Connected Vehicle Pilot Deployment Program Phase 2

Data Management Plan – New York City

www.its.dot.gov/index.htm

Final Report — October 17, 2017
Updated: April 15, 2021

Publication Number: FHWA-JPO-17-454

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.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-JPO-17-454		

4. Title and Subtitle	5. Report Date
Connected Vehicle Pilot Deployment Program Phase 2, Data Management Plan	April 15, 2021
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Drew Van Duren, OnBoard Security; Robert Rausch, David Benevelli, TransCore ITS – New York	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
New York City Department of Transportation (NYCDOT) Traffic Operation, ITS Management division 34-02 Queens Boulevard, Long Island City, NY 11101	11. Contract or Grant No. DTFH6116H00026

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
ITS-Joint Program Office 1200 New Jersey Avenue, S.E., Washington, DC 20590	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 management plan that delineates all of the data types and data treatment throughout the New York City Connected Vehicle Pilot Deployment (NYC CVPD). This 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 researcher use. Data management processes include data categorization, data processing, data handling by administrators and custodians (including anonymization and privacy protection), data transmission (communication), storage, retrieval, purging and record keeping.

17. Keywords	18. Distribution Statement
Data, Connected Vehicle, Privacy, Management, New York City	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	96	

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 vehicle fleets, NYC Department of Citywide Administrative Services (DCAS) vehicle fleets, MTA/NYCT, and Pedestrians for Accessible and Safe Streets (PASS) have expended considerable resources participating in the development of the overall Concept of Operations as well as this document.

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

Table of Contents

Chapter 1. Introduction 1
Purpose of the Plan 1
Organization of the Plan 1
Chapter 2. Connected Vehicle Pilot Deployment Overview 2
Chapter 3. Data Management Approach 3
Data Sharing and Re-Use 3
Independent Evaluator (IE) 4
Data Availability to External Researchers 6
Data Privacy and Intellectual Property Data Risks 8
Risks 8
Privacy Management 9
Data Access: Privacy and Security in Data Management 9
Security and Privacy Procedural Controls by Data Management Activity 9
Protecting Access to NYU Participant PII During Collection and Storage 13
Access Control During Device Installation and Management 14
Roles and Responsibilities in Data Management 23
Personnel Qualifications and Training 25
Data Quality Control 26
Data Quality Facets 26
Reproducibility and Data Processing Integrity 27
Data Preservation and Archiving 27
TMC Originated Data 27
Pedestrian Data (NYU) 28
Chapter 4. New York City Pilot Data Plan 29
Data Types, Sources and Destinations 30
Participant Information Collection 30
ASD and RSU Operations and Maintenance (O&M) Data 31
Mobility (Performance) Data 32
Event Data 33
ASD and RSU Device Management and Connected Vehicle Data 36
PID Data 38

Data Quantities and Collection Constraints....38
CV Data Quantities....38
RSU Communications and NYCWiN Network Constraints....44
ASD Device DSRC Data Upload and Download....45
Data Collection, Processing and Storage....45
Privacy Condition Impacts on Data Collection....47
ASD O&M Data Collection....47
RSU O&M Data Collection....50
Mobility Data Collection....51
Event Data Collection....53
Operations Logging....54
PII Data Collection – NYU....54
TMC Data Lifecycle: Post-Collection Processing, Treatment Analysis and Storage....55
NYU Data Lifecycle: Pedestrian and Driver Information....63
Process and Tools for Creating, Processing and Visualizing ASD and RSU-Originated Data....64
Documenting Data Collection....65
Data Organization, Documentation and Metadata....66
Project Data (Raw and Processed)....66
Project Documentation....67
Metadata....70
Directory and File Naming Conventions....72
Project Identifiers....76
Chapter 5. References....77
Chapter 6. Glossary....79
Appendix A. Log Formats....83
ASD Sighting RSU Log Format....83
V2V Encounter Log Format....84
RSU Sighting ASD Log Format....85
Probe Data Log Format....86
Event Data Log Format....87

List of Tables

Table 1. Sharing with the Independent Evaluator 4
Table 2. Data Sharing with External Researchers 6
Table 3. Data Management Related Risks and Mitigations 8
Table 4. Security and Privacy Controls by Data Management Activity 10
Table 5. Management and Use of Cryptographic Keys in the NYC CVPD 19
Table 6. Data Management Activities by Role 23
Table 7. Data Quality Facets 26
Table 8. Security and Privacy Controls by Data Management Activity 31
Table 9. Operations and Maintenance Data Types 32
Table 10. Mobility Data Types 32
Table 11. ASD Event Data 34
Table 12. Example Safety-Related Event Data for the IE 35
Table 13. ASD and RSU Device Management and Connected Vehicle Data 36
Table 14. Quantity of Vehicles by Type 38
Table 15. BSM J2735 Core Data Fields 39
Table 16. BSM J2735 Part II Crumb Data Fields 39
Table 17. BSM Core Data Fields 40
Table 18. J2735 SPaT Message Fields 41
Table 19. J2735 MAP Message Fields 42
Table 20. Total Data Originating from ASDs 44
Table 21. ASD File Extensions 50
Table 22. Project Documentation Folder Structure 67
Table 23. NYU Project Directory Structure 69

List of Figures

Figure 1. Data Management Activities – Approximate Timeline 3
Figure 2. Network Connectivity Architecture 30
Figure 3. ASD Event and Breadcrumb Data Collection and Processing Sequence 46
Figure 4. Event Data Collected by ASDs and Sent to TMC 47
Figure 5. Data States Applicable to Data Management at the TMC 56
Figure 6. ASD and RSU Data Collection, Processing and Storage Strategy 58
Figure 7. Event Log Upload Process 59
Figure 8. Overall Performance Data Flow and Processing 61
Figure 9. Suspect Bin Processing 62
Figure 10. Participant Data Collection and Storage 63
Figure 11. Metadata Types and Functions 71
Figure 12. V2V Encounter Log 84

Chapter 1. Introduction

Purpose of the Plan

This Data Management Plan (DMP) describes how data will be collected, managed, integrated, and disseminated before and during Phase 3 of the New York City Connected Vehicle Pilot Deployment (NYC CVPD). This includes real-time and archived data that are inputs to and outputs from systems managed by the CV Pilot team and its partners.

This DMP discusses the team's plans for managing data as a strategic asset and making open, machine-readable data available to the public – subject to applicable privacy, security and other safeguards – to fuel entrepreneurship and innovation to improve citizens' lives, create jobs, and spur economic development.

Additional details regarding data collection will be made available in the NYC CVPD Interface Control Document (ICD) and System Design Document (SDD).

Organization of the Plan

This data management plan is structured as follows:

- Chapter 2 provides a brief overview of the NYC CVPD
- Chapter 3 provides the data management approach, consisting of:
  - Data sharing and re-use
  - Data privacy and intellectual property issues and risks
  - Data privacy and security controls
  - Quality control
  - Data preservation and archiving
- Chapter 4 provides the NYC data plan, consisting of:
  - A full description of data types, sources, destinations
  - Data quantity as well as data collection constraints
  - A detailed description of data collection, processing and storage processes
  - A description of how data is documented and organized, including metadata

The Appendix contains a listing of each log file emanating from Aftermarket Safety Devices and Roadside Units, including the log file's structure, contents and format.

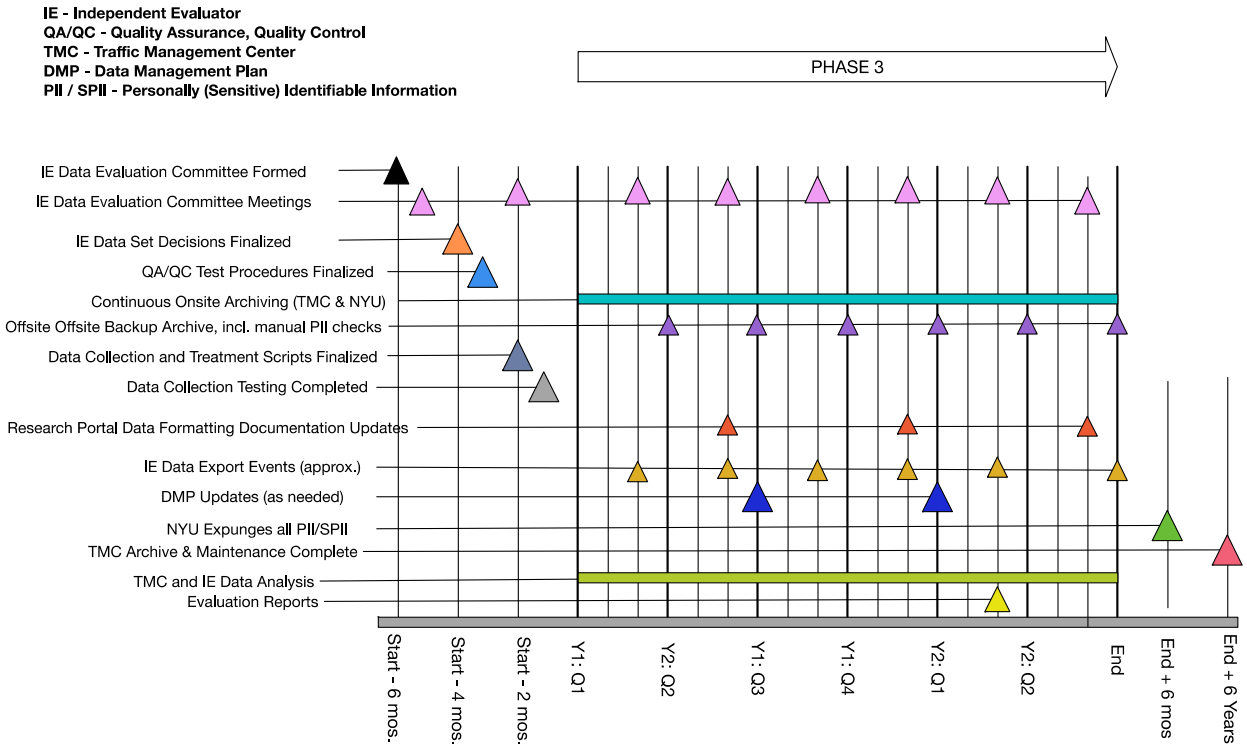
Chapter 2. Connected Vehicle Pilot Deployment Overview

The NYC Connected Vehicle Pilot Deployment (NYC CVPD) 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 connected vehicle (CV) Pilot 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 to provide a comparative sample that can be used to verify benefits against those for locations that are not instrumented.

Chapter 3. Data Management Approach

This chapter provides the data management approach to data collected during the New York pilot. The major discrete and iterative data management activities and milestones described in this document will be performed according to the approximated timeline indicated in Figure 1.



Source: NYCDOT, 2017

Figure 1. Data Management Activities – Approximate Timeline

Data Sharing and Re-Use

Once collected and stored – short or long term – some of the data collected will be made available to research entities as described in this section. The evaluation data will be provided to the Independent Evaluator (IE) and researchers at agreed, pre-defined levels of frequency, precision, specific units and aggregation levels. Given the potential to unintentionally expose Personally Identifiable Information (PII) at certain aggregation and resolution levels, these performance measures will be continuously re-evaluated during the pilot and checked prior to external sharing.

Independent Evaluator (IE)

Selected portions of the collected data will be shared on an adjustable, periodic basis with the Independent Evaluator via the USDOT Connected Vehicle Performance Evaluation Platform / Secure Data Commons (CV PEP/SDC). The Independent Evaluator (IE) and select researchers will use data collected before and during deployment to derive quantitative and qualitative measures of system impact. Table 1 indicates the requirements and processes for sharing data with the Independent Evaluator.

Table 1. Sharing with the Independent Evaluator

Requirement	Process(s)
All sanitized, normalized and merged CV data shall be made available to the Independent Evaluator (IE)	The NYC TMC will upload all binned data (data that has been fully normalized, sanitized and aggregated into data 'bins') to the IE in accordance with US Department of Transportation (USDOT) plans (see Chapter 4, TMC Data Flows and Systems, for more details on the data binning process). CV PEP/SDC details will be made available to the NYC TMC for data migration planning. The NYC CVPD Data Steward will periodically upload the latest binned data to the IE over prescribed CV PEP/SDC interfaces. Note that the time between uploads will be determined during the initial data collection to ensure that the privacy of the data is not compromised.
All CV data sent from NYC CVPD to the IE shall be protected from PII and SPII disclosures and the potential to expose privacy-related tracking information	All data will be collected at the Traffic Management Center (TMC), and the data will be sanitized, normalized, anonymized and merged prior to exporting into releasable data bins. Data binning scripts, databases and file system directories are to be human-reviewed prior to IE and researcher release. Data anonymization, normalization and data merges will be performed on raw data sets using data processing scripts. When new tests or analysis needs to be conducted on raw data sets, new processing scripts will be created or existing scripts revised, i.e., new data actions will be created. When this occurs, privacy engineering best practices will be followed to review new and revised "data actions" for potential privacy implications as described in NISTIR 8062. Data Actions therein, are *"…any system operations that process PII. Processing can include, but is not limited to, the collection, retention, logging, analysis, generation, transformation or merging, disclosure, transfer, and disposal of PII."* (NISTIR 8062: An Introduction to Privacy Engineering and Risk Management in Federal Systems (http://nvlpubs.nist.gov/nistpubs/ir/2017/NIST.IR.8062.pdf)

Requirement	Process(s)
Data formats in all shared data shall be published and made available to the Independent Evaluator (IE)	The NYC Data Steward (see Chapter 3, Roles and Responsibilities in Data Management) will be responsible for providing to the Independent Evaluator and USDOT documentation on the data formats for all relevant data sets. The documentation will be made available at the CV PEP/SDC. Data documentation to the IE will include 1) an overall explanation about the data organization, sources and content, 2) attribute description, including attribute data types/units, source, measurement method and measurement frequency, 3) data attribute aggregations, default values and known accuracy limitations (based on measurements) and 4) methods of data sanitization and aggregations. Documentation will additionally include a description of the data definitions and relationships as they pertain to different sources.
All data submitted to the IE shall be reviewed and approved by the NYC pilot team prior to release.	Final checks for PII or potential PII will be made by both the New York City Department of Transportation (NYCDOT) data steward and New York University (NYU) data steward (see Chapter 3, Roles and Responsibilities in Data Management) in conjunction with other pilot researchers. When data is confirmed to contain no PII (or potential PII if merged with other data sets), data will be released.
All data formatting documentation will be provided for the IE to ingest and analyze the data.	The IE will be provided with any necessary data format documentation, data parsing or query tools necessary to perform analytics work. Some tools may be provided by USDOT at the CV PEP/SDC. These tools may consist of various application scripts, database schemas and queries, as necessary.
Data sharing and re-use to the IE shall contain restrictions, as appropriate, to ensure that some types of data (as determined by the NYC CVPD and USDOT) are not re-used without specific policies in place	Once data is made available to the IE for analysis, not all the data will necessarily be sharable or accessible by all IE analysts or other researchers. Only a sub-set may be made available to the researchers, for example. Additionally, any restrictions on re-use will be made explicitly in documentation provided to the IE at the time (or prior) data is provided. The IE, via its CV PEP/SDC portal, will maintain its copy of uploaded data and control researcher access to specific data sets.
All data sharing policies and provisions shall be controlled by an authorized entity	Data sharing to the IE (and at the IE) will be controlled by a group of representatives, at least one from NYCDOT, the Institutional Review Board (IRB) and IE. Each representative will be versed in data privacy concepts and policies and will be designated at the discretion of the representative's organization.

Requirement	Process(s)
Data shared to USDOT shall be for restricted use only	Data provided to the IE (Texas A&M Transporation Institute (TTI)) will be stored and Volpe Center under a restricted use policy if or when any the provisioned data may contain PII (note: storage location and facility will be updated in subsequent revisions of this document as plans are matured). Some of this data may be provided to other researchers (see Chapter 3, Data Availability to External Researchers). Data stored at the IE's CV PEP/SDC shall be protected in accordance with government sensitive information protection policies consistent with NIST SP800-122 and SP800-53 controls for Sensitive PII.
Privacy-protected data shall be protected and authorized when shared between the IE and NYU.	It may be necessary to share between NYU and the IE as the NYC CVPD evolves. Data privacy protection procedures will be upheld in the process by not exporting any Pedestrian PII from NYU to any other organization.

Data Availability to External Researchers

This section describes the site's plans for transmitting data to the research community. Note that these plans are still in development at the time of this writing, and additional information will be provided by USDOT and the NYC CVPD.

Sanitized and obfuscated vehicle, mobile device, and infrastructure data captured during deployment are planned to be collected and shared with the research community via USDOT's ITS DataHub (formerly known as Research Data Exchange (RDE)). This is a separate portal compared to USDOT's Secure Data Commons (SDC). Details of these sharing agreements to researchers outside of the IE are pending as of this revision of the Data Management Plan, however Table 2 provides high-level requirements and processes for sharing data with outside researchers.

Table 2. Data Sharing with External Researchers

Requirement	Process(s)
A sub-set of IE-available data shall be made available to researchers	A group of representatives of NYC CVPD, the IE and USDOT will be established prior to Phase 3 to provide guidance and make decisions on which IE data subsets will be made available researchers. This group will convene quarterly (or in response to community or stakeholder requests) to make ongoing decisions throughout the life of the pilot, and communicate those decisions to the NYC CVPD Data Steward (see Chapter 3, Roles and Responsibilities in Data Management). The IE will make the data sub-sets available through the USDOT-provided ITS DataHub portal.
All CV data from NYC CVPD shall be protected from PII and SPII disclosures and the potential to expose privacy-related tracking information	All data is collected at the TMC, sanitized, normalized, anonymized and merged prior to exporting into releasable data bins. A subset of the sanitized, binned data is expected to be provided to researchers via the ITS DataHub.

Requirement	Process(s)
Data made available to researchers shall be discoverable by researchers	NYCDOT will provide data to USDOT via the ITS DataHub. USDOT may publish in transportation forums and websites the availability and web location of archived research data as future decisions are made about sharing it, methods of sharing via the ITS DataHub, formats and needed access controls if any.
Data formats for all shared data shall be published and made available to the public	The NYC Data Steward will be responsible for providing USDOT and researchers documentation on the data formats for all relevant data sets. Documentation applicable to shared data sets are expected to be made available via ITS DataHub portal. Any formatting deviations made by the IE from those provided to the ITS DataHub by NYC are expected to be documented and furnished by the IE.
All data provided to researchers shall contain no restrictions on re-use	This policy will be made available at the portal location. The sharing and re-use policy are expected to be clearly indicated by the IE.
All data sharing policies and provisions shall be controlled by an authorized entity	Data sharing to researchers will be controlled by a group of representatives from NYCDOT, IRB, USDOT and the IE.

The NYC CVPD and IE data sharing will abide by all terms of use (as yet to be defined by USDOT). These terms are expected to be provided by USDOT in preparation for Phase 3 of the pilot.

All metadata and data documentation provided to researchers is expected to be in ASTM 2468-05 (https://www.astm.org/Standards/E2468.htm) standard format, though this requirement from USDOT is pending as of the date of this document. NYC CVPD plans to include requisite information about the data environment and metadata concerning data elements in the data files. Anticipated information includes:

- Description of data collection procedures
- Time and location of collected data (this will be obscured for privacy reasons)
- Contact Information
- Data Elements (each)
    - Type
    - Units,
    - Field length
    - Max/min values
    - Code definitions

Data Privacy and Intellectual Property Data Risks

This section describes data privacy issues, risks and relevant data management approaches. A more thorough treatment description and treatment are provided in the Data Privacy Plan (DPP) (Connected Vehicle Pilot Deployment Program Phase 2 Data Privacy Plan – New York City, December 27, 2016).

Risks

This section and Table 3 list and describe the top-level data management related risks associated with the NYC CVPD.

Table 3. Data Management Related Risks and Mitigations

Data Management Activity	Risks	Mitigation Control IDs (See Table 4)
Collection, Storage and Use of Participant PII and SPII	• Unauthorized access to participant PII data at NYU	• AC-3,5,6 • AU-2,6 • IA-2 • MP-2,3,4,6 • SC-8,9 • SC-19,28 • SI-4
Collection, Storage and Use of Device Raw Data (Event data, mobility [breadcrumb] data and ASD/RSU sighting data)	• Correlation of non-PII with other data to produce PII and violate participant privacy • Unauthorized access to Aftermarket Safety Devices (ASD) and Pedestrian Information Devices (PID) resulting in disclosure of PII or adversarial tracking • Integrity losses of raw data on ASD and PID platforms • Insufficient device storage to record and maintain availability and upload of ASD or PID data	• AC3 • IA-2 • SE-8,9 • AU-2,6 • SC-19 • PE-2,3 • NIST SP800-122 (4.2.4)
Management of Vehicle and Vehicle Owner-Operator Data	• Unauthorized access to vehicle telematics data allowing data correlation that violates participant (driver) privacy • Vehicle ASD or telematics unit faults corrupt data • Insufficient device storage amounts needed to record and maintain availability and upload of vehicle data	• PE-2,3 • AU-2,6 • SC-28
Upload of Raw ASD/PID Data to TMC	• Confidentiality or integrity losses during upload of raw data to TMC • Compromise of TMC private key material allowing intercept of raw ASD/PID data	• SC-8,9,12

Data Management Activity	Risks	Mitigation Control IDs (See Table 4)
Processing and storage of data at the TMC	• Corruption of sanitization or normalization application during processing • Corruption or failure of system storing short-term raw data • Corruption or failure of system storing processed data • Law enforcement attempts to retrieve raw CV and telematics data held by TMC • Unauthorized access to TMC networks, servers and applications	• AU-2,6 • AC-3,5,6 • MP-4,6 • SC-28 • PE-2,3
Distribution and Storage of Pilot data to Researchers (IE portal, ITS DataHub)	• Integrity losses of archived/distributed data • Loss or failure of data archive system	• AC-21 • MP-4 • SC-28 • SI-4
Cryptographic Key Management Risks	• Unauthorized disclosure and compromise of TMC key material, including SCMS, TLS, SSH and SNMP keys • Compromise of RSU and ASD cryptographic keys • Compromise of file and database encryption keys	• SC-8,9,12

Each risk cited is addressed through the policies and procedures in this Data Management Plan.

Privacy Management

In addition to the existing Data Privacy Plan (Connected Vehicle Pilot Deployment Program Phase 2 Data Privacy Plan – New York City, December 27, 2016), Chapter 3, Data Access: Privacy and Security in Data Management, identifies relevant privacy and security controls mapped to pertinent data management activities.

Data Access: Privacy and Security in Data Management

This section provides the requisite data management activities that must be performed by designated, authorized individuals during the NYC CVPD. Each of these activities serves a vital role in mitigating the afore-mentioned risks.

Security and Privacy Procedural Controls by Data Management Activity

NIST Special Publications 800-53 (NIST Special Publication 800-53, Appendix J (Security and Privacy Controls for Federal Information Systems and Organizations) and 800-122 (NIST SP 800-122: “Guide to

Protecting the Confidentiality of Personally Identifiable Information (PII)") provide recommended security controls pertinent to the protection of PII. Controls are those listed in Table 4.

Table 4. Security and Privacy Controls by Data Management Activity

Control ID	Systems Affected	Applicable Data Management Activities and Actions
AC-3 Access Enforcement	• All TMC systems processing PII • All NYU systems processing PII • All RSU, ASD and PIDs • Data archive systems	• All activities
AC-5 Separation of Duties	• All	• All activities
AC-6 Least Privilege	• All	• All system and CV device service accesses
AC-17 Remote Access	• ASD and RSUs • TMC • RSU	• All system and CV device service accesses
AC-21 User-based Collaboration and Sharing	• TMC • ITS DataHub as research portal	• Administer data delivery systems • Maintain sanitized data uploads to the research portal • Verify data tagging indicating sanitized state of uploaded data
AC-19 Access Control for Mobile Devices	• PIDs	• All interactions/activities with PIDs
AU-2 Auditable Events	• All TMC Systems • All NYU Systems • RSU • ASD • PID	• All device, host and network monitoring related activities
AU-6 Audit Review, Analysis and Reporting	• All systems	• All host and network monitoring related activities • All privacy policy adherence activities • Internal PII audits
IA-2 Identification and Authentication	• All systems	• All activities requiring device and system access

Control ID	Systems Affected	Applicable Data Management Activities and Actions
MP-2 Media Access	• All systems	• All activities requiring media access
MP-3 Media Marking	• All systems	• Assign identifiers and label ASDs • Assign identifiers and label PIDs • Perform data custodial functions for pedestrian participant PII/SPII data • Maintain and archive collected privacy consent forms
MP-4 Media Storage	• All systems, including network drives, tape drives, and information hard copies	• Manage physical inventory of devices • Assign and maintain unique identifiers to devices • Collect participant hard copy and electronic PII/SPII data • Coordinate participant surveys, interviews and polls • Perform data custodial functions for pedestrian participant PII/SPII data • Maintain privacy and consent forms • Oversee NYC CVPD adherence to privacy protection policies • Maintain and archive collected privacy consent forms • Administer privacy and security training and related materials to NYC CVPD program personnel • Backup and archive data • Facilitate participant redress of privacy-related issues
MP-6 Media Sanitization	• All systems containing PII, SPII, or data that in concert with other data (specifically, public data sources, search engines, etc.) may become PII when subject to contemporary data mining techniques	• Verify sanitized state of uploaded data • Wipe and purge raw data from ASDs and PIDs • Use robust paper shredding equipment in all document hard-copy destruction

Control ID	Systems Affected	Applicable Data Management Activities and Actions
SC-8, SC-9 Transmission Confidentiality and Integrity (Note: SC-8 not in SP800-122)	• All wired and wireless links	• Transmission, upload of all NYC CVPD data, including participant PII, vehicle telematics, Basic Safety Messages (BSM) and other event data • Encrypt network connections from RSU to Traffic Signal Controller • Encrypt network connections from roadside environment to TMC • Cryptographically authenticate and integrity protect all links
SC-12 Cryptographic Key Establishment and Management	• ASDs • RSUs • PIDs • NYCDOT Servers	• Manage Security Credential Management (SCMS) Keys • Manage Transport Layer Security (TLS) Keys • Manage Secure Shell (SSH) Keys • Manage Simple Network Management Protocol (SNMP) Keys
SC-28 Protection of Information at Rest	• All electronic storage devices, servers, and equipment for storing PII	• All activities involving storage of any NYC CVPD data • All activities involving storage of participant PII (see Chapter 3, Protecting Access to NYU Participant PII During Collection and StorageRoles and Responsibilities in Data Management) for a discussion of collected PII)
SI-4 Information System Monitoring	• All electronic and networked systems and devices	• Perform diagnostics, collect and analyze host log files • Manage security configuration of RSU units • Device and servers verify software/firmware integrity • Monitor local and remote access to TMC systems • Evaluate and monitor data drives and database integrity systems • Monitor firewalls • Monitor TMC host-based Intrusion Detection Systems
PE-2, PE-3 Physical and Environmental Monitoring	• Physical security and access at all facilities	• Physical access of vehicle to install and maintain ASD • Physical access and control at fleet owner facilities

Control ID	Systems Affected	Applicable Data Management Activities and Actions
NIST SP800-122 Anonymization Methods (SP800-122, Section 4.2.4)	• Raw data collection systems	• Raw data collection from devices • Raw data processing, normalization and anonymization

Allocation of the above SP800-122 and SP800-53 controls and adherence to New York City CityNet security policies (New York City Citywide Security Policies (DOITT) - https://www1.nyc.gov/site/doitt/business/it-security-requirements-vendors-contractors.page) will be maintained during all data management activities addressed in this plan.

Protecting Access to NYU Participant PII During Collection and Storage

This section is pertinent only to NYU collection and storage of participant PII. Note that anything related to pedestrian participant PII is handled by NYU/IRB, including consent forms. No pedestrian PII will be sent to the NYCDOT TMC.

All pedestrian participant PII and SPII data will be marked as 'PII Sensitive.' Data marked as such will be:

1. Stored in locked safe, or
2. Electronically stored in encrypted format
3. Strictly access controlled

Hard Copies: All PII collected directly from pedestrians will be collected by the Participant Liaison only. Hard copies that are stored in a locked safe will only be accessible to the Participant Liaison or other researchers who have completed data privacy training. Keys to safes will be distributed only to those granted access to safes. Physical copies of data may be destroyed 1) once electronic copies are created, backed up and encrypted, or 2) when the data is no longer needed. Physical copies of hard-copy data will be treated like the originals. Scanned data will be stored encrypted like other electronic copies of PII. The process of destroying physical copies shall make use of a paper shredder conforming to DIN 6639, level P-5 (http://www.din-66399.com/index.php/en/securitylevels) and NIST SP800-88 (NIST SP 800-88r1 "Guidelines for Media Sanitization") descriptions of media destruction.

Electronic Copies: File servers, databases and other electronic host artifacts shall be stored in locked, secured rooms with keyed access granted only to approved researchers. Server configurations and access will follow best practices as outlined in NIST SP800-122.

Participant Data updates (NYU): In accordance with the data protection plan (DPP), pedestrian participants will be provided the ability to redress (update and correct) information during the pilot. Pedestrian participants may coordinate this information through the New York Participant Liaison who is responsible for all participant coordination and participant record keeping. Either hard copy updates or secure web access will be provided to participants.

Participant Surveys and Polls (NYU): Some polls may be performed by physical form and others electronically. All polling and surveying of NYC CVPD pedestrian participants (and drivers, via their fleet

points of contact) will be coordinated by the New York participant liaison through a designated Fleet Operator point of contact. Once collected, all pedestrian poll data containing sensitive PII will be stored in encrypted form and/or locked in a secure safe (no driver poll data will contain PII).

Participant Devices (PIDs): PID information that is PII is the mapping of which pedestrian participant has possession of which PID. This information will be stored encrypted like other electronic PII material. The PID identifier (Serial number) by itself is not PII and will be shared (along with PID operational status information) with the TMC. The PID identifier in isolation is not mapped to any individual. Only NYU will contain this mapping that makes the information PII. NYU will protect this information as PII.

Access Control During Device Installation and Management

ASD and RSU installers will have physical and electronic access to devices during installation and maintenance activities. All installer/maintainer personnel will receive training on device management, data management and data access procedures.

ASD and RSU device inventories consisting of all device serial numbers will be monitored throughout the pilot. Lost or stolen devices will be identified flagged to prohibit such devices from connecting and uploading data to the TMC. Prior to installation and deployment, RSUs and ASDs will be maintained in a secure, locked facility.

Device software/firmware images will be updated throughout the NYC pilot. The TMC device management servers (one for the ASD and one for the RSU) – using secured SNMPv3 - will be used to indicate to fielded devices when firmware updates are available. All software and firmware made available to the ASDs and RSUs will be digitally signed by the device vendor. Devices will successfully verify the digital signatures prior to storing or loading the updated firmware. Failed integrity checks will be noted and identified via SNMP status blocks to the TMC. This mechanism prevents unauthorized access to ASD and RSU processing logic through corruption of the software/firmware supply chain.

Access, Integrity and Availability During ASD and RSU Data Collection and Storage

Data access and integrity will be maintained during collection, upload and storage of device data, as follows:

Logs and Logging Functions in ASD and RSU devices: Each ASD and RSU device will support sufficient memory for log file short term storage. Additionally, log file generation and management functions will possess fault detection indication and reporting features to identify and correct possible faulty data collection. Log files on devices will be purged after 48 hours regardless of whether they've been uploaded. Additionally, physical and logical access to the device will be strictly controlled by the TMC and Fleet Owners.

ASD and RSU devices will support log file integrity value generation and checks. Log files will be integrity protected by trusted processes within each device and any unauthorized log file accesses or modifications will be indicated to NYCDOT ASD and RSU monitoring personnel.

Metadata (Confounding Factor Data) Collection and Use: Metadata such as weather and other metadata sources (confounding data such as weather data, volume data, TMC logs, etc.) will be collected from existing, trusted internal and external portals. This data will be genericized and sanitized of strongly

correlatable details to prevent privacy losses stemming from offline data analysis that merges such metadata with ASD and RSU device data.

Protecting Data-in-Transit: ASD Data Upload Communications with TMC

DSRC communications will make use of IEEE 1609.2 integrity mechanisms to digitally sign all datagrams emanating from ASDs. These digital signatures will ensure the integrity of logged data sent to the TMC (all will be verified by the TMC Collection server).

Some data sent from ASDs to the TMC will be SNMPv3 data. SNMP sessions will be performed over TLS 1.2 (over wireless IP connection (NYCWiN) or fiber) –Each protocol provides endpoint authentication and data origin authentication on each management session datagram.

All ASD data in transit will be encrypted, authenticated and integrity protected when uploading to the TMC. The process will be initiated when ASDs are in range of RSUs and the RSUs transmit a centrally signed (TMC) WSA message requesting specific data types. In other words, it is the TMC that authenticates the request for the data via the signed WSA. Upon the ASD successfully verifying the TMC signature on the WSA, the pre-processed BSMs telematics and other 'event' data will be sent from ASDs to a collection service (the "Collection server") located at the TMC. ASDs will prepare the event records by encrypting all event data using the monthly-refreshed TMC encryption key. The TMC encryption key shall be provided to each device via SNMP-provisioned configuration file updates. The encrypted data will be additionally authenticated via the ECDSA signature generated over the message prior to uploading via the RSU. ASDs will sign their TMC-destined raw data files using a pseudonym certificate. Data uploads will be encrypted to the RSU. There they will be staged and securely transmitted in a store-and-forward manner to the TMC.

Once uploaded to the TMC, the buffered, encrypted ASD raw data will remain in encrypted form (data at rest protection) until the Treatment server (server that performs the normalization and sanitization, as described in Chapter 4, TMC Data Flows and Systems, is ready to decrypt, normalize, analyze and then sanitize the data. Once processed and analyzed – barring detection of any errors – the plaintext raw data will first be zeroized (actively overwritten at least one time in memory) by the Treatment server, and then de-referenced in memory. The Treatment server will then wipe the original raw data copy created by the Collection server. This process will also use an active overwrite mechanism to purge the data. Such a process is necessary to prevent any recovery of sensitive data through hard drive forensics or volatile memory dumps.

The TMC's Data Steward will monitor the collection and sanitization application logs for any reported errors. Due to the sensitivity of the operation, the application will log errors and actively notify operators to minimize the troubleshooting and resolution time. No raw data in either encrypted or plaintext form will be stored for a period longer than 24 hours on the Collection or Treatment servers (in total). Note that on the originating devices, raw data is limited to 48 hours.

Protecting Communications from RSU to the TMC

Access to data will be restricted for uploads from the RSU to the TMC by encrypting, authenticating and integrity-protecting all data in transit between the RSU and TMC.

RSU as Data Gateway: The RSU, acting as gateway, will allow the encrypted payloads to be staged from ASDs to the TMC Collection server. Though the data is already encrypted by the ASDs, the RSU will

relay it to the TMC over an authenticated TLS tunnel either 1) traversing the NYCWiN network or 2) traversing fiber-optic cable to the TMC (depending on the connectivity available at the RSU). The RSU will have no access to sensitive ASD information, as it will not have the private TMC key material needed to decrypt it.

RSU as Data Origin: The RSU will also communicate its own data with the TMC using one of the following protocols:

1. *SNMPv3:* Management data GETs, SETs and TRAP messages needed by the TMC to configure and monitor the RSU. This channel is encrypted and authenticated by a 128-bit AES key derived via the method specified in the SNMPv3 User Security Model (USM).
2. *SSH Version 2:* This channel is also used by the TMC's RSU administrator to connect to the RSU for the purposes of configuring and loading data, as needed. It is mutually authenticated using pre-loaded certificates, and encrypted using one-time use session keys.
3. *TLS:* TLS 1.2 will be used for communicating with the RSU. All SNMPv3 traffic (already protected) will traverse a TLS 1.2 session. The TLS tunnel is mutually authenticated, encrypted and integrity protected.

All data and communications sent between the RSU and TMC are therefore encrypted, authenticated and integrity protected to control access.

Protecting Communications from PID devices

PID devices will be under the control of NYU and the pedestrians who carry the devices. PIDs will only upload event-related data to the TMC over a protected TLS link. PID data will also be collected on the Collection server like ASD and RSU data. All device and application management on the PID will be handled through device firmware updates and/or secure TLS connections between NYU and the PID. NYU will set up and operate a dedicated PID management server similar to the TMC's operation of ASD and RSU management servers.

Access During Short-Term TMC Storage and Data Treatment

Data access during this process is particularly sensitive because it is possible for collection and sanitization processes to fail, resulting in unintentional PII leakage. Additionally, failure to correctly process and make available needed data sets can reduce the analytical effectiveness of the NYC CVPD and independent researchers. This section describes various servers at the TMC used to collect, cleanse and analyze data. Figure 6 on page 58 provides an illustration of these TMC-internal elements.

Collection Processes – *TMC*: The TMC Collection server will be responsible for the upload and temporary storage of raw data. The Collection server will physically reside at the TMC with only authorized individuals allowed physical access. Logical access to the server and its applications will be restricted to the TMC IT staff and Data Steward. Server access logs will be monitored for unauthorized access. No data read access will be possible on the Collection server due to the encrypted state of the data. Once purged, even encrypted data on this server will not be accessible. ***NYU:*** NYU will similarly operate a collection server for PID-originating pedestrian data. It will be protected and monitored similarly.

Treatment Processes: Data treatment occurs on a machine termed the Treatment server (see Chapter 4, TMC Data Flows and Systems, for a full depiction of this process). This server is to be collocated with the data Collection server (also described in Chapter 4, TMC Data Flows and Systems) and placed on the same subnet. Physical and logical access to the sever will be strictly controlled and monitored because the encrypted logs are all decrypted on this machine. The integrity of the data treatment processes – dependent on error-free data reads, parsing, removal and data write functions - will be monitored by the Data Steward who will routinely check processing logs for failures. Once purged, neither encrypted or plaintext raw data will be accessible on this server.

Access During Backup and Recovery

Short-Term Data Backup: All short and long-term data backups will be encrypted and access controlled. Machine backup images will be maintained for the Treatment, Analytics and Bin servers (see Chapter 4, TMC Data Flows and Systems, for more details on the treatment and binning processes). These backup images only pertain to the machines (operating systems and executables), not the short-term data that has not been fully processed. This data is only allowed to age 24 hours on these machines. Data is backed up on the Bin server, however. Once bins are fully populated on the Bin server, they are exported to the onsite archive and queued for selective, periodic export to the IE research portal (CV PEP/SDC). Partially populated bins are not backed up as these can still potentially (though unlikely) constitute PII due to lack of full aggregation and anonymization of the data sets. Based on this approach, no sensitive data is backed up even for short term storage. All backups are encrypted and integrity protected, stored on-premises, physically protected.

Data Recovery: Data and system recovery, if necessary, will be made from the short-term backup solution in the event of system or host application corruption or other failure. TMC administrators in conjunction with the TMC Data Steward will perform system and data recoveries. No other individuals will have access to the backup system restore interfaces.

If a short-term data restoration fails, some data may be lost (partially completed bins) and a server may need to be restored from a backup image maintained by the TMC's IT staff.

Access During Archive (Long Term Storage), Export and Recovery

TMC Onsite Archive: All archived data stored onsite will be integrity protected and access-controlled. Only the TMC's Data Steward and one alternate will be given access to input, configure or read from the TMC's data archive system.

TMC Offsite Archive: The NYC CVPD will support a contracted service to handle and manage archiving of any physical media (backup tapes, disks), as necessary. The backup service provider will support a Service Level Agreement conforming to the requirements of this data management plan and NYC's Citywide information technology policies (New York City Citywide Security Policies (DOITT) - https://www1.nyc.gov/site/doitt/business/it-security-requirements-vendors-contractors.page).

NYU Archives: NYU archiving requirements are in Chapter 3, Pedestrian Data (NYU). The NYU Data Steward will similarly enforce technical and procedural access controls to all NYU-stored electronic data based on NIST SP800-122.

In some cases, NYU and TMC archive material may be exported, imported and merged between NYU and the TMC. In these cases, the data recipient will apply the same procedural access controls to the merged data sets and maintain them as distinct from original archive data.

Intellectual Property Management

All normalized, obfuscated and aggregated data (in the Binned state) is sharable with USDOT, the IE. Relevant participants Intellectual Property (IP) rights for data collected and managed during the NYC CVPD is as follows:

Raw Data Emanating from RSUs:

- *Ownership:* This data is owned by NYCDOT
- *Transfer:* Some RSU raw data will be archived and will continue to be NYCDOT's IP
- *Copyright:* This data is not under copyright protection

Raw Data Emanating from ASDs and PIDs:

- *Ownership:* Data ownership is multi-party, established by Memorandum of Understanding (MOU) agreement between NYDOT and NYC CVPD participating organizations. Note that fleet owners own the ASD devices and can access raw ASD data. While the raw ASD data is encrypted on the device, fleet owners have the option of procuring equipment and the requisite cryptographic key material to decrypt the data if they so wish.
- *Transfer:* Outside of NYC CVPD consumers, ASD raw data will only be retrievable by vehicle Fleet Owners, if they choose to retrieve it. PID raw data will never be transferred outside the TMC.
- *Copyright:* This data is not under copyright protection

NYU Archived Data:

- *Ownership:* This data is jointly owned under MOU agreement
- *Transfer:* Data may be shared/transferred to the NYCDOT TMC and/or to the IE CV PEP/SDC portal under terms of a MOU
- *Copyright:* This data is not under copyright protection

NYDOT TMC and IE Archived Data:

- *Ownership:* This data is jointly owned under MOU agreement
- *Transfer:* Data may be transferred to the NYU and/or CV PEP/SDC (removed of potential PII)
- *Copyright:* This data is not under copyright protection

CV PEP/SDC Exported Data:

- *Ownership:* This data is jointly owned under MOU agreement
- *Transfer:* Data is expected to be available to external researchers under to-be-defined terms of use (this information will be published when the CV PEP/SDC research portal is defined). This is expected to include licensing under either the Creative Commons Attribution 4.0 International license or the Creative Commons Attribution-ShareAlike 3.0 license (https://creativecommons.org/licenses/by-sa/3.0/us/legalcode)
- *Copyright:* This data is not under copyright protection

Cryptographic Key Data Management

Proper cryptographic key management processes are a critical prerequisite to all data security and privacy protections, both in transit and at rest. Compromise of any key material is a substantial risk and implies the potential compromise of PII. All secret and private key material will be protected from unauthorized disclosure, modification and substitution. All public key material will be protected from unauthorized modification and substitution.

In accordance with NYC Citywide Encryption Policy (http://www1.nyc.gov/assets/doitt/downloads/pdf/encryption.pdf),

1. All private keys shall be kept confidential
2. Key lifecycle management shall be implemented and adhered to
3. Keys shall be encrypted in transit and at rest
4. Keys shall be chosen randomly from the entire key space
5. Encryption keys shall be available for retrieval for administrative or forensic use

Table 5 lists the key management data that will be managed according to the New York City Citywide policy, NIST SP800-53, NIST SP800-57 and Security Credential Management System requirements, as appropriate:

Table 5. Management and Use of Cryptographic Keys in the NYC CVPD

Keys	Description
ASD and RSU Enrollment Keys	These are the public and private 1609.2 keys that authenticate a device's certificate provisioning requests to the SCMS **Generation:** Generated at the ASD and RSU manufacturers, and signed by the SMCS **Distribution:** Distributed pre-embedded in the ASD and RSU to the NYC CVPD **Update**: These keys will not be updated during the NYC CVPD. If an enrollment key update is necessary, the device will be returned to the RSU or ASD vendor for re-outfitting. **Zeroization:** Zeroized upon command, or if the ASD or RSU is tampered
ASD Pseudonym Credentials	These keys are used to sign Basic Safety Messages, mobility and event data uploads from the ASD to the TMC. **Generation:** Pseudonym certificates will be generated by the SCMS in responds to a pseudonym provisioning request generated and signed by the ASD enrollment key (on the device) **Distribution:** The SCMS will provide a retrieval URL in a certificate response message. The ASD will retrieve the encrypted pseudonym credentials when ready. **Update:** These keys are short-lived and continuously updated using the initial provisioning method. **Zeroization:** Keys are zeroized because of tamper or local command on the ASD

Keys	Description
RSU Application Certificates	These keys are used by RSUs (roadside) to authenticate application services without the need for anonymity **Generation:** Generated by the SCMS in response to a RSU application certificate request sent to the TMC (and signed by the RSU's Enrollment key) **Distribution:** Distributed to the RSU as a certificate response message from the SCMS **Update:** RSU application certificate updates are performed by sending a new certificate request to the SCMS. **Zeroization:** Zeroized by local or remote command (SNMP) to the RSU
TMC Enrollment Certificate	This key is used by the TMC to sign requests for TMC Application certificates **Generation:** Generated by the SCMS RA during an Enrollment request emanating from the TMC **Distribution:** Distributed securely over TLS to the TMC **Update:** This key will not be updated during the life of the NYC CVPD **Zeroization:** Zeroized upon command on the TMC server
TMC Application Certificates	1609.2 public and private keys used by the TMC application server to centrally sign messages such as connected vehicle MAP and RTCM GPS correction messages destined for RSUs and ASDs. **Generation:** Generated by the SCMS in response to a TMC server application certificate request sent to the SCMS (signed by the TMC Enrollment key) **Distribution:** Distributed to the TMC signing service Hardware Security Module (HSM) using an SCMS certificate response. **Update:** TMC application certificate updates are performed by sending a new certificate request to the SCMS. **Zeroization:** Zeroized by local command at the TMC signing server HSM.
TMC Encryption Keys	The public key is used by ASDs and RSUs to encrypt event data for upload to the TMC. The TMC encryption keys will be made available to ASDs via SNMP-directed configuration file updates. **Generation:** Generated by the TMC and signed by the TMC Application certificate **Distribution:** The private key component is distributed securely to the TMC Treatment server which decrypts the queued data off of the TMC Collection Server. Public keys are distributed through configuration batch files to ASDs and RSUs **Update:** Updated monthly using the same generation/distribution process **Zeroization:** The private key can be zeroized locally by command on the TMC Treatment server

Keys	Description
TMC TLS Root Certificate (Certificate Authority)	Managed by the NYC TMC, this key is used to digitally sign public keys of TMC TLS servers and connected vehicle end entities (RSU). **Generation:** Generated by the NYC CA **Distribution:** Transmitted out of band to the RSU and ASD manufacturer/installers for initial installation and embedment in device firmware **Update:** N/A NYC CVPD assumes no TLS CA updates. If this occurs, full compromise recovery procedures will be initiated. **Zeroization:** NYC CVPD assumes no TLS CA zeroization. If this occurs, full compromise recovery procedures will be initiated.
TMC Collection Server TLS Keys	Client and Server TLS keys signed by TMC Certificate Authority and installed by TMC IT personnel. These keys are used to authenticate the TMC Collection server communicating via TLS to RSUs and ASDs. **Generation:** Generated on the TMC Collection Server and routed through the TMC CA for CA signature **Distribution:** Private keys (N/A) – Collection server public keys sent in the TLS handshake to RSUs and ASDs which subsequently authenticate them using the TMC TLS Root Certificate's public key **Update:** Re-generated on the server and new public key signed by TMC TLS Root CA **Zeroization:** Local command on the TMC collection server
RSU TLS Keys	RSU Client TLS keys used for secure TMC connection with the RSU (SNMP traffic is additionally tunneled over TLS). **Generation:** Generated on the RSU, then sent to the TMC Certificate Authority for signature on the RSU's TLS public key **Distribution:** N/A (though returned from the TMC CA in the form of a certificate response message) **Update:** Same as generation method **Zeroization:** Local or remote command to the RSU
Fleet Owner Encryption Keys	Fleet Owners technically have the capability of retrieving raw ASD data from their vehicles (since they own the ASD devices) by decrypting the data using the Fleet Owner's key. This key is used to encrypt event data on board the ASD prior to uploading to the TMC or the Fleet Owner's system. As such, both the TMC and Fleet Owner can decrypt raw ASD data. Each Fleet Owner's ASD encryption key is unique to ensure only that Fleet Owner has access to its data. Fleet Owner keys are to be managed by the Fleet Owner. NYC CVPD will not be responsible for the secure provisioning and storage of Fleet Owner public and private encryption keys. Fleet Owner encryption keys are out of scope of the NYC CVPD and are solely provisioned by Fleet Owners using their own equipment and software.

Keys	Description
TMC RSU Management Host SSH Keys	Public key certificates and private keys required by the RSU management host to securely log into remote RSUs **Generation:** Generated on the Management Host. **Distribution:** The management host's SSH public key is provisioned over a secured SNMPv3/TLS connection to the RSU. **Update:** Same process as initial provisioning **Zeroization:** Zeroized by command on the RSU management host
RSU SSH Public and Private Keys	Installed by NYCDOT IT staff and used to authenticate the RSU to the TMC and allow TMC authentication to the RSU **Generation:** Initial SSH keys generated and pre-installed on the RSU by the RSU vendor. **Distribution:** N/A, however new RSU SSH public keys will be distributed over SSH or SNMPv3 GET commands to the RSU management host **Update:** Updates to RSU SSH keys can only be performed by the RSU administrator who has an active, secured shell (SSH) with the device. The SSH key generation may be performed, followed by a fresh public key export to the RSU management server. **Zeroization:** Zeroized by command on the device, or zeroized via key update
SNMP Key Material	SNMP keys are AES session keys used to securely encrypt and decrypt SNMP communications between the RSU Management server and the RSUs. They are derived from a shared secret string entered on both SNMP Manager and SNMP agent. **Generation:** These keys are derived from SNMP passphrases through a password-based key derivation function **Distribution:** Initial SNMP passphrases will be provisioned to devices by the device vendor. SNMP best practices **Update:** Subsequent SNMP passphrase updates will be managed between the TMC's RSU/ASD management hosts and end devices via an encrypted configuration file update **Zeroization:** Zeroized by creating a new passphrase and key on both the RSU and ASD Manager hosts and the end devices.

Roles and Responsibilities in Data Management

Table 6 lists the data management activities and the authorized roles that perform them.

Table 6. Data Management Activities by Role

Role	Data Management Activities
CISO (NYC role)	Corporate Information Security Officer within the NYCDOT – This role performs the following functions regarding data management: • Reporting IT security incidents (including privacy breaches) • Overseeing security training • Overseeing data classification and policies
CV Device Manufacturer-DCM (Connected Vehicle Device manufacturer for RSUs and ASDs)	• Perform initial software/firmware loads • Key Management: Install SCMS enrollment keys • Key Management: Install initial RSU SSH keys • Configure SCMS network & reachability information • Generate and distribute software/firmware patches to TMC operators / device managers Note: This role exists uniquely for ASDs and RSUs
Fleet ASD Installer	• Install and configure ASDs in vehicles • Perform diagnostics, collect and analyze host log files • Administer software/firmware patches for ASDs
NYCDOT ASD Monitoring Personnel (Aftermarket Safety Device Installer/Manager)	• Assign ASD identifiers • Verify configurations • Update certificates • Initiate or verify data purges • Key Management: Provision TMC Encryption Keys (via RSUs)
PID Configuration Manager (Personal Information Device Manager at NYU)	• Assign PID identifiers • Provide usage instructions, troubleshooting • Administer software/firmware patches for PIDs • Collect and analyze PID logs • Update certificates • Initiate or verify data purges
NYCDOT RSU Monitoring Personnel (Roadside Equipment Administrator at NYCDOT)	• Perform SNMP/HTTPS management of Roadside Units • Perform collection of application and host log files • Manage security configuration of RSU units • Load software/firmware patches • Verify software/firmware integrity • Collect and monitor RSU logs • Key Management: Update certificates

Role	Data Management Activities
NYCDOT Device Custodian (within NYCDOT)	• Manage physical inventory of connected vehicle RSUs • Label and track devices based on assigned identifiers (serial numbers)
NYCDOT TMC System Administrator (Traffic Management Center System Administrator role within DoITT)	• Maintain data collection service • Manage data sanitization process • Ensure sanitization state of data prior to transferring data to USDOT CV PEP/SDC • Initiate or verify data purges in TMC applications • Monitor local and remote access to TMC systems • Key Management: Manage cryptographic keys in TMC IT systems • Manage anti-virus on connected vehicle servers: validate version and signature files for stand-alone computers that are not connected to the network • Manage the software component inventory and associated software licenses
TMC Network Security Monitor (NYCDOT and DoITT personnel)	• Monitor Network Intrusion Detection Systems (NIDS), including firewalls and other sensors • Monitor host-based Intrusion Detection Systems (HIDS) and Anti-Virus
TMC Data Steward (Role within NYC Department of Information Technology and Telecommunications [DoITT])	• Ensures all information assets are valued, classified and labeled as either public, sensitive, private or confidential • Manage TMC databases • Maintain NYC CVPD project documentation • Perform data purging functions on all databases and application servers • Evaluate and monitor data integrity functions • Archive host and system logs • Purges data manually or ensures automated purges are completed prior to IE upload or sending to the USDOT CV PEP/SDC portal
IE (Independent Evaluator)	• Analyze normalized data sets • Consent and advise on data analytical methods • Consent and advise on survey models and content • Interview and survey participants
NYU Pilot Participant Liaison	• Designated point of contact for pedestrian participants in the NYC CVPD • Provide consent forms and literature to participants • Collect participant hard copy and electronic PII/SPII data • Coordinate participant surveys, interviews and polls • Coordinate with Fleet Owner designated points of contact for collecting anonymous driver polling data

Role	Data Management Activities
NYU Security and Privacy Manager	• Secures accesses to electronic systems at NYU that contain participant PII and SPII.
NYU Data Steward	• Ensures all information assets are valued, classified and labeled as either public, sensitive, private or confidential • Manages NYU databases and networked data storage devices • Maintains NYC CVPD project documentation for NYU • Performs data purging functions on all databases and application servers • Evaluates and monitor data integrity functions • Archives host and system logs • Purges data manually or ensures automated purges are completed prior to IE upload or sending to the CV PEP/SDC
Privacy and Security Plan Administrator	• Perform data custodial functions for pedestrian participant PII/SPII data • Maintain privacy and consent forms • Oversee NYC CVPD adherence to privacy protection policies • Maintain and archive collected privacy consent forms • Administer privacy and security training and related materials to NYC CVPD program personnel • Facilitate participant redress of privacy-related issues • Notify and coordinate with USDOT on any privacy-related breaches and other concerns
IRB (Institutional Review Board)	• Approve privacy and data plans • Approve participant consent forms
USDOT (US Department of Transportation)	• Provide data analysis analytics goals and oversight • Provide program oversight

Personnel Qualifications and Training

All NYC CVPD TMC and field personnel shall receive and maintain security privacy protection training in accordance with the New York Citywide Information Personnel Security Policy (http://www1.nyc.gov/assets/doitt/downloads/pdf/Personnel_Policy.pdf). All personnel with access to PII will receive training in the policies and procedures that protect PII.

All TMC personnel training records will be managed and maintained by the CISO in accordance with New York City policy.

All NYU personnel training records will be managed and maintained by the NYU Security and Privacy Manager.

Data Quality Control

Data quality assurance throughout the data management lifecycle ensures that data collected and produced during the NYC CVPD results in reliable analytical results.

Data Quality Facets

Six facets of data quality are addressed in this data management plan, as shown in Table 7.

Table 7. Data Quality Facets

Data Quality Facet	Data Quality Management Approach
Relevance	Addresses the degree to which NYC CVPD data products meet the analytical and record-keeping goals of NYCDOT, USDOT and researchers. This plan provides the data collection, data treatment, integrity and handling processes that provide the machinery to provision relevant connected vehicle analytical data and metadata. Data relevance will also be enhanced through a) customized experiments and b) steadily improving device behavior. These processes are facilitated by the ability to dynamically update the entire ASD fleet or portions of the fleet with custom firmware and/or custom threshold settings by custom groups. To this end, data relevance will be continuously improved. Firmware version and settings groups comprise custom metadata that will be collected and managed for control group based analysis.
Accuracy	Aberrant data is identified from errors and other conditions (e.g., GPS outages, lack of GPS corrections, device errors, violation of data thresholds, etc.) logged in ASDs and RSUs. Corrupt or otherwise bad data is identified at the source, allowing data uploaded to the TMC to be flagged as needing to be rejected or placed in a suspect or reject bin. These mechanisms ensure data accuracy and quality, preventing downstream corruption of intermediate and final data sets. Some aberrant data will be identifiable in the source (lack of data or data outside of normal ranges). These items will be caught in early data processing). Other aberrations may be caught via other means such as from device log files. For example, loss of GPS will appear in an audit log and will be retrieved by the device manager at the TMC when auditing the devices over SNMP. Some data accuracy items will be identifiable by both means. The NYCDOT team will institute Q/A processes that 1) trace erroneous data back to source, 2) identify root cause(s), 3) update audit processes to better capture such errors at the source and 4) update data processing scripts to better flag and/or filter known data items that are found to be erroneous. Example data thresholds that can be caught using these techniques include 1) vehicles that appear to travel at hundreds of miles per hour due to GPS ‘urban canyon’ problems in NYC, and 2) devices that have not been able to obtain fresh GPS correction updates (thus, whose positions do not appear to be on known roads or parking lots).

Data Quality Facet	Data Quality Management Approach
Timeliness	The potential for stale data (needing to be processed)) is avoided by the ASD requirement to upload all mobility and event data to the TMC within 48 hours of collection. Wave Service Advertisements (WSA) will be sent from a sufficient number of fielded RSUs to enable timely (within 48 hours) ASD data uploads. All real-time and near real-time data needed for the CV applications to function correctly will be available in a timely basis per the design and configuration of the RSUs, ASDs and the TMC monitoring applications.
Accessibility	Sanitized, normalized and binned data sets will be made available to researchers
Interpretability	Data and metadata formatting descriptions, encodings, and parsing rules will be published in a project repository for researchers planning to read and interpret data sets. Data interpretation documentation will be provided to the IE's CV PEP/SDC enclave.
Transparency	All published data sets will be clearly purged of all PII and as a result, end researchers will not have access to raw data. To facilitate open and transparent research, a description of the data collection, treatment (normalization and sanitization) processes will be made available to researchers who need to qualify their source data sets.
Utility and Results	While not all descriptors are yet available for specifying the results of the NYC CVPD data analysis and studies, these will become available in the first quarter of Phase 3 as the analysis is underway.

Reproducibility and Data Processing Integrity

Due to inherent need to protect the privacy of participants in the NYC CVPD, discrete data sets collected will not be 'reproducible' as such, however data management integrity processes will be sufficiently strong to ensure a consistent, reproducible data collection and management function. Data will be collected, sanitized, normalized and purged as necessary in one-way functions to simultaneously facilitate connected vehicle data analytics and the protection of the participants. Robust, integrity-enforcing controls shall be instituted throughout the data management lifecycle, as described in Chapter 3, Security and Privacy Procedural Controls by Data Management Activity.

Data Preservation and Archiving

This section describes the policies and processes necessary to preserve and process NYC CVPD data.

TMC Originated Data

1. All treated and exportable data (completed data bins fully normalized, sanitized, aggregated and placed in the Binned state) produced by the NYC CVPD will be archived and maintained at the TMC for a minimum 6-year period (per New York State Transportation and Engineering archive requirements (http://www.archives.nysed.gov/records/retention_mu-1_transportation-engineering)).

2. Data will be reviewed for privacy-related content prior to export and archiving
3. All archived data will be in CSV format per RFC 4180
4. All data archives will include data archive format descriptions
5. TMC-generated data will be archived continuously at the TMC. This data will consist of the completed data 'bins' as described in Chapter 4, TMC Data Lifecycle: Post-Collection Processing, Treatment Analysis and Storage.
6. The TMC Archive updates will be periodically replicated to a backup system and stored offsite.
7. The offsite archive backups will be updated no less frequently than once per quarter
8. The offsite archive will be available for a period of 6 years following the conclusion of the 18-month NYC CVPD period.
9. All TMC archive data will be cryptographically integrity protected and signed (may use the backup system's integrated integrity tools to accomplish this)
10. If PII is later found to be resident in archived data, a sanitization and re-archive process (and toolset) will be created to cleanse the archive of PII
11. All backups will be made to a secured, environmentally controlled facility
12. The external backup provider will support a Service Level Agreement consistent with the NYC CVPD archive requirements

Pedestrian Data (NYU)

1. NYU will expunge all encrypted pedestrian participant PII and SPII data following a 6-month period after data analysis has completed.
2. NYU analytical results may be published and archived after analytical results have been evaluated to contain no PII. The NYU Data Steward will coordinate a final PII check to ensure no candidate archive data has PII before going to archive.
3. Pedestrian or driver survey and poll raw data will be archived based on a determination of future needs for that raw data. This determination will be made by NYU researchers under IRB oversight and be made in accordance with NYU policies. Any archived raw data will be stored in encrypted form.
4. If unintended PII is later found to be resident in NYU-archived data, a sanitization and re-archive process (and toolset) will be created to cleanse the archive of PII
5. All NYU-related NYC CVPD data archives consisting of pedestrian study results (metrics, statistics) and select raw data will use a media-based backup system or a secured cloud backup system that can provide a Service Level Agreements that meet documented retention periods either established by this plan or periods later determined to be necessary in accordance with NYU policy.

Chapter 4. New York City Pilot Data Plan

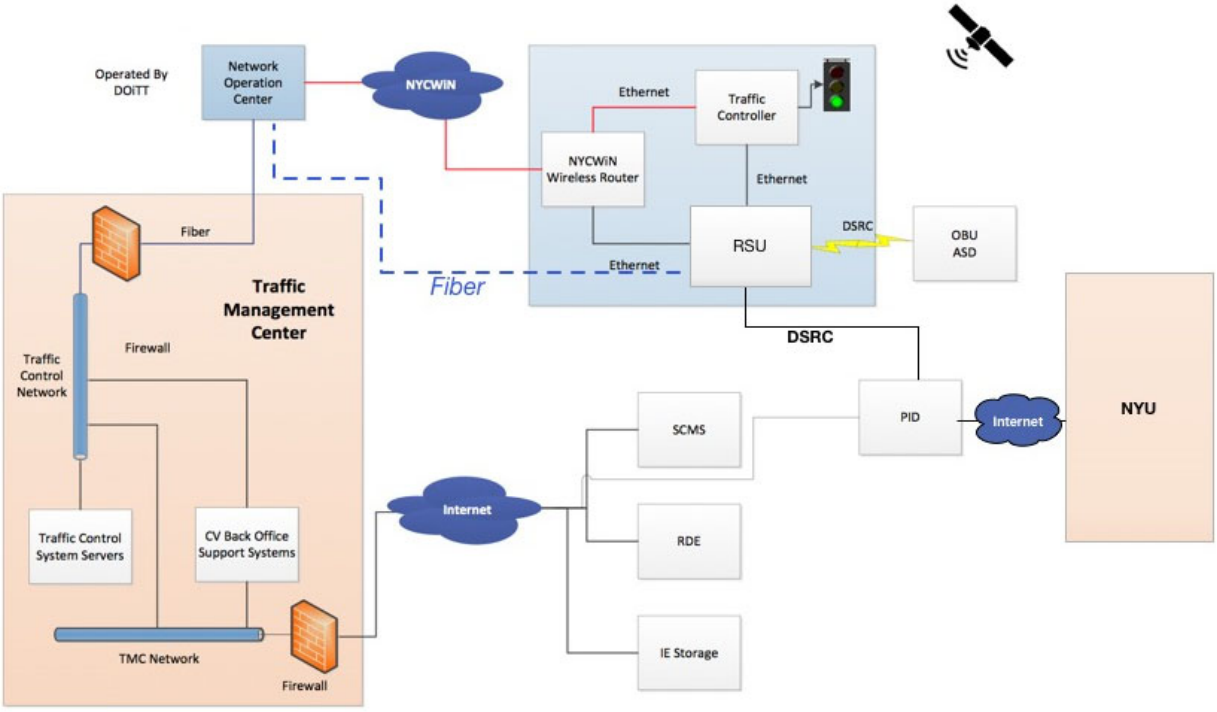
This chapter delineates the types of data generated and/or used in the New York City Connected Vehicle Pilot Deployment, including its treatment, storage and quality assurance.

The data collection is divided into 5 general areas:

1. Participant Data (including Pedestrian and Fleet Owner Information)
2. Device Operations and Maintenance (O&M) data
3. Mobility (performance) data
4. Event data
5. Device Management and Connected Vehicle application data

This data management plan describes data management with respect to the above categories rather than by each individual CV application. This is due to the substantial overlap in data types and sources between the various CV applications.

The bulk of all collected data at the New York City Traffic Management Center and New York University is 'Evaluation data.' Evaluation data consists of different types collected to support New York City (NYC) and the US Department of Transportation's (USDOT) evaluation of the safety benefits from connected vehicle technology. Accordingly, the potential for helping reduce crashes, crash severity and other safety hazards poses enormous public implications on the value of this data. Researchers evaluating data identified in this Data Management Plan (DMP) will be able to better interpret and configure applications and devices throughout the connected vehicle ecosystem to help save lives. Figure 2 depicts the networked endpoints involved in collecting data.



Source: NYCDOT, 2017

Figure 2. Network Connectivity Architecture

Data Types, Sources and Destinations

The following tables provide, based on data management activity, the data types and data flows involved in the NYC pilot.

Participant Information Collection

The participant data, shown in Table 8, will be collected by the NYC CVPD.

Table 8. Security and Privacy Controls by Data Management Activity

Data Type	Description	Source	Destination
Pedestrian Participant Data, including PII and SPII (This data is only managed by and stored at New York University with Institutional Review Board [IRB] oversight)	Includes the following: • Full name (Last, Middle, First) • Address • Phone and Email contact information • Demographics • Pedestrian traveling trends (typical travel frequency, distances, etc.) • Unique pilot-specific identifier(s) assigned by NYC CVPD (Equipment) • Interviews, survey and polling response data (Note: Pedestrian Participant data does NOT include raw data emanating from Pedestrian Information Devices that is sent to the Traffic Management Center – by definition, such data contains no participant-identifying information or PII)	Participant, via population of physical and electronic forms	**Physical:** Access-controlled safe at New York University (NYU) (See Chapter 4, New York University) **Logical:** Human Participant tables in Participant Database at NYU
Fleet Vehicle Information	Vehicle Information, including: • Vehicle Identification Number (VIN) • Vendor information • Vehicle type • Vehicle role • Equipment (Aftermarket Safety Device) identifier markings (e.g., 32-character serial numbers) Fleet vehicle info stored by the TMC will not include driver information, driver schedules, etc.	Fleet Owner/Operators and CV Equipment Maintenance Personnel (manages vehicles' Aftermarket Safety Devices (ASD) serial numbers in conjunction with vendors)	**Physical:** TMC project repository server at TMC (see Chapter 4, TMC) **Logical:** Vehicle tables in Participant Database (Traffic Management Center)

ASD and RSU Operations and Maintenance (O&M) Data

Operations and maintenance data will be collected from Aftermarket Safety Devices (ASD) and Roadside Units (RSU) during the New York City Connected Vehicle Pilot Deployment (NYC CVPD). This data will facilitate Radio Frequency performance and adjustment and other statistical measures pertaining to ASD-ASD and ASD-RSU 'sightings.'

Table 9. Operations and Maintenance Data Types

Data Type	Description	Source	Destination
ASD Sighting RSU Logs	Information collected from ASDs indicating sighting of a RSU. (Note: A sighting is a specific event in which one device, an ASD comes, across another ASD or RSU. This information is useful in the analysis of the radio performance, antenna placement and orientation, and in probability studies regarding frequency, duration and location of CV device-to-device interaction)	ASD	TMC, via RSU
ASD Vehicle-to-Vehicle (V2V) Encounter (Sighting) Log	Information collected from a vehicle's ASD indicating sighting of another vehicle's ASD	ASD	TMC, via RSU
RSU Sighting ASD Logs	Information collected from a RSU indicating it has sighted a vehicle ASD	RSU	TMC

Mobility (Performance) Data

Mobility data pertains to information collected by the NYC CVPD to analyze travel time and overall mobility performance.

Table 10. Mobility Data Types

Data Type	Description	Source	Destination
ASD Probe Data "Breadcrumb" Logs	A configurable number of BSMs - indicating vehicle location and pertinent metadata - that the ASD will log in a given unit of time and store within a single breadcrumb log. Note that the breadcrumb collection rate is on-demand configurable over SNMP. Raw ASD Probe data breadcrumb logs are processed and analyzed at the TMC. They are not forwarded in raw format to any other party and are only sent to the TMC. While the raw breadcrumb lobs are not shared, some of the raw Basic Safety Messages (BSM) contained within the logs may be shared with the Independent Evaluator.	ASD	TMC, via RSU

Data Type	Description	Source	Destination
Processed ASD Mobility Data	Analyzed mobility data (sourced from breadcrumb data) may be sent to the IE, but no raw breadcrumb data logs. This would include a list of travel times by block in certain time periods (likely 5-15 minutes, but possibly as low as 90 seconds to match cycle lengths and Midtown-in-Motion data)	TMC	IE
RSU Travel-Time Source Data	This data, consisting of a single, passing ASD's BSM (the one BSM that is closest to a demarcated point in each instrumented intersection) is collected by the RSU to facilitate travel time computations for the traffic control system. This data will be limited and used by the TMC to develop link travel times to compare the Connected Vehicle (CV) data with the data currently being collected by the RFID toll tag readers strategically placed for the Midtown in Motion (MIM) adaptive control area.	RSU	TMC
Analyzed RSU-Originated Travel-Time Data	RSU Travel-Time source data will be analyzed at the TMC and made available to the IE. It will consist of link travel times only.	TMC	IE

Event Data

ASD Event Data pertains to data collected regarding safety-related events, collisions or other hazards. Event data collection is triggered by specific data patterns detected by the ASD in certain applications.

Fundamentally, events are triggered whenever an alert is generated by each onboard CV application.

Example event triggers based on CV application types include:

- Speed Compliance application events (driver receives an alert for these):
  - If
    - Time of day and day of week (from ASD clock) within schedule-constrained speed schedule (for a given roadway segment), AND
    - Vehicle location within roadway segment per speed schedule, AND
    - Vehicle speed greater than current regulatory speed for roadway segment at given time in the schedule

    - Then
        - Trigger alert to driver and collect Event data
- Oversize vehicle application event (driver receives alert for these)
    - If
        - ASD location (latitude/longitude) is on roadway segment for which height restriction exists, AND
        - ASD-recorded vehicle height within configurable threshold to roadway's height restriction (receive from RSU MAP messages), AND
    - Then
        - Trigger alert to driver and collect Event data
- Pedestrian in signalized intersection warnings (driver and/or pedestrian receives alert)
    - If
        - If ASD location on roadway segment with signalized intersection (from RSU MAP message), AND
        - Vehicle approaching signalized intersection, AND
        - ASD receives SPaT with 'pedestrian in roadway' flag set (originates from signal controller sensor feedback to RSU)
    - Then
        - Trigger alert to driver and collect Event data

Event data triggers (application alert thresholds) will be defined and adjusted throughout the NYC CVPD based on lessons learned. Note that some events may be configured to trigger that do not correspond to a driver alert. For example, a speed compliance application event may be configured to trigger at a threshold that is below the driver alerting threshold.

Table 11. ASD Event Data

Data Type	Description	Source	Destination
ASD Event Data	Fine resolution Basic Safety Messages (BSM), Signal Phase and Timing (SPaT) and/or MAP messages collected by vehicles involved in a custom-defined 'event.' Event data can range from several seconds or minutes prior to and succeeding the event.	ASD	TMC, via RSU Processed ASD Event data (not raw event data logs) will be sent to the IE.

Safety-related event data that has been fully sanitized and obfuscated will be uploaded to the IE via USDOT's Connected Vehicle Performance Evaluation Platform/Secure Data Commons (CV PEP/SDC). Once managed in the CV PEP/SDC, data sets will be controlled and made accessible to associated, authorized IE (TTI) and the Volpe Center. Examples of safety-related event to be provided via the CV PEP/SDC include those listed in Table 12.

Table 12. Example Safety-Related Event Data for the IE

Event Type	Event Metadata	Sanitized Message Data
Speed-Compliance Event	Timestamp, control parameters, event logging parameters, alert verification, threshold values	MAP Message with speed limit Host Vehicle BSM Pre-trigger (X seconds) Post-trigger (Y seconds)
WorkZone Speed Compliance Event	Timestamp, control parameters, event logging parameters, alert verification, threshold values	TIM Message with speed limit and work zone limits Host Vehicle BSM Pre-trigger (X seconds) Post-trigger (Y seconds)
Curve Speed Compliance Event	Timestamp, control parameters, event logging parameters, alert verification, threshold values	TIM Message with speed limit Host Vehicle BSM Pre-trigger (X seconds) Post-trigger (Y seconds)
Red-Light Violation Warning Event	Timestamp, control parameters, event logging parameters, alert verification, threshold values	MAP Message for two nearest intersections SPaT Messages for two nearest intersections Pre-trigger (X seconds) Post-trigger (Y seconds) Host Vehicle BSM Pre-trigger (X seconds) Post-trigger (Y seconds) Remote Vehicle 1 BSM Pre-trigger (X seconds) Post-trigger (Y seconds) Remote Vehicle N BSM Pre-trigger (X seconds) Post-trigger (Y seconds)
PedInCrosswalk Event	Timestamp, control parameters, event logging parameters, alert verification, threshold values	MAP Message for two nearest intersections SPaT Messages for two nearest messages Pre-trigger (X seconds) Post-trigger (Y seconds) Host Vehicle BSM Pre-trigger (X seconds) Post-trigger (Y seconds) Remote Vehicle 1 BSM Pre-trigger (X seconds) Post-trigger (Y seconds) Remote Vehicle N BSM Pre-trigger (X seconds) Post-trigger (Y seconds)

ASD and RSU Device Management and Connected Vehicle Data

Field connected vehicle devices will require periodic firmware patches and reconfiguration. The following data types pertain to ASD and RSU endpoint management.

Table 13. ASD and RSU Device Management and Connected Vehicle Data

Data Type	Description	Source	Destination
Device Command Messages	Command messages sent from TMC to roadside devices. Command messages may be SNMPv3-based (GET, SET, TRAP) or SSH-based (shell) commands.	TMC Management Applications	End devices (RSU)
ASD Firmware	Firmware and firmware patches for the in-vehicle Aftermarket Safety Devices. Firmware is digitally signed by the vendor. New images and patches are downloaded by devices upon receiving a specialized WSA (Wave Service Advertisement) from select RSUs (likely at the fleet barns).	ASD Vendor	ASD, via TMC and RSU
ASD Application Configuration Data	Configuration parameters for ASD processed applications	TMC ASD Management Server	ASD, via RSU
RSU Firmware	Firmware and firmware patches for RSUs. RSU firmware is updated by command from the TMC. The download location is provided to the RSU.	RSU Vendor	RSU, via TMC's RSU Management Server
RSU Application Configuration Data	Configuration parameters for RSU processed applications	TMC RSU Management Server	RSU
ASD Operations Logs	ASD log files pertaining to monitored activities of ASD applications. This data comes in the form of status blocks to the TMC.	ASD	TMC ASD Management Server
RSU Operations Logs	RSU log files pertaining to monitored activities of RSU applications. This data comes in the form of status blocks to the TMC.	RSU	TMC RSU Management Server

Data Type	Description	Source	Destination
MAP messages	MAP messages conform to SAE J2735 and are transmitted by RSUs to indicate intersection or roadway geometry. Signal Phase and Timing (SPaT) messages are used in conjunction with map messages to provide Vehicle to Infrastructure (V2I) application warnings to drivers. MAP messages are generated at the TMC, signed at the TMC and distributed by SNMPv3 to the respective RSUs.	TMC Mapping system	ASD (via RSU)
RTCM Messages	These represent the Radio Technical Commission for Maritime Services (RTCM) different GPS update messages that originate from RTCM equipment fielded around NYC. These are sent over a secure channel from RTCM reference stations to the TMC. These messages are used by ASDs and RSUs to correct for atmospheric effects as well as ephemeris-related position errors from GPS signals. RTCM reference station will be assigned and map to each RSU. RTCM messages are signed (vouched for) by the TMC's signing system (used to sign MAP, TIM and RTCM data)	RTCM reference stations	ASDs and RSUs (via the TMC and its central signing mechanism)
TIM Messages	Traveler Information Messages transmitted by RSUs to convey miscellaneous road-related information (e.g., work zones). TIM messages are signed (vouched for) by the TMC's signing system (used to sign MAP, TIM and RTCM data)	TMC	ASDs (via RSU)
IEEE 1609.2 and other Cryptographic Keys	IEEE 1609.2 keys include ASD and RSU device enrollment keys, RSU and TMC application keys, ASD pseudonym credentials (for signing BSMs) as well as the TMC and vehicle fleet encryption keys. See Chapter 3, Cryptographic Key Data Management, for more information on cryptographic keys and cryptographic key data management.	Originate from device-to-SCMS interactions	ASD RSU TMC Servers Fleet vehicle servers

PID Data

NYU will separately manage the full suite of pedestrian participant and PID data. NYU will maintain (and encrypt) a mapping of which pedestrian participants maintain possession of which PIDs. The subset of PID data collected and used by the TMC will only be an inventory of what PID devices are fielded and which are operational. Only a list of operational PIDs and their identifiers will be provided to the TMC (no pedestrian-identifying information).

Data Quantities and Collection Constraints

This section provides estimates on the amount of field data needing to be collected, along with conditions and constraints that impact that collection.

CV Data Quantities

Most data in the NYC CVPD will be comprised of the raw data emanating from ASDs and RSUs. This section provides estimates for raw and processed data to be generated in the NYC CVPD. These estimates will impact device storage requirements, frequency of upload to the TMC, and storage volume at the TMC. The magnitude of raw and processed data volume will be closely monitored over time to anticipate and respond to any needed data storage needs, for example by increasing storage at the TMC or changing the frequency at which devices upload data.

Vehicles (ASD)

The NYC CVPD fleet is expected to be 3,000 vehicles among three (3) different vehicle types. Table 14 lists vehicle types and anticipated, daily operational hours.

Table 14. Quantity of Vehicles by Type

Vehicle Type	Quantity	Approximate Daily Operating Hours	Days per Week
NYCDOT Fleet Vehicles	1230	8	5
Department of Citywide Administrative Services (DCAS) Fleet Vehicles	1759	8	5
Buses	11	12	7

Correcting for anticipated days per week, the average, weighted number of vehicle-hours per day is approximately ***14 hours***.

Two categories of data will be collected from each participating vehicle: mobility-related data (breadcrumbs) and event data pertinent to ASD-generated warnings or other defined events based on ASD sensor data.

Mobility – Probe "Breadcrumb" Data: Each vehicle ASD will collect and store a configurable amount of all its self-generated J2735 Basic Safety Messages (BSM), approximately one per second out of the 10

per second generated. These 'breadcrumbs' will be stored as the *Probe Data Log*. Each entry will consist of vehicle location, heading, speed and path history. In other words, breadcrumb data will consist of the J2735 BSMCoreData elements indicated in Table 15.

Table 15. BSM J2735 Core Data Fields

BSM CoreData: Size=80 Bytes	Description
msgCnt	Message Count
id	Temporary ID of vehicle
secMark	Message Time
lat	Latitude of vehicle
long	Longitude of vehicle
elev	Elevation of vehicle
accuracy	ASD estimation of location sensor accuracy
transmission	State of the vehicle's transmission
speed	Vehicle speed
heading	Vehicle heading
angle	Steering wheel angle
accelSet	Vehicle acceleration state in all three axes
brakes	Vehicle braking status
size	Vehicle size

Additionally, the J2735 Part II path history data element call 'crumbData' may be included, depending on configuration. Crumb data allows for finer path history resolution within the time window of each breadcrumb BSM. The 'crumbData' element includes the repeating sub-fields (one for each path history crumb) indicated in Table 16.

Table 16. BSM J2735 Part II Crumb Data Fields

J2735 Part II optional crumbData	Description
elevationOffset	the elevation offset from the BSM's CoreData:elev field
heading	vehicle heading at the time offset of this crumb
latOffset	the latitude offset from the BSM's CoreData:lat field
lonOffset	the longitude offset from the BSM's CoreData:lon field
posAccuracy	position accuracy with regard to multiple axes
speed	vehicle speed at time = BSM time + this crumb's timeOffset
timeOffset	the time offset from the time of the parent BSM

The BSM's core data (and if configured, crumb information) will be uploaded daily to the TMC daily through RSUs. Breadcrumb data structures are stored on the ASD and uniquely tagged and processed separately at the TMC for mobility analytics. An average BSM (breadcrumb) size of 80 Bytes results in approximately *4MB (1x/sec*60sec/min*60min/hr*14 hr*80 B/BSM = 4,032,000 Bytes) of breadcrumb data per vehicle per day.* This equates to 32.26 GB breadcrumb-related BSMs per day for all 3000 vehicles in the NYC deployment. No other message types other than BSMs are included in breadcrumb data logs.

Note that breadcrumb data will be purged from the vehicle probe data log for a pseudo-random period of time within a configurable range prior to and after the occurrence of an "event" such as an alert or warning issued to the driver. This same action will be taken whether or not the audible alarm is actually presented to the driver since there will be a control group and a period of tuning when audible alarms will be suppressed even when detected. This step is being taken to ensure that the breadcrumb data cannot be used in conjunction with other data to identify a specific vehicle. Further, the vehicle ID stored with the BSM will be the pseudo random vehicle ID which changes periodically further limiting the ability to use this data to track specific vehicles (raw breadcrumb data logs containing ASD serial number will not be sent beyond the TMC. Breadcrumb data will be sanitized of vehicle identifier). Presence of stored breadcrumb data during an interval of time implies no event has been triggered during such period.

Event - Safety Data: Event data constitutes BSM, SPaT and MAP messages collected during ASD generated warnings or other defined trigger events based on ASD warnings. This data is sourced both from the host vehicle, remote vehicles and nearby RSUs.

BSM Messages

Each BSM core data record will average approximately 80 Bytes in size.

Event BSM data will constitute the following SAE J2735 CoreData fields.

Table 17. BSM Core Data Fields

BSM CoreData: Size=80 Bytes	Description
msgCnt	Message Count
id	Temporary ID of vehicle
secMark	Message Time
lat	Latitude of vehicle
long	Longitude of vehicle
elev	Elevation of vehicle
accuracy	ASD estimation of location sensor accuracy
transmission	State of the vehicle's transmission
speed	Vehicle speed
heading	Vehicle heading

BSM CoreData: Size=80 Bytes	Description
angle	Steering wheel angle
accelSet	Vehicle acceleration state in all three axes
brakes	Vehicle braking status
size	Vehicle size

Event assumptions and subsequent BSM message volumes as contained in Event data are conservatively approximated as follows:

- 1 Event per 5-minute interval per vehicle
- 60 seconds of collection per event
- Event captures one host vehicle and 10 other vehicle BSMs
- Total Event BSMs per vehicle-event = 528K
- Total Event BSMs per vehicle per day = 88.7 MB
- Total Event BSMs for all 3000 vehicles per day = 266 GB

SPaT Messages

SPaT messages are generated and transmitted by RSUs every 100 milliseconds and may be include in a vehicle's Event log, if captured. SPaT message data as recorded in the Event log are average approximately 120 Bytes in size (depending on the size and complexity of the intersection and its movements). Event log SPaT data will include J2735 SPaT fields indicated in Table 18.

Table 18. J2735 SPaT Message Fields

J2735 SPaT Field	Description
Timestamp	The indicated time when the SPaT was generated by the signal controller's RSU
IntersectionState	A single intersection's state
Name	Name of intersection
Id	Intersection Reference Identifier
revision	Message count
status	Intersection status
moy	Minute of the year
timestamp	Time as measured by this intersection controller
enabledLanes	Lanes enabled in the intersection
States (one or more)	A list of movement lists, each containing a lane and its signal phase and timing state

J2735 SPaT Field	Description
signalGroup	Group ID used to associate to lists of lanes. Lane descriptors match those in accompanying MAP messages (described next) that describe lane/roadway topology
state-time-speed	
signal phase state	Signal phase of the lane (11 possible phases)
time change details (phase timing information)	Indicator when phase state will change
advisory speeds	(Optional) Advisory speed for the lane

Event assumptions and subsequent SPaT message volumes as contained in Event data are conservatively approximated as follows:

- 1 Event per 5-minute interval per vehicle
- 60 seconds of collection per event
- Event recorder captures 16 SPaT messages
- Total Event SPaT data size = 1920 Bytes
- Total Event SPaTs per vehicle per day = 323 KB
- Total Event SPaTs for all 3000 vehicles per day = 0.97 GB

MAP Messages

Event logs containing MAP messages are used principally for understanding the geometry of the intersection to which each recorded SPaT message pertains. MAP messages – each ranging from approximately 100 to 300 Bytes in size (dependent on the complexity of the described roadway or intersection) as stored in the log file - are generated at the TMC and downloaded to each RSU for transmission. MAP messages will be transmitted at a rate of once per second. Table 19 lists the MAP data that will be collected from events.

Table 19. J2735 MAP Message Fields

J2735 MAP Message Field	Description
time	When the map message was generated by the TMC
msgIssueRevision	Message count data
****** One or more intersections ******	
intersections	List of one or more intersections referenced by this MAP), each with fields indicated below:
id	Intersection reference ID. This will pertain to the SPaT's intersection state data.
revision	Message Count (for the given intersection)

J2735 MAP Message Field	Description
speedLimit	List of speed limits, one for each lane
laneSet	List of lane descriptors
(if MAP is describing one or more road segments)	
roadSegments	(list of roadway segments, each with fields indicated below)
id	Road segment's unique reference ID
revision	Revision of the road segment definition
refPoint	Reference point used for offsets that describe the road segment
laneWidth	Width of the lane in the roadsegment
speedLimits	List of speed limits, one for each lane in the road segment
roadLaneSet	A list of lanes in the road segment, and descriptors for each

Event data includes MAP and SPaT messages transmitted by RSUs and recorded to the ASD. An event can consist of a single vehicle or multiple vehicles. Note that the thresholds that can trigger event data collection are configurable in addition to the amount of time before and after the event. The quantity of event data is based on certain assumptions regarding the frequency of hazardous events. Assuming an approximate 3 to 12 events per hour per vehicle, a variable 16 to 60 seconds time range of collection, and variable grid sizes (impacting number of surrounding vehicle BSMs), event data collection estimates (per ASD) for NYC CVPD *range from 1.1MB to 54MB per vehicle per day*. Given the wide potential range, actual data volumes will be closely monitored in early phases of the deployment to determine impacts to data management and storage approaches.

Event assumptions and subsequent MAP message volumes as contained in Event data are conservatively approximated as follows:

- 1 Event per 5-minute interval per vehicle
- 60 seconds of collection per event
- Event recorder captures 2 MAP messages per event
- Total Event MAP data size = 2000 Bytes
- Total Event MAPs per vehicle per day = 336 KB
- Total Event MAPs for all 3000 vehicles per day = 1.01 GB

RSU

RSUs will also collect raw data about nearby ASDs. This information is needed by TMC systems to compute travel times. Mobility application data collected at the RSU will be approximately one vehicle BSM, or 80 bytes, for each vehicle-RSU encounter. RSU sighting info will be near real-time (estimated < 2 seconds) and then purged. This will result in a dynamically updated set of ASD sightings to the TMC for

use in travel time calculations. The RSU will purge this information once successfully transmitted to the TMC.

PID

Only trajectory-related event data will be collected from PIDs. Like ASDs, PIDs will have a configurable set of 'event triggers' causing the PID to begin cataloging connected vehicle message data (BSMs, MAP and SPaT) to its event record. Due to the small number of PID devices, the total daily amount from all devices is not expected to be high. The per-device daily total will be assumed to be similar to ASD daily totals, or 1.1MB to 54MB per device per day. This data will be collected and encrypted on the PID and transmitted securely to the NYU database on a daily basis using the 4G/LTE wide area communications feature of the PID.

Evaluation Data

The bulk of all data transferred to the TMC is in the form of mobility (breadcrumb) and event (safety/hazard) evaluation data. Data totals are based on 1) 3000 vehicles and 2) 548 operational days in the NYC CVPD. Combining daily anticipated data collection per vehicle results in the following quantities:

Table 20. Total Data Originating from ASDs

ASD Data Type Max	Quantity (per Vehicle / Day)	Daily Totals for Mobility and Safety Analysis (3000 Vehicles)	Project Totals (365 Days)
ASD-Originated Mobility (Breadcrumb) Data	4MB	12GB	12TB
ASD-Originated Safety (Event Data) Collected Data	54MB	203GB	158TB

RSUs will also contribute to the evaluation data quantities in the form of travel time source data (mobility data) and RF data from ASD-equipped vehicles. Based on each intersection passing of 3 BSMs per passing (two for RF and one for mobility), approximately *24MB/day* of O&M and mobility data will be collected from each RSU and uploaded to the TMC Collection server.

RSU Communications and NYCWiN Network Constraints

There are ~12,300 signalized intersections that are wirelessly connected (via NYCWiN) and provide real-time data to the TMC. The system uses dynamically configured exception based reporting (NTCIP 1103) and polled data retrieval. Historical collection mechanisms like removable storage are not practical due to the fleet size and costs, therefore all data will be wirelessly collected through "casual" over-the-air communications using NYCWiN. The available backhaul bandwidth is limited despite needing to support a dense collection of field devices (RSUs) and interacting vehicles (ASDs). Additionally, there are variable latencies and other limitations given that this is a shared media with other users (i.e. police and fire departments). Latencies and delays are on the order of 200-750 milliseconds on average, with

significant deviations under “busy” conditions for all backhaul exchanges. Where connections are readily available (more likely at vehicle fleet barns), select RSUs (such as some located at Fleet Owner ‘barns’) will transmit data via fiber instead of via NYCWiN, but this is expected to be a limited number of RSU locations.

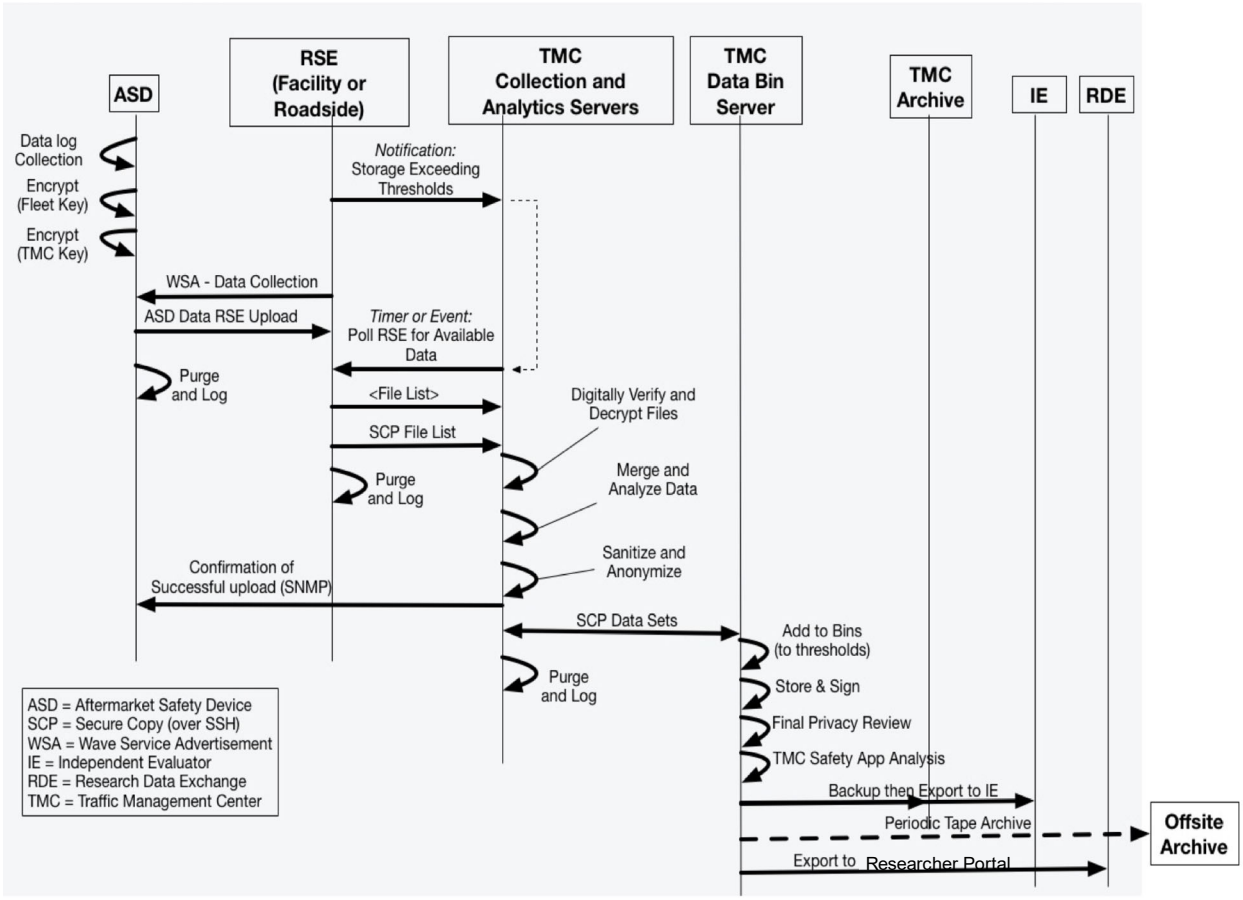
ASD Device DSRC Data Upload and Download

All ASD communications (infrastructure to vehicle and vehicle to infrastructure communications) will use 5.9 GHz DSRC. The DSRC service channels will be used by all edge devices to upload data collected on the vehicles and to download applications and operating parameters.

All NYC CVPD project vehicles are “fleet” owned – and hence return to a “barn” typically once per day – sometimes several times per day. Note that we estimate the average vehicle runtime (ignition on, engine running) is between 13-14 hours per day due to the nature of the fleet operations. Fleet barns will include fixed RSUs that are dedicated DSRC upload and download “hotspots.” ASDs will also upload and download data from roadside RSUs that are configured for specific data operations advertised using Wave Service Advertisements (WSA).

Data Collection, Processing and Storage

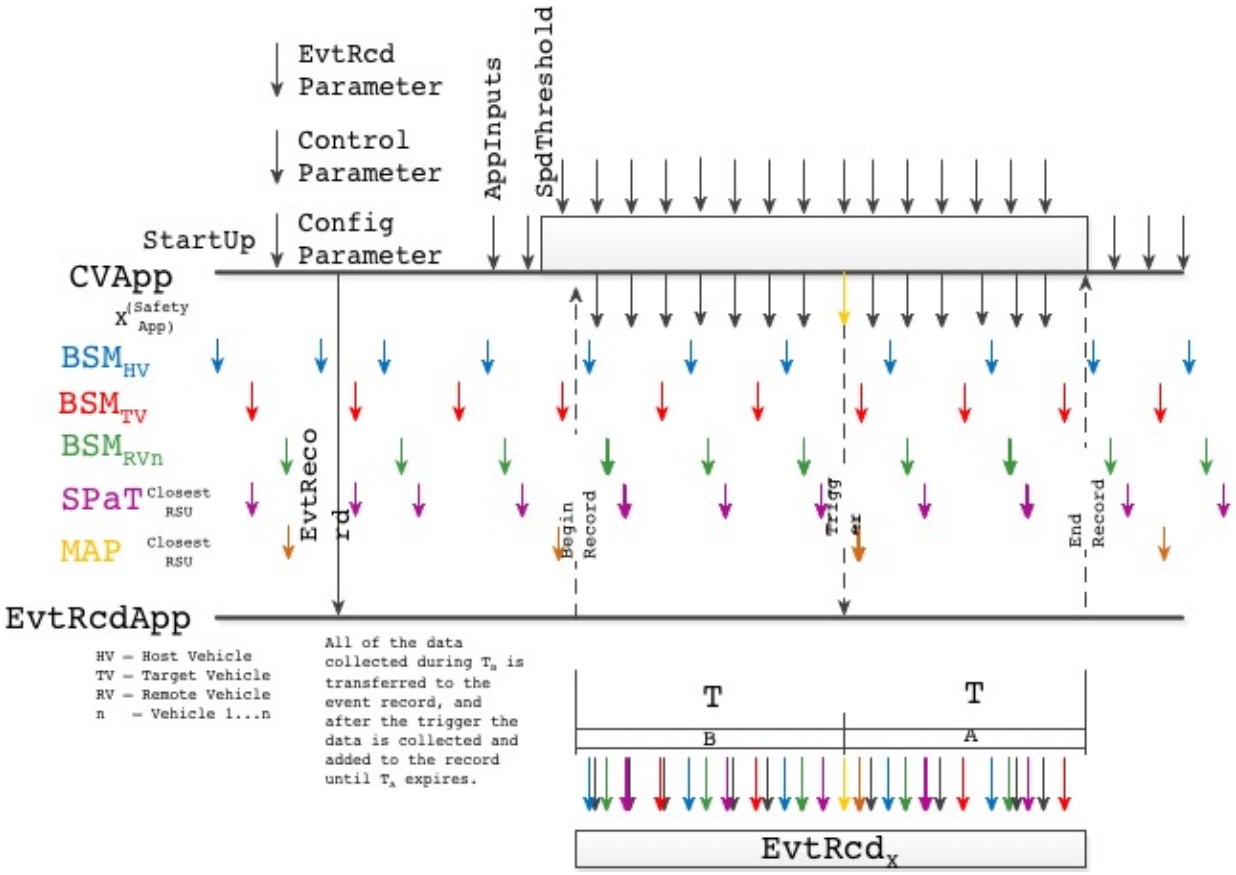
This section provides a lifecycle description of how all NYC CVPD pilot Event data will be collected, processed, analyzed and stored. Figure 3 provides a high-level view of where and how data flows from ASDs and RSUs to the TMC for subsequent analysis, sanitization, analysis, backup, and lastly archive and export.



Source: NYCDOT, 2017

Figure 3. ASD Event and Breadcrumb Data Collection and Processing Sequence

Figure 4 depicts at a granular level the constitution of Event data that is collected by different CV applications running on the ASD and later sent to the TMC Collection, Treatment and Analytics Servers (EvtRcdApp). The diagram depicts an example time window (time is left to right) in which host, target and remote vehicle BSMs as well as SPaT and MAP messages (from the nearest RSU) are sent from CV applications to the Event Record App (EvtRcdApp). At time=Trigger (when the Event trigger has fired due to exceeding the speed threshold, 'SpdThreshold'), all of the data starting at time 'Begin Record' before the trigger through time 'End Record' after it are collected by the EventRcdApp and placed in a log file.



Source: NYCDOT, 2017

Figure 4. Event Data Collected by ASDs and Sent to TMC

Privacy Condition Impacts on Data Collection

It is possible that the CV data collected can and will be subpoenaed and subject to Freedom of Information Act (FOIA) requests for tort cases and "investigations." For this reason (and that privacy must be protected when data is made available to researchers and the public) data will be aggregated, normalized, and obfuscated before it is stored and archived. It is important to note that time and location data (BSM) have the potential to become PII if it can be tied to other records such as police reports and used in legal, disciplinary, and insurance proceedings. Hence, all such data will be immediately processed as noted in this data management plan.

Regarding pedestrian participant data that is Personally Identifiable Information (PII), this information is entirely collected, managed and stored by New York University with privacy oversight from the Institutional Review Board (IRB). Such data is only available to the NYU researchers.

ASD O&M Data Collection

Operations and maintenance data refer to data required by the NYC CVPD engineers to analyze radio transmission strength and ranging as it impacts frequency of ASD-ASD and ASD-RSU 'sightings.' This

section describes the process of collecting various types of O&M data. Like Event data, All O&M data undergoes the normalization, sanitization and binning process at the TMC (e.g., to remove attributes such as ASD Serial number).

ASD Logs and Encrypts RSU Sightings

The first time an ASD receives and authenticates a SPaT message from a RSU it has not yet encountered, the ASD records the following:

- Its own BSM contents (location, time, speed, etc.)
- SPaT and MAP message content from the subject RSU
- RF power level of the received signals into a "first" and "last" entry of an "ASD encounter record"

It continues to receive and replace the "last" entry until a configurable timeout occurs. When this happens, it adds this "ASD encounter record" to the RF "ASD sighting RSU log". The tentative timeout duration is 10 minutes to conservatively accommodate the time expected to approach an RSU, remain stationary at the RSU-equipped traffic intersection, and then egress from the RSU's range. The timeout duration will be adjusted as RSU ranging and sighting characteristics are analyzed and better understood. The "ASD sighting RSU log" format is shown in the Appendix.

The ASD will similarly record the first and last MAP message it receives from each RSU (recording the content of the MAP message and its own BSM message) and continue to replace the "last" entry until a configurable timeout occurs. It then adds this "encounter record" to the RF "ASD sighting RSU log.' The ASD performs this same action for each separate [new] RSU encountered (both for SPaT and MAP messages) so that engineers can develop an effective communications radius for the RSU and message type – as determined by the ASD. These activities are vital to 'tuning' the V2I/I2V deployments.

Note that all "ASD Logs" and "RSU sighting logs" additionally identify 1) Vehicle Unique ID (serial number) and 2) ASD Firmware Version and revision date. These items are needed to determine if a given ASD is reporting data at all and determine if any errors or other issues are attributable to a given version of firmware.

ASD Logs and Encrypts V2V Encounters

Each ASD will also keep a log of the other vehicle's BSM and its own BSM at their closest point of encounter. The time-out for "new encounters" will be configurable from a default value of approximately 20 minutes (to accommodate possible collinear travel patterns), the goal of which is to ascertain where vehicles encounter each other during the life of the project. Each encounter with the same vehicle will be considered a new encounter, from a logging perspective, if the configurable timeout has expired. For example. if two ASD-equipped vehicles are within DSRC range of each other, each will log the encounter. If they're still within range of each other after the timeout, each will log it as a fresh encounter even if the other vehicle's temporaryID or pseudonym certificate has not changed. This information is useful for determining the breadth of their dispersion and coverage of the CV deployment. The host vehicle's ASD serial number is included in the uploaded log header which must be encrypted prior to transmission and can only be requested with the proper authorization keys. This data constitutes the "V2V Encounter Log" wherein each record contains two BSMs (host vehicle and target vehicle) for each unique encounter.

When the vehicles change pseudonym identifier, it will appear as a new encounter [for both vehicles] in the log entry, but because each vehicle's log contains its ASD serial number, the TMC will be able filter out extraneous encounters in the post processing (because it collects each vehicles log). This log is called the "V2V Encounter Log" and each record is a "V2V Encounter Record". The "V2V Encounter Log" format is shown in the Appendix.

Note that all V2V encounter logs additionally identify 1) Vehicle Unique ID (serial#) and 2) ASD Firmware Version and revision date. These items are needed to determine if a given ASD is reporting data at all and determine if any errors or other issues are attributable to a given version of firmware.

ASD Encrypts All O&M Logs/Records

Once an O&M log buffer is either 1) full or 2) the ASD encounters a support RSU, it is immediately encrypted using the TMC's encryption key. This process protects the data-at-rest and only allows the TMC to decrypt it. The *TMC Encryption Key* is rotated monthly and distributed to ASDs via ASD configuration updates. A description of provisioning of this key is provided in Chapter 3, Cryptographic Key Data Management.

ASD Uploads RSU Sightings and V2V Encounter Logs

Whenever the ASD encounters an RSU that collects the "ASD sighting RSU" or "V2V Encounter" logs (using a properly signed WSA), it uploads the data on the identified service channel to the RSU where it is stored and then forwarded to the TMC when backhaul bandwidth is available. Once the data is received and authenticated at the TMC, the TMC will initiate an acknowledgement to the ASD (through the RSU) which then purges its copy of the data. The ASD continues to record (log) this data – which must be buffered so that no data is lost and only complete records are transmitted.

Each "ASD encounter record" is not considered complete until the timeout has occurred. The software will handle the various race conditions (moments when the ASD processor needs to perform multiple, simultaneous tasks) so that no data is corrupted or lost and data collection is continuous.

The TMC application will retrieve this data from selected "support" locations that have frequent access to fleet vehicles. The TMC will also have the ability (through the OTA update process) to activate this data collection capability at any designated RSU.

All ASDs will include this O&M data collection feature, and will be configurable, activated and deactivated through software that is loaded and/or updated using the OTA configuration management process. The ASD's "ASD sighting RSU log" will contain the serial number of the ASD in the header such that it can be analyzed and correlated with other O&M data. Therefore, the "ASD sighting RSU log" will be encrypted on the ASD when stored, and transmitted in encrypted form to the TMC. The TMC server will then decrypt, process and purge the data.

The BSM data recorded in the vehicle includes the temporary vehicle ID which will be used to match the data received by the RSU as identified below. The exact mechanism to retrieve the log data is still under development; it may use an IP connection from the vehicle to the TMC whereby it establishes a secure file transfer (e.g. SFTP, SCP) with an acknowledgement that signals it is time to purge the existing buffer. In either case, all of this data is stored and sent encrypted. It remains encrypted – end-to-end - until it reaches the TMC.

ASD Purges O&M Logs

Following confirmation from the TMC of a successful upload (this will be provided to the ASD upon its next call-back to the ASD SNMP Manager), the ASD will purge the encrypted O&M data making the memory available for additional O&M data logging.

Data purges are accomplished by actively overwriting memory with 1's, 0's or random data. At least one single overwrite (zeroization) will be performed to prevent legal seizures of devices resulting in memory analysis and subsequent privacy losses. Simply dereferencing of variables in memory is not sufficient.

ASD Data Types

The encrypted data, collected on the ASD as described above, fits into one of the following categories.

Table 21. ASD File Extensions

File Name.Extension	Contents Description	Encoding
***.01**	RF Log (BSM, MAP, SPaT, TIM)	UPER
***.02**	Event Log (Event Type)	UPER
***.03**	OTA Status Log	UPER
***.04**	Bread Crumb Log	UPER
***.10**	System Status Log	Raw Text

RSU O&M Data Collection

The Roadside Unit also collects data usable by traffic engineers in monitoring and configuring roadside unit performance parameters.

RSU Logs and Encrypts O&M Data

The RSU collects the first and last BSM it receives from a vehicle encounter and authenticates (along with the RF level) each "new" & unique vehicle it sees. The logging includes the BSM contents and the time stamp along with the RF level of the received BSM. The TMC server can then match the temporary Vehicle Identification (VID) contained in the recorded BSM and match it with a BSM from a vehicle's "ASD sighting RSU log" (time will not be identical, but likely close; vehicle ID will match). Matching these two data flows in post-processing, the TMC can confirm the RF communications pattern around each vehicle and RSU – with sufficient statistical data to determine if the RSU is operating within accepted criteria and if the ASD/Vehicle system is operating within its accepted criteria. Traffic engineers will also be able to construct a history and look for significant changes to alert the TMC operators to the need for proactive maintenance (e.g. foliage, obstructions). This file is called the RF "RSU Sighting ASD Log" and each

record in the log is called a “RSU Encounter Record.” The “RSU Sighting ASD Log” format is shown in the Appendix.

These log file entries contain only the temporary vehicle ID and only for initial and final contact; they cannot be used for any tracking or monitoring of the actual vehicle. Once a RSU’s O&M log buffer is full (like ASDs), it is immediately encrypted using the TMC's encryption key. This process protects the data-at-rest and only allows the TMC to decrypt it. The *TMC Encryption Key* is rotated monthly and is distributed to RSUs via a configuration file update. The encryption process ensures that the file contents, when transmitted over-the-air or over the network, are not visible to anyone. A description of provisioning this key is provided in Chapter 3, Cryptographic Key Data Management.

RSU Uploads O&M Logs

RSU O&M data will be continuously uploaded to the TMC over a SNMPv3 over a Transport Layer Security (TLS) connection established between the TMC and RSU. This interface will use NTCIP 1103 and 1201.

RSU Purges O&M Logs

The RSU purges logs that have been successfully uploaded to the TMC. The TMC indicates to the RSU a successful transfer. Once the data has been uploaded, it is purged on the RSU. Once it has been completely analyzed at the TMC, the raw data copy is destroyed there as well.

Mobility Data Collection

Mobility data consists of ASD location history data (vehicle “breadcrumbs”) as well as RSU-collected mobility source data (RSU Sighting ASD logs). The collection of this data from ASDs and RSUs is described in this section.

ASD Logs and Encrypts Probe Data (Breadcrumbs)

All ASDs will have the capability to collect mobility data, termed ‘breadcrumbs.’ The ASDs will be controlled and configured over-the-air for mobility evaluation purposes and for future use in Intelligent Traffic Signal System (I-SIG) applications which are not part of the current CV Pilot project. The data collection intervals will be configurable in terms of the interval and distance traveled (or both) for each entry within the data collection scheme. This is not intended to be full 10 Hz resolution, rather the data collection interval will be configurable based on distance traveled or time traveled or both - whichever occurs first. This data is termed the *“Probe data log,”* each entry consisting of vehicle location (scaled for NYC), heading, speed, and path history.

The “Probe Data Log” format is shown in the Appendix.

Note that all probe data logs additionally identify 1) Vehicle Unique ID and 2) Recording Resolution. These items are needed to determine if a given ASD is reporting data at all and determine if any errors or other issues are attributable to a given recording rate.

This breadcrumb data recording will be buffered and continuously stored on the ASD so that no data is lost by the infrequent transmissions. Note that in the CV instrumented section of the City, this data will be

relatively short, however because the instrumented infrastructure is a very small part of the vehicle range, the system must be able to receive relatively large files from the vehicles that have traveled outside the instrumented area. The data will include configurable portions of the BSM data and the temporary vehicle ID and "rotating" pseudonym certificates. The vehicle ASD will not be allowed to change its ID during its exchange until it either receives the request to purge the data, or it times out. Techniques will be employed (using sequence numbers and time-outs) to manage overlaps and anomalies without loss of data where possible. This data may be used at the TMC for measuring segment travel times where there is no instrumentation. Any data received at the TMC will be purged after processing. Thus, the data cannot be used to link the vehicle segments together, nor will this data be passed to USDOT's ITS DataHub (IDH) or Secure Data Commons (SDC). This data will not be sent directly to the IE, rather it will first be sent to the TMC where it will be 1) sanitized of vehicle identifiers and 2) analyzed before being sent to the SDC. Time and location data will be obfuscated.

Once a log buffer is full on the ASD, it is encrypted using a 256-bit AES key that is encrypted using both the Fleet Owner's public encryption key along with the TMC's public encryption key. This process allows the Fleet Owner to decrypt information from its own vehicles if it wishes. Before this data is uploaded over DSRC to the RSU (then TMC), the ASD encrypts one more time the entire package with the TMC encryption key to conceal the public key identifier of the fleet owner. This protects the privacy of the Fleet Owner. The encrypted BSM contents will incorporate the ASD serial number in the log header to facilitate post-collection analytics.

ASD Uploads Probe Data Logs

ASD-encrypted breadcrumb data will be sent to either the RSU or the TMC when the ASD receives the appropriate authenticated WSA instructions to send the data to a specific destination (either RSU as intermediary, or the TMC as direct connection, depending on whether the RSU is NYCWIN or fiber-connected). All analysis will be performed at the TMC (none on the RSU). The potential file sizes along with the anticipated latencies and limitations of NYCWiN may force the RSU in some cases to be the intermediary.

ASD Purges Probe Data Logs

Breadcrumb purging is performed under three conditions:

1. Whenever the breadcrumbs have been successfully uploaded to the TMC (upon reception of a confirmation message.
2. Whenever the "event" mechanism (See Chapter 4, ASD Logs and Encrypts Event Data) is triggered, all breadcrumb BSM data collected for a configurable amount of time preceding and following the event record will be purged to avoid the possibility of compromising PII.
3. All probe data logs will be purged by the ASD after 48 hours, regardless of whether it has been uploaded to the TMC (note: the TMC has a separate 24-hour purge window for its raw data)

Data purges are accomplished by actively overwriting memory with 1's, 0's or random data. At least one single overwrite (zeroization) will be performed to prevent legal seizures of devices resulting in memory analysis and subsequent privacy losses. Simply dereferencing variables in memory is not sufficient.

RSU Sighting ASD Mobility Data

This data is collected as part of the RSU's O&M data collection described in Chapter 4, RSU O&M Data Collection, and its subsections.

Event Data Collection

Event data is comprised of a vehicle's BSM, surrounding vehicles' BSM data and RSU SPaT and MAP message data surrounding key events. The collection and analysis of this safety-related data is one of the key goals of the NYC CVPD.

ASD Logs and Encrypts Event Data

The NYC CVPD is handling safety oriented "event" data in a very different manner than other data types. The ASD contains a continuous, rotating log wherein it temporarily records:

1. its own 10 Hz BSM data
2. all BSMs for remote vehicles within a *configurable* distance
3. all SPaT and MAP messages heard from the nearest RSUs (configurable)
4. the time stamp

This log is a five-minute rotating buffer that replaces the oldest data as new data is acquired. (This will be evaluated to determine the longest recording period and the impact on the storage within the ASD.)

When a configured "event" is detected (the criteria and rules for each event type will be configurable on the ASD), the data collected immediately prior to the event (for a configurable amount of time) is copied from this temporary log to the "event record" and the data collected during and after the event for a configurable amount of time is added to the record. In addition to the collection of event record data, at the same time the ASD will purge mobility 'breadcrumb' data relative to the demarcated event record time. The amount of breadcrumb data that will be purged before and after the event time is configurable, but will typically vary between one and three times the total event duration (before and after).

The event record is then closed and encrypted using both the TMC encryption public key and the Fleet Owner's public key (if it exists) and stored in the *"Vehicle event log."* The ASD uses the same TMC encryption key as it uses to encrypt other data (mobility and O&M).

The "Event Data Log" format is shown in the Appendix.

In the event a vehicle and/or its ASD are damaged, inoperable or otherwise not network-reachable following an event (e.g., a vehicle accident), the Fleet Operator will inform TMC personnel who will retrieve the device and attempt to retrieve the data directly, if needed.

Note that all ASD Event Data logs additionally identify:

1. Vehicle Unique ID
2. ASD Firmware Version.

3. An optional device 'configuration code' or group identifier (used to indicate a specific configuration of the device as needed for a specific set of tests or other analysis). The configuration code can be used to indicate membership in a test or control group.

These items are needed to determine if a given ASD is reporting as it should, and additionally determine if any data or logic impacting errors are attributable to a given version of firmware.

ASD Uploads Event Data

Assuming no damage to the vehicle systems or ASD, vehicle event logs (containing the device serial number in the header) will be uploaded when it encounters a RSU broadcasting a properly authorized request [WSA] for this data.

ASD Purges Event Data

Once uploaded via the RSU, the ASD receives a confirmation of upload, after which it immediately purges the encrypted event data. Data purges are accomplished by actively overwriting memory with 1's, 0's or random data. At least one single overwrite (zeroization) will be performed to prevent legal seizures of devices resulting in memory analysis and subsequent privacy losses. Simply dereferencing of variables in memory is not sufficient.

Operations Logging

Operations logging refers to the TMC's management and control of RSUs and monitoring of ASDs. This information is provided from RSUs over a TLS-secured SNMP session used to convey 'status blocks.' Status blocks and other Simple Network Management Protocol (SNMP)-based attribute collection is distinct from the O&M, mobility and event data collection from fielded devices.

Operations logging of ASDs is similar to that of RSUs, however the ASD network connections are intermittent. ASDs are configured to reach back irregularly (but still daily) to the TMC to establish management sessions for configuration updates and monitoring. Reach-back configurations can be for a certain number of times per day or whenever within sight of a supporting RSU. These configurations will likely change over the course of the NYC CVPD as data volumes and RSU reach-back patterns are better established.

PII Data Collection – NYU

Pedestrian Data

The IRB for New York University (NYU) will oversee NYU's collection and management of all pedestrian-related analysis outside of the TMC. This information will include PII such as:

- Participant information (Name, address, phone number, email address, date of birth and other demographics)
- Completed pedestrian surveys, polls and other feedback (some of these may be anonymous)
- Privacy policy and consent forms

Note that NYU may coordinate, if requested, with the IE on the design and objectives of pedestrian surveys.

This data will only be collected and managed at NYU; it will not flow to the TMC. In addition, only NYU researchers approved by the IRB will have direct contact with Pedestrian participants. Coordinating through the 'New York Participant Liaison,' NYU researchers will complete the consent collection process, inventory and distribute the PIDs, train users and collect any ongoing or discretely solicited feedback during the project.

Driver Survey Data

Driver privacy is equally important, and there will be instances in which polling or surveying of drivers is necessary. No project personnel at NYU, including IRB personnel, will have direct access to driver participants (only to their data, anonymously provided), i.e., no entity will directly collect data from them. All communication with drivers will be performed anonymously via request and coordination through the Fleet Owners. This process will be performed, as follows:

1. NYU researchers will coordinate with the IRB (and if requested, the IE) to develop and finalize surveys and polls relevant to drivers
2. NYU will coordinate with designated Fleet Owner representatives to distribute website URL information (for digital surveys) or hard-copy survey forms.
3. Approved surveys and polls will not contain any requests for PII
4. Drivers will submit digital poll responses via website (anonymously) or hard-copy results.

Online polling of drivers will be anonymous to promote unbiased feedback. While NYU will track IP addresses and time of response, there will be no additional controls to limit multiple responses from drivers.

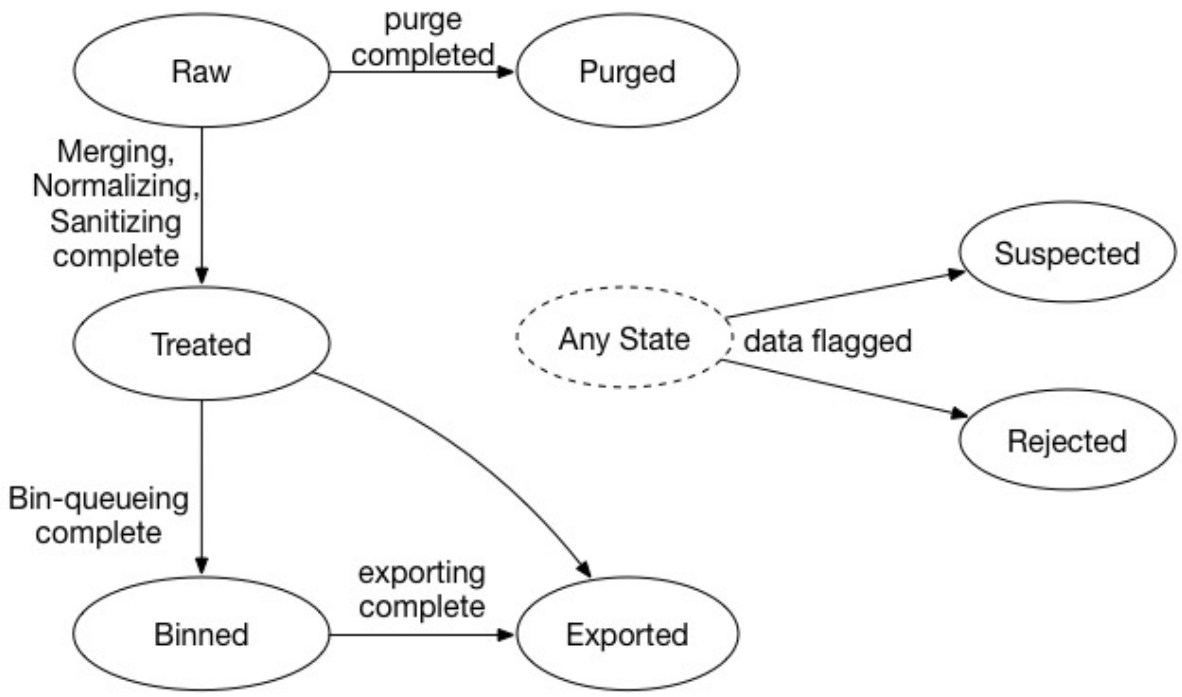
This process will be followed to ensure that any driver-related research data collected from Fleet Owners contains no PII for NYU to manage.

TMC Data Lifecycle: Post-Collection Processing, Treatment Analysis and Storage

This section provides a description of the NYC CVPD data storage, processing, backup, export and archive policies, processes and systems.

Data States for Data Collected by the TMC

All ASD-related data collection and processing is performed relative to well defined data states. Data states are tracked throughout the data processing lifecycle (via automated inspection of application processing logs) to ensure the proper post-processing and security treatment of data items in each state. Figure 5 depicts data states applicable to data managed by the TMC.



Source: NYCDOT, 2017

Figure 5. Data States Applicable to Data Management at the TMC

Raw: Raw ASD and RSU data that is in the original log or event file format, as it arrives at the TMC. This data is privacy sensitive (directly or indirectly contains PII) in terms of needing to be protected from tampering and unauthorized access that could lead to vehicle tracking. Raw data emanating from vehicles is collected by the TMC (by its Collection server); access to the Collection server is strictly limited to the Data custodian and IT management personnel. While in this state, raw ASD and RSU data may be comingled with other evaluation confounding factors to facilitate initial analytics that cannot be performed on treated data sets. This data is strongly privacy sensitive.

Treated: The Treated data state refers to data that has been expunged of any discretely identifying, geographic, time or other easily correlatable data that might link vehicles and participants to historical movements. Data in this state has also been normalized and merged with external confounding factor data sources (e.g., weather data, volume data, TMC logs, etc.) Data in this state can potentially be sensitive if collected for too short a period of time or too small a geographic area.

Purged: Data in the Purged state is data that has been actively expunged from all systems and memory devices on which it was resident. This data is no longer retrievable.

Binned: Data in this state is treated data that has completed the process of being amassed into larger data 'bins' and is ready to be exported for the Independent Evaluator (IE) analysis and archive. A subset of binned data will be provided to researchers as data in this state are no longer considered sensitive and can be made public. To reach this state, the following will first be verified: 1) the sanitization state and 2) the quantity and geo-temporal spread of data in the bins. Bins are designed to be geo-temporally large

enough (geographic area and/or time) to ensure the infeasibility of analytical methods to track specific users and vehicles in the data sets. Until it is in the Binned state, data is still potentially sensitive and prone to privacy-violating data analytics. See [12] for further information on data formats and binned data cleansing.

Exported: This state represents data that has completed the archive or export process. Data in this state can be fully published to the IE.

Suspected: This state represents data that is potentially corrupted or unusable. To enter the Suspected state, data must be flagged as such.

Rejected: This state represents data that is rejected for any identified reason (e.g., corruption, data gaps, etc.). To enter the Rejected state, data must be flagged as such.

Data States for Data at NYU

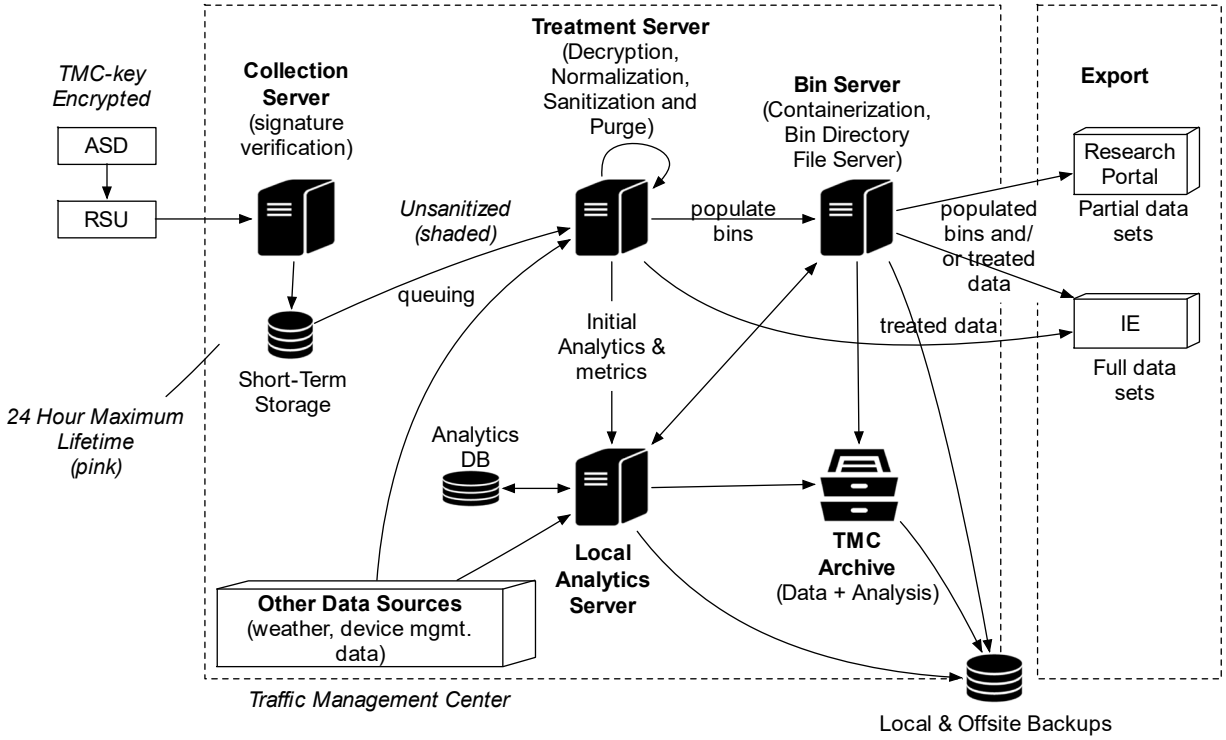
NYU will also collect data that needs to be managed appropriately to certain states. Data such as pedestrian participant information PII, consent forms, etc. at NYU will be classified simply as:

Non-PII: This information does not need to be access controlled

PII: This information must be strictly access controlled

TMC Data Flows and Systems

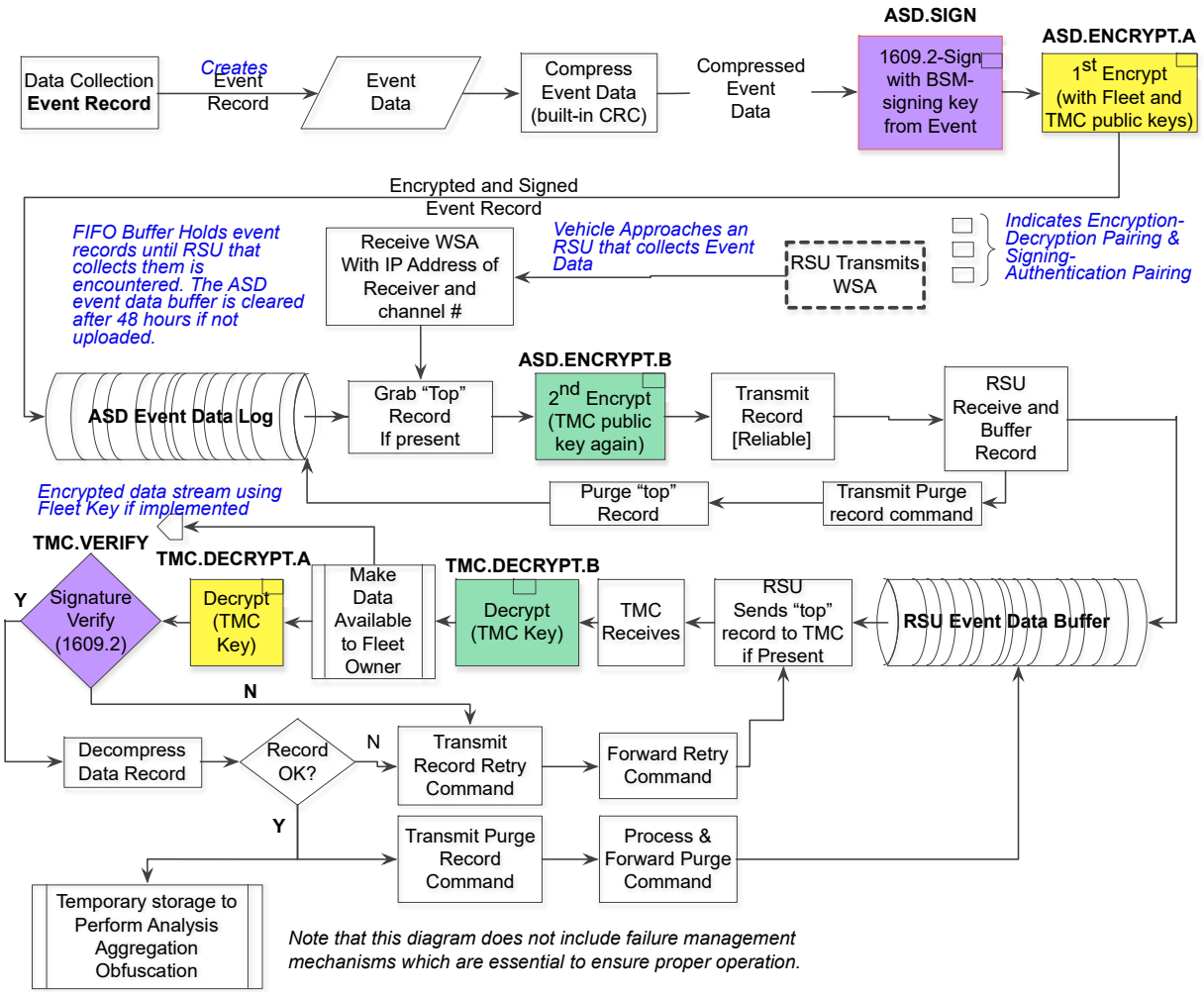
Collection of all ASD and RSU originated data is described in Chapter 4, Data Collection, Processing and Storage. Depicted in Figure 6, the data migrates through the afore-mentioned data states as it is collected, processed, stored, backed up and processed. This process is described next.



Source: NYCDOT, 2017

Figure 6. ASD and RSU Data Collection, Processing and Storage Strategy

Collection Server and Data Treatment Server Processing: Once data is retrieved from field devices into the 'Collection' Server, it is placed into temporary disk storage awaiting a queued retrieval from the Data Treatment server. The Collection server will first decrypt the log data using the TMC's encryption key. The resulting data is still encrypted using two public keys: 1) the TMC's encryption key and the fleet owner key (if the fleet owner requests data access). In this state, the data – if so requested – may be ported to a short-term distribution server for fleet owners to download (note that only a given fleet owner with the matching private encryption key can decrypt its data). Once the TMC has performed a final decryption using its private decryption key, it will perform a digital signature verification of each uploaded file. The movement of the log file from ASD to the TMC is depicted in Figure 7. In addition, the Collection server will accept a RSU-generated token that maps to the upload file. This token will be used later to indicate out-of-band to the ASD or RSU that a specific file has successfully uploaded. Upon receiving and successfully verifying that token, the ASD or RSU knows to purge its own copy of the encrypted data.



Source: NYCDOT, 2017

Figure 7. Event Log Upload Process

Following each successful digital signature verification on uploaded files, the Collection server will perform the following:

1. Log the upload along with the received token
2. Transmit the received token to the external RSU or ASD manager application (for subsequent SNMP relay of 'upload confirmation' to the device)
3. Await Treatment server queuing of the data from the disk

The signature verification process will make use of a connected vehicle security library that has provisioned the relevant TMC application public key. The files will remain in encrypted form on the Collection server disk, however, and will not be decrypted using the TMC encryption private key until they have been queued to the Treatment server for sanitization, normalization and merge.

Due to the privacy sensitivity of raw ASD data, ASD data directories and files on the Collection server disk are limited a 24-hour longevity period. All ASD data files and directories are either purged 1) following confirmation from the Treatment server that a batch job has completed initial analysis (required to be performed on raw data) and sanitization or 2) by the Collection Server's continuously running file wiping service (job schedule for wiping files and directories that have aged 24 hours). Normal operations will not see copies of the raw ASD data purged from short-term Collection server storage until the Treatment server's sanitization and normalization is completed.

Treatment and Analytics: Some local analytics may be performed on the Treatment server, only those that require full access to raw data. Before any data merging, sanitization and analytics are performed, the Treatment server first decrypts the encrypted ASD and RSU logs. When decrypted, the Treatment server must subsequently 1) expand the logs by Base64 decoding and 2) ASN.1 decoding some of the log file fields (e.g., BSM, MAP and SPaT message binary data). Decoded data can then be processed, exchanged to database tables, normalized and then merged with other data sets. To ensure that raw data was sensible, data engineers will create and manage filters to flag data as being suspect, missing information, possessing out-of-range values or exceeding some other thresholds and tolerances. Once these have been processed, results will be allowed to pass to a dedicated Local Analytics server for additional analysis and metrics that need to be computed. The Local Analytics server will have its own dedicated Analytics Database (DB) for data requiring extensive memory or fast-accessibility needs to analyze. No processing of raw data is performed on the Local Analytics server, however. Data in the raw state is not allowed to leave the Treatment server.

Local analytics will consist of the processing of event data, O&M data (RSU RF ranging), mobility and other performance data. The overarching data flow related to performance data is depicted in Figure 8.

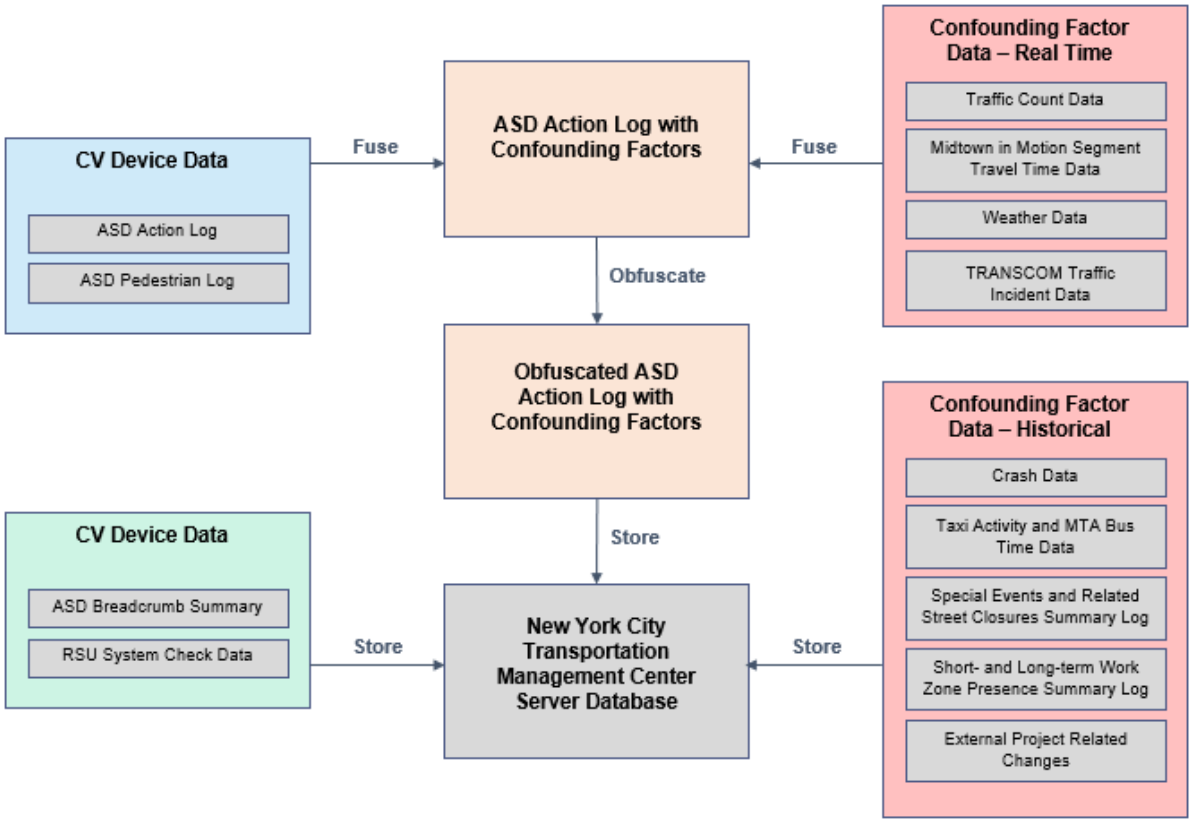
Bin server: The Bin server amasses data until geo-temporal thresholds are reached for each bin and it contains enough data to prevent privacy-threatening analytics. Bin server data is available for export to the Independent Evaluator, researchers and TMC Archive.

It is expected there will be times when specific data sets emanating from ASDs are corrupt (do not decode correctly, have missing information, or other data validation rule) or suspected of being corrupted. In other cases, data may need to be outright rejected for various reasons. When this occurs, the Data Steward and data engineers will identify the metadata ranges and values corresponding to the suspect or reject data (e.g., log date time ranges, serial number information, collection time ranges, etc.) and input these to the Bin server and the servers upstream of it (note: the TMC Data Steward and data engineers do not have access to NYU-based PII and do not need to be covered under the IRB). When this occurs, data sets that match the flagged metadata values will be put into separate 'suspect' directories or tables and finally input to a suspect bin on the Bin server. The process is illustrated in Figure 9 and detailed data bin descriptions and formats are described in [12].

Collection, Treatment and Analytics Server Backups and Recovery Processes and Tools: Each server's backed up image (consisting of operating system as well as installed and configured executables) shall be maintained as it may be necessary to rebuild and redeploy a server in the event of a full restore.

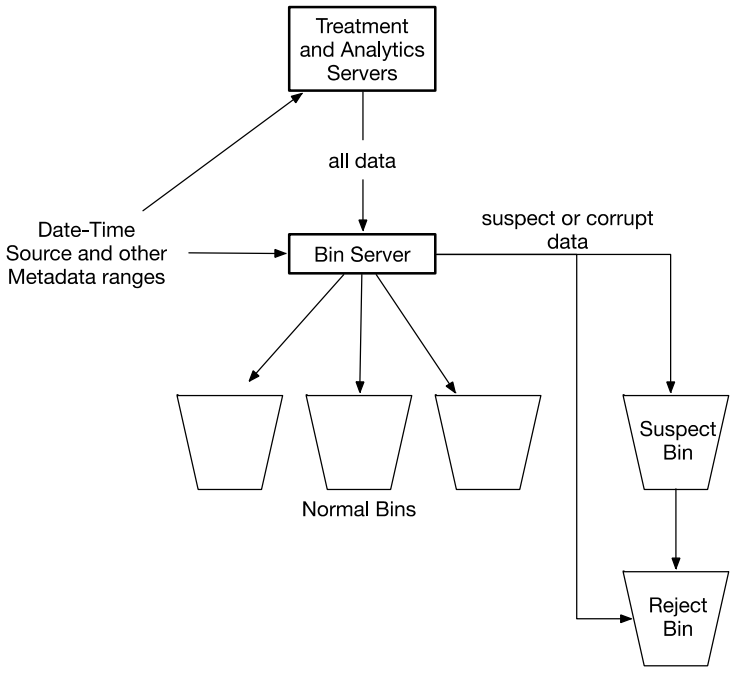
Normal restores will be performed from daily backups, however normal, daily backups will be generated of each machine (or Virtual Machine) to preserve each server's operating system configuration, host firewall settings and applications in the event of a server loss. There will be no PII or potential PII

contained in these backups as they are independent of the data backups. Daily server backups will be performed using the existing SQL Server tools at the NYC TMC.



Source: NYCDOT, 2017

Figure 8. Overall Performance Data Flow and Processing



Source: NYCDOT, 2017

Figure 9. Suspect Bin Processing

Data Backup and Recovery Processes and Tools: Data backups will be performed differently on data that is stored in normal file systems vs. stored in databases.

File system-resident data stored on the Collection server will not be backed up at all as this data is highly ephemeral and has a lifetime of just a few hours. File system data on the Treatment server will only be subject to daily backups if the data is in the *Treated* state, i.e., it contains no raw data and fully completed its normalization and merging. This data will be backed up daily and recovered using the SQL Server backup utility included with the TMC archive.

Database-resident data offloaded from the Collection, Treatment and Local Analytics servers to the TMC's SQL Server database will be backed up and encrypted using SQL Server backup. Backups will be recorded onto backup media with copies additionally maintained offsite.

Archiving and Recovery: The New York City TMC's existing archive utility (at the TMC) will be used for archiving all processed and binned data. The current system includes an archive system to support the long-term storage of the traffic control system's data. In addition to binned data, the system will continue to archive the following types of data:

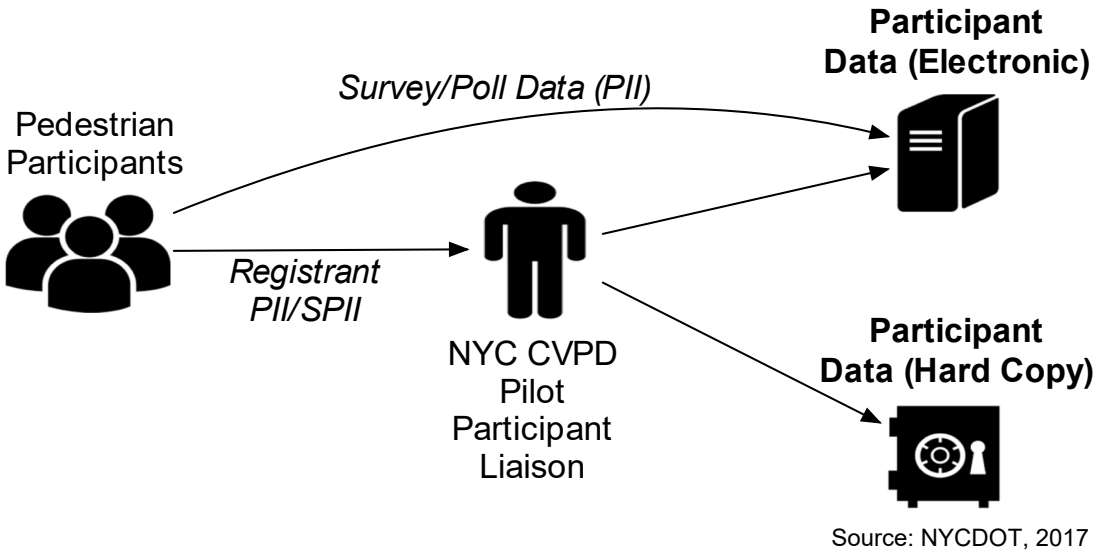
- Traffic signal database (restored SQLServer production database backup)
- Traffic measurement data (sensor data for volumes, occupancy, speeds)
  - o raw detector archive
  - o detector archive – smoothed data
  - o travel time reader archive

- Traffic system operations (daily event logs)
  - o event log archives
  - o conflict monitor log archives
  - o transit signal priority archives
- Incident histories

SQLServer is installed and operates on the archive server to manage the various archive databases. The archive system hardware consists of an HP DL360 series server managing a HP Storageworks D2700 disk system. The current capacity of the disk system is 25 TB which can be expanded to a total of 240 TB with additional enclosures, drives, and interfaces. Depending on data volume trends, this system should suffice for archive and recovery, however it will be replaced with another commercial solution, if needed.

NYU Data Lifecycle: Pedestrian and Driver Information

This section contains the rules and process for NYU to store and manage pedestrian data (PII) as well as anonymous driver participant information that does not contain PII. Figure 10 depicts how the survey data from pedestrian participants will be collected, sent, and stored.



Source: NYCDOT, 2017

Figure 10. Participant Data Collection and Storage

While coordinating with pedestrian participants, NYU will maintain a Pedestrian Information Device (PID) inventory and manage device checkouts to the participants. Device inventory and checkout records will contain mapping of device serial numbers to participant names. When this data is maintained together, it will be stored in encrypted form on a server created and securely managed by NYU. NYU will protect this information according to this document and the NYC Data Privacy Plan (DPP) (Connected Vehicle Pilot Deployment Program Phase 2 Data Privacy Plan – New York City, December 27, 2016).

All collected PII and SPII will be strictly controlled. The NYU 'Participant Liaison,' will be responsible for interfacing with individuals participating in the pedestrian studies. This liaison will collect any completed hard-copy forms and store them in a locked safe at NYU to which only the liaison has access. This liaison is the approved NYU researched approved by the IRB as mentioned in Chapter 4, PII Data Collection – NYU. In addition to maintaining hard copy records, this individual will be responsible for:

1. Transcribing and entering any hard copy form data into electronic form
2. Providing secure web links for pedestrian participants and drivers to provide poll-related feedback, survey data, etc.
3. Coordinating all consent gathering activities with participants, in adherence to the Data Privacy Plan
4. Manage all interactions with the Pedestrian Participants, such as providing them devices (PIDs) and helping manage any device issues

All pedestrian data will be stored in encrypted form on a NYU-managed database server that 1) maintains strict access controls in accordance with NIST SP800-122 and 2) is backed up daily. NYU data engineers will author and maintain analysis scripts that will perform data analytics, collect metrics and perform relevant studies by connecting to this database. Following analysis, raw data will be purged with the resultant data archived. Fully analyzed data sets (purged of PII) will be archived by NYU in encrypted form. Certain data sets from this archive may be exported and merged with the TMC's data sets. Collection, storage and analysis of the individual and merged data sets will be re-evaluated for PII any time the data management processes change. This is described in Chapter 4, Metadata.

Process and Tools for Creating, Processing and Visualizing ASD and RSU-Originated Data

Data Collection: Data collection will be performed on the Collection server using the Secure Copy Protocol (SCP) or Secure File Transfer Protocol (SFTP), which is a secure file transfer call over a Secure Shell (SSH). File-level authentication and integrity of the file upload is achieved via the sending device's IEEE1609.2 digital signature performed over the log data files and the subsequent signature verification by the Collection server. Files are also 1609.2-encrypted. Local file, directory and upload management processes will be managed by an application running on the Collection server. This application will be responsible for certain signaling actions to the RSU and ASD management hosts when uploads have completed. Data purge logs will also be collected and analyzed by this application.

Data Processing and Treatment: Data will be accomplished on the Treatment and Analytics servers using a variety of processing scripts. The Treatment server scripts will execute 1) merges with other data sources, 2) raw data analysis and intermediate data exports to the Analytics server and 3) sanitization of data.

The Analytics server will perform analytics on much larger data sets populated from the Treatment server (and at times pulled from the Bin server); it will also make use of a fast access database (SQL Server) when performing computation over large data sets that would otherwise exhaust the memory resources of the analytics and processing applications. Results that are cleansed of all PII (or potential PII) may be exported to a variety of external systems or visualization tools.

Data Purges: Tools for data purging will consist of 1) Data purge application and 2) a scheduling service running on both the Collection and Treatment servers. The scheduling service is a backup mechanism to ensure data purges if data fails to process within the 24-hour window.

Cryptographic Tools: All 1609.2 security processing on ASDs, RSUs and various TMC applications (Collection server for signature verification of uploads, Treatment server decryption of uploads and TMC data signing utilities) is performed using the 1609.2 compliant cryptographic security library, Aerolink, by Security Innovation. TLS implementations will be run on TMC management application servers as well as ASDs and RSUs. Device vendors will furnish TLS implementations in accordance with NYC CVPD device specification requirements.

Documenting Data Collection

Documenting NYC CVPD data collection activities is vital to ensuring the full integrity, quality and availability of data the pilot requires. This section describes the process of documenting the data collection activities at relevant points in the NYC CVPD system architecture.

Data Collection Event Logging

Data collection related events are events of interest concerning per-device and system-level data collection functions. They consist of data collection metadata that is logged in each device as part of normal post-collection auditing, correlation and archiving. Logged data collection events transmitted back to the TMC in the form of 'status blocks' will contain a variety of data collection relevant status. Events such as errors in log file creation and update, errors in upload, purge errors, etc. will be included. These types of errors indicate abnormal or dysfunctional data collection activities pertinent to identifying suspect and reject data.

Device status blocks from RSUs are sent periodically to the TMC over SNMP. Device status blocks from ASDs are also sent over SNMP to the TMC when devices make periodic contact.

Data Collection Records

The device status blocks comprise the audit source material from which data collection records and reports are generated at the TMC. The data records are required for archiving, proof of data collection integrity and used as inputs for computing data collection metrics and statistics. Like event and probe data, records associated with data collection events must similarly be purged of any possible time and geo-location information, especially after merging and analyzing them with other data sets. Once used for identifying suspect and reject data (necessary for downstream isolation and treatment), they will be purged of any possible PII and then archived at the TMC.

Data Collection Metrics

Data collection and upload activities are initially recorded on the ASD and RSU devices, and are uploaded and catalogued on the TMC's ASD management server and RSU management server to identify any data collection anomalies. This data is needed for post-collection correlation and determination of any gaps in data collection (e.g., devices that have failed to collect or upload data). The sanitized data collection records (based on each device's data collection logs) will be used to compute data collection metrics such as:

- Periodic minimum, average and maximum time between data collection events (per device and device family), i.e., data upload behavior
- Quantity of devices that fail to reach back to RSU or TMC within prescribed thresholds
- Frequency of exceeding prescribed data collection thresholds
- Data collection and success based on device and device family (to identify which devices are failing to meet data collection objectives)

Data Organization, Documentation and Metadata

This section documents the plan for organizing, documenting, and using descriptive metadata to assure quality control and reproducibility of the data.

Project Data (Raw and Processed)

Project data evolving through its lifecycle will be managed by applications, file system managers and databases. Original data emanating from devices through log file uploads will be structured as depicted in the Appendix.

These files will only be resident short-term on fielded devices and on the TMC collection server.

All raw and processed data will be structured and named on various TMC servers consistent with the file and directory naming conventions identified in Chapter 4, Metadata Structure

The metadata will be structured as defined in the Connected Vehicle Pilot Deployment Metadata for DMP v1.5 document for subsequent loading in the SDC along with the obfuscated data exported from the system.

Metadata Update Process

When metadata need to be updated, initial updates will be made to the Performance Measurement Evaluation Support Schedule (PMESS) and the System Operation and Maintenance Schedule (SOMS). Thereafter, the TMC resident metadata on the Collection server will be updated and submitted to the SDC using the standard upload processes already in place. USDOT is working on making the PMESS docs more broadly available at this time.

Directory and File Naming Conventions.

Data imported from other sources (e.g., confounding data elements such as weather) will be:

1. imported via the Local Analytics server and
2. sanitized on the Treatment server, if needed
3. associated to ASD event data on the Local Analytics server

The data re-organization will be performed by various applications that ingest data from one CSV file format and either generate new CSV files (file-based intermediate data) or input the data into a database

for faster access. Part of this process will involve the base64 decoding and extraction of raw ASN.1 data (i.e., from BSMs, SPaT and MAP messages) from an ASN.1 parser and message schema.

Analytics performed at the TMC will make use of databases (SQL Server). Some database schemas (e.g., for travel time data) may be re-used from other TMC projects. Others will be created, defined and documented per the demands of the NYC CVPD.

Data fully processed into data bins will be stored in a conventional file system ready for archiving, export, or both. All data exports to the IE will consist of fully populated 'binned' data or modified subsets thereof, respectively. Please see the NYC CV Pilot's System Design Description (SDD) [12] for a detailed description of the data bin format and contents.

The binned data is made available to the IE in the format described in [12]. The general structural strategy for the binned data is designed to facilitate:

- Simplified data import into various data processing tools, scripts, databases, etc.
- Indexing the data sets to enable analytical queries based on connected vehicle application, and application event type and custom studies
- Identification of flagged data sets where data quality may be suspect
- Factoring of data measurement accuracy and other parameters into IE analysis

Project Documentation

This section describes the management of NYC CVPD project-related documentation.

TMC

All official project documentation will be maintained by the TMC Data Steward in a configuration-managed, version-controlled repository server located at the TMC. Periodic updates to certain documents will be added to the repository server such that a version and time-stamped history is maintained. Microsoft SharePoint will be used to manage all project documentation using the directory folders specified in Table 22.

Table 22. Project Documentation Folder Structure

Directory	Description & Subfolders
/Specifications	Documents such as RSU and ASD device acquisition specifications, device standards, Systems Architecture Document (SAD), Systems Design Document (SDD) and software specifications are stored in this folder.
/Plans-Policies	Stores all policy and planning materials such as the Data Privacy Plan (DPP), Data Management Plan (DMP), Acquisition Plan, Installation Plan, Maintenance and Operations Plan, Performance measurement and evaluation support plans and outreach plans. Policies and plans related to protecting participant privacy will also be stored in this folder.

Directory	Description & Subfolders
/Training	Used to manage documents related to training records for personnel involved with the NYC CVPD.
/Change Management	Used to manage documents related to transitioning individuals, organizations, resources, plans, objectives and processes that will necessarily change over the course of the pilot.
/Inventory Management	Used to store all inventory-related information related to the acquisition, installation, maintenance and tracking of connected vehicle equipment such as RSUs, ASDs, PIDs and ancillary equipment, software, servers and data management systems.
/Project Management	Used for general project management documentation such as the Project Management Plan (PMP), lists of participating personnel, project deliverables, etc. This folder contains a sub-folder /schedule for storing all schedule-related information.
/Communications	Records related to specific internal communications, external communications and announcements
/Financials	Stores all financial information related to the pilot deployment
/Useful Information	Miscellaneous folder used to store information useful to the pilot deployment, test scenarios and studies of interest, reference documents such as standards and others
/Reports	Used to store periodic and ad-hoc internal (NYC) and external (e.g., USDOT) reports related to the ongoing progress of the NYC CVPD. Reports related to project-wide and specific tests may be stored in this folder, along with documents such as system performance reports. This folder also has a sub-folder, '/Data Specifications,' used to describe the content and structure of data shared with specific parties at specific times.
/Tests	Stores information related to specific tests and analyses that are undertaken as part of the NYC CVPD

Internal CV deployment participants such as New York University, USDOT and the IE will be given access to the repository, or will be provided notices and updates from the TMC Data Steward for managing on their own systems. The Data Steward may restrict certain files or directories based on need-to-know. Such decisions will be coordinated by the CISO in conjunction with NYCDOT and USDOT.

New York University

Electronic storage of pedestrian PII and SPII will be secured on NYU servers in accordance with industry and NIST SP 800-122 best practices (NIST SP 800-122: "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)").

The NYU project repository will contain the folders indicated in Table 23.

Table 23. NYU Project Directory Structure

Directory	Description & Subfolders
/Specifications	Stores documents related to PID specifications, traffic system specifications, etc. This folder also has a sub-folder called '/Data Specifications' that is used to document project metadata such as data sources, structure, data collection metrics, etc.
/Participants	Stores all participant-related data (this folder is SPII and access controlled as such). This folder also stores electronic copies of completed consent forms.
/Plans-Policies	Stores all policy and planning materials such as the Data Privacy Plan (DPP), Data Management Plan (DMP), Acquisition Plan, Installation Plan, Maintenance and Operations Plan, Performance measurement and evaluation support plans and outreach plans. Policies and plans related to protecting participant privacy will also be stored in this folder.
/Training	Used to store all training materials, training records and training plans related to NYU personnel participating in CV pedestrian studies. Training will address protection of PII and SPII.
/PID-Inventory	Used to store all inventory-related information related to the acquisition, installation, maintenance and tracking PID devices deployed to participants. PID management, hardware and application data version information will be stored here in addition to which participant has possession (start date, end date) of it. This folder is PII-related due to it containing a map of what person carries which PID.
/Project Management	Used for general NYU project management documentation. This folder contains a sub-folder /schedule for storing all schedule-related information.
/Communications	Records related to specific internal communications, external communications and announcements
/Financials	Stores NYU-related financial information related to the pilot deployment
/Useful Information	Miscellaneous folder used to store information useful to the pilot deployment, test scenarios and studies of interest, reference documents such as standards and others
/Surveys	Stores /draft and /final poll and survey forms (templates) for pedestrian and driver participants. Survey results are access-controlled appropriately and stored in the following sub-folders: /results-anonymous /results-non-anonymous.
/Reports	Used to store periodic and ad-hoc internal (NYCDOT) and external (e.g., USDOT) reports related to the ongoing progress of the NYC CVPD pedestrian studies. Reports related to project-wide and specific tests may be stored in this folder, along with documents such as system performance reports.

Directory	Description & Subfolders
/Tests	Stores information related to designation, tracking and data collection for specific PID-related tests and other analyses that are undertaken as part of the NYC CVPD

The /Participants folder and any other directory sub-folders that related to PII will be strictly accessible only by the following roles indicated in Chapter 3, Roles and Responsibilities in Data Management:

- NYU Participant Liaison
- NYU Data Steward

An individual with the role 'NYU Security Manager' (an NYU researcher approve by the IRB) will be responsible for maintaining servers and controlling access to applications and folders on those servers.

Metadata

Metadata includes certain data types contained within the CV application messages and other data sources as well as information about the data itself.

Metadata Types

Message-related metadata will be collected and stored by ASDs, collected/uploaded, parsed and analyzed based on the log file type:

1. Basic Safety Messages, per data elements and format described in SAE J2735
2. MAP messages, per data elements and format described in SAE J2735
3. SPaT messages, per data elements and format described in SAE J2735

Useful metadata from these messages includes:

- Location
- Time information
- GPS status and accuracy
- Intersection-ID
- Intersection configuration and geometry
- Signal phase information (important for analyzing certain event types)

The above raw data is stored in the Event Log and ASD RF Logs, These logs are subject to obfuscation prior to transfer to the Secure Data Commons (SDC), Additional metadata that will be used to collect, organize, analyze, and protect data include the following:

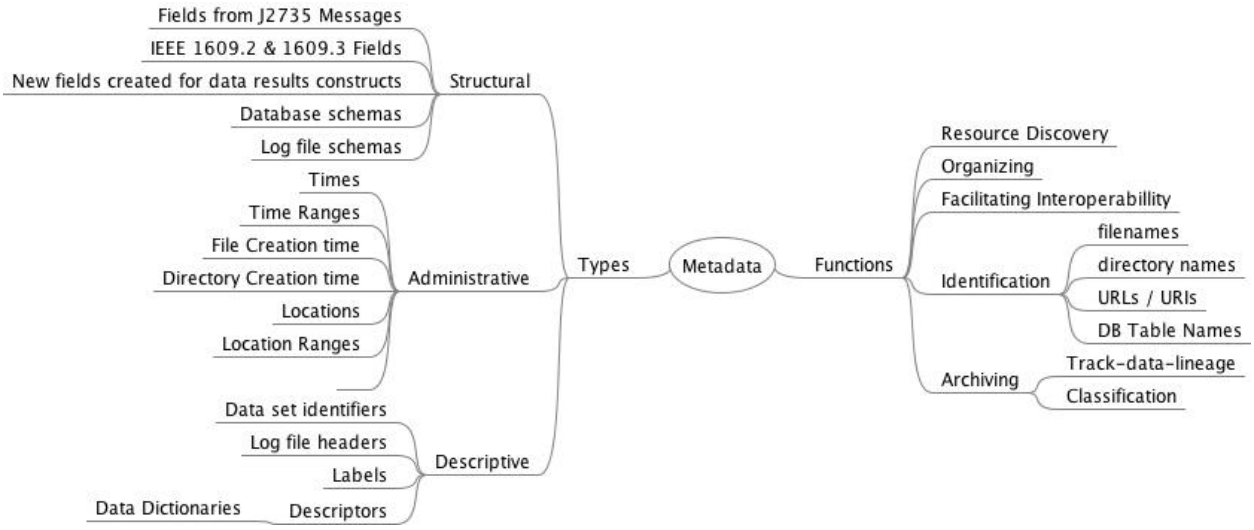
- ASD Serial Number and RSU Serial Number – This information will be used to compute travel time (mobility) data

- ASD and RSU firmware version information and release dates – This information will be source from the log files create by ASD and RSU devices
- Configuration ID – This information will be a unique, coded value that can be added to raw data files and logs (e.g., ASD and RSU log file headers) to help track and organize data pertinent to specific test cases and control groups. Configuration ID, in some cases, will not originate from raw data, but will be a label added to intermediate and final data sets. In some cases, multiple configuration IDs may be need to be deployed simultaneously to perform certain tests (i.e., ascertain interactions between vehicles and roadside units with distinctive configurations)

Figure 11 illustrates the generic metadata types and functions pertinent to the New York Pilot.

Device configuration metadata (e.g., configuration threshold setting, configuration baselines as set in device MIBs) will also be set and tracked by the ASD and RSU management hosts to facilitate data cleansing and organization operations. Some analysis will require discrete thresholds and configurations to be set in devices such that device data collection parameters can be modulated and tracked for specific tests (i.e., map to Configuration ID parameters set in log files). Collection of statistics about these data will be stored as metadata in the SDC however the raw data will remain confidential.

Metadata parameters will be represented in log file schemas, MIB tables, device status blocks, database schemas (labels) and in some cases file and directory names.



Source: NYCDOT, 2017

Figure 11. Metadata Types and Functions

To minimize the risk of metadata exposing PII, an abbreviated ***privacy impact assessment*** (PIA) will be performed any time changes take place to the following:

1. Data actions (any collection, storage, processing and archive processes)
2. Database table design
3. Log file designs and schemas

4. Data sanitization and purge scripts
5. New or modified data sources (used in analysis)

This abbreviated PIA will evaluate:

1. whether any privacy rules with respect to the established the data states are being violated
2. if any data treatment changes need to be made to accommodate the modified data actions, structures or applications
3. a proposed design to handle any data treatment changes, as necessary

Metadata documentation will leverage ASTM 2468-05 (https://www.astm.org/Standards/E2468.htm) to ensure the following data information is available to researchers:

1. Data publication information
2. Data collection procedures and processes
3. Time/location information (except for anonymized data sets)
4. Contact information for data-related inquiries
5. Data element metadata, namely data types, units, field lengths, maximum and minimum values (as applicable) and data code definitions
6. Known data errors or gaps

Metadata Structure

The metadata will be structured as defined in the Connected Vehicle Pilot Deployment Metadata for DMP v1.5 document for subsequent loading in the SDC along with the obfuscated data exported from the system.

Metadata Update Process

When metadata need to be updated, initial updates will be made to the Performance Measurement Evaluation Support Schedule (PMESS) and the System Operation and Maintenance Schedule (SOMS). Thereafter, the TMC resident metadata on the Collection server will be updated and submitted to the SDC using the standard upload processes already in place. USDOT is working on making the PMESS docs more broadly available at this time.

Directory and File Naming Conventions

This section describes the directory and file naming conventions of NYC CVPD data collection and processing at the TMC and at NYU. Given the discrete data states that are maintained throughout data processing, directory and file-naming provides a mechanism for establishing data state and organizing test-specific data.

Collection Server

The Collection server is the first stop for all ASD and RSU data collected at the TMC.

RSU Raw Data

O&M Data collected from RSUs are uploaded to the Collection server. Files are created and named by the RSU using the following naming convention:

RSU-O&M-Raw-<ID>-<Type><FileCreationTime>.csv

<*ID*> denotes the concatenated RSU serial number and intersection ID

<*Type*> denotes the raw data type

<*FileCreationTime*> denotes the time the log file data was exported to its own file just prior to RSU signing (prior to upload to the TMC)

RSU O&M data files are entered into Collection server directories named as follows:

/RSU-O&M-Raw-<*DateTimeStart*>|<*DateTimeEnd*>/<*files*>

<*DateTimeStart*> *and* <*DateTimeEnd*> denote the Collection server's datetime range for when the raw data files were collected.

RSU Mobility (travel time) data files are created on the RSU and named as follows:

RSU-Mobility-Raw-<ID>-<FileCreationTime>.csv

<*ID*> denotes the concatenated RSU serial number and intersection ID

<*Type*> denotes the raw data type

<*FileCreationTime*> denotes the time the log file data was exported to its own file just prior to RSU signing (prior to upload to the TMC)

RSU Mobility raw data files are uploaded to the TMC Collection server into directories named as follows:

/RSU-Mobility-Raw-<*DateTimeStart*>|<*DateTimeEnd*>/

<*DateTimeStart*> *and* <*DateTimeEnd*> denote the Collection server's datetime range for when the raw data files were collected.

This data is continuously pulled from the Collection server, travel times computed and then entered real-time into a database.

The frequency at which the collection server will create and populate new directories will be configurable and tuned to the queuing strategy between the Collection and Treatment servers.

ASD Raw Data

O&M Data collected from ASDs is stored in a short-lived directory. Files are created and named by the ASD using the following naming convention:

<Hash>.10

<*Hash*> denotes the ASD created hexadecimal hash code. Serial numbers that are not detected within a configurable datetime range will be investigated for malfunction or connectivity problems.

.10 denotes the raw System Status Log data type

<Hash>.03

<*Hash*> denotes the ASD created hexadecimal hash code.

.03 denotes the raw Over-The-Air Status Log data type

<Hash>.01

<*Hash*> denotes the ASD created hexadecimal hash code.
.10 denotes the raw Radio Frequency Log data type

ASD Mobility (travel breadcrumb) data collected from ASDs is stored in a short-lived directory. Files are named using the following naming convention:

<Hash>.04

<*Hash*> denotes the ASD created hexadecimal hash code.
.04 denotes the Bread Crumb Log raw data type

ASD Event data collected from ASDs are stored in short-lived directories. Event data files will be created on ASDs from an assortment of internal and external (other vehicle) connected vehicle messages that are logged. The event files will be named using the following naming convention:

<Hash>.02

<Hash> denotes the ASD created hexadecimal hash code.
.02 denotes the raw Event Log data type (event type)

Following decryption and decoding, the ASD raw data files are placed in Collection Server directories named as follows:

<*Type*>_<CabAdr>-<*Hash*>.txt

<*Type*> denotes the raw data type replacing the coded file extension above with either SSL, OTA, BC_, RF_, or EVT.

<*CabAdr*> denotes the hexadecimal cabinet address of the RSU receiving the ASD data files. The RSU Inventory lists these codes.

Timestamps of file and directories on the Collection Server are used to execute purge operations and also manage queuing of the raw data to the Data Treatment server.

Data Treatment Server

This section describes file and directory naming conventions for data stored and processed on the Data Treatment server.

Operations and Management (O&M) and Mobility Data

Whether originating from ASDs or RSUs, O&M and mobility data will be processed and then the raw data thrown out.

O&M Data such as RF ranging information will be output files and directories commensurate with the external systems used to perform the analysis.

Mobility, travel-time data will not preserve vehicle information; it will consist only of a travel time record across segments of interest. Travel time data will be exported in the file format of the consuming system.

Event Data

Data queued to the Treatment server is in the same file and directory structure as that on the Collection server. Once the data has been treated (sanitized, normalized and merged), the raw data files and directories on the Treatment server are purged and a signal is sent to the Collection server to purge its copy.

The Treatment server is also used to perform initial analysis that must be executed over raw data and external data sources (e.g., weather data). Files and directories on the Treatment server are named as follows.

Event Files: *Event*-<*ID*>-<*EventCategory*>-<*EventDate*>-<*EventLocation*>.csv
<*ID*> denotes a unique, non-sequential identification number assigned by the Treatment server to the event
<*EventCategory*> denotes type of event that occurred
<*EventDate*> denotes the date of the event
<*EventLocation*> denotes the ASD's location (latitude/longitude) at the time it recorded the event

Event File Directories: /EventData-Processed-<*DateTimeStart*>|<*DateTimeEnd*>/
The <*DateTimeStart*> *and* <*DateTimeEnd*> denote the Treatment server's datetime range indicating the date range of the processed events. These datetime ranges will be vital to establishing date and geospatial thresholds for ensuring participant privacy.

The EventDate and EventLocation information in the filename will not be preserved as the processed data is queued to the Bin server. This information will be used by the Bin server to fill data bins and ensure that temporal and geospatial thresholds are sufficiently wide to preserve privacy.

Special Test Cases

In some cases, special tests are to be performed using specific ASD and RSU configurations, firmware versions and other test criteria. In these instances, impacted files will be copied to a separate test directory and their filenames appended with the following string: "testXYZ', where XYZ indicates a unique test code given for the analysis. This filename suffix will be retained throughout the processing and storage of this data.

Bin Server

The Bin server will be configurable to specific time and geospatial thresholds to ensure that event data is difficult to reverse engineer and be correlatable to specific vehicles or individuals.

As data is queued into bins and the bins reach geo-temporal thresholds that guarantee privacy among a sufficient quantity of events, data files will be renamed. Specific time and location information will be expunged from file and directory names. The data bins themselves consist of the processed files renamed to exclude the event date and location information. Once the bin has been fully populated to the configured threshold, the bin directory name will be renamed with the prefix 'NYCCVPD-Bin' and appended with a unique bin code (e.g., 'ABCDEF').

Archive (Export) Server

By definition, all files that are exported to the Archive server or exported to the IE are fully populated bins (or to-be-defined, additionally sanitized subsets thereof). The directories and files retain the obscured names from the Bin server and are zipped up into individual files as follows: <*NYCCVPD-Bin-ABCDEF*>.zip.

Project Identifiers

The official project identifier for the TMC is *NYCCVPD*. This identifier will be used in project, file, repository and database references.

Sub-projects will be created, as necessary, to address customized tests or other studies. These special cases will include the main project identifier *NYCCVPD* concatenated with a sub-project or test name starting with *TMC* and ending with a sub-project name created, assigned and managed by the TMC Data Steward (e.g., *NYCCVPD-TMC-TESTCASE123*).

Sub-project names for studies that are performed and managed solely by NYU will be created, assigned and managed by the NYU Data Custodian using a similar convention (e.g., *NYCCVPD-NYU-PROJECTX*).

Chapter 5. 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 SP 800-122: " Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)”
4	Comma Separated Value (CSV) format standard: RFC 4481 (https://tools.ietf.org/html/rfc4180)
5	Connected Vehicle Pilot Deployment Program Phase 2 Data Privacy Plan – New York City, December 27, 2016
6	NISTIR 8062: An Introduction to Privacy Engineering and Risk Management in Federal Systems (http://nvlpubs.nist.gov/nistpubs/ir/2017/NIST.IR.8062.pdf%20)
7	NIST Big Data Interoperability Framework: Volume 1 Definitions
8	NIST Big Data Interoperability Framework: Volume 4, Security and Privacy
9	NIST SP 800-88r1 “Guidelines for Media Sanitization”
10	New York City Citywide Security Policies (DOITT) –(https://www1.nyc.gov/site/doitt/business/it-security-requirements-vendors-contractors.page)
11	New York State Archives MU-1: Transportation and Engineering Data Retention policy (http://www.archives.nysed.gov/records/retention_mu-1_transportation-engineering%20)
12	Connected Vehicle Pilot Deployment Program Phase 2 System Design – New York City, May 26, 2017

Chapter 6. Glossary

Term	Definition
Aggregated	Data that has been combined with other data of a similar or disparate type to facilitate analysis
ASD	Aftermarket Safety Device – The vehicle installed connected vehicle equipment
CV	Connected Vehicle
CV PEP/SDC	Connected Vehicle Performance Evaluation Platform / Secure Data Commons: An access portal and data management enclave for the Independent Evaluator to collect, assess, and manage NYC CV data.
Data bin	A typically large quantity of processed data that is sufficiently aggregated to reduce the potential for data correlation-related privacy losses.
Identifier	a label (numeric, alphanumeric, or other) that is used within a specific context to uniquely identify an entity
IDH	ITS DataHub: USDOT database that was formerly known as Research Data Exchange (RDE) for sharing sanitized and obfuscated CV data with researchers and other parties interested in CV. No PII data will be shared in this database.
IE	Independent Evaluator – The evaluator entity performing and approving studies within the New York City Connected Vehicle Pilot Deployment
IRB	Institutional Review Board – governing board overseeing NYC CVPD research under guidance from US Department of Transportation's Human Use Approval process
Normalization or 'Normalized'	(used in "normalizing," "data normalization"); the process of converting raw data into structured data either in a database or structured, attribute-delimited files. The process involves removing data redundancies, converting attributes to consistent data types and removing any anomalies that may negatively impact post-normalization data analytics.
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.
PID	Pedestrian Information Device – A mobile, handheld device sending or receiving connected vehicle data.

Term	Definition
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.
Procedure:	Actions or activities required of an individual or organization to satisfy PII-related policies.
NWS	National Weather Service
NYC CVPD	New York City Connected Vehicle Pilot Deployment
NYU	New York University – Performing all pedestrian connected vehicle related research
Obfuscated	Data that has either had correlative data items removed from it (to reduce the potential for privacy loss) and/or been aggregated with a large volume of similar data to reduce the potential for a data analytical process to correlate an individual driver or pedestrian with the data.
RSU	Roadside Unit – the roadside mounted connected vehicle device used to communicate with vehicle ASDs. RSUs connect over the network to the Traffic Management Center.
Sanitized	Data that has had potentially privacy-compromising information filtered from it.
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.
SCP	Secure Copy (Copy over Secure Shell [SSH])
SDC	Situation Data Clearinghouse
SDW	Situation Data Warehouse
SFTP	Secure File Transfer Protocol
SNMP	Simple Network Management Protocol

Term	Definition
SSH	Secure Shell – A secure, authenticated, integrity protected and confidentiality protected network channel
TLS	Transport Layer Security – A cryptographically-secured network transport protocol.
TMC	Traffic Management Center
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
WSA	Wave Service Advertisement (IEEE 1609.3 and 1609.2 security) used to advertise roadside application services from roadside equipment

Appendix A. Log Formats

ASD Sighting RSU Log Format

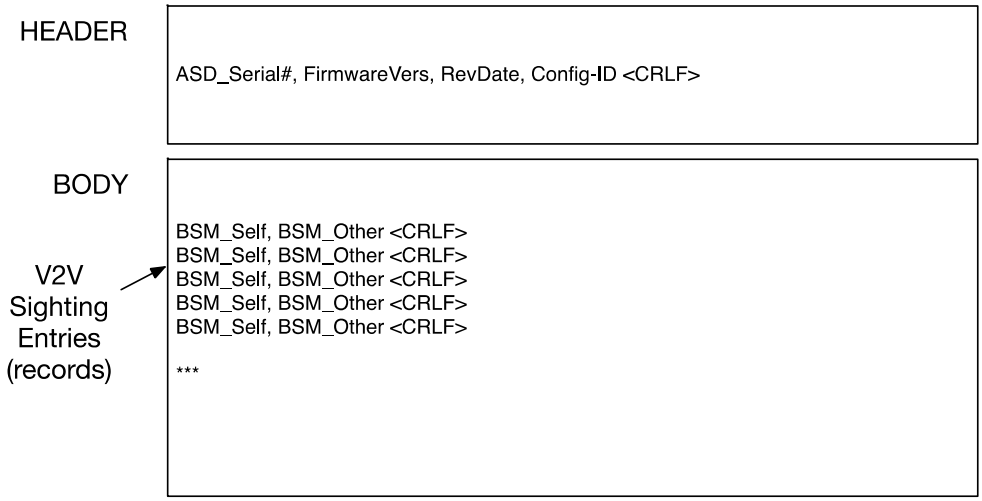
HEADER: The ASD Sighting RSU Log header is structured as MSG_NycCvpdASDRF in the NYC CVPD Event Format document.. It contains the following information:

- **ASD_Serial#** indicates the unique serial number assigned to this ASD

BODY: Each record of this log file consists of a sequence of first and last MAP, SPaT, or TIM messages received from external RSU devices along with the ASD's location and time when receiving these first/last messages.the following information:

There is a "first entry" and a "last entry" indicating the first and last sighting, respectively, of a given RSU in each encounter. All fields will be comma-separated and adhere to RFC 4180.

V2V Encounter Log Format



Source: NYCDOT, 2017

Figure 12. V2V Encounter Log

HEADER: The V2V Encounter Log is structured as MSG_NycCvpdASDRF in the NYC CVPD Event Format document. It contains the following information:

- **ASD_Serial#** indicates the unique serial number assigned to this ASD

BODY: Each subsequent record of this log file is structured as the log in the previous section using BSM records from the external vehicle.

RSU Sighting ASD Log Format

HEADER: The RSU Sighting ASD Encounter Log header is contained in the first line of the CSV file. It contains the following information:

- **RSU ID** indicates the unique serial number assigned to theRSU reporting
- **Temp ID** indicates the sighted ASD's transmitted temporary identity
- **Timestamp** is the reported time the RSU sighted the ASD
-

The log file will created daily by the server.

BODY: Each subsequent row of this comma-separated value (CVS) log file will consist of the following information (records):

- Rsu ID fields contain the reporting RSU's permanent identifier assigned in the RSU inventory.
- **Temporary-ID** fields indicate the pseudonym from the sighted vehicle ASD's pseudonym certificate. This information is also in the BSM, but is extracted for log file indexing
- **Timestamp** fields will contain the full date and time of the observation
- **<CRLF>** is the newline symbol in the CSV file indicating to start the next line and its new record

Probe Data Log Format

HEADER: The Probe Data log is structured as MSG_NycCvpdASDMobility in the NYC CVPD Event Format document. It contains the following information:

- **vehID** indicates the Vehicle Identification Number (VIN) of the vehicle. As this information was not available in all vehicles, it lists the string “3935313130”.
- **asdSerialNumber** is the unique value assigned to each ASD.
- **timeRecordResolution** contains the data collection rate as defined in the Event Format document.

BODY: Each subsequent data frame of this log file will consist of the following vehicle location records of the locList:

- **Datetime-Stamp** fields contain the datetime that the vehicle location was logged (by the ASD). Each datetime entry (record) will be approximately 1 second apart, but this frequency will be configurable. Additionally, the time range from the beginning of the file to the end of the file will be configurable (may range from several minutes to a couple of hours).
- **Latitude, Longitude, and Elevation** fields will identify the location of the subject vehicle..
- **<CRLF>** is the newline symbol in the CSV file indicating to start the next line and its new record

Event Data Log Format

Each event log will be used for a single event, only as defined in the NYC CVPD Event Format document describing the contents of the MSG_NycCvpdEvent.

HEADER: The Event Data log header is contained in the eventHeader (DF_EventHeader) . It contains the following information:

- **asdSerialNumber** is the unique serial number assigned to the ASD
- **eventType** is the Connected Vehicle application identifier associated with the Connected Vehicle application that triggered the creation of this log file. The eventType will be one of the following enumerations:
  - **"FCW"** (1) if the Forward Crash Warning application triggered the event
  - **"EEBL"** (2) if the Emergency Electronic Brake Lights application triggered the event
  - **"BSW"** (3) if the Blind Spot Warning application triggered the event
  - **"LCW"** (4) if the Lane Change Assist application triggered the event
  - **"IMA"** (5) if the Intersection Movement Assist application triggered the event
  - **"VTRW"** (6) if the Vehicle Turning Right in Front of Bus application triggered the event
  - **"SPDCOM"** (7) if the Speed Compliance application triggered the event
  - **"CSPDCOMP"** (8) if the Curve Speed Compliance application triggered the event
  - **"SPDCOMPWZ"** (9) if the Work Zone Speed Compliance application triggered the event
  - **"RLVW"** (10) if the Red Light Violation Warning application triggered the event
  - **"OVCTURNPROHIBIT"** (11) if the Oversize Vehicle Compliance Turn Prohibition application triggered the event
  - **"OVCCLEARANCE"** (12) if the Oversize Vehicle Compliance Clearance Limit application triggered the event
  - **"EVAC"** (13) if the Emergency Communications and Evacuation Information application triggered the event
  - **"PEDINXWALK"** (14) if the Pedestrian in Signalized Crosswalk application triggered the event
  - **"PED-SIG"** (15) if the Mobile Accessible Pedestrian Signal System application triggered the event

- **asdFirmwareVersion** indicates the ASD firmware version at the time the specific log entry (record) was created
- **grpId** contains a unique test or control group number that maps to a specific device configuration
- **EventTime** contains the reference event time, the time when all of the criteria for triggering the event episode were satisfied (as recognized by the ASD). Some records in the log will be message data received leading up to the EventTime; some will be message data received after the official EventTime (as configured by event type in the ASD).
- **parameters** contains a list of values controlling the application and data collection for the event.

After this event log is created and written to, it is immediately closed and encrypted for upload.

- BODY: Sequences of BSM, MAP, SPaT, and/or TIM messages constitute the body of the event log as defined in the NYC CVPD Event Format document.
- **<CRLF>** is the newline symbol in the CSV file indicating to start the next line and its new record

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-17-454

U.S. Department of Transportation