

Phase 1 Data Management Plan (DMP)

Atlanta Regional Commission: ITS4US Deployment Project

www.its.dot.gov/index.htm

Final Report — August 23, 2021
FHWA-JPO-21-865



U.S. Department of Transportation

Produced by Atlanta Regional Commission (ARC)
U.S. Department of Transportation
Intelligent Transportation Systems (ITS) Joint Program Office (JPO)
Office of the Secretary of Transportation (OST)
Federal Highway Administration (FHWA)
Federal Transit Administration (FTA)

Notice

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

The U.S. Government is not endorsing any manufacturers, products, or services cited herein and any trade name that may appear in the work has been included only because it is essential to the contents of the work.

Technical Report Documentation Page

1. Report No. FHWA-JPO-21-865		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Phase 1 Data Management Plan (DMP) Atlanta Regional Commission: ITS4US Deployment Project			5. Report Date August 23, 2021		
			6. Performing Organization Code 075863845		
7. Author(s) Kofi Wakhisi (ARC), Maria Roell (ARC), Randall L. Guensler (Georgia Institute of Technology), Angshuman Guin (Georgia Institute of Technology), Polly Okunieff (GO Systems and Solutions), Natalie Smusz-Mengelkoch (Kimley-Horn & Associates)			8. Performing Organization Report No.		
9. Performing Organization Name and Address Atlanta Regional Commission – Georgia Ste Gov Atlanta RGL COM 229 Peachtree St NE, Ste 100 Atlanta, GA 30303-1601			10. Work Unit No. (TRAIS)		
			11. Contract or Grant No. 693JJ321C000008		
12. Sponsoring Agency Name and Address U.S. Department of Transportation ITS Joint Program Office 1200 New Jersey Avenue, SE Washington, DC 20590			13. Type of Report and Period Covered		
			14. Sponsoring Agency Code 693JJ3		
15. Supplementary Notes Elina Zlotchenko (Program Manager), Amalia Rodezno (Contracting Officer), Karen Timpone (Contracting Officer Representative)					
16. Abstract The Atlanta Regional Commission Complete Trip - ITS4US Deployment project, Safe Trips in a Connected Transportation Network (ST-CTN), is leveraging innovative solutions, existing deployments, and collaboration to make a positive impact using transportation technology to support safety, mobility, sustainability, and accessibility. The ST-CTN concept is comprised of an integrated set of advanced transportation technology solutions (connected vehicle, transit signal priority, machine learning, predictive analytics) to support safe and complete trips, with a focus on accessibility for those with disabilities, aging adults, and those with limited English proficiency. The Data Management Plan (DMP) provides an inventory of the datasets and their characteristics related to the Atlanta Regional Commission (ARC) ITS4US project – Safe Trips in a Connected Transportation Network (ST-CTN). The inventory includes datasets that are ingested, generated, processed and exported by the ST-CTN system including static, real-time, and archived datasets. The plan includes information on data governance, management, security and privacy policies, storage and access.					
17. Keywords ITS4US; Complete Trip; Deployment; ITS; Intelligent Transportation Systems; Safe Trips in a Connected Transportation Network; ITS4US; Data Management Plan			18. Distribution Statement		
19. Security Classif. (of this report) None		20. Security Classif. (of this page) None		21. No. of Pages 90	22. Price N/A

Revision History

Name	Date	Version	Summary of Changes	Approver
ARC Team	23 August 2021	1.0	Version 1	USDOT

Table of Contents

Revision History	i
Table of Contents	ii
1. Introduction	1
1.1. Project Background	1
2. Project Overview	5
2.1. Change Control	5
2.2. Relevant Sources.....	6
2.3. Data Schedule.....	7
3. Data Overview	11
3.1. Data Needs Summary	11
3.2. Data Overview.....	13
4. Data Stewardship	33
4.1. Data Owner and Stewardship.....	34
4.2. Access Level	39
4.2.1. Private Datasets.....	41
4.2.2. Access Request.....	51
4.2.3. Related Tools, Software and/or Code	51
4.2.4. Relevant Privacy and/or Security Agreements.....	51
4.3. Re-Use, Redistribution, and Derivative Products Polices	53
4.4. Data Storage and Retention.....	58
4.4.1. Storage Systems	58
4.4.2. Data Storage System Description	64
4.4.3. Cybersecurity Policies	65
4.4.4. Data Security Policies and Procedures.....	66
4.4.5. Back-up and Recovery Policies and Procedures.....	66
5. Data Standards	67
5.1. Data Standards	67
5.2. Versioning.....	69
5.3. Metadata.....	69
5.3.1. Metadata Types	69
5.3.2. Metadata Structure	70
5.3.3. Metadata Update Process	71

Appendix A. Acronyms and Glossary.....73
 Acronyms 73
 Glossary 76
Appendix B. Comprehensive List of Standards Used79
Appendix C. Dataset Impact Log.....81

List of Tables

Table 1: Project Area Demographics3
 Table 2: Project Summary4
 Table 3: References6
 Table 4: Data Schedule7
 Table 5: Critical ST-CTN Connection Descriptions (related to Context Diagram with Data Storage Systems) 12
 Table 6: Data Overview Table 14
 Table 7: Dataset Roles in Data Governance Organization.....35
 Table 8: List of Privacy and Access Levels Assigned to Datasets42
 Table 9: Relevant Privacy and Security Policies and Agreements52
 Table 10: Data Owner License Applied to Datasets54
 Table 11: Storage Systems60
 Table 12: Cybersecurity Policies and Procedures65
 Table 13: Data Archiving and Retention Policies and Procedures66
 Table 14: Dataset Standards and Rationale67

List of Figures

Figure 1. Phases of the Complete Trip – ITS4US Deployment Program (Source: USDOT)2
 Figure 2. ST-CTN Deployment Site Map (Source: ARC)4
 Figure 3. Safe Trip in a Connected Transportation Network Data Exchange Flow Diagram with Data Storage Systems (Source: ARC)..... 11
 Figure 4. Data Governance Roles for the ST-CTN ITS4US Project (adapted from source: ARC Data Governance, p. 17).....33
 Figure 5. Metadata Update Process (source: ARC)72

1. Introduction

This Data Management Plan (DMP) provides an inventory of the datasets and their characteristics related to the Atlanta Regional Commission (ARC) ITS4US project – Safe Trips in a Connected Transportation Network (ST-CTN). The inventory includes datasets that are ingested, generated, processed, and exported by the ST-CTN system including static, real-time, and archived datasets. The plan includes information on data governance, management, security and privacy policies, storage, and access.

During the Phase 1 effort, the DMP covers the ARC team’s preliminary approach to managing the datasets including identifying security, privacy and governance policies related to existing datasets and describing proposed system datasets and their dependent standards. As the system design and testing phases progress, additional details on application interface (API) and protocol interface conformance specifications (PICS) will be added.

This document is intended for technical reviewers, independent evaluators (IE), designers, researchers, and other persons interested in the datasets and their derivatives that drive or are produced by the ST-CTN system.

1.1. Project Background

Access to a robust multimodal transportation system is crucial to providing equitable access for all communities within the Metro Atlanta region and around the United States. Housing, employment, education, and healthcare choices are all dictated by access to a transportation system that allows users to make complete trips [ICS]. A complete trip is defined by an individual’s ability to go from origin to destination reliably, spontaneously, confidently, independently, safely, and efficiently without gaps in the travel chain regardless of location, income, or disability [BAA]. Stakeholders, particularly those within underserved communities, need more traveler information and support to complete their trips safely, reliably, and efficiently.

The Complete Trip - ITS4US Deployment Program is a multimodal effort – led by the Intelligent Transportation Systems (ITS) Joint Program Office (JPO) – and supported by the Office of the Secretary (OST), Federal Highway Administration (FHWA), and Federal Transit Administration (FTA) – to identify ways to provide more efficient, affordable, and accessible transportation options for underserved communities that often face greater challenges in accessing essential services. The program aims to solve mobility challenges for all travelers with a specific focus on underserved communities, including people with physical or cognitive disabilities, older adults, low-income individuals, and limited English proficiency (LEP) travelers. This program seeks to enable communities to build local partnerships, develop and deploy integrated and replicable mobility solutions to achieve complete trips for all travelers.

The Complete Trip – ITS4US Deployment Program will be executed in three phases. As depicted in **Figure 1**, deployment sites are expected to go through three phases:

- **Phase 1. Concept Development**

- **Phase 2.** Design and Testing
- **Phase 3.** Operations and Evaluation

Post deployment, sites are expected to sustain operations for a minimum period of five years without supplementary federal funds.

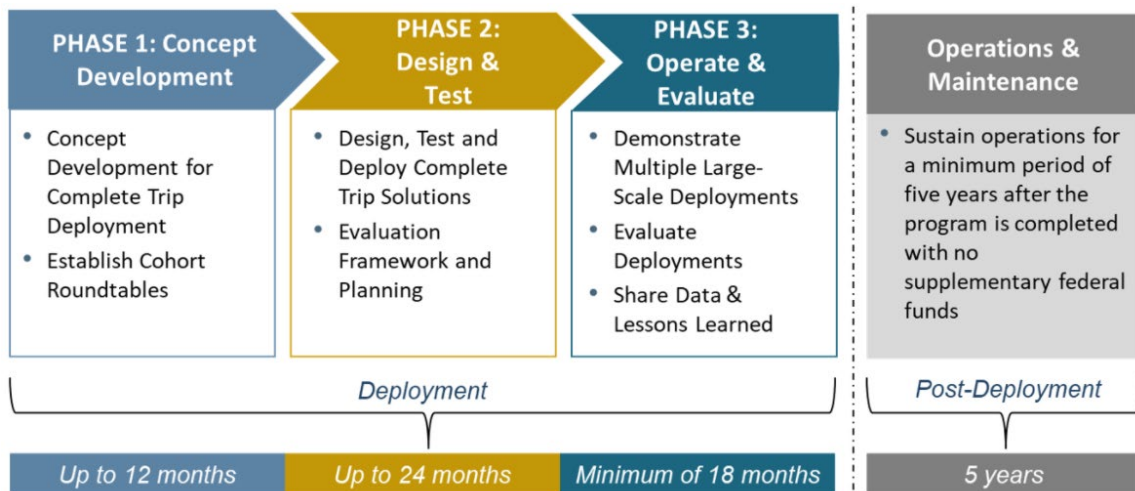


Figure 1. Phases of the Complete Trip – ITS4US Deployment Program (Source: USDOT)

The ARC was selected by U.S. Department of Transportation (USDOT) as one of the Phase 1 projects to showcase innovative business partnerships, technologies, and practices that promote independent mobility for all travelers regardless of location, income, or disability. The project team intends to address multiple aspects of the Complete Trip by integrating multiple technological innovations. The ST-CTN system will integrate connected vehicle (CV) data with an open-sourced web-based and mobile application. The application will provide users with the ability to create a personalized trip plan with information regarding the navigation of physical infrastructure, the ability to resolve unexpected obstacles, and ensure users visibility to vehicle operators throughout the trip. The proposed deployment will provide all users with the ability to dynamically plan and navigate trips based on their personal needs and preferences. Consistent with the ITS4US Program goals, the ST-CTN project is specifically focused on supporting the following underserved communities:

- **People with Physical Disabilities.** People with physical disabilities are limited in independent, purposeful physical movement of the body or of one or more extremities, and substantially limits one or more major life activities.
- **People with Cognitive Disabilities.** People with cognitive disabilities have a condition that makes it more difficult to interact or participate in the environment around them. Cognitive disabilities may affect a person’s thinking, remembering, learning, communicating, mental health, sensory processing, or social interactions.
- **Aging Adults.** Aging adults may have trouble performing specific tasks within a set time (e.g., crossing a road or boarding a transit vehicle), standing for an extended period of time, or be more sensitive to the elements (e.g., waiting for transit in excessive heat). Aging adults are people (typically 60 years of age or older) who have physical or cognitive limitations that impact their ability to perform daily activities.

- **Limited English Proficiency (LEP) Communities.** A person with LEP refers to a person who is not fluent in the English language. Users who have LEP may have trouble understanding directions and alerts when delivered in their non-native language, may have different culture norms that make it difficult to follow directions others would feel are standard, or may have difficulty understanding wayfinding signs.
- **Low Income Communities.** Users who fall into the low-income category may be single or no-vehicle households, may have trouble accessing different forms of technology (i.e., cellphone or personal computer), may be on reduced payment or fixed payment transit plans, may be unbanked (e.g., not have access to a bank account or credit card), or may use transit as their sole means of transportation. A person who has low income has a median household income that is at or below the Department of Health and Human Services poverty guidelines. Poverty guidelines designate \$26,500 as the threshold for a household of four in the state of Georgia in 2021.

Table 1 below describes the demographics of each of these populations in the project area as a whole and in relation to Gwinnett County according to the 2017 American Community Survey. However, it must be noted that these populations are not mutually exclusive as many individuals are members of multiple of the below communities.

Table 1: Project Area Demographics

Population Type	Project Site Population	% Pop in Project Site	Gwinnett County Population	% Gwinnett Pop. in Project Site
People with Disabilities (non-institutional)	16,802	6.0%	32,032	52.5%
Aging Adults (Age 65+)	19,435	7.0%	78,898	24.6%
LEP Households	14,098	15.1%	24,069	58.6%
Low-Income (Individual Poverty)	53,223	19.1%	107,267	49.6%

The ST-CTN project will be implemented in Gwinnett County, which was chosen partially due to its representative nature. It faces many of the same challenges as much of Metro Atlanta, including suburban land-uses; wide, high-speed roadways; and inconsistent pedestrian infrastructure. This area also was chosen to leverage its implementation readiness and the CV planning work recently completed. A map of the project area can be found in **Figure 2**.

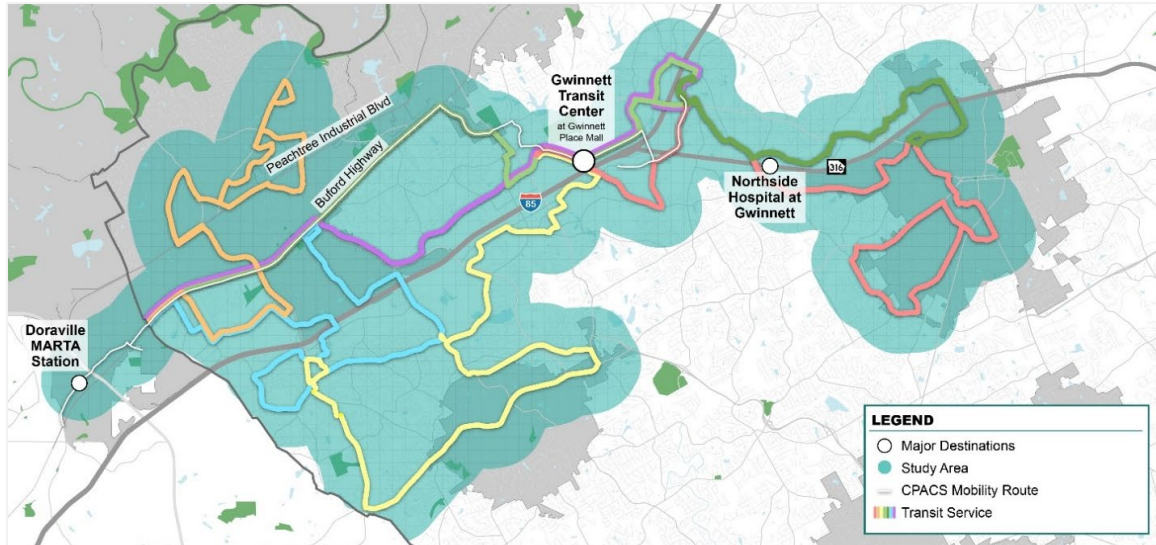


Figure 2. ST-CTN Deployment Site Map (Source: ARC)

Table 2 provides a summary of the ST-CTN project – project goal, brief project description, and reference to performance measures.

Table 2. Project Summary

Project Summary
<p>Project Title: Safe Trip in a Connected Transportation Network (ST-CTN)</p>
<p>Project Goal: The goal of the ST-CTN system is to leverage existing advanced transportation technology solutions to support safe, reliable, accessible, complete trips for all, particularly underserved communities, including people with disabilities, aging adults, people with LEP, and low-income travelers.</p>
<p>Project Description: The ST-CTN can be thought of as a <i>system of systems</i>; the scope of work required to develop, design, and deploy ST-CTN is focused on the expansion or enhancement of current systems and added connectivity between those systems. The ST-CTN concept is comprised of an integrated set of advanced transportation technology solutions (CV, transit signal priority (TSP), machine learning, predictive analytics) to support safe and complete trips, with a focus on accessibility for those with disabilities, aging adults, and those with LEP. The ST-CTN system is focused on the integration and enhancement of three primary subsystems: the ATL RIDES complete trip planner application, Space Time Memory (STM) Platform which uses enhanced network data to provide safe, accessible routes, and the Connected Vehicle (CV) deployment applications, including TSP and broadcast messaging.</p>
<p>Project Performance Measures: Performance measures will be developed and documented in the forthcoming <i>Performance Measurement and Evaluation Support Plan (PMESP) – FHWA-JPO-21-875</i>. A summary of the measures and their associated datasets will be described in the PMESP and any data required to support that plan is included within this document.</p>

2. Project Overview

The proposed ST-CTN system is being developed to address the needs and challenges that have been expressed by end users, particularly those representing underserved communities. The new system is a result of end user, infrastructure owners and operators (IOO), and system needs and desired capabilities as described in the Concept of Operations (ConOps).

The ST-CTN concept will integrate five programs currently existing or underway with regional commitments into a single system. These programs are leading the Atlanta region towards providing all travelers with a suite of innovative mobility solutions, which will be leveraged to support the ST-CTN system. ST-CTN will merge these separate initiatives through data fusion and communications network integration. ARC intends to leverage the successes of the infrastructure, tools, and capabilities of these programs and apply them to support trip planning and wayfinding for all travelers, particularly underserved communities.

*The goal of the ST-CTN system is to leverage existing advanced transportation technology solutions to support **safe, reliable, accessible, complete** trips for all, particularly underserved communities, including people with disabilities, aging adults, people with LEP, and low-income travelers.*

ST-CTN system aims to upgrade and integrate existing technologies and services to assist underserved populations with completing their trip successfully, safely, and reliably. The vision of the project is to provide users complete trip functionality with directions, conditions, and status on the links between trip legs that are personalized based on the user's profile, while connecting the user to CV infrastructure to provide safer trips and more transportation network awareness.

2.1. Change Control

This DMP is a living document that is updated during each phase of the three-phase project development – concept development (Phase 1), design and test (Phase 2), operate and evaluate (Phase 3).

During this concept development phase, a preliminary set of datasets, related access levels, storage and rules of curation and standards are described. As the project development and implementation phases are rolled out, this plan will be refined, updated and version controlled to reflect the most up-to-date information available. Each version number will be incremented numerically. The whole number, starting with version 1.0, will reflect the phase, and the decimal will reflect incremental updates during a phase, as necessary.

Specifically, the change to the document will be reflected not only in the **Revision History** table, but will be captured in **Dataset Impact Log** included in **Appendix C**. The list contained in **Table 16** includes the state of the dataset by DMP version, and the reason for the change or removal (if applicable). Several rules are implemented to ensure the integrity of the comprehensive list:

- No identifier will be reused or replaced for a different dataset. Currently, each dataset is associated with a unique identifier.
- References to the specific DMP table and row will reference where the change was made. For example, if the Data Custodian was replaced or standard compliance changed due to an update of the standard specification.
- DMP version changes will include the previous and current versions from which the change was made. This ensures that changes may be made each time the document is reissued.
- When datasets undergo multiple changes between versions, only one entry will be inserted into the table, and all the reasons for the change shall be listed in the reason column.

2.2. Relevant Sources

This section lists references cited in this document or relevant data specifications or policy provisions associated with a dataset.

Table 3. References

ID	Referenced Documents
[ARC Governance]	Atlanta Regional Commission. ARC Data Governance Best Practices and Recommendations Report: Transportation System Management and Operations (TSMO) Vision and Regional Intelligent Transportation Systems (ITS) Architecture Update. Atlanta.: Atlanta Regional Commission. (2019).
[ATL DMP]	The ATL. Atlanta-Region Rider Information and Data Evaluation System (ATL RIDES) Data Management Plan. Atlanta.: The ATL. (2021).
[BAA]	U.S. Department of Transportation, Federal Highway Administration. ITS4US Broad Agency Announcement. Washington D.C.: U.S. Department of Transportation. (2020).
[ConOps]	Atlanta Regional Commission. Deliverable Task 2.3 Concept of Operations (Report No. FHWA-JPO-21-857). Atlanta: U.S. Department of Transportation. (2021).
[GTFS]	GTFS. General Transit Feed Specification Reference. Washington D.C.: GTFS. (2019).
[HUA]	Atlanta Regional Commission. Deliverable Task 8 Human Use Approval Plan (Report No. TBD), TBD.
[ICS]	Rusnak, Sean. How an Insufficient Public Transportation System Decelerates Economic Mobility. Washington D.C.: Institute for Child Success. (2019).

ID	Referenced Documents
[IPFP]	Atlanta Regional Commission. Deliverable Task 10 Institutional, Partnership and Financial Plan (IPFP) (Report No. TBD), TBD.
[OSM]	OpenStreetMap Foundation. Open Street Map. Washington D.C.: UCL, ByteMark. (2021).
[PMESP]	Atlanta Regional Commission. Deliverable Task 5 Performance Management and Evaluation Service Plan (Report No. FHWA-JPO-21-875), TBD.
[SAE-J2735]	SAE J2735-2020 C2X Message Set. Warrendale: SAE International (2020).
[SAE-J2945]	SAE J2945-2017 On-Board System Requirements for V2V Safety Communications. Warrendale: SAE International.
[VPFP]	Guensler, R., Y. Xu, and V. Elango. Value Pricing Fellowship Project. Atlanta: Georgia Department of Transportation. (2013).

2.3. Data Schedule

The Data Schedule will extend through all three phases of the project and into the operational and maintenance period. The DMP will be updated to reflect the most current inventory of datasets (and their characteristics) needed to operate and continually evaluate system operations and performance.

For Phase 1, only limited information is available for determining the timing of data milestones. **Table 3** provides a schedule of data milestones. The times and dates are intended to provide a general expectation of when data milestones are expected to occur.

Table 4. Data Schedule

ID	Event Title	Description	Date
1	Draft Phase 1 DMP is delivered to USDOT	<ul style="list-style-type: none"> Draft Phase 1 DMP with basic information known at the time of writing. 	July 2021 (Phase 1)
2	Final Phase 1 DMP	<ul style="list-style-type: none"> Phase 1 DMP is updated with USDOT comments addressed. 	August 2021 (Phase 1)
3	Enterprise Data Governance (EDG) Data Committees	<ul style="list-style-type: none"> Establish EDG data committees to establish governance data for integrated datasets. The EDG data committees will include USDOT and IE 	March 2022 (Phase 2)

ID	Event Title	Description	Date
		representatives as observers.	
4	Select data specification for data flows	<ul style="list-style-type: none"> Identify and profile data standards for use in project; or Develop draft specifications for each subsystem (with no applicable standard); Acquire and reconcile data dictionaries from each subsystem. 	April 2022 (Phase 2)
5	Data specification profiles and curation processes	<ul style="list-style-type: none"> Describe specific schemas and profiles (i.e., PICS, interface control documents (ICD)) and common dictionary / semantics used for current systems (Atlanta-Regional Rider Information and Data Evaluation System (ATL RIDES), CV, space time memory (STM)). Describe curation and metadata processes for shared datasets. Describe method of access. Submit to Institutional Review Board (IRB) for data with personally identifiable information (PII). 	April 2022 (Phase 2)
6	Performance Measure Requirements	<ul style="list-style-type: none"> Describe datasets needed to generate performance measurement (and methods for collecting data) 	April/May 2022 (Phase 2)
7	Initial data samples provided to USDOT	<ul style="list-style-type: none"> Initial Data samples (e.g., SidewalkSIM, General Transit Feed Specification (GTFS)) are created, validated, and submitted to USDOT for review. Note: The datasets will evolve due to the Agile 	May/June 2022 (Phase 2)

ID	Event Title	Description	Date
		approach to developing ATL RIDES and the ongoing deployment of CV applications.	
8	Initial meeting with USDOT data team to review data	<ul style="list-style-type: none"> Meeting to review data with USDOT and walkthrough the data schema and DMP. 	July 2022 (Phase 2)
9	Baseline data collection starts	<ul style="list-style-type: none"> Initial collection of data on current conditions starts. Collect Gwinnett Sidewalk Data: Organize and train staff to collect Gwinnett County sidewalk datasets. Collect, review, and generate into STM format. Collect GTFS-Pathways data: Organize and train staff to collect Metropolitan Atlanta Rapid Transit Authority (MARTA) Doraville Station datasets. Collect, review, and generate sidewalkSim format. Collect Facility Data: Organize and train staff to collect a public building dataset. Collect, review, and generate into sidewalkSim format. These collection periods will include public right of way (PROW) and other datasets. No PII will be shared. See Table 10 for when data storage is ready for initial input. 	August 2022 (dependent on weather) (Phase 2)
10	DMP Update	<ul style="list-style-type: none"> DMP updated based on changes that have occurred during design and collection. 	Starting at Minimum Viable Project (MVP) of integrated system deployment and annually (August 2023; annually)

ID	Event Title	Description	Date
			(Phases 2/3)
11	Baseline data provided to USDOT	<ul style="list-style-type: none"> Baseline datasets and metadata files are made available for the USDOT and the IE to access. 	Starting at MVP of integrated system deployment (Phases 2/3)
12	Month of testing of applications begins	<ul style="list-style-type: none"> Initial upload of “after datasets” are collected and stored on project research data storage systems through testing. 	Starting at MVP of integrated system deployment (Phases 2/3)
13	Data accessed by USDOT	<ul style="list-style-type: none"> Daily updates of <i>after case data</i> are available to USDOT and IE to access. 	TBD (Phase 2/3)
14	Data Review	<ul style="list-style-type: none"> Data Review conducted with USDOT and IE to ensure datasets and metadata files are complete. 	TBD (Phase 2/3)
15	Draft Final Test Report submitted (test report will include a section on data fidelity)	<ul style="list-style-type: none"> Draft Final Task Report submitted to USDOT. 	TBD (Phase 2/3)
16	Final Test Report submitted	<ul style="list-style-type: none"> Draft Final Test Report submitted to USDOT. 	TBD (Phase 2/3)

3. Data Overview

The following sections provide an overview of the data that will be required to support the functionality of the ST-CTN system.

3.1. Data Needs Summary

The ST-CTN can be thought of as a *system of systems*. The scope of work required to develop, design, and deploy ST-CTN is focused on the expansion or enhancement of current systems and added connectivity between those systems. **Figure 3** provides a context diagram of the proposed system – indicating the system of interest and added subsystem connectivity. Each subsystem is indicated by color and icon: Sidewalk Inventory Collection Tools is burnt orange; STM Platform is peach; CV is green; ATL RIDES is turquoise; and Gwinnett County Transit (GCT) is teal. The STM Platform, ATL RIDES, and CV subsystems will each require expanded capability and added connectivity to support the proposed ST-CTN system. The Sidewalk Inventory Collection Tools and GCT existing independent systems will serve to support the proposed ST-CTN system. Data exchanges between subsystems are denoted by a gray or black line. A gray line indicates an existing and unchanged data exchange between subsystems. A black line indicates a new or upgraded data exchange between subsystems.

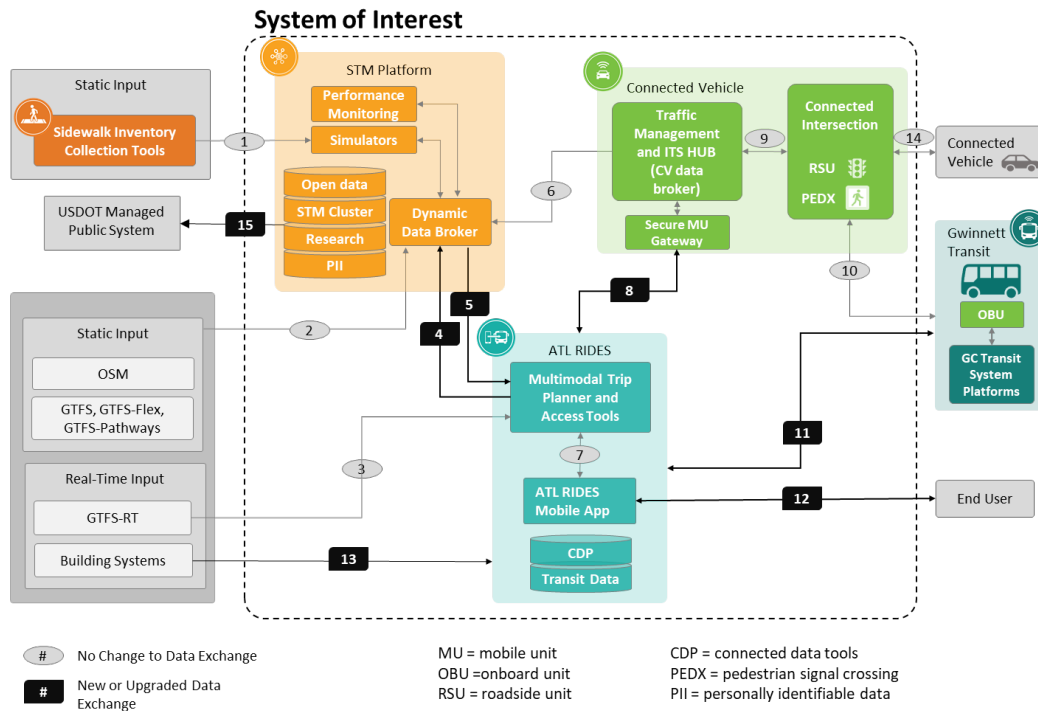


Figure 3. Safe Trip in a Connected Transportation Network Data Exchange Flow Diagram with Data Storage Systems (Source: ARC)

Critical ST-CTN data exchanges are identified by number in the context diagram above and described in **Table 5**. The grey oval labels indicate existing data exchanges that will be utilized with no change to the current data exchange. Black rectangular labels indicate data exchanges that will be new or upgraded to support the ST-CTN system.

Table 5. Critical ST-CTN Connection Descriptions (related to Context Diagram with Data Storage Systems)

Data Exchange ID (EX ID)	Description	Status
1	Sidewalk inventory data, including accessibility features to the STM Platform simulators	Updated content (not format)
2	Static and dynamic data from various existing sources to the STM Platform dynamic data broker	Updated content (not format)
3	Static and dynamic data from various existing sources to the ATL RIDES multimodal trip planner and access tools	No change
4	Mobile App logs and trip feedback	New
5	STM Network Impedance API	New
6	CV and Traffic Operations Messages: signal phasing and timing (SPaT), MapData (MAP), CV automatic terminal information service (ATIS) broadcast data, NaviGator ITS, road characteristics, traffic data	Updated
7	Open Trip Planner (OTP) APIs and ATL RIDES APIs	No change
8	Mobile Accessible Pedestrian Signal System (PED-SIG) / Pedestrian Safety Message (PSM)	New
9	CV messages	No change
10	Transit signal priority and other CV application messages	No change
11	CV application transactions for transit applications including Transit Stop Request (TSR)	New
12	ATL RIDES and Traveler exchange – profile, trip plan, settings, notifications, feedback, etc.	Updated content (not format)
13	Static and dynamic information from building facilities to the ATL RIDES	New
14	CV Data	No change
15	Project data for USDOT-managed Public System	New

3.2. Data Overview

The datasets that meet the needs of the ST-CTN are listed in **Table 6**. The table lists each dataset, referenced with a unique identifier for each dataset and subset of dataset. Each dataset type, scope and reference to the context diagram are listed in the table. The datasets are described by the following columns:

- **Dataset Identifier (ID):** Unique identifier related to every dataset or dataset subset
- **Data Exchange ID (EX ID):** References interface flow(s) in the Context Diagram (see **Figure 3**)
- **Dataset Type:** The category associated with the dataset content. Values include (Asset, Asset Condition, Crowdsourcing, CV, Land Use, Network, Network Operating Conditions, System-Customer Performance, Transit, Vulnerable Road User (VRU) Modes, Weather, Mobility Service API)
- **Dataset Name:** A name or title for the dataset.
- **Dataset Description:** Describes a short description of the dataset, its purpose (particularly which subsystem ingests, uses or generates it), its general content and current status.
- **Dataset Subset Description:** If a portion of the dataset is extracted for special use, the subset is described with its purpose and use.
- **Data Collection Methods:** Describes the method by which data is collected including input sources. Collection methods include:
 - External – acquired as input from external sources,
 - Derived – summarized, fused or integrated data generated from multiple datasets, or
 - Collect / Forward – created, collected, forwarded and stored data from system. The *collect / forward* method includes user input transactions (e.g., between APIs), web forms, user tracking methods (e.g., trace data from mobile phones).

Information on data file format will need to comply with Data Standards and profile provisions, so data file formats are included in **Table 14**. The collection method may change over time and impact the quality of the data. Hence data collection methods will be described in the technical metadata description (see **Section 5.3**).

Table 6. Data Overview Table

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
1	2	Land Use	Parcel-level Land Use Data	Polygon data for sidewalk network development, land use classification, geographic research, etc. used to develop simulation network models in STM.	Data subsets are employed in: <ol style="list-style-type: none"> 1. Demographic analysis, coupled with licensed demographic data 2. Sidewalk asset allocation to parcel boundaries 	External Input
2	2	Land Use	Building Address and Landmark Data	Address directory with addresses in the geographic region (referenced to parcel-level land use data) for use in navigation and pathfinding apps. Full data set is used in research and geographic performance metrics related to land use.	Data subsets are employed in: <ol style="list-style-type: none"> 1. Wayfinding routines 2. Interior pathway referencing 	External Input
3	2	Network	Whole Road Network	Comprehensive roadway network for Metro Atlanta, including all facility type roadway links and intersection nodes. The network is mapped to, and reconciled with, all other network data sources (serving as the master network). Full data set serves as underlying disaggregate link-node structure for all roadway networks in the region. Includes nodes needed for future Activity Based Model (ABM) and simulation model application (e.g., large parking lots that input/absorb demand). Basis for link-to-link mapping between multi-provider roadway networks. Working data sets are generated for pathway and impedance analyses. Link-and-nodes collapsed to improve algorithm processing	Data subsets are employed in: <ol style="list-style-type: none"> 1. STM Network structures 2. Connections between sidewalk, transit, and road networks 3. Updating the OSM network 4. Connecting data across travel demand and simulation models 	Derived

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
				time. Research analysis subsets created for case studies.		
4	2	Network	ABM Network	<p>Road network employed by the ARC regional activity-based travel demand model (condensed link and node structure from the whole road network). ABM outputs are used to provide STM data for conditions that have not been observed (e.g., freeway closure) and for research purposes (e.g., comparing predicted vs observed traffic).</p> <p>ABM networks are generated for each ARC modeling scenario (e.g., RTP vs. TIPA1) and calendar year for scenario analysis (e.g., 2030 model network containing roads that will be in place in 2030). Standardized naming convention ABM2020-TIPA1-2030 (150k-link network).</p>	<p>Data subsets are employed in:</p> <ol style="list-style-type: none"> 1. STM Network structures 2. Connections between sidewalk, transit, and road networks 3. Connecting data across travel demand and simulation models 	External Input
5	5	Network	STM Network	<p>Road and pathway network employed by the STM for impedance calculations and shortest path analyses. Includes all ABM links and as many links from the whole road network as deemed necessary to support mode and pathway analyses.</p> <p>Network is employed with full data set (Historic STM contains link-by-link over time for research and machine learning), working data set for current conditions (previous two hours), and forecast conditions from machine learning projections (future one hour).</p>	<p>Data subsets are employed in:</p> <ol style="list-style-type: none"> 1. Connections between sidewalk, transit, and road networks 2. Updating the OSM network 3. Connecting data across travel demand and simulation models 	Derived

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
6	2	Network	NaviGator Network	<p>Road network employed by NaviGator for reporting on-road operating condition data. Will be referenced to the whole road network to enrich basic simulation network in STM.</p> <p>The road network provides underlying network basis for Georgia Department of Transportation (GDOT) on-road working data sets for lane-by-lane and corridor speed, volume, and vehicle class splits.</p>	<p>Data subsets are employed in:</p> <ol style="list-style-type: none"> 1. Spatial referencing between the STM and other networks 2. Establishing data flow connections between connected infrastructure elements, the STM, and other databases 	External Input
7	2	Network	SRTA Managed Lane Network	<p>Road network employed by State Road and Tollway Authority (SRTA) of Georgia for reporting on-road operating condition data. Will be referenced to the whole road network to enrich basic simulation network in STM. The network will provide underlying network basis for SRTA on-road working data sets for lane-by-lane and corridor speed, volume, and vehicle class splits.</p>	<p>Data subsets are employed in:</p> <ol style="list-style-type: none"> 1. Referencing data collection locations with the STM and other networks 	External Input
8	2, 3	Network	OpenStreetMap (OSM) Network	<p>OpenStreetMap network needed to support ATL RIDES OTP engine and STM simulator component. OSM serves as the basis for all routing processes in ATL RIDES app. The extract OSM network will be updated to reflect the whole road, STM, and sidewalk networks to ensure data compatibility.</p> <p>APIs will provide update linkages and transfer of path impedance costs from the STM to the OSM format for full compatibility with the routing app.</p>	<p>Data subsets are employed in:</p> <ol style="list-style-type: none"> 1. ATL Rides Wayfinding 2. Connectivity between STM and OSM reference network 3. Processes designed to update OSM spatial accuracy 	External Input and Derived during operations

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
				Note: the updated OSM data will not be published during operations; rather the project team will work with OSM consortium to update permanent changes to the network when necessary.		
9	2	Network	Licensed Networks	Licensed subscription road network data from private data provider(s). Data flows for subscription speed and volume data will be mapped to all other roadway networks via the whole road network. These networks serve as the basis for the import of licensed operating condition data sets are used in machine learning predictions, performance metrics, and research. Working data for lane-by-lane and corridor speed, volume, and vehicle class splits.	Data subsets are employed in: <ol style="list-style-type: none"> 1. Spatial referencing to improve ABM an OSM network data 2. Spatial referencing for licensed STM speed and volume data into machine learning 	External Input
10	1	Network	Sidewalk Network	Link and node structure for all sidewalks and potential sidewalks developed from parcel-level land use and roadway link data. Full network includes sidewalks that do not yet exist (coded as width=0 and high link impedance).	Data subsets are employed in: <ol style="list-style-type: none"> 1. Referencing between STM and OSM wayfinding network 2. Impedance calculations 3. Shortest path planning analyses 	Derived
11	1, 13	Network	Indoor Pathways	The description of indoor pathways including the location and description of vertical conveyances and planned or current obstructions. Includes connectivity to the sidewalk network. Data will be formatted in OSM structure for use in ATL RIDES app wayfinding.	Data subsets are employed in: <ol style="list-style-type: none"> 1. Referencing between STM and OSM wayfinding network 2. Impedance calculations 3. Shortest path planning analyses (with impedance) 	External Input

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
12	2, 3	Network	GTFS Transit Network	General Transit Feed Specification standard structures for routes serving the study area. Coupled with GTFS stops, schedule and other GTFS files for use in the ATL RIDES app and to develop the TransitSim shortest path research tool.	Data subsets are employed in: 1. ATL RIDES wayfinding 2. TransitSim network connectivity for mode transfers	Derived
13	5	Network	TransitSim Network	Link and node structure for transit pathways used in server-based shortest path planning analyses.	Data subsets are employed in: 1. Impedance cost calculations for wayfinding	Derived
14	5	Network	BikewaySim Network	Link and node structure for bicycle-accessible facilities (includes roadway links as well as dedicated and shared use paths).	Data subsets are employed in: 1. Impedance cost calculations for wayfinding	Derived
15	2, 3	Network Operating Conditions	NaviGator Data	Roadway facility volume and speed data mapped to whole road network and STM links. GDOT roadway operating condition data sets are integrated into the STM for machine learning predictions, performance metrics, and research. Working data for lane-by-lane and corridor speed, volume, and vehicle class splits.	Data subsets are employed in: 1. STM speed and volume data for machine learning	External Input
16	2, 3	Network Operating Conditions	SRTA Data	SRTA machine vision and loop count data provide traffic volumes and speeds for managed lane facilities.	Data subsets are employed in: 1. STM speed and volume data for machine learning	External Input

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
17	2	Network Operating Conditions	SRTA Tolling Data	Spatial and temporal toll rates by toll segment (which impact route selection and travel times). These data are used in network performance predictions for future time steps in machine learning.	Data subsets are employed in: 1. Machine learning for impact of variable tolls on mode choices and route decisions	External Input
18	2	Network Operating Conditions	Subscription Roadway Operating Condition Data	Subscription roadway facility volume and speed data procured from commercial data provider(s) and mapped to an aggregated whole road network and STM network. The team will evaluate and select data providers in Phase 2. A DMP element will be prepared for the data from each provider. Licensed operating condition data sets are imported into the STM and used in machine learning predictions, performance metrics, and research. Working data for lane-by-lane and corridor speed, volume, and vehicle class splits.	Data subsets are employed in: 1. STM estimation of real-time route impedance factors 2. Machine learning processes 3. Benchmark performance assessment	External Input
19	2	Network Operating Conditions	Historic Roadway Operating Condition Data	Historic roadway facility volume and speed data procured from commercial data providers and mapped to an aggregated whole road network and STM network. The team will evaluate and select data providers in Phase 2. A DMP element will be prepared for the data from each provider. Licensed historic operating condition data sets by vendor are imported into the STM and used in machine learning predictions, performance metrics, and research.	Data subsets are employed in: 1. STM speed and volume data for machine learning	External Input

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
20	2	Network Operating Conditions	Modeled Future Operating Conditions	Model-predicted (regional ABM and simulation) future roadway facility on-road spatial and temporal operating conditions (e.g., volume and speed data) mapped to STM links for conditions that may not have been encountered in observational data for use in machine learning analyses. Each modeling run generates subsets of data employed in the STM.	Data subsets are employed in: 1. STM speed and volume data from model predictions for machine learning	Derived
21	2	Network Operating Conditions	Waze Alerts	Alerts about traffic, incidents, and work zones from Waze via the Connected Citizens Program.	Data subsets are employed in: 1. STM speed, volume, and alert triggers for machine learning	External Input
22	14	Network Operating Conditions	GDOT traffic management center (TMC) Incident Data	GDOT TMC incident real-time reports (NaviGator data set).	Data subsets are employed in: 1. STM speed, volume, and incident severity triggers for machine learning	External Input
23	14	Network Operating Conditions	GDOT TMC Special Event Data	GDOT TMC pre-planned and real-time special event reporting (NaviGator data set).	Data subsets are employed in: 1. STM speed, volume, and incident severity triggers for machine learning	External Input
24	14	Network Operating Conditions	GDOT TMC Work-Zone Data	GDOT TMC work-zone planned and real-time reporting (NaviGator data set).	Data subsets are employed in: 1. STM speed, volume, and incident severity triggers for machine learning	External Input

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
25	5	Network Operating Conditions	Network Impedance API	New data exchange to communicate changes in network impedance values for complete paths to the ATL RIDES app using OSM/OTP data structures. The API will be developed during the agile development cycles in collaboration with STM and ATL RIDES platform developers.	Data subsets are employed in: <ol style="list-style-type: none"> 1. Machine learning 2. Performance measurement 	Derived
26	1	Assets	Roadway Design and Condition Data	Roadway characteristics typically carried with planning models and operating characteristic tracking (number of lanes, lane width, speed limit, design capacity, etc.). STM carries all available design elements from each vendor data source for use in machine learning analyses.	Data subsets are employed in: <ol style="list-style-type: none"> 1. Machine learning 2. Performance measurement 	Derived
27	1	Assets	Roadway Intersection Design and Condition Data	Intersection design and operations data for vehicle operations (intersection lane design, bay length, lane-by-lane signal technology and configuration, sensors, timing plans, etc.).	Data subsets are employed in: <ol style="list-style-type: none"> 1. Machine learning 2. Performance measurement 	Derived
28	1	Assets	Pedestrian Pathway Asset Design and Condition Data	Sidewalk ramps, curb cuts, crossings, pedestrian signals, and signage. Referenced to sidewalk network links and used in impedance calculations.	Data subsets are employed in: <ol style="list-style-type: none"> 1. Pedestrian impedance calculations 2. Wayfinding via shortest path 	Derived
29	1	Assets	Pedestrian Intersection Asset Design and Condition Data	Sidewalk ramps, curb cuts, crossings, pedestrian signals, and signage at signalized intersections. Referenced to sidewalk crossing network links and used in impedance calculations. Subsets are generated by asset	Data subsets are employed in: <ol style="list-style-type: none"> 1. Pedestrian impedance calculations 2. Wayfinding via shortest path 	Derived

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
				type for performance reporting and scenario analysis for accessibility improvement scenarios (ramps, curb cuts, crossings, etc.).		
30	13	Assets	Building Pathway Asset Design and Condition Data	Building interior pathway assets such as door access, thresholds, ramps, push-button activations, signage, etc. Referenced to sidewalk approach links and interior pathway links and used in impedance calculations.	Data subsets are employed in: 1. Pedestrian impedance calculations 2. Wayfinding via shortest path	Derived
31	13	Assets	Building Wayfinding Asset Design and Condition Data	The location of wayfinding signs and announcements in facilities including transit hubs and stations. Includes status of current obstructions and vertical conveyances status (e.g., operating, out of order, under maintenance). Used in impedance calculations and shortest path generation.	Data subsets are employed in: 1. Pedestrian impedance calculations 2. Wayfinding via shortest path	Derived
32	2, 3	Transit, Assets	Transit Stop Asset Design and Condition Data	Bus stop shelters, landing pads, benches, approaches, door access points, ramps, signage, etc. Referenced to sidewalk network links and used in impedance calculations. Not currently within GTFS or OSM features but can be used in server-side impedance calculations.	Data subsets are employed in: 1. Pedestrian impedance calculations 2. Wayfinding via shortest path	External Input
33	2, 3	Transit	Transit Vehicle Asset Design and Condition Data	Information about the accessibility of specific transit vehicles (lift presence/configuration/design, lift operational status, etc.) for which real-time automatic vehicle location (AVL) data are employed. Not currently a GTFS or OSM features but can be used in server-side impedance calculations.	Data subsets are employed in: 1. Pedestrian impedance calculations 2. Wayfinding via shortest path	External Input

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
34	2, 3	Transit	GTFS (GCT)	General Transit Feed Specification data files including accessibility attributes for GCT.	Data subsets are employed in: 1. Wayfinding for ATL RIDES 2. Development of the TransitSim network for multi-modal impedance calculations	External Input
35	2, 3	Transit	GTFS (MARTA)	General Transit Feed Specification data files including accessibility attributes for MARTA.	Data subsets are employed in: 1. Wayfinding for ATL RIDES 2. Development of the TransitSim network for multi-modal impedance calculations	External Input
36	2, 3	Transit	GTFS Realtime (GCT)	GTFS-RT API for GCT. Data on where transit vehicles are located in real-time, event data, and data on when transit vehicles will arrive and depart from a stop.	Data subsets are employed in: 1. Wayfinding for ATL RIDES 2. Development of the TransitSim network for multi-modal impedance calculations	External Input
37	2, 3	Transit	GTFS Real-time (MARTA)	GTFS-RT API for MARTA transit service. Data on where transit vehicles are located in real-time, event data, and data on when transit vehicles will arrive and depart from a stop.	Data subsets are employed in: 1. Wayfinding for ATL RIDES 2. Development of the TransitSim network for multi-modal impedance calculations	External Input
38	3	Transit	GTFS-Flex	General Transit Feed Specification Flex data files for Gwinnett paratransit services. This is typically a combination of GTFS and GTFS Flex.	Data subsets are employed in: 1. Wayfinding for ATL RIDES 2. Development of the TransitSim network for multi-modal impedance calculations	External Input

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
39	15	CV	BSM	The basic safety message (BSM) is used in a variety of applications to exchange safety data regarding vehicle state and location. The BSM data will be used in this application to enhance the network operations state information in the STM.	Data subsets are employed in: <ol style="list-style-type: none"> 1. STM traffic operations state updates 2. Machine learning 3. Mobility performance measure computations 	External Input
40	8, 15	CV	PSM	Personal safety message is used to broadcast safety data regarding the kinematic state of various types of VRUs, such as pedestrians, cyclists, or road workers. Archived PSM messages will be used in the STM for performance monitoring. This includes messages sent from the CV subsystem warnings about emergency vehicles preemptions at specific intersections and railroad crossing gate closing.	Data subsets are employed in: <ol style="list-style-type: none"> 1. Safety Performance measure computations 2. Machine learning 	External Input
41	8, 15	CV	SPaT	Signal phase and timing is a message type which describes the current state of a signal system and its phases and relates this to the specific lanes (and therefore to movements and approaches) in the intersection. It is used along with the MAP message to describe an intersection and its current and future control states. The SPaT data will be used in this application to enhance the network operations state information in the STM.	Data subsets are employed in: <ol style="list-style-type: none"> 1. STM traffic operations state updates 2. Machine learning 3. Mobility performance measure computations 4. Pedestrian impedance calculations 	External Input

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
42	8,15	CV	MAP	The MAP message is used to convey many types of geographic road information. At the current time, its primary use is to convey one or more intersection lane geometry maps within a single message. The map message content includes such items as complex intersection descriptions, road segment descriptions, high speed curve outlines (used in curve safety messages), and segments of roadway (used in some safety applications). A given single MAP message may convey descriptions of one or more geographic areas or intersections. The contents of this message involve defining the details of indexing systems that are in turn used by other messages to relate additional information (for example, the signal phase and timing via the SPaT message) to events at specific geographic locations on the roadway.	Data subsets are employed in: <ol style="list-style-type: none"> 1. STM traffic operations state updates 2. Machine learning 3. Mobility performance measure computations 	External Input

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
43	11	CV	Signal Status Message Exchange	The signal status message (SSM) and signal request message (SRM) are messages exchanged between the onboard units (OBU) and a roadside unit (RSU) at a signalized intersection. The SSM is issued by the RSU while the SRM is sent by the OBU (or mobile unit (MU)). The SSM is used to relate the current status of the signal and the collection of pending or active preemption or priority requests acknowledged by the controller. It is also used to send information about preemption or priority requests which were denied. This in turn allows a dialog acknowledgment mechanism between any requester and the signal controller. The data contained in this message allows other users to determine their "ranking" for any request they have made as well as to see the currently active events. When there have been no recently received requests for service messages, this message may not be sent. While the outcome of all pending requests to a signal can be found in the SSM, the current active event (if any) will be reflected in the SPaT message contents.	Data subsets are employed in: <ol style="list-style-type: none"> 1. STM traffic operations state updates 2. Machine learning 3. Mobility performance measure computations 4. Pedestrian impedance calculations 	External Input
44	8	CV	Ped-X	A series of messages associated with pedestrian signal control including change interval, clearance time, phase, and walk interval. Archived Ped-X messages will be used in the STM for performance monitoring.	Data subsets are employed in: <ol style="list-style-type: none"> 1. STM traffic operations state updates 2. Machine learning 	Collect/Forward

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
					<ul style="list-style-type: none"> 3. Mobility performance measure computations 4. Pedestrian impedance calculations 	
45	12	Mobility Service API	Trip options	Trip options calculated from OTP Routing Engine. Calculated itinerary results when inputting an origin and destination in the OTP engine based on the personalized OSM network.	<p>Data subsets are employed in:</p> <ul style="list-style-type: none"> 1. Benchmark performance assessment 2. Trip destination and purpose research 3. Route adherence research to improve impedance factors 	Collect/Forward
46	2	VRU Modes	VRU categories	List of categories and their default edge impedance values. The enumerated list will correspond to the list of disabilities and assistive devices offered in the ATL RIDES preference menu.	<p>Data subsets are employed in:</p> <ul style="list-style-type: none"> 1. Benchmark performance assessment by VRU category 2. Trip destination and purpose research by VRU category 3. Route adherence research to improve impedance factors by VRU category 	Derived
47	2	Weather	Weather data	Open weather information from NOAA, including precipitation and temperature data as well as severe weather alerts for use in impedance calculations and routing decisions. Precipitation subset, temperature subset, severe weather alert subset	<p>Data subsets are employed in:</p> <ul style="list-style-type: none"> 1. Impedance calculations 2. Machine learning related to mode and route choice 	External Input

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
48	2	Demographics	Customer Demographic Data	User demographic data (standard transportation planning categories) for use in research analyses.	Data subsets are employed in: 1. Benchmark performance assessment 2. Socioeconomic impact assessment research	External Input
49	2	Demographics	Household Level Licensed Demographic Data	Licensed household-level demographic data for use in research analyses licensed from marketing firms.	Data subsets are employed in: 1. Benchmark performance assessment 2. Socioeconomic impact assessment research	External Input
50	2	Demographics	Household Level Vehicle Registration Data	Licensed household-level vehicle ownership data (not available from licensed demographic providers) for use in research analyses.	Data subsets are employed in: 1. Benchmark performance assessment 2. Socioeconomic impact assessment research	External Input
51	4	System-Customer Performance	Mobile App Logs	ATL Mobile app log files which include all the trips, trip preferences and travel results as well as user's app usage logs will be forwarded to the STM dynamic data broker for analysis and aggregation into performance measures.	Data subsets are employed in: 1. Benchmark performance assessment 2. Socioeconomic impact assessment research	Collect/Forward
52	4	System-Customer Performance	Traverse Data	Customer traverse data through the system (in space and time at highest practical resolution) for use in performance assessment (response times, wait times, travel times, etc.) and that can be compared to recommended routes for	Data subsets are employed in: 1. Benchmark performance assessment	Collect/Forward

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
				use in refining impedance calculations and route recommendations.	2. Socioeconomic impact assessment research	
53	4	System-Customer Performance	Trip Feedback Reports	ATL RIDES trip reports and feedback from app users including survey data from app users.	Data subsets are employed in: 1. Benchmark performance assessment 2. Socioeconomic impact assessment research	Collect/Forward
54	Flow from 12 to 4	Crowdsourcing	Trip Crowdsourcing Reports	Crowdsourced updates on asset design and condition data for use in verified updating of the asset data sets. Extracted subsets from ATL RIDES Trip Feedback reports include information on impacts to sidewalks, intersections, bus stops, ramps, curb cuts, signage, etc.	Data subsets are employed in: 1. Proposed updates to network feature attributes (to be reviewed and approved) 2. Network impedance calculations 3. Alert messaging by data type	Collect/Forward
55	3	Mobility Service API	Geocode.earth API	Address information that is used to generate locations in the ATL RIDES trip planner. The API is used to convert addresses/landmarks into spherical coordinates and back for trip planning purposes.	Data subsets are employed in: 1. ATL RIDES navigation routines 2. Trip destination and trip purpose research	External Input
56	2	Demographics	Business Level Licensed Facility Data	Business demographic data (standard transportation planning categories) for use in research analyses.	Data subsets are employed in: 1. Simulating changes in network accessibility by mobility mode 2. Socioeconomic impact assessment research	External Input
57	2	Energy/Emissions	MOVES-Matrix Energy Consumption	Energy and emission rate matrix from the United States Environmental Protection Agency's motor vehicle emissions simulator (MOVES) model. The matrix contains rates	Data subsets are employed in energy and emissions analyses employed as performance metrics	External Input

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
			and Emission Rates	per vehicle source type, on-road operating condition, environmental conditions, calendar year, etc., for the metro area.		
58	-	System-Customer Performance	ST-CTN Performance Measures Data	Ongoing random sample data collection conducted through the ATL RIDES app will gather customer opinion data on system performance. Standardized questions on a Likert scale and open comment fields will be used to collect data.	Likert scale values will be collected to gauge changes in individual satisfaction with specific system features and outcomes.	Derived
59	-	STM Performance Logs	STM Communication Logs	The STM will continuously track inbound and outbound communications with time-stamps for use in assessing latency.	Logged time stamps will be used to continuously track and quantify latency along each communications leg.	Derived
60	-	STM Performance Logs	STM Impedance Calculation Logs	The STM will continuously track the time at which impedance calculations begin and are completed to assess computational speed.	Logged time stamps will be used to continuously track and quantify algorithm speeds.	Derived
61	-	CV Performance Logs	Secure MU Gateway event logs	The CV will collect and store the transactions received and transmitted to the ATL RIDES through the SMUG	Separated by CV application.	Derived
62	-	Incident Data	Pedestrian Crash Data	GDOT retains a database (queriable) of all individual crash reports (i.e., data from all form-fillable crash report fields). These reports can be screened to eliminate PII; however, the data cannot be released without review to ensure that the officers left no data in the sketch or comments fields that could be used to identify individuals.	The crash database can be employed to assess changes in the numbers and types of crash events in the deployment zone over time. The team anticipates that there will be insufficient numbers of events to infer impacts on causality, so a case control analysis of pedestrian events will also be conducted using crash data coupled with police reports.	External Input

Data ID	EX ID	Dataset Type	Dataset Name	Dataset Description	Dataset Subset Description	Collection Method
					Furthermore, many pedestrian and bicycle incidents are not reported into the database.	
63	-	Incident Data	Pedestrian Incidents Police Reports	Individual police reports for each crash are retained by GDOT (sent by police agencies to the GDOT clearinghouse). The police reports contain the original information used by officers to populate the crash database fields for each event. The police reports contain officer descriptions of the incidents, drawings, and other details.	Police descriptions and drawings are used in case-control analysis of individual crash events to assess potential causal factors.	External Input
64		Transit	Ridership: Fixed Route	Transit vehicle ingress and egress counts collected by automated passenger count (APC) equipment.	Pedestrian count data will be used to assess changes in vehicle occupancy and passenger throughput for transit metrics.	External Input
65		Transit	GCT Complaint Log	GCT maintains an electronic incident log that contains the records of individual passenger complaints that reach the call center.	Subset of complaints associated with transit service, routes, stop locations, navigation, and other factors employed in user-related performance metrics.	External Input

4. Data Stewardship

A data governance framework establishes a formal organizational structure on the capture, curation, maintenance and dissemination of data. Because the ST-CTN project is driven by exchanging and analyzing data for traveler trip generation and execution, data curation, quality, access, privacy and security, the project will establish a formal data governance organization to manage and oversee integration and access to the data. Governance organizational hierarchy, roles and responsibilities are described in **Figure 4**. This governance organization will be described and the schedule for convening and facilitating the Board will be described in the [IPFP] and in the ARC Program Management Plan for Phase 2.

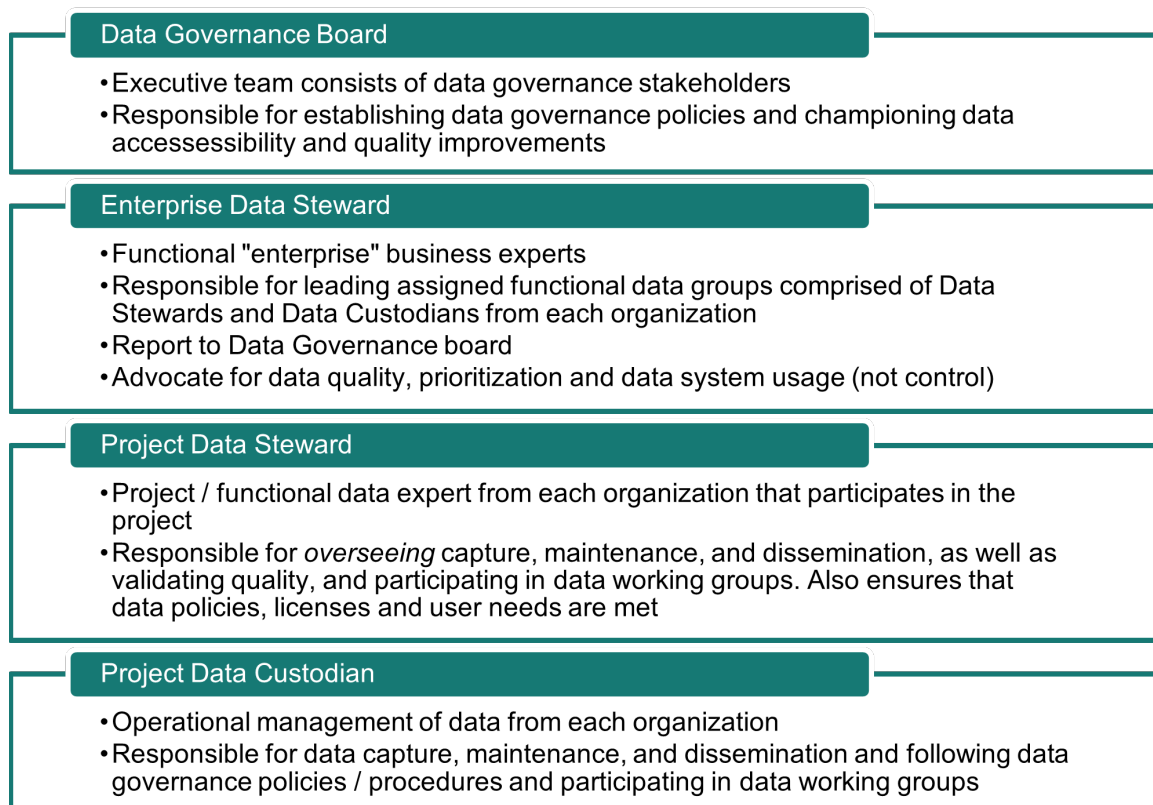


Figure 4. Data Governance Roles for the ST-CTN ITS4US Project (adapted from source: ARC Data Governance, p. 17)

In general, the *project data custodians* are subconsultants of public agencies that fund and operate subsystems of the ST-CTN project. The *project data custodians* process, manage, maintain and validate a dataset that is captured, used or generated by one or more of the subsystems. The *project data steward*, typically staff of a public agency, operates the project and oversees the data custodian(s) activities. The *enterprise data steward* (EDS) is a committee that

addresses crosscutting issues, policies and quality standards needed to ensure an interoperable environment to exchange data. Finally, the *data governance board* is an executive team of public sector stakeholders that champion the data governance policies.

A *data owner* may be a data steward if the data is generated from a system owned by the stakeholder, however, it may be the person about whom the data was captured, for example, a trip trace or account information stored in an account management system. It is then the responsibility of the data steward to enforce policies to protect PII and delete data for which the owner does not wish to be saved. In addition, data may be acquired from a third-party organization, in which case their license agreement governs usage and distribution rights.

The Project Data Steward role assigned to specific datasets is listed in **Table 7**.

The Project Data Custodian role is currently under contract to a data steward organization, as listed in **Table 7**.

The EDS and the EDG committee will be held by ARC for the duration of this project (the role will not be listed in **Table 7**). Details of the roles and responsibilities of the EDS and EDG will be described in the [IPFP] and will be included in future iterations of this DMP.

The ARC data governance framework, described above, will be implemented in parallel with the BAA/NOFO governance roles described in **Table 6**. As the Federal Sponsor, USDOT will become the dataset owner of data where the capture, generation or derivation is within the scope of this project.

4.1. Data Owner and Stewardship

Roles are assigned to each dataset that is captured, ingested, generated, used and disseminated in the ARC ST-CTN project that is consistent with the USDOT-managed Data Store governance model. **Table 7** lists roles associated with each dataset with the following columns:

- **Dataset Identifier (ID)** – unique identifier assigned to each dataset (only full datasets are included) from **Table 6**.
- **Dataset Title** – name of the dataset from **Table 6**.
- **Generated by Project** – indicates datasets that are generated by the project and not proprietary or subject to data sharing restrictions.
- **Data Owner** – owner of the dataset, organization that licenses or grants access to the dataset. The data owner is the person or organization with the authority, ability, and responsibility to access, create, modify, store, use, share, and protect the data. Data owners have the right to delegate these privileges and responsibilities to other parties. During project testing and initial operations, the dataset owner is the organization designated in Dataset Owner column.
- **Data Steward** – responsible for overseeing the capture, curation, maintenance and dissemination of the dataset, as well as enforcing the data policies in compliance with license agreements. Specifically, the data steward, at the direction of the data owner, is

the person or organization that is delegated the privileges and responsibilities to manage, control, and maintain the quality of a data asset throughout the data lifecycle, including delegating the curation and operations to the data custodian (as described in the governance process). The data steward may also apply appropriate protections, restrictions, and other safeguards depending on the nature of the data, subject to the direction of the data owner.

- **Federal Sponsor** – the project sponsor of datasets where the capture, generation or ownership is within the scope of this project. The federal sponsor will assume the role of Data Owner once the dataset(s) are provided to them per BAA/NOFO requirements later in the project. The federal sponsor only appears when the dataset is generated by the project and not subject to data sharing restrictions.

Table 7. Dataset Roles in Data Governance Organization

ID	Dataset Title	Generated by Project	Data Owner	Data Steward	Federal Sponsor
1	Parcel-level Land Use Data	No	Gwinnett County	ARC	Out of scope
2	Building Address and Landmark Data	No	ARC	ARC/Georgia Institute of Technology (GA Tech)	Out of scope
3	Whole Road Network	Yes	ARC	ARC/GA Tech	USDOT ITS JPO
4	ABM Network	No	ARC	ARC	Out of scope
5	STM Network	Yes	STM	STM (GA Tech)	USDOT ITS JPO
6	NaviGator Network	No	GDOT	GDOT	Out of scope
7	SRTA Managed Lane Network	No	SRTA	STM	Out of scope
8	OSM Network	No	OSM	Atlanta-Regional Transit Link Authority (ATL) (IBI) and ARC (GA Tech)	Out of scope
9	Licensed Networks (private data provider)	No	By Source	STM (GA Tech)	Out of scope
10	Sidewalk Network	Yes	ARC	GA Tech	USDOT ITS JPO

ID	Dataset Title	Generated by Project	Data Owner	Data Steward	Federal Sponsor
11	Indoor Pathways	Yes	ARC	GA Tech	USDOT ITS JPO
12	GTFS Transit Network	Yes	ARC	GA Tech / IBI	USDOT ITS JPO
13	TransitSim Network	Yes	GA Tech	GA Tech	USDOT ITS JPO
14	BikewaySim Network	Yes	GA Tech	GA Tech	USDOT ITS JPO
15	NaviGator Data	No	GDOT	GDOT	Out of scope
16	SRTA Data	No	SRTA	SRTA	Out of scope
17	SRTA Tolling Data	No	SRTA	SRTA	Out of scope
18	Subscription Roadway Operating Condition Data	No	GDOT	GDOT/STM (GA Tech)	Out of scope
19	Historic Roadway Operating Condition Data	No	GDOT	GDOT/STM (GA Tech)	Out of scope
20	Modeled Future Operating Conditions	Yes	STM	ARC/GA Tech	USDOT ITS JPO
21	Waze Alerts	No	Waze	STM (GA Tech) / IBI	Out of scope
22	GDOT TMC Incident Data	No	GDOT	GDOT	Out of scope
23	GDOT TMC Special Event Data	No	GDOT	GDOT	Out of scope
24	GDOT TMC Work-Zone Data	No	GDOT	GDOT	Out of scope
25	Network Impedance API	Yes	STM	STM (GA Tech)	USDOT ITS JPO
26	Roadway Design and Condition Data	Yes	STM	STM (GA Tech)	USDOT ITS JPO

U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology
Intelligent Transportation System Joint Program Office

ID	Dataset Title	Generated by Project	Data Owner	Data Steward	Federal Sponsor
27	Roadway Intersection Design and Condition Data	Yes	STM	STM (GA Tech)	USDOT ITS JPO
28	Pedestrian Pathway Asset Design and Condition Data	No	Gwinnett County	Gwinnett County	Out of scope
29	Pedestrian Intersection Asset Design and Condition Data	Yes	Gwinnett County	Gwinnett County	USDOT ITS JPO
30	Building Pathway Asset Design and Condition Data	Unknown	Facility Owner (TBD)	Facility Owner	unknown
31	Building Wayfinding Asset Design and Condition Data	Unknown	Facility Owner (TBD)	Facility Owner	unknown
32	Transit Stop Asset Design and Condition Data	Yes	Gwinnett County	Gwinnett County	USDOT ITS JPO
33	Transit Vehicle Asset Design and Condition Data	Yes	GCT	ATL (IBI)	USDOT ITS JPO
34	GTFS (GCT)	No	GCT	ATL (IBI)	USDOT ITS JPO
35	GTFS (MARTA)	No	MARTA	ATL (IBI)	USDOT ITS JPO
36	GTFS Realtime (GCT)	No	GCT	ATL (IBI)	USDOT ITS JPO
37	GTFS Real-time (MARTA)	No	MARTA	ATL (IBI)	USDOT ITS JPO
38	GTFS-Flex	Yes	GCT	ATL (IBI)	USDOT ITS JPO
39	BSM	Yes	GDOT/GCDOT	GDOT/ GCDOT	USDOT ITS JPO
40	PSM	Yes	GDOT/GCDOT	GDOT/GCDOT	USDOT ITS JPO

ID	Dataset Title	Generated by Project	Data Owner	Data Steward	Federal Sponsor
41	SPaT	Yes	GDOT/GCDOT	GDOT/GCDOT	USDOT ITS JPO
42	MAP	Yes	GDOT/GCDOT	GDOT/GCDOT	USDOT ITS JPO
43	SSM Exchange	Yes	GDOT/GCDOT	GDOT/GCDOT	USDOT ITS JPO
44	Ped-X	Yes	ATL/GCDOT	ATL/GCDOT	USDOT ITS JPO
45	Trip Options	No	ATL	ATL (IBI)	Out of scope
46	VRU categories	Yes	ARC	ARC (GA Tech)	USDOT ITS JPO
47	Weather data	No	NOAA	STM (GA Tech)	Out of scope
48	Customer Demographic Data	Yes	GA Tech	GA Tech	USDOT ITS JPO
49	Household Level Licensed Demographic Data	No	ARC	GA Tech	Out of scope
50	Household Level Vehicle Registration Data	No	GA Department of Revenue (DOR)	Georgia Tech Research Institute (GTRI)	Out of scope
51	Mobile App Logs	Yes	ATL	ATL (IBI)	USDOT ITS JPO
52	Traverse Data	Yes	GA Tech	GA Tech	USDOT ITS JPO
53	Trip Feedback Reports	Yes	ATL	ATL	USDOT ITS JPO
54	Trip Crowdsourcing Reports	Yes	Owner of Asset Data Sets	Owner/User of Asset Data Sets (e.g., STM, ATL Rides, ARC, etc.)	USDOT ITS JPO
55	Geocode.earth API	No	Geocode.earth	Geocode.earth	Out of scope
56	Business Level Licensed Facility Data	no	ARC	GA Tech	Out of scope
57	MOVES-Matrix Energy Consumption and Emission Rates	no	GA Tech	GA Tech	Out of scope

ID	Dataset Title	Generated by Project	Data Owner	Data Steward	Federal Sponsor
58	ST-CTN Performance Measures Data	yes	ATL / GA Tech	GA Tech	USDOT ITS JPO
59	STM Communication Logs	yes	ATL / GA Tech	GA Tech	USDOT ITS JPO
60	STM Impedance Calculation Logs	yes	ATL / GA Tech	GA Tech	USDOT ITS JPO
61	Secure MU Gateway event logs	yes	GDOT/GCDOT	GCDOT	USDOT ITS JPO
62	Pedestrian Crash Data	no	GDOT	GDOT	Out of scope
63	Pedestrian Incidents Police Reports	no	GDOT	GDOT	Out of scope
64	Ridership: Fixed Route	no	GCT	GCT	Out of scope
65	GCT Complaint Log	no	GCT	GCT	Out of scope

4.2. Access Level

This project includes a large number of datasets used in a wide variety of applications. Dataset access levels are differentiated by proprietary data which are restricted by license or usage agreements versus protected datasets that require access controls to preserve privacy and security. Data license restrictions and provisions are discussed in **Section 4.3**. This section describes access levels at four levels. Data access levels are defined as:

- **Open** –Data that can be used by the public with no or limited licensing restrictions. This data is available to the public without needing to request permissions and will be provided to the USDOT-managed Public System. These will be anonymized or aggregated version of private datasets to protect PII.
- **Private** – Data that cannot be shared with external users. Access to these data is limited and only granted with IRB and Project Team approvals. Subcategories:

- **Operational** – real-time and other data used in the applications and operations of the system. The data may contain licensed data restricted by usage agreements.
- **Proprietary** – Licensed data from third parties or commercial business interests (CBI). This data may be used for planning or operational purposes. Any access to the data is determined by usage agreements between the parties.
- **Research** – Data that is available for research, but users of the data must meet IRB requirements before gaining access to the data. These datasets may have PII. These datasets are compiled for machine learning and research purposes that do not contain PII.
- **PII Certification** – Data that has PII included in the data set. The access to this data should be as restrictive as possible to protect the PII based on IRB-approved processes. Data in this category should have an operational purpose that justifies its storage.

Raw data are stored and archived to ensure data integrity. Data used in active implementation of the system will appear in online working datasets that are updated and queried in real time.

PII data protection is governed by GA Tech secure sever systems data management protocols:

- Transmission of encrypted PII data to the GA Tech interface server.
- Decryption of data on the GA Tech Secure Server.
- Elimination of remote access to the Secure Server.
- Direct hard-wire connections between the Secure Server and the data analysis terminals located in the Secure Data Lab on the first floor of the SEB Building.
- Secure Data Lab door security that requires cardkey access.
- Login ID and password protection to terminals and the secure server.
- Layered protection within the secure server data storage system to limit access to different datasets to specific users.

Access to protected data is governed by non-disclosure agreement (NDA) and approved IRB protocols:

- IRB approval is required for any access to PII data (amendments may be submitted to facilitate supplemental data access and uses while ensuring continued privacy protection)
- Individual access to the data requires card key access to the Secure Data Lab, login ID and password access to the Secure Data Lab terminals, and pre-approved login ID and password access to the secure server, login ID and password access to each data folder.

- Access to PII data can only be accomplished through dedicated direct-connect terminals located in the Secure Data Lab on the first floor of the Sustainable Education Building (no remote access)

4.2.1. Private Datasets

The restriction of datasets occurs on two levels: proprietary datasets constrained by license agreements and protected datasets that are restricted because they contain PII. **Table 8** describes the classification of every dataset as Protected, Proprietary, or Open and identifies the basis for private data classification. All proprietary data are stored on the secure server and working datasets are generated to which access is limited to project-specific modeling and data processing routines. Licensed data may be made available by the data owners to third parties. Most agencies that limit access to their data require the execution of a data user agreement to ensure data integrity and track data usage (at no fee). Data licensed from private companies generally contain intellectual property and commercial value. Access to these data typically requires the execution of a data license and fees are likely required to obtain data access. Use of private data is minimized in this project, with the focus being on the use of open data.

The datasets are described by the following columns:

- **Dataset Identifier (ID):** unique identifier related to every dataset or dataset subset
- **Dataset Name:** the name or title for the dataset.
- **Access Levels:** describes the how individuals may gain access or how the datasets may be used. Values include: PII Certified, Research, Proprietary, Operational and Open.
- **Reason(s) Data is Private:** describes why the data is designated as proprietary or protected. General reasons are stated above.
- **Safeguarding Methods and Processes:** describes the method used to secure and gain access to the data.

Table 8. List of Privacy and Access Levels Assigned to Datasets

ID	Dataset Name	Access Level	Reason(s) Data is Private	Safeguarding Methods and Processes
1	Parcel-level Land Use Data	Research / Proprietary	NA	Access to licensed data may be granted by the data owner. Third-parties are usually required to execute a user agreement to use the data.
2	Building Address and Landmark Data	Operational	The data is restricted by license	NA
3	Whole Road Network	Open	NA	NA
4	ABM Network	Open	NA	NA
5	STM Network	Open	NA	NA
6	NaviGator Network	Open	NA	NA
7	SRTA Managed Lane Network	Open	NA	NA
8	OSM Network	Open	NA	NA
9	Licensed Networks (CBI data, etc.)	Research / Proprietary	The data is restricted by the license.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data.
10	Sidewalk Network	Open	NA	NA

ID	Dataset Name	Access Level	Reason(s) Data is Private	Safeguarding Methods and Processes
11	Indoor Pathways	Open	NA	NA
12	GTFS Transit Network	Open	NA	NA
13	TransitSim Network	Open	NA	NA
14	BikewaySim Network	Open	NA	NA
15	NaviGator Data	Open	Subject to GDOT usage agreement.	Although data is open, access to data is restricted by data usage agreement with GDOT. Third-parties typically must execute a user agreement with the data owner to access data.
16	SRTA Data	Research / Proprietary	Subject to SRTA data license and usage restrictions	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data.
17	SRTA Tolling Data	Research / Proprietary	Subject to SRTA data license and usage restrictions	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data.
18	Subscription Roadway Operating Condition Data	Research / Proprietary	The data is restricted by the license.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data. Fees are likely required to obtain data access.

ID	Dataset Name	Access Level	Reason(s) Data is Private	Safeguarding Methods and Processes
19	Historic Roadway Operating Condition Data	Research / Proprietary	The data is restricted by the license.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data. Fees are likely required to obtain data access.
20	Modeled Future Operating Conditions	Research / Proprietary	Derived from proprietary data sources.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data.
21	Waze Alerts	Operational	The data is restricted by the license.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data. Fees are likely required to obtain data access.
22	GDOT TMC Incident Data	Operational	The data is restricted by a usage agreement.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data.
23	GDOT TMC Special Event Data	Operational	The data is restricted by a usage agreement.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data.
24	GDOT TMC Work-Zone Data	Operational	The data is restricted by a usage agreement.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data.

ID	Dataset Name	Access Level	Reason(s) Data is Private	Safeguarding Methods and Processes
25	Network Impedance API	Operational; Open	NA	NA
26	Roadway Design and Condition Data	Open	NA	NA
27	Roadway Intersection Design and Condition Data	Open	NA	NA
28	Pedestrian Pathway Asset Design and Condition Data	Operational	Subject to agency data license and usage restrictions	NA
29	Pedestrian Intersection Asset Design and Condition Data	Operational	Subject to agency data license and usage restrictions	NA
30	Building Pathway Asset Design and Condition Data	Open	NA	NA
31	Building Wayfinding Asset Design and Condition Data	Open	NA	NA
32	Transit Stop Asset Design and Condition Data	Open	NA	NA
33	Transit Vehicle Asset Design and Condition Data	Operational	Subject to agency data license and usage restrictions	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data.
34	GTFS (GCT)	Open	NA	NA

ID	Dataset Name	Access Level	Reason(s) Data is Private	Safeguarding Methods and Processes
35	GTFS (MARTA)	Open	NA	NA
36	GTFS Realtime (GCT)	Open	NA	NA
37	GTFS Real-time (MARTA)	Open	NA	NA
38	GTFS-Flex	Open	NA	NA
39	BSM	Open	NA	NA
40	PSM	Open	NA	NA
41	SPaT	Open	NA	NA
42	MAP	Open	NA	NA
43	SSM	Open	NA	NA
44	Ped-X	Open	NA	NA
45	Trip options	Operational (real-time, transmitted, deleted), PII Certification (raw data)	Customer names and other traveler destinations such as home/work locations, and daycare/school locations, all constitute PII. Turn-by-turn directions reveal address locations.	Data protection governed by Georgia Tech secure sever systems data management protocols. Access conditions governed by NDA and approved IRB protocols.

ID	Dataset Name	Access Level	Reason(s) Data is Private	Safeguarding Methods and Processes
46	VRU categories	Open	NA	NA
47	Weather data	Open	NA	NA
48	Customer Demographic Data	PII Certification	Customer names, home/work locations, and daycare/school locations, all constitute PII and cannot be revealed.	Data protection governed by Georgia Tech secure sever systems data management protocols. Access conditions governed by NDA and approved IRB protocols.
49	Household Level Licensed Demographic Data	Research / Proprietary	The data is restricted by a usage agreement.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data. Fees are likely required to obtain data access.
50	Household Level Vehicle Registration Data	PII Certification	The State classifies vehicle registration as PII. Polk and others send people to County courthouses to manually collect vehicle transaction data because these data are PII. Data cannot be released as the names and addresses can be	Data protection governed by GA Tech/DOR protocols. Access conditions governed by NDA and approved GA Tech/DOR protocols. Third-parties may not access the data.

ID	Dataset Name	Access Level	Reason(s) Data is Private	Safeguarding Methods and Processes
			inferred from other data sets in the marketplace.	
51	Mobile App Logs	PII Certification (raw data); Open (when PII is removed)	Customer names and other sensitive trip related data (e.g., home/work locations, healthcare visits and daycare/school locations), all constitute PII. Trip logs reveal address locations.	Data protection governed by GA Tech secure sever systems data management protocols. Access conditions governed by NDA and approved IRB protocols.
52	Traverse Data	PII Certification	Customer names, home/work locations, and daycare/school locations, all constitute PII. Traverse data reveal address locations.	Data protection governed by GA Tech secure sever systems data management protocols. Access conditions governed by NDA and approved IRB protocols.
53	Trip Feedback Reports	PII Certification (raw data); Open (when PII is removed)	Customer names, home/work locations, and daycare/school locations, all constitute PII. Trip	Data protection governed by GA Tech secure sever systems data management protocols. Access conditions governed by NDA and approved IRB protocols.

ID	Dataset Name	Access Level	Reason(s) Data is Private	Safeguarding Methods and Processes
			logs reveal address locations.	
54	Trip Crowdsourcing Reports	Open	NA	NA
55	Geocode.earth API	Proprietary	Data is restricted by license agreement.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data. Fees are likely required to obtain data access.
56	Business Level Licensed Facility Data	Research / Proprietary	Data is restricted by license agreement.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data. Fees are likely required to obtain data access.
57	MOVES-Matrix Energy Consumption and Emission Rates	Open Data	NA	NA
58	ST-CTN Performance Measures Data	Open Data	NA	NA
59	STM Communication Logs	Open Data	NA	NA
60	STM Impedance Calculation Logs	Open Data	NA	NA
61	Secure MU Gateway event logs	Research	The raw data includes transactions from	Data protection governed by GA Tech secure sever systems data management protocols.

ID	Dataset Name	Access Level	Reason(s) Data is Private	Safeguarding Methods and Processes
			individual requests for transit pedestrian indication (TPI), PED-X and may contain PSM.	Access conditions governed by NDA and approved IRB protocols.
62	Pedestrian Crash Data	Research	The data contains data that needs to be secured.	Data protection governed by GA Tech secure sever systems data management protocols. Access conditions governed by NDA and approved IRB protocols.
63	Pedestrian Incidents Police Reports	Research	The data contains data that needs to be secured.	Data protection governed by GA Tech secure sever systems data management protocols. Access conditions governed by NDA and approved IRB protocols.
64	Ridership: Fixed Route	Open Data	NA	NA
65	GCT Complaint Log	Research	The data contains data that needs to be secured.	Data protection governed by GA Tech secure sever systems data management protocols. Access conditions governed by NDA and approved IRB protocols.

4.2.2. Access Request

Access to protected datasets requires conformance with all PII data protection security policies and approval of the IRB. The analysts on the project team will be able to access the PII datasets on the secure server at the GA Tech (as will be delineated in the forthcoming Human Use Approval [HUA] Plan). Access for third-party members to the protected data (such as the ITS4US IE or researchers who desire to use PII data in research) may be granted through the implementation of a NDA and approval of an IRB human subjects protocol amendment developed specifically for the data access purpose and need. This method was implemented in the Commute Atlanta study to allow external researchers to travel to the Georgia Tech secure data lab to use instrumented vehicle data for development of new energy, emissions, and travel behavior models [VPFP]. Datasets with Access Levels of “PII Certification” in **Table 7** are subject to this restriction.

4.2.3. Related Tools, Software and/or Code

The project has focused on the use of open data and open-source tools for data processing. No specialized tools are required to use any of the data in the system. As indicated in the previous sections, some proposed data may require the execution of a license agreement and open distribution of original or processed data may be limited by a license agreement.

4.2.4. Relevant Privacy and/or Security Agreements

Formal privacy and security agreements will be developed among the stakeholder partners and data owners prior to implementation in Phase 2. The [IPFP] will describe the plan for generating these memoranda of understanding (MOU) and usage agreements among the stakeholder organizations operating and sourcing datasets. All dataset user and license agreements will be compiled when the system “goes live” and the team will make every effort to convince the data owners to make these datasets available in an open-source format that does not require the execution of a user agreement or payment of any fees. Several stakeholders (data owners) have already established data security, data privacy, and protected data access policies. The project dataset privacy and security agreements will incorporate appropriate agency dataset policy into these agreements.

The policies are described by the following columns:

- **Data Owner:** the owner and stakeholder responsible for issuing license or agreement.
- **Data Owner Type:** the type of organization issuing agreement or license. Values include: academic, public (public agency), commercial, other.
- **Policy Type:** the type of policy or agreement issued by the data owner. Although the values vary by organization, the types of agreements contained in this table relate to dataset access, usage, security and privacy.

Table 9: Relevant Privacy and Security Policies and Agreements

Data Owner	Data Owner Type	Policy Type	Policy Reference
ATL	Public	TBD, will be developed during Phase 2.	More details on the scope and type of policy and user agreements will be included in the IPFP
GA Tech	Academic	License and usage policies (copying, using, redistribution, etc.)	<u>Usage Policy</u>
GA Tech	Academic	Data privacy and access policies (1)	<u>Data access policy</u>
GA Tech	Academic	Data privacy and access policies	<u>Data Privacy</u>
GA Tech	Academic	Cybersecurity	<u>Cybersecurity policy</u> https://policylibrary.gatech.edu/information-technology/cyber-security-policy
GA Tech	Academic	Protected data practices	<u>Protected data practices</u> https://policylibrary.gatech.edu/information-technology/cyber-security-policy
GDOT - smart corridor policies	Public	TBD, will be developed as part of the Connected Corridor project (Spring 2022)	TBD
SRTA/ ATL (for ATL RIDES)	Public	Network security policy	See [ATL DMP]
SRTA/ ATL (for ATL RIDES)	Public	Systems security policy	See [ATL DMP]

4.3. Re-Use, Redistribution, and Derivative Products Policies

The project team is making every effort to employ open access data to the greatest extent practicable in this project. The resulting transportation networks, sidewalk and roadway infrastructure design and condition data, and the space-time memory (STM) will all be open access. There are some datasets for which the team will negotiate with third party data providers to allow the input data to flow into the open access outputs. For example, the team has proposed using high-quality private map data vendor nodes in the development of the whole road network that will be employed in the STM and the update to the OSM network. If the private vendor elects not to license the node data to the team, the team will proceed with development of the whole road network using OSM nodes and a manual QA/QC process. With respect to agency data that traditionally require the execution of a data management agreement to ensure data integrity and track data usage, the team will execute agreements with the partners to allow the data to flow into the STM, at which time they become open access and can be accessed by the public.

Table 10 lists the current data license provisions for each of the datasets that are being employed in the project (data user agreement, data license agreement, and whether fees are currently associated with data use). Open access data will be managed through an open access license agreement that will be selected and implemented in the first months of Phase 2 of this project. Different open access license models may be applied to these datasets, depending upon the model that is currently being used and the data types within the dataset. At this point, the team has identified all current agency user agreements that must be executed to access agency data. However, as noted above, the project goal is to eliminate the need for all of these license agreements and to make the agency data available through the open access STM. License agreements for private traffic volume and speed data feeds that will be mapped to the STM whole road network will be evaluated at the start of Phase 2 as part of provider selection. As noted above, the goal will be to integrate only those speed and volume data sources that the data provider agrees can be made available through the open access STM.

The columns in **Table 10** include:

- **Dataset Owner:** owner of the dataset
- **Dataset ID(s) and Dataset Title(s):** group of datasets that are subject to the same license
- **Access Level:** Only the high level Open and Private are used in this table.
- **License Used:** the license applied to the dataset(s).
- **Reasons for a Non-Open License:** the reason why no license is applied or a non-open license is used for a dataset.

Table 10. Data Owner License Applied to Datasets

Data Owner	Dataset IDs and Titles	Access Level	License Used	Reason(s) for Non-Open License
ARC	2. Building Address and Landmark Data 3. Whole Road Network 4. ABM Network 10. Sidewalk Network 11. Indoor Pathways 12. GTFS Transit Network 46. VRU categories	Open	TBD	N/A
ARC	49. "Household Level Licensed Demographic Data" 56 Business Level Licensed Facility Data	Private	License agreements executed with individual data providers.	Data constitute intellectual property and commercial value
ATL	51. Mobile App Logs (raw) 53. Trip Feedback Reports (raw)	Private	See ATL RIDES DMP Version 1 and requires IRB Protocol Approval	Protection of PII
ATL	51. Mobile App Logs (removed PII) 53. Trip Feedback Reports (removed PII)	Private	Will be covered by usage agreement	Contains trace data, but may be used for research purposes
ATL	45. Trip options	Private	IRB Protocol Approval	Protection of PII. The turn-by-turn data reveal home and work locations and constitutes PII Even subsets of data that occur repeatedly, such as travel distance,

Data Owner	Dataset IDs and Titles	Access Level	License Used	Reason(s) for Non-Open License
				reveal home location
ATL / GA Tech	58. ST-CTN Performance Data	Private	IRB Protocol Approval	Protection of PII. The turn-by-turn data reveal home and work locations and constitutes PII Even subsets of data that occur repeatedly, such as travel distance, reveal home location
ATL / GA Tech	59. STM Communications Logs 60. STM Impedance Calculation Logs	Open	TBD	N/A
ATL / GCDOT	PED-X	Open	TBD	N/A
By Source (or commercial data provider)	9. Licensed Networks21. Waze Alerts 55. Geocode.earth API	Private	By source	License agreements executed with individual data providers
Facility owner	30. Building Pathway Asset Design and Condition Data 31. Building Wayfinding Asset Design and Condition Data	Unknown	TBD	N/A
GA DOR	50. Household Level Vehicle Registration Data	Private	GA DOR User Agreement and IRB Protocol Approval	Protection of PII
GA Tech	13. TransitSim Network 14. BikewaySim Network	Open	TBD	N/A

Data Owner	Dataset IDs and Titles	Access Level	License Used	Reason(s) for Non-Open License
	57. MOVES-Matrix Energy Consumption and Emission Rates			
GA Tech	48. Customer Demographic Data 52. Traverse Data	Private	IRB Protocol Approval	Protection of PII
GCT	33. Transit Vehicle Asset Design and Condition Data 34. GTFS (GCT) 36 GTFS Realtime (GCT) 38 GTFS-Flex 64. Ridership: Fixed Route	Open	TBD	N/A
GCT	GCT Complaint Log	Private	IRB Protocol Approval	Protection of PII
GDOT	15. NaviGator Data 22. GDOT TMC Incident Data 23. GDOT TMC Special Event Data 62. Pedestrian Crash Data 63. Pedestrian Incidents Police Reports	Private	GDOT User Agreement	User agreement ensures data integrity, tracks data use, and facilitates secure data access
GDOT	6. NaviGator Network 24. GDOT TMC Work-Zone Data	Open	TBD	N/A
GDOT/GC DOT	39. BSM 40. PSM 41. SPaT	Open	License and usage agreements will be developed as part of the Connected Corridor Project	N/A

Data Owner	Dataset IDs and Titles	Access Level	License Used	Reason(s) for Non-Open License
	42. MAP 43. SSM			
GDOT / GCDOT	61. Secure MU Gateway Event Logs	Private	IRB Protocol Approval	Protection of PII
Gwinnett County	28. Pedestrian Pathway Asset Design and Condition Data 29. Pedestrian Intersection Asset Design and Condition Data 32. Transit Stop Asset Design and Condition Data	Open	TBD	N/A
Gwinnett County	1. Parcel-level Land Use Data	Private	License agreements executed with individual data providers.	Access to licensed data may be granted by the data owner. Third-parties typically must execute a user agreement with the data owner to access data. Fees are likely required to obtain data access.
MARTA	35. GTFS (MARTA) 37. GTFS Real-time (MARTA)	Open	TBD	N/A
NOAA	47. Weather data	Open	TBD	N/A
OSM	8 OSM Network	Open	CC BY-SA 2.0	N/A
SRTA	16. SRTA Data 17. SRTA Tolling Data	Private	SRTA User Agreement	User agreement ensures data integrity, tracks data use, and

Data Owner	Dataset IDs and Titles	Access Level	License Used	Reason(s) for Non-Open License
				facilitates secure data access
SRTA	7. SRTA Managed Lane Network	Open	TBD	N/A
STM	20. Modeled Future Operating Conditions	Private	License agreements executed with owners of models or individual model scenario runs.	User agreement ensures data integrity, tracks data use, and facilitates secure data access.

4.4. Data Storage and Retention

The storage and retention provisions applied to the datasets acquired, collected, managed, derived and distributed by the project will be designed during Phase 2.

This section of the DMP identifies the servers and the systems that will be used to manage, store, and maintain data for the project. The sections that follow provide details for each dataset employed in the project about where the data will be hosted (system type), how data will be managed, how long data will be retained, what data storage and retention policies will be implemented, and how security provisions will be managed.

When these external datasets arrive, the system reviews and validates the quality of the data prior to its usage. Many of the datasets are used and discarded unless the source system retains the dataset; when relevant, the project team will archive the dataset after its use and retain the data through the project period. Although external data storage and retention provisions are out of scope of this project, the project team will review the retention provisions to ensure that key datasets that are needed for performance measurements are retained in project data storage systems.

4.4.1. Storage Systems

The storage systems that are used in the project among the three subsystems are identified in **Table 11**. The datasets are described by the following columns:

- Data Storage System Name:** the name of the data storage system the dataset will be stored in. Only one data system in the project system will be identified as the single source of primary data.
- Data Storage System Subsystem and Status:** denotes the subsystem steward organization of the data storage system and status of the dataset. Status values include: under development, implemented, operational. Note: operational implies that the data store may be implemented and operational in another instance of the system. In most

cases, the data store will be augmented to include new data elements generated from this project.

- **Dataset IDs and Title(s):** list of Dataset IDs and Titles contained in the data store.
- **Initial Storage Date:** The initial date that data will be available in each data storage system. The date is reference to the Phase (P) and its related Quarter (Q). Phase 2 has 8 quarters in the two-year period, and Phase 3 has 6 quarters in the 18 month period.
- **Frequency of Update:** Describes how frequently the data will be updated in the data storage system once ingestion begins (i.e., "Continually," "Daily," "Weekly," "Monthly," "Annually," "Unknown," "As needed," "Irregular," or "None planned").
- **Archiving and Preservation Period:** The duration for which the dataset will be maintained in each data storage system.

Table 11. Storage Systems

Data Storage System Name	Data Storage System Subsystem	Dataset ID and Title	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
ATL RIDES Connected Data Platform Module	ATL RIDES; under development (expected operational by early 2022)	45. Trip Options (raw) 51. Mobile App Log (raw) 53. Trip Feedback Reports (raw)	Phase 2, Q5 to Phase 2, Q6	Continuous	Through project period (see [ATL DMP])
External (not stored in project data stores)	Public/ commercial; operational	21. Waze Alerts 22. GDOT TMC Incident Data 23. GDOT TMC Special Event Data 24. GDOT TMC Work-Zone Data 55. Geocode.earth API	Phase 2, Q4	N/A	Compliant with Data Provider archiving policy
Open Data Server	STM; operational	4. ABM Network 10. Sidewalk Network 12. GTFS Transit Network 13. TransitSim Network 39. BSM 42. MAP	Phase 2, Q3 to Phase 2, Q6	Continuous	Through project period
Open Data Server	STM; under development Note: although most of the datasets originate in the GDOT/GCDOT ITS Data Hub, they are not archived.	2. Building Address and Landmark Data 3. Whole Road Network 8. OSM Network 14. BikewaySim Network (currently sits outside of the STM) 40. PSM 41. SPaT 43. SSM Exchange 44. Ped-X 46. VRU categories	Phase 2, Q4 to Phase 2, Q6	Continuous	Through project period

Data Storage System Name	Data Storage System Subsystem	Dataset ID and Title	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
PII Server Processed Data on Research Server	STM; operational	48. Customer Demographic Data 49. Household Level Licensed Demographic Data 50. Household Level Vehicle Registration Data 51. Mobile App Logs 52. Traverse Data 53. Trip Feedback Reports 56. Business Level Licensed Facility Data 58. ST-CTN Performance Measures Data 59. STM Communication Logs 60. STM Impedance Calculation Logs 61. Secure MU Gateway event logs	Phase 2, Q4 to Phase 2, Q6	Continuous	Through project period

Data Storage System Name	Data Storage System Subsystem	Dataset ID and Title	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
Research Server	STM (subset on STM Cluster); operational	1. Parcel-level Land Use Data 6. NaviGator Network 7. SRTA Managed Lane Network 9. Licensed Networks 11. Indoor Pathways 15. NaviGator Data 16. SRTA Data 17. SRTA Tolling Data 18. Subscription Roadway Operating Condition Data 19. Historic Roadway Operating Condition Data 20. Modeled Future Operating Conditions 26. Roadway Design and Condition Data 27. Roadway Intersection Design and Condition Data 28. Pedestrian Pathway Asset Design and Condition Data 29. Pedestrian Intersection Asset Design and Condition Data 30. Building Pathway Asset Design and Condition Data 31. Building Wayfinding Asset Design and Condition Data 32. Transit Stop Asset Design and Condition Data 33. Transit Vehicle Asset Design and Condition Data 47. Weather data	Phase 2, Q3 to Phase 2, Q6	Continuous	through project period and/or compliant with data provider archiving policy

Data Storage System Name	Data Storage System Subsystem	Dataset ID and Title	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
Research Server	STM (subset on STM Cluster); operational	51. Mobile App Logs (processed) 53. Trip Feedback Reports (processed) 54. Trip Crowdsourcing Reports (processed)	Phase 2, Q5 to Phase 2, Q6	Continuous	through project period and/or compliant with data provider archiving policy
STM Server Cluster	STM; operational	5. STM Network	Phase 2, Q4	Continuous	through project period
STM Server Cluster	STM; under development	25. Network Impedance API	Phase 2, Q5	Continuous	through project period
TRANSIT-data-tools	ATL; operational	34. GTFS (GCT) 35 GTFS (MARTA) 36 GTFS Realtime (GCT) 37 GTFS Real-time (MARTA) 38 GTFS-Flex	Phase 2, Q3 to Phase 2, Q6	Continuous	Compliant with SRTA archiving and retention policy; configurable for each dataset (see [ATL DMP])

4.4.2. Data Storage System Description

For this project, data storage and retention procedures are required to host and manage static network and demographic data (traditional file storage), user interface systems and the space-time memory (database and cloud database systems), and to handle PII and research activities (dedicated secure research servers). To implement these systems, the team will manage a mix of servers for this project:

- Private cloud servers will be employed in the provision of user interface systems
- A server farm will host the space-time memory systems
- Open source servers will handle open source data for operational systems, public access and to push data to the U.S. DOT-managed – Public System
- Third-party servers will feed data from outside sources into the project systems
- Research servers accessible only to the project team will be used in system development, development of APIs and key performance indicators (KPIs), and management of PII.

During Phase 2, the data storage systems will be designed, tested and deployed. The ATL RIDES and STM currently have operational data storage systems. These will be replicated and deployed in an ST-CTN instance of their systems and updated to comply with additional system requirements and refined in the design and implementation phases.

TRANSIT-data-tools is a ATL RIDES data store, with managed and controlled access by the IBI Group (data custodians). The data storage is configurable for each GTFS dataset. Realtime transit information is not archived or retained (it is available from GCT upon request).

ATL RIDES Connected Data Platform Module is a ATL RIDES data store, with managed and controlled access by the IBI Group (data custodians). The system is currently under development through the design process for the Connected Data Platform. Anticipated deployment will be by early 2022.

Open Data Server will be an open data portal where the public may access the data (dependent on a project user agreement). The design, structure, and user interfaces for the open data portal will be undertaken in Phase 2.

STM Hosted Data Stores: The real-time data and historic data archives hosted by the STM will be hosted on a Georgia Tech server farm dedicated to this project. These servers will manage the open source data for operational systems, provide public access to open data, and push project data to the U.S. DOT-managed – Public System.

- **STM Server Cluster:** A cluster of compute and storage servers that store and preprocess historic operation and mobility information and translates them into information. The STM ties in with the Partnership for an Advanced Computing Environment (PACE) cluster at Georgia Tech to leverage high-performance computing for processing the archival data while it uses the STM cluster resources for integrating real-time information with the historical conditions.

- **Research Server:** Multiple high-resource computers (e.g., 96 cores, 512 GB RAM) are used for the development, training and execution of the Machine Learning processes.
- **PII Server:** Separate resources are maintained for handling PII data at Georgia Tech. The PII servers reside on a separate secure network that restricts access to other networks even within the GA Tech Civil Engineering network. The access to the server requires physical presence in the secure access lab (badge access controlled) in the GA Tech Civil Engineering building.

4.4.3. Cybersecurity Policies

Similar to privacy and security policies, each organization hosting or managing a datastore will drive the cybersecurity policies for their datasets. A set of agreements on the overall security processes to be put in place will be developed as part of the EDG committees described earlier in this section.

The policies are described by the following columns:

- **Data Owner:** the owner and stakeholder responsible for the policy.
- **Data Owner Type:** the type of organization issuing policy. Values include: academic, public (public agency), commercial, other.
- **Policy Type:** the type of policy issued by the data owner. The values may include network, system and cybersecurity policies and procedures
- **Policy Reference:** a link or reference to the policy provisions

Table 12: Cybersecurity Policies and Procedures

Data Owner	Data Owner Type	Policy Type	Policy Reference
SRTA/ATL	Public	Network security policy	[ATL DMP]
SRTA/ATL	Public	System security policy	[ATL DMP]
GDOT	Public	Connected Corridor Security and IT policies (TBD)	TBD
GA Tech	Academic	Security procedures and standards	Cyber-security-policy https://policylibrary.gatech.edu/information-technology/cyber-security-policy

4.4.4. Data Security Policies and Procedures

The data security policies and procedures are governed by each dataset owner, and a systemwide policy and procedures will be developed as part of the EDG committee.

4.4.5. Back-up and Recovery Policies and Procedures

Similar to privacy and security policies, each organization hosting or managing a datastore will implement existing policies and procedures for back-up and recovery. A set of agreements on the overall retention and recovery processes to be put in place will be developed as part of the EDG committees described earlier in this section.

The current individual dataset owner policies are described by the following columns:

- **Data Owner:** the owner and stakeholder responsible for the policy.
- **Data Owner Type:** the type of organization issuing policy. Values include: academic, public (public agency), commercial, other.
- **Policy Reference:** a link or reference to the policy provisions

Table 13: Data Archiving and Retention Policies and Procedures

Data Owner	Data Owner Type	Policy Reference
SRTA/ ATL	Public	[ATL DMP]
GA Tech	Academic	Data Archiving and Retention Policy https://gatech.service-now.com/technology?id=kb_article_view&sysparm_article=KB0022866

5. Data Standards

5.1. Data Standards

Each dataset is specified and encoded using an open, standards-based or proprietary format. This section identifies the standards used to specify each dataset. **Table 14** lists the dataset, specified standard, and the rationale for using the standard.

Columns listed in **Table 13** include:

- **Dataset Identifier (ID):** unique ID that corresponds to **Table 5**.
- **Dataset Name:** dataset title that corresponds to **Table 5** dataset name.
- **Data Standard (Standard ID):** the name(s) of the data standard(s) in which the data are made available to the USDOT. The data standard is notated with a Standard ID that references entries in the **Comprehensive List of Standards (Appendix B)**. **Appendix B** contains the details of the standard, standard profile, (PICS, or ICD used to describe the data) and on-line reference to obtain the standard.
- **Data Standard Rationale:** explanation for use of the chosen data standard and preferred format.

Table 14. Dataset Standards and Rationale

ID	Dataset Title	Data Standard(s) Name	Data Standard Rationale
8	OSM Network	OSM	Required for use with the OTP
12	GTFS Transit Network	GTFS + other formats	Industry standard for exchanging transit schedule data
21	Waze Alerts	WAZE	Provided by WAZE
24	GDOT TMC Work-Zone Data	WZDX	Industry standard for work zone data exchange
25	Network Impedance API	OSM + TBD	Required for use with the OTP and becoming an industry standard for exchanging attribute and feature information on ROW and PROW
34	GTFS (GCT)	GTFS	Industry standard for exchanging transit schedule data

ID	Dataset Title	Data Standard(s) Name	Data Standard Rationale
35	GTFS (MARTA)	GTFS	Industry standard for exchanging transit schedule data
36	GTFS Realtime (GCT)	GTFS Realtime	Industry standard for exchanging transit event, real time and arrival prediction data
37	GTFS Real-time (MARTA)	GTFS Realtime	Industry standard for exchanging transit event, real time and arrival prediction data
38	GTFS-Flex	GTFS-Flex	Industry standard for exchanging flexible and demand-responsive transit schedule data
39	BSM	SAE J2735 + J2945 Part 1	Industry standard for CV message sets and produced by existing GDOT deployment
40	PSM	SAE J2735 + J2945 Parts 1, 2 and 9	Industry standard for CV message sets
41	SPaT	SAE J2735	Industry standard for CV message sets and produced by existing GDOT deployment
42	MAP	SAE J2735	Industry standard for CV message sets and produced by existing GDOT deployment
43	SSM	SAE J2735	Industry standard for CV message sets and produced by existing GDOT deployment
44	Ped-X	SAE J2735 + J2945 Parts 2 and 9	Industry standard for CV message sets

A comprehensive list of data standards and specifications are listed in **Appendix B**. The Data Standard names refer to the Standard Name in **Table B1**. Additional information contained in **Table B1** includes a link to the standard document, the open or proprietary nature of the standard (open, licensed, proprietary), and encoding formats (delimited file, JavaScript Object Notation (JSON), extensive markup language (XML), representational state transfer (REST), etc.).

5.2. Versioning

Datasets will be generated on a timely basis as previously specified in **Table 11**, frequency of update. Each file or database storage system will include version information embedded in the storage container. Depending on the dataset, the data will include published date, activation date, and deactivation or expiration date information. The versioning information will be included in metadata files that are associated with the dataset.

Derived datasets such as aggregated data or summary data, will include the duration for which the datasets span as well as references for the datasets. In addition, the publication date will be identified as the version for the dataset.

The system cannot version datasets that are imported from other sources, however, the system will log the date and time the dataset was ingested or when a transaction occurs.

5.3. Metadata

The USDOT requires metadata with each dataset submission and project information to support research and enable search and discovery. This section describes the definitions for entries into the Metadata files. These metadata types describe information about the datasets, their source(s), processes, quality, restrictions, and usage.

5.3.1. Metadata Types

Several datasets contain formats that describe their source, ownership, usage, processing, quality and lineage. For example, the GTFS specification includes two files (`feed_info.txt` and `attributions.txt`, see [GTFS]) that contain appropriate metadata information. The metadata types described in this section will be used for datasets derived from the ST-CTN project that do not have existing metadata or versioning formats.

1. **Discovery** – Metadata that is used to allow other users to find and work with the data. The types of metadata will include:
 - a. **Responsibilities:** Who publishes, owns, and produces values for the datasets.
 - b. **Permissions:** What user role is required to view the dataset to ensure privacy and security of the data is maintained.
 - c. **Access:** Where the dataset is stored and how is it accessed.
 - d. **Abstract:** What is included in the dataset and what is its intended purpose.
 - e. **Lineage:** How the dataset was created, derived or processed (summary level). This information points to the technical metadata file that describes the processes in more detail.
 - f. **Versioning:** When the dataset was created, produced or derived.
 - g. **Publication:** When the dataset was published (optional).
 - h. **Currency:** What is the currency of the dataset (when was the data active, from activation to deactivation).
 - i. **Licensing:** What are the restrictions on the data use including copying, publishing, distributing, transmitting, citing or adapting the data. This field may contain a link to the licensing terms and agreements document.

2. **Technical Metadata** – Data that is used to provide technical details on the data.
- a. **Data Schema** – Metadata that documents the exact fields, data elements or objects in the data including, field name, description, data type, and notes. This information will be at the asset level, describing the contents of the data or dataset generally, in accordance with the Project Open Data Metadata Schema another appropriate standard.
 - b. **Data Collection Method** – Metadata that documents the data collection method used to acquire the data. The information includes the quality and accuracy of the collection devices and methods.
 - c. **Data Processing** – Metadata that documents any data processing that was done to the data from the data inception (when the data was produced) to when it was processed for the USDOT.
 - d. **Data Impact Log** – Metadata that provides information on any changes to data during the collection period. Any time the data changes in a unique way that is not expected in the experimental design either by internal or external forces is documented here. Some examples of possible triggers for updates include:
 - i. Internal Events
 1. Data collection start and end
 2. Testing being done
 3. Changes in the deployment scope
 4. Software updates that modify data fields
 5. Sensor disruptions either communication based or hardware based
 - ii. External Events
 1. Volume increase or reductions
 2. Construction and/or work zones
 3. Special events in area
 4. Changes in roadway or sidewalk configurations

This impact log will include information on the date and time of the impact, type of trigger categories particularly those related to schema/data field or data collection method changes or gap/disruptions to collection.

- If a schema change, the previous and current data field and replacement impact will be described.
- If a data collection or processing change, the impact to the dataset quality or data collection update frequency will be described.
- If gap or disruption, the duration of the disruption will be described.

5.3.2. Metadata Structure

The Metadata Structure will be delivered in a flat file format using JSON or comma-separated values (CSV) encoding methods.

Note: *the specific format and tags will be published in a subsequent version of this DMP during the Phase 2.*

A compressed set of files that are composed of the following embedded folders will be provided:

ST-CTN project file – contains a summary of the project information including:

- Discovery information (as defined in **Section 5.3.1**, limited to project file publication)

U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology
Intelligent Transportation System Joint Program Office

- Folders containing metadata and the related metadata formats used (this will identify if the data uses a standard format rather than the DMP Metadata Types)
- Link to published DMP

Dataset metadata file (folders) – Discovery and technical metadata information for each dataset. The core datasets will be included in this set of files because they may be collected or derived using different collection and processing methods. The core dataset files may include sidewalk and pathway information, signal system and RSU asset data and stop information (from GTFS). The associated files per dataset will include:

- Discovery information (as defined in **Section 5.3.1** or using native specification format)
- Schema or Standard Profile (if not an open specification or standard) – includes the version used to generate the referenced dataset
- Data collection and processing information
- Impact Log

5.3.3. Metadata Update Process

A new metadata file will be generated for each dataset that is published. The update process will follow the data curation process. As depicted in **Figure 5**, as each curation process occurs, the methods will be verified and changes documented to relevant metadata fields, as well as logged in the Impact Log. This update process includes:

- Updating collection methods during and following the data collection stage.
- Updating quality and processing methods following the data quality checking and processing stage.
- Updating data storage including formats, organization and storage methods. This also includes updating information on permission levels (note: the permissions are generated at this level, because the database management system will enforce a major portion of the data security and privacy provisions).
- Updating distribution and publication methods (e.g., APIs or flat file formats) for datasets.

Metadata Update Process

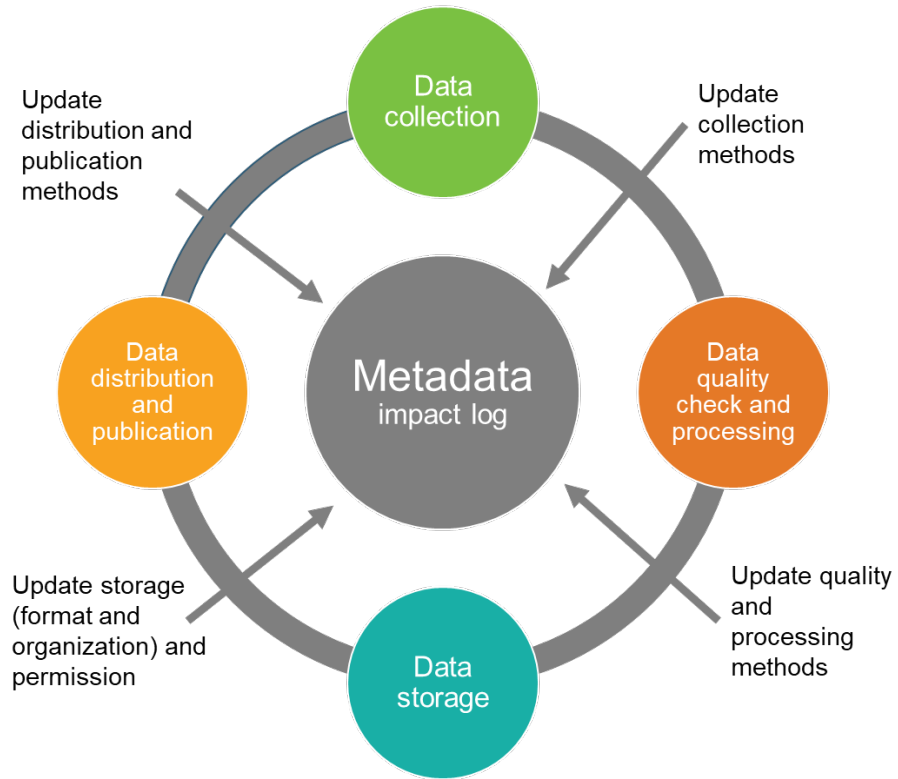


Figure 5. Metadata Update Process (source: ARC)

Appendix A. Acronyms and Glossary

Acronyms

ABM - Activity Based Model

APC – automated passenger count

API – application programming interface

ARC – Atlanta Regional Commission

ATIS – automatic terminal information service

ATL – Atlanta-Region Transit Link Authority

ATL RIDES – Atlanta-Region Rider Information and Data Evaluation System

AVL – automatic vehicle location

BSM – basic safety message

CBI – commercial business interest

ConOps – Concept of Operations

CSV – comma-separated value

CV – connected vehicle

DMP – Data Management Plan

DOR – Department of Revenue

EDG – Enterprise Data Governance

EDS – Enterprise Data Steward

EX ID – Data Exchange ID

FHWA – Federal Highway Administration

FTA – Federal Transit Administration

GA Tech – Georgia Institute of Technology

GCT – Gwinnett County Transit

GDOT – Georgia Department of Transportation

GTFS – General Transit Feed Specification

GTRI – Georgia Tech Research Institute

ICD – interface control document

IE – independent evaluator

IOO – infrastructure owner and operator

IRB – Institutional Review Board

ITS – Intelligent Transportation Systems

JPO – Joint Program Office

JSON – JavaScript Object Notation

KPI -- key performance indicators

LEP – limited English proficiency

MAP – MapData

MARTA – Metropolitan Atlanta Rapid Transit Authority

MOU – memoranda of understanding

MOVES – motor vehicle emissions simulator

MU – mobile unit

MVP – Minimum Viable Project

NDA – non-disclosure agreement

OBU – onboard unit

OSM – OpenStreetMap

OST – Office of the Secretary

OTP – Open Trip Planner

PACE – Partnership for an Advanced Computing Environment

PED-SIG – Mobile Accessible Pedestrian Signal System

PICS – protocol interface conformance specification

PII – personally identifiable information

PROW – public right of way

PSM – pedestrian safety message

REST – representational state transfer

RSU – roadside unit

SPaT – signal phasing and timing

SRM – signal request message

SRTA - State Road and Tollway Authority

SSM – signal status message

ST-CTN – Safe Trips in a Connected Transportation Network

STM – space time memory

TMC – traffic management center

TPI – transit pedestrian indication

TSP – transit signal priority

TSR – transit stop request

USDOT – U.S. Department of Transportation

VRU – vulnerable road user

XML – extensive markup language

Glossary

Table A1. Glossary

Category	Term	Definition	Table Reference
Dataset Type	Assets	Asset information about facility and ATMS devices-- sensors, signals, comm including electronic signs, PED-X signals, etc.	Table 6
Dataset Type	Crowdsourcing	Data generated by the "crowd"	Table 6
Dataset Type	CV	Connected vehicle produced dataset	Table 6
Dataset Type	Demographics	Data about the structure of populations	Table 6
Dataset Type	Land Use	Human use of land	Table 6
Dataset Type	Mobility Service API	Application programming interfaces (API) or services that are pushed or pulled by an application.	Table 6
Dataset Type	Network	Infrastructure characteristics and topological connectivity including right of way, PROW (sidewalk, crosswalk, bike lanes/paths)	Table 6
Dataset Type	Network operating conditions	Condition and status of network infrastructure including planned and unplanned events: incidents, special events, work zones, and other impacts.	Table 6
Dataset Type	System-Customer Performance	Data sets that require archiving and use by performance measures	Table 6
Dataset Type	Transit	Transit network (routes) and real time conditions/event data	Table 6
Dataset Type	VRU Modes	Types of VRUs; used to describe categories of default impedance values	Table 6
Dataset Type	Weather	Weather data	Table 6

Category	Term	Definition	Table Reference
Access Level	PII Certification	Data that has PII included in the data set. The access to this data should be as restrictive as possible to protect the PII based on IRB-approved processes. Data in this category should have an operational purpose that justifies its storage.	Table 8
Access Level	Private	Data that cannot be shared with external users. Access to these data is limited and only granted with IRB and Project Team approvals. Private data include four subcategories: operational, proprietary, research and PII Certification.	Table 8
Access Level	Proprietary	Licensed data from third parties or CBIs. This data may be used for planning or operational purposes. Any access to the data is determined by usage agreements between the parties.	Table 8
Access Level	Operational	Real-time and other data used in the applications and operations of the system. The data may contain licensed data restricted by usage agreements.	Table 8
Access Level	Research	Data that is available for research, but users of the data must meet IRB requirements before gaining access to the data. These datasets may have PII. These datasets are compiled for machine learning and research purposes that do not contain PII.	Table 8
Access Level	Open	Data that can be used by the public with no or limited licensing restrictions. This data is available to the public without needing to request permissions and will be provided to the USDOT-managed Public System. These will be anonymized or aggregated version of private datasets to protect PII.	Table 8
Collection Method	Derived	Data is derived from one or more sources for summary or fusion purposes	Table 6
Collection Method	External Input	Data is ingested from a third party source (the ingestion process may be through a digital interface or through manual processes.)	Table 6

Category	Term	Definition	Table Reference
Collection Method	Collect / forward	Data created, collected, forwarded and stored. These include user input transactions (e.g., between APIs), web forms, user tracking methods (e.g., trace data from mobile phones)	Table 6

Appendix B. Comprehensive List of Standards Used

The Comprehensive list of standards used provides the set of profiles, message sets, data dictionaries that are used by the datasets to organize and encode the datasets, and define the data meaning, formats and enumerated values of the data.

- **Standard Identifier:** unique identifier assigned to the identifier in this document.
- **Standard Profile:** the name of the profile that customizes provisions of one or more message or data standards.
- **Standard Name(s):** the name of the data standard in which the data are made available to the USDOT.
- **Data Standard Digital Object Identifier(s) (DOI):** the DOI of the standard for the data. A URL to the data standard(s). "N/a" indicates that a reference is not available on-line.
- **Open or Proprietary:** Indicates whether the data standard is "Open" or "Proprietary." Open includes specifications published using the creative commons license or licensed by a consensus based standards organization.
- **File Format Used:** the encoding and file format approach used. Values include CSV, JSON, XML, or other.

Table B1. Comprehensive List of Standards

Std ID	Standard Profile	Std Name (s)	DOI / URL	Open / Proprietary	File Format Used
STD-1	GTFS	GTFS	https://gtfs.org/reference/static	open	CSV
STD-2	GTFS realtime	GTFS realtime	https://gtfs.org/reference/realtime/v2/	open	GTFSrealtime.proto
STD-3	GTFS Pathways	GTFS pathways	bit.ly/gtfs-pathways	open	CSV
STD-4	GTFS-Flex	GTFS flex + GTFS	https://github.com/MobilityData/gtfs-flex/blob/master/spec/reference.md	open	CSV

Std ID	Standard Profile	Std Name (s)	DOI / URL	Open / Proprietary	File Format Used
STD-5	OSM	OSM	https://www.openstreetmap.org/about	open	Geographic format (can be exported to multiple standard formats including Shape Files)
STD-6	BSM	SAE J2735 and J2945/1	J2735SET_202007 -- https://www.sae.org/standards/content/j2735set_202007/ J2945_201712 -- https://www.sae.org/standards/content/j2945_201712/	open	XML
STD-7	PSM	SAE J2735 and J2945 parts 1 and 9	J2735SET_202007 -- https://www.sae.org/standards/content/j2735set_202007/ J2945_201712 -- https://www.sae.org/standards/content/j2945_201712/	open	XML
STD-8	SPaT	SAE J2735	J2735SET_202007 -- https://www.sae.org/standards/content/j2735set_202007/	open	XML
STD-9	MAP	SAE J2735	J2735SET_202007 -- https://www.sae.org/standards/content/j2735set_202007/	open	XML
STD-10	SSM	SAE J2735	J2735SET_202007 -- https://www.sae.org/standards/content/j2735set_202007/	open	XML

Appendix C. Dataset Impact Log

This section includes a log of datasets that are specified in any version of the DMP, their identification and name, current state (initial, updated, removed, replaced), the version in which they were last specified, and information on changes to the dataset or its characteristics, when and what was changed (e.g., table and item in table).

For example, if the GCT GTFS dataset was changed because a new version of the GTFS standard was used, then the table would include two entries for the GCT GTFS, the first with a state of “initial” and the Phase 1 version (1.0), and the second with the same ID and name, but with a state of “updated”, DMP Version of 1.0, DMP Update of 2.0 (for Phase 2), and Reason would include **Table 13** row # and a statement that states “dataset now conforms to current standard version”.

Table C1. Example of Dataset Impact Log

Dataset ID	Dataset Name	Status	DMP Version (last version)	DMP Update (version / date)	Reason

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-21-865



U.S. Department of Transportation