1 “Guidelines for Media Sanitization”
6	NIST SP 800-122: " Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)”

APPENDIX A List of Definitions and Acronyms

A.1 Definitions

The definitions from NIST SP 800-122 [6], and the following, apply.

Term	Definition
Identifier	a label (numeric, alphanumeric, or other) that is used within a specific context to uniquely identify an entity
Personally Identifiable Information (PII)	information which can be used to distinguish or trace an individual EXAMPLE: PII that may be used to distinguish an individual may include their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc.
Sensitive Personally Identifiable Information (SPII)	information that if lost, compromised, or disclosed could result in substantial harm, embarrassment, inconvenience, or unfairness to an individual EXAMPLE: PII that may be classified as SPII includes Social security numbers, Bank account numbers, Passport information, Healthcare related information, Medical insurance information, Student information, Credit and debit card numbers, Driver's license and State ID information NOTE: In addition to static or de facto information that is classifiable as PII some contextual data may be linked to PII and become, in combination, SPII. For example, this may include making assertions regarding the individual from places visited, times of travel and similar.
Recipient:	a person or organization to which PII or SPII is sent.
Collection:	the process of collecting and storing PII or SPII
Physical Control:	the physical maintenance and storage of PII information – Physical control may pertain to hard copies of information or physical control of electronic resources that store such information.
Policy:	Rules or sets of rules that pertain to: 1) access rules for PII within a system, 2) PII retention schedules and procedures, 3) PII incident response and data breach notification, 4) privacy in the system development life cycle process, 5) limitation of collection, disclosure, sharing and use of PII and 6) consequences for failure to follow privacy rules of behavior. [6]
Procedure:	Actions or activities required of an individual or organization to satisfy PII-related policies.

A.2 Acronyms

Acronym / Abbreviation	Definition
MTA	Metropolitan Transportation Authority
DSNY	City of New York Department of Sanitation
DPP	Data Privacy Plan
DMP	Data Management Plan
RDE	Research Data Exchange
PII	Personally Identifiable Information
SPII	Sensitive Personally Identifiable Information
ASD	Aftermarket Safety Device

U.S. Department of Transportation
ITS Joint Program Office-HOIT
1200 New Jersey Avenue, SE
Washington, DC 20590

Toll-Free “Help Line” 866-367-7487
www.its.dot.gov

FHWA-JPO-17-453

U.S. Department of Transportation