Phase 1 Data Management Plan (DMP)

Buffalo NY ITS4US Deployment Project

www.its.dot.gov/index.htm

Final Report — September 13, 2021
FHWA-JPO-21-868

COMPLETE TRIP
ITS4US

U.S. Department of Transportation

Produced by ICF
U.S. Department of Transportation
Intelligent Transportation Systems Joint Program Office
Federal Highway Administration
Office of the Assistant Secretary for Research and Technology
Federal Transit Administration

Notice

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-JPO-21-868		

4. Title and Subtitle	5. Report Date
Phase 1 Data Management Plan (DMP) – Buffalo NY ITS4US Deployment Project	September 13, 2021
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Deepak Gopalakrishna (ICF), Nayel Urena Serulle (ICF), Cindy Peck (ICF), Adel Sadek (UB), Robert Jones (NFTA), Jordana Maisel (UB), Victor Paquet (UB), Edward Stanfield (UB), Polly Okunieff (ICF), Chunming Qiao (UB), Stephen Still (UB), Rahul Dagli (ICF), Katie O'Sullivan (ICF)	Task 3 - DMP

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
ICF International, 9300 Lee Highway, Fairfax, VA 22031 University at Buffalo, Amherst, NY 14228 Open Doors Organization, 8600 W. Catalpa Avenue, Chicago, IL 60656 RSG, 55 Railroad Row, Suite 101, White River Junction, VT 05001 ETCH, 4696 Smothers Road, Westerville, OH 43081 BNMC, 640 Ellicott Street, Buffalo, NY 14203 NFTA, 181 Ellicott Street, Buffalo, NY 14203	11. Contract or Grant No. 693JJ321C000005

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
United States Department of Transportation 1200 New Jersey Ave., SE Washington, DC 20590	Data Management Plan
	14. Sponsoring Agency Code

15. Supplementary Notes
Elina Zlotchenko (USDOT ITS-JPO) is the Contracts Officer Representative (COR), and Amalia Rodezno (USDOT) is the Contracting Officer (CO).

16. Abstract
The Buffalo NY ITS4US Deployment Project seeks to improve mobility to, from, and within the Buffalo Niagara Medical Campus by deploying new and advanced technologies with a focus on addressing existing mobility and accessibility challenges. Examples of the technologies to be deployed are electric and self-driving shuttles, a trip planning app that is customized for accessible travel, intersections that use tactile and mobile technologies to enable travelers with disabilities to navigate intersections, and Smart Infrastructure to support outdoor and indoor wayfinding. The deployment geography includes the 120-acre Medical Campus and surrounding neighborhoods with a focus on three nearby neighborhoods (Fruit Belt and Masten Park) with underserved populations (low income, vision impaired, deaf or hard of hearing, wheeled mobility device users and older adults).

This document is the Data Management Plan, which details the underlying data-related needs and processes of the Buffalo, NY ITS4US deployment concept—such as data collection, analysis, protection of users' privacy, storage and sharing.

17. Keywords	18. Distribution Statement
ITS4US; Complete Trip; Deployment; ITS; Intelligent Transportation Systems; Complete Trip; Self-Driving Shuttle; Open Trip Planner; Systems Integration; Data Management	This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
None	None	56	N/A

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

Revision History

Name	Date	Version	Summary of Changes
ICF	July 26, 2021	0.1	Initial Draft
ICF	Aug 23, 2021	0.2	Updated based on comments received from USDOT
ICF	Sept 13, 2021	0.3	Revised version updated based on comments from USDOT

Table of Contents

List of Tables

List of Figures

1 Introduction

Buffalo, New York (NY) is one of five sites selected for U.S Department of Transportation (USDOT) Complete Trip - Intelligent Transportation Systems for Underserved Communities (ITS4US) Deployment Program, which seeks to integrate innovative technologies to improve mobility and accessibility. The Buffalo, NY project plans to deploy an integrated set of travel support services and systems within neighborhoods surrounding Buffalo Niagara Medical Campus (BNMC).

This document, the Phase 1 Data Management Plan (DMP), describes the data expected to be collected as part of this project and how the data will be managed (i.e., processed, stored, and shared).

1.1 Project Background

Buffalo is striving toward a sustainable future at all levels of society, incorporating actions in the community, government, and private entities in the area. Enabling community mobility and access to jobs, healthcare, and services to traditionally underserved populations is the primary motivation for all the regional partners involved in this deployment.

The Complete Trip - ITS4US Deployment Program is an effort led by the Intelligent Transportation System (ITS) Joint Program Office (JPO) and supported by Office of the Secretary, 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 disabilities, older adults, low-income individuals, rural residents, veterans, and limited English proficiency travelers. This program will enable communities to build local partnerships, develop and deploy integrated and replicable mobility solutions to achieve complete trips for all travelers.

As one of the selected sites, the Buffalo, NY ITS4US deployment concept addresses:

1. **Providing transit access to healthcare and jobs** to underserved residents including persons with disabilities and allowing them to share in the economic development in downtown Buffalo.
2. **Leveraging technology to work in support for accessible transportation**, integrating accessible transportation technology, transit, and connected automation to solve a transportation need.
3. **Developing a scalable model** for considering accessibility and universal design in transportation technology projects.

The Buffalo, NY ITS4US project will be completed in three phases: Phase 1- Concept Development, Phase 2- Design and Test, and Phase 3- Operation and Evaluation.

1.2 Document Organization

This document is organized following the guidance provided by FHWA, which are based on the USDOT's National Transportation Library's guidance for Creating Data Management Plans for Extramural Research, the ITS4US Broad Agency Agreement (BAA), and provide additional clarification for the ITS JPO's expectations for DMPs.

The remainder of the plan is as follows:

- Section 2 provides a brief overview of the ITS4US project.
- Section 3 summarizes the data to be used as part of this project.
- Section 4 describes the data stewardship and management approach.
- Section 5 details the data standards, including the metadata.
- Appendix A lists the acronyms used in this document.

1.3 System Overview

The Greater Buffalo-Niagara Regional Transportation Council (GBNRTC) stablished its vision of the region for 2050 in its "Moving Forward 2050 – A Regional Transportation Plan for Buffalo Niagara" (GBNRTC; University at Buffalo Regional Institute, The SUNY at Buffalo School of Architecture and Planning; Cambridge Systematics; TyLin International, 2018). The plan seeks to guide transportation investments to:

1. Raise the region's standard of living
2. Support efficient freight movement
3. Maximize infrastructure resiliency
4. Support focused growth in communities (urban, suburban, and rural)
5. Ensure access to opportunities and services
6. Support healthy and safe communities through targeted transportation investment
7. Strengthen the fiscal health of local governments
8. Preserve and protect a healthy environment and accessible open spaces and waterways
9. Create a fully integrated and seamless transportation environment

The Buffalo ITS4US project goals directly align with GBNRTC's goals 1, 4, 5, 6, and 9 by providing innovative tools and services to better enable travelers to make complete trips in and around the BNMC. Furthermore, the proposed system focuses on providing transit access to

healthcare and jobs to underserved citizens and allow them to share in the economic development in downtown Buffalo.

To achieve these goals, the proposed system of interest is made of four major subsystems and a variety of data interfaces between them. The four major subsystems include:

- **Complete Trips Platform** – The complete trip platform is the integrated trip planning function for travelers. It includes various modules that allow users to personalize their trip planning, execution, and navigation experience. Specific modules in this subsystem include:
    - User Profiles
    - Trip Booking
    - Trip Planning
    - Trip Monitoring and Notifications
    - Geolocation and Mapping
    - Navigation
    - Real-time situation monitor
    - Performance metrics
    - Trip history/ledger
    - User Interface (UI): Mobile application
    - UI: Web and Interactive voice response **(**IVR)
- **Community Shuttle Subsystem** – The Community Shuttle subsystem provides demand-responsive transit services within a specified zone of operations, using a mix of vehicles, including both human-driven and self-driving shuttles (SDS). The SDS will operate on a predefined route(s), consisting of a set of streets within the zone and pick-up and drop-off locations, but will be responsive to travelers' demand (e.g., it can skip certain pick-up/drop-off locations if there is no demand). The human-driven vehicles will provide door-to-door on demand service within the zone of operation. Modules within this subsystem include both types of vehicles, as well as a Shuttle Operations Center (SOC).

- **Smart Infrastructure Subsystem** – The smart infrastructure subsystem includes wayfinding and orientation for indoor and outdoor, provision of navigation and destination finding through information kiosks (Transportation Information Hub, TIH), augmented communications technologies (Smart Signs), and intersection treatment for hands-free, pedestrian signal requests.

- **Performance Dashboard Subsystem** – This subsystem measures and presents the performance of the system to the agency operating the system. The dashboard will be used to monitor the current condition of the system and assess historical information.

Figure 1 provides the context diagram for the system. The interfaces are numbered and described later in the document. The following subsections provide more detail on each component's functions and capabilities.

The reader is referred to the Phase 1 Concept of Operations (FHWA-JPO-21-860) and Performance Measurement and Evaluation Support Plan (FHWA-JPO-21-878) for more details on the system's components and performance assessment approach, respectively.

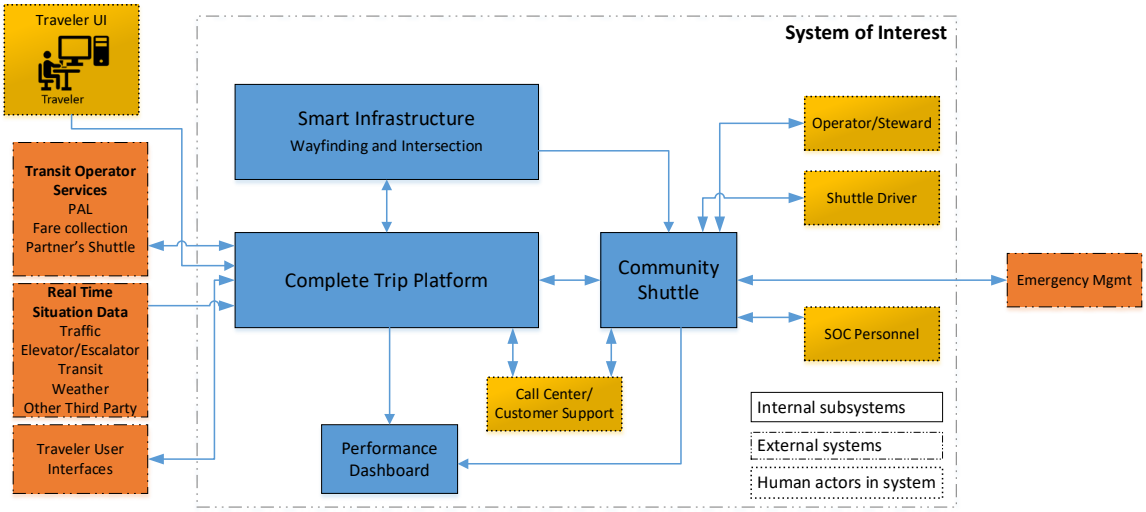

Figure 1. Buffalo, NY ITS4US Deployment Context Diagram with Interfaces.
Source: Buffalo, NY ITS4US

1.3.1 Complete Trip Platform

The Complete Trip Platform (CTP) provides trip planning and travel functions for travelers. The tool is available for registered and non-registered account users. Account holders will be able to interact with other mobility partners for which they have accounts (e.g., Niagara Frontier Transportation Authority (NFTA) paratransit and community shuttle services), personalize their trip preferences and customize hands-free turn-by-turn notifications, and access to wayfinding assets using components specified in the smart infrastructure subsystem. Non-registered travelers will be able to use the trip plan and travel tools to view accessible paths, transit services and alerts about asset status (e.g., elevator / escalator operations). The functions are described in the following sections.

1.3.2 Community Shuttle Subsystem

The Community Shuttle (CS) subsystem will provide demand-responsive transit services within the Fruit Belt neighborhood. The CS consists of the following three components:

1. The SDS component, which will be a demand-responsive shuttle constrained to operate over a pre-defined route (i.e., a set of streets that satisfy the SDS Operations Design Domain) and pre-designated pick-up/drop-off locations.

2. The human-driven shuttle (HDS), which will provide door-to-door on-demand service.

3. The Shuttle Operations Center (SOC), which will receive all calls for services and will track the status of each vehicle in the CS fleet.

The envisioned service area for the proposed community shuttle fleet is shown in Figure 2.

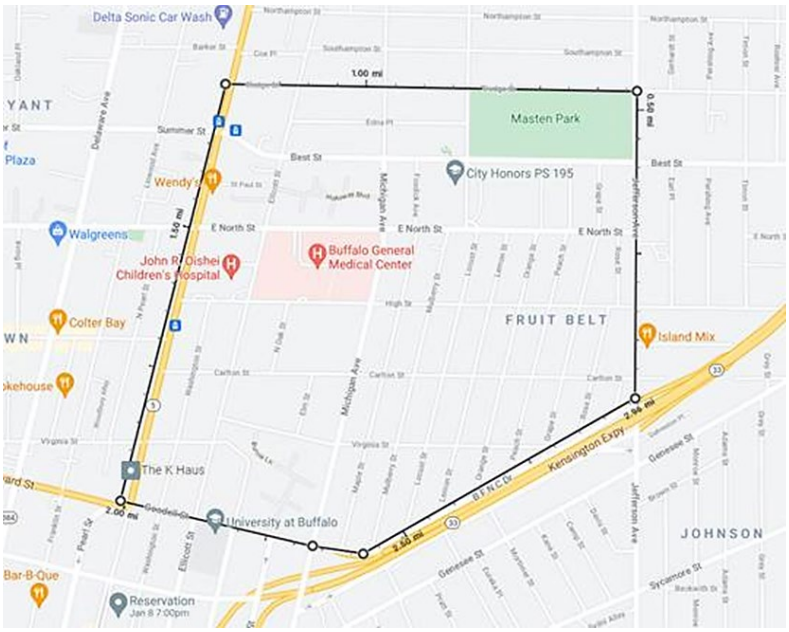

Figure 2. Proposed Service Area for the Community Shuttle.
Source: Buffalo, NY ITS4US & Google Maps.

1.3.3 Smart Infrastructure Subsystem

This subsystem includes “smart” wayfinding components used for indoor and outdoor navigation at public right of ways, parking lots, building/facilities, bus shelters and stations, as well as requests for pedestrian crossing at selected signalized intersections—Main St. & Best St. and Ellicott St. & High St. Additionally, transitioning between indoor and outdoor environments will be seamless. Trip planning and navigation functions will be provisioned through the CTP, with the smart infrastructure assets and communications technologies offering travelers with similar turn-by-turn and hands-free functions. The CTP will also drive the interaction between the travelers’ mobile app and the pedestrian signal request.

Many of these smart components and traveler information products (e.g., trip plans, notifications) are supported by planned physical infrastructure improvements and accessible physical features, to be implemented during the next few years to complement or support the proposed system. These improvements will be combined with the digital components to create a “Smart Infrastructure” subsystem.

1.3.4 Performance Measure Dashboard Subsystem

The Performance Measure Dashboard subsystem monitors, integrates, analyzes, and displays performance measures from other subsystems and external sources. The subsystem will include functions to ingest log files from the subsystems and external data sources, storage, analytic and visualization tools to display and access current and historic data sets produced from the integrated system.

The details of the performance measures are described in the Performance Measurement and Evaluation Support Plan (PMESP).

2 Document Overview

2.1 Change Control

The DMP uses information from internal documents developed as part of this project, namely the Concept of Operations (ConOps) and the PMESP (currently in development, first draft due in late August 2021 and final version due in late October 2021). This document is the **revised** version of DMP (dated as September 13), with edits based on the feedback received by USDOT.

The DMP also feeds other key internal documents, as shown in Figure 3. As such, the DMP will be updated periodically and as needed throughout the project to correct any inconsistencies that may arise as other project documents become available and as the concept continues to mature.

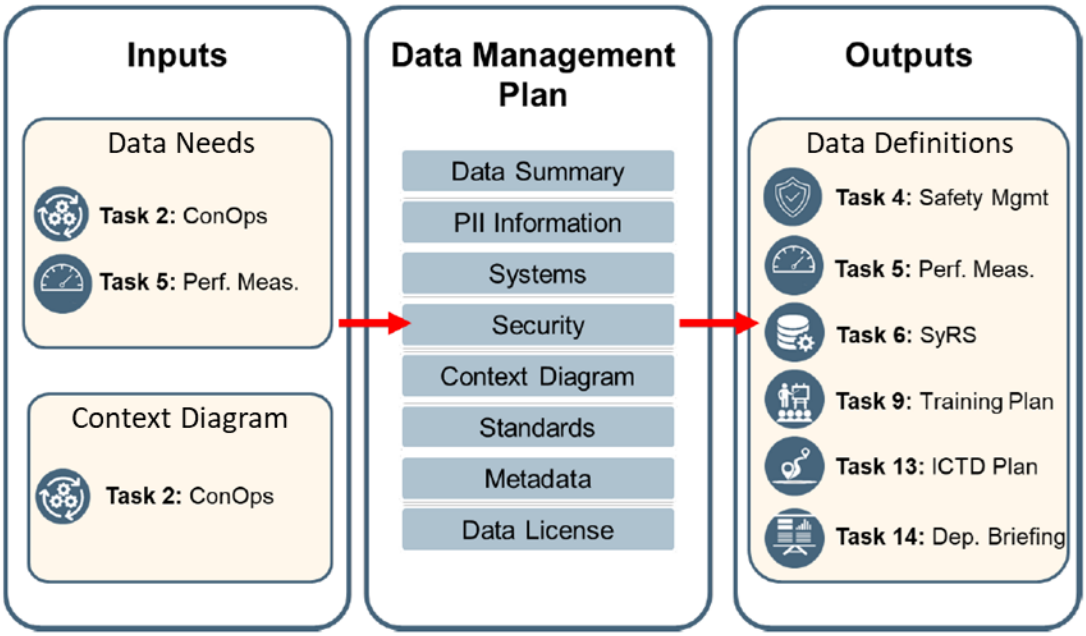

Figure 3. Inputs and outputs of the Data Management Plan.
Source: FHWA

The DMP will be stored and version controlled through the project's Teams Site. It is expected that multiple versions will be published by USDOT throughout the project lifecycle. All edits to the current operating version of the DMP will need to be approved by FHWA before being officially integrated into the "new" version. All previous versions of the DMP will be archived within the Teams Site and will be available to USDOT and the Independent Evaluator (IE). Changes to the data will follow a similar process—i.e., logged by the Buffalo, NY team on a Data Impact Log and presented to USDOT for approval.

2.2 Relevant Sources

The following are documents and sources used to develop this document.

- Phase 1 Concept of Operations (ConOps) – Buffalo NY ITS4US Deployment Project.
- Phase 1 Performance Measurement and Evaluation Support Plan – Buffalo NY ITS4US Deployment Project (draft version completed on Aug 23, 2021).
- Standards listed in Section 5.

2.3 Data Schedule

Limited data collection, restricted to qualitative interviews, will take place during Phase 1 of this project. Collection of the baseline and other performance- and system-related data will be explained in the PMESP (currently in development). While there is limited information at this stage to determine the timing of data milestones, potential milestones are listed in Table 1.

Table 1. Expected Data Milestones.

ID	Event Title	Description	Date
1.	Phase 1 - Draft DMP	Initial Draft DMP with basic information known at the time of writing.	July 28, 2021
2.	Phase 1 - Final DMP	DMP is updated with USDOT comments addressed	Aug 23, 2021
3.	DMP Update #1	DMP is updated at the end of Phase 1 to account for any edits based on other key deliverables (see Figure 6).	End of Phase 1
4.	Institutional Review Board (IRB) Approval	DMP will be updated once the project obtains the IRB approval.	Phase 2
5.	Initial Data Samples	Initial Data samples (based on the provided metadata) are created, validated and submitted to USDOT for review.	Phase 2
6.	Data management meeting with USDOT data team	Meeting to review data with USDOT and walkthrough the data schema and DMP. The IE may participate in this meeting.	Phase 2
7.	Baseline data collection starts	Initial collection of data on current conditions starts. May leverage existing surveys from local agencies and partners.	Phase 2

ID	Event Title	Description	Date
8.	Baseline data provided to USDOT	Complete Baseline data sets are uploaded to USDOT and the IE.	Phase 2
9.	Testing of applications begins	Initial upload after datasets are collected through testing. Interim dataset collections will be provided following major Agile development sprints and epics.	Phase 2
10.	Data transferred to USDOT	Monthly updates of data are provided to USDOT and IE. This will be done following completion of Minimum Viable Product (MVP) for each subsystem.	Phase 2.
11.	DMP Update #2	DMP updated with any changes coming from the Agile process, testing and sample data schema.	Phase 2
12.	Data Review	Data Review conducted with USDOT and IE to ensure datasets are complete. Note, the Agile development process will commence early in Phase 2 and continue mid-way through Phase 3 when the complete system will be deployed and operational.	Phase 2/3
13.	Draft Final Analysis Report submitted	Draft Final Analysis Report submitted to USDOT. Content to be determined (TBD) based on USDOT guidance for Phase 2/3.	Phase 3
14.	Final Analysis Report submitted	Draft Final Analysis Report submitted to USDOT.	Phase 3
15.	DMP Update #3 (If needed)	If needed, a third update of the DMP updated with any changes coming from the Agile process, testing and sample data schema.	Phase 2

3 Data Overview

3.1 Data Needs Summary

The data needs are derived from the subsystem functions and interfaces as described in the Section 1.3 System Overview. A detailed system concept diagram (see Figure 4) shows the major subsystem components, interfaces, personnel and data stores. The data stores support the storage, security and access of datasets that are used for planning, operations, and evaluation of the system. The datasets from each subsystem will be stored and archived in the Performance Measurement Dashboard subsystem, and the output will feed the USDOT Public Data Store.

Data flows between subsystems provide insight into input and output datasets. The data flows identified in Figure 4 represent the flows between subsystems, ingested and disseminated to external data sources. The data flows are described in Table 2, Table 3, and Table 4 below. Note that flows 7 and 11 are no longer part of this project and are therefore omitted from the concept diagram and the tables.

The first column of each table represents the identifier listed in the Context Diagram (Figure 4). Related tables describe the internal and external (inputs and outputs) data flows. They defined as follows:

- Table 2 “subsystem” columns identify a subsystem in the system of interest (SOI).
- Table 3 describes the inputs to the SOI, so the column labeled “input flows” lists the external sources providing the data.
- Table 4 describes the data output by the SOI, so the column labeled “output flow” lists the destination of the dataset, while “source” lists the SOI subsystem that disseminates the dataset.

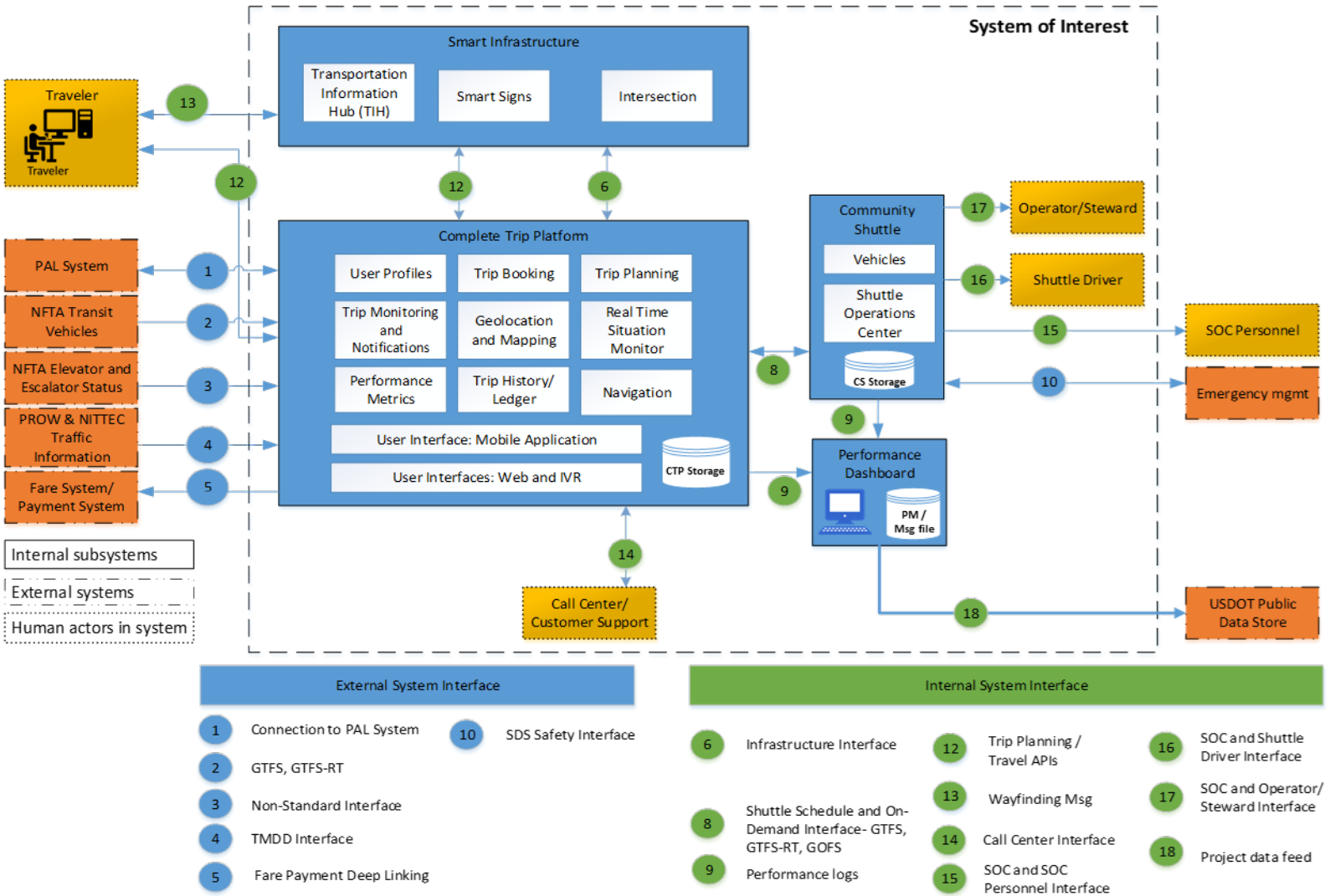

Figure 4: Buffalo, NY ITS4US Full System Deployment Context Diagram with Interfaces and Data Stores
Source: ICF Buffalo ITS4US

Table 2. Internal Data Flows related to Buffalo ITS4US Context Diagram System of Interest (Figure 4).

Flow #	Subsystem	Subsystem	Description
6	Complete Trip Platform	Intersection (Smart Infrastructure)	**Bidirectional** communication of the request-response interaction between the CTP mobile app making a request to the Intersection for pedestrian signal crossing and the intersection responding to the request.
8	SOC (Community Shuttle)	Complete Trip Platform	**Bidirectional** communication that provides static and dynamic information on shuttle schedules and operations, and trip booking transactions. These include: • Fixed Route Shuttles: GTFS, GTFS-RT • Micro-transit Shuttles: GTFS-Flex (optionally GOFS)
9	Complete Trip Platform / SOC (Community Shuttle)	Performance Dashboard	Performance data shared **from** the CTP and SOC subsystems **to** the dashboard. Types of data include trip plans and executions, user preferences and other log files that provide insight into traveler usage and utilization of services. r
12	Complete Trip Platform	Traveler UI / Wayfinding (Smart Infrastructure subsystem)	**Bidirectional** interface that includes: Static and dynamic information and interactive forms that support the provision of trip planning, wayfinding and navigation using the TIH and Smart Signs using CTP platform functions. Crowdsourced information provided by users of the CTP access channels (web, mobile and call center) on trip conditions and obstacles to travel along indoor and outdoor pathways. The data flow provides transaction exchanges between the CTP and end users (travelers and caregivers). The UIs including exchanges using: • TIH (at bus stops, shelters, parking, and other facility locations) • Call Center and IVR • Mobile app • Web browser The TIH, call center and web browser use a subset of the mobile app APIs. The APIs will include the following types of transactions: • Trip Planning • Selection of preferences and notifications • Selection of navigation processes and UI preferences

Flow #	Subsystem	Subsystem	Description
			• Booking (PAL and Shuttle) processes • Actualization of notifications and wayfinding (indoor and outdoor) messaging Requests for services (e.g., pedestrian signal request, shuttle and transit vehicles, facility wayfinding connections)
13	Wayfinding (Smart Infrastructure)	Traveler / Traveler UI	**Bidirectional** interface that provides the orientation and wayfinding signals and messages provisioned by the Smart Sign components in the Wayfinding sub-subsystem. It is important to note that while there is a hand-shaking protocol, at this time it is yet unclear what the solution will look like. At this stage, we will keep the bidirectional interface and plan to update/refine the diagram if needed once a solution is designed.
14	Call Center	CTP	Interface to connect with a customer support person to support phone access to CTP services, and by proxy, the community shuttle (i.e., interface with the shuttle takes place through the CTP). The call center may be part of existing entities or new.
15	SOC (Community Shuttle)	Shuttle Operations Personnel	Practices and policies around the interface of shuttle operations personnel and the SOC.
16	SOC (Community Shuttle)	Shuttle Driver	Communications, practices, and policies between the SOC and the driver of the human-driven shuttle.
17	SOC (Community Shuttle)	Operator/Steward	Communications, practices, and policies between the SOC and the driver of the self-driven shuttle.
18	Performance Dashboard	USDOT Data Store	The data feed **from** the dashboard provides an upload of project data **to** the USDOT Data Store.

Table 3. External Interfaces of the System of Interest (Input Flows).

Flow #	Input Flows	Description	Subsystem Destination
1	PAL System	A set of transaction APIs to book ADA paratransit trips that include messages that support CTP functionality for booking, booking confirmation and status update messages. Additional messages may include cancellation requests. The	CTP

Flow #	Input Flows	Description	Subsystem Destination
		APIs will be developed and provisioned by the external paratransit vendor.	
2	NFTA Transit Vehicles	Data feeds using GTFS and GTFS RT (and maybe GTFS Flex) that describe scheduled and real time NFTA data. Additional data feeds for the Shuttle subsystem flex service are described in the Shuttle Trip Interface.	CTP
3	NFTA Elevator, Escalators Status	Data feed that provides information on NFTA station pathways including vertical paths (elevators, escalators). The data will be restricted to the stations in the project area. Additional real time status (e.g., operational/under maintenance, direction of travel, current floor) information will be provided as well. The format of the static pathways data has not yet been determined.	CTP
4	PROW & NITTEC Traffic Information	The data feeds provide: • Static information (see #4a). • Dynamic information (see #4b).	CTP
4a	PROW & NITTEC Traffic Information	The static network data includes right of way (ROW), public right of way (PROW) and parking capacity data feeds. The following data feeds are currently identified: • Road network or right of way includes rules and regulations will be downloaded from OpenStreetMap. • Public right of way data includes features and conditions of sidewalk, crosswalk, curb cuts, and walking paths. • Parking lot data (location, entrance/exit, and accessible paths through the lot to the sidewalk or facility). • Facility entrances (including accessible entrances), pathways and vertical pathways (elevators/escalators/stairs). • Location of wayfinding assets -- indoor and outdoor. • Location of traffic signal systems with pedestrian crossing functionality.	CTP
4b	PROW & NITTEC Traffic Information	The dynamic network data includes incidents on the road network in the project area, obstacles affecting traversal of PROW and parking fill data feeds.	CTP

Flow #	Input Flows	Description	Subsystem Destination
		Broadcast information using standardized formats for traffic, incidents, and work zone on right of way (arterials near BNMC). • PROW real time obstacles (work zones, weather related conditions). • Parking lot occupancy. • Incident and weather data. Several organizations in the area collect and disseminate situational awareness data through Niagara International Transportation Technology Coalition (NITTEC). This data is based on the Traffic Management Data Dictionary (TMDD).	

Table 4. External Interfaces of the System of Interest (Output Flows).

Flow #	Output Flow	Description	Source
1	PAL System	Complementing the input flow, this entails a set of transaction APIs to book ADA paratransit trips that include messages that support CTP functionality for booking, booking confirmation and status update messages. Additional messages may include cancellation requests. The APIs will be developed and provisioned by the external paratransit vendor.	CTP
5	Fare System / Payment System	Deeplink that brings up a specific page on a supported fare payment app already stored on a traveler's mobile device.	CTP
10	Emergency Management: SDS Safety Interface	The data feed provides information about the SDS vehicles to emergency management center (EMC). The information enables the EMC to monitor compliance with traffic regulations and passenger safety. In case of an incident involving a shuttle or a traveler onboard the shuttle, the SOC will communicate to EMC: (1) the location of the shuttle; (2) the nature of the incident; and (3) any additional information that would support the EMC in determining and applying the correct response.	Shuttle subsystem

Additional datasets are needed to monitor and measure performance of the system. Many of these datasets will be described in more detail in the PMESP that is currently under development and will be added to this document as new datasets are identified in future versions.

3.2 Data Overview

The ConOps incorporates a diverse group of datasets that are needed to implement and evaluate the system of interest. The datasets are listed in Table 5 at the end of this section. For each dataset, the table provides the following information:

- **Dataset ID:** A unique identifier used to reference the dataset.
- **Context Diagram Flow #:** A reference to the flow number in the ConOps Context Diagram (see Figure 4 and Tables 3 through 5).
- **Dataset Title**: A title for the dataset.
- **Description**: A concise yet informative description of the dataset.
- **Type**: The type of data that is generated. The type contains information on whether the data is structured, semi-structured or unstructured, and its velocity (static or dynamic), or unknown. Specific types include:
    - **Structured** – formatted with schema defining data entities, their syntax and semantics. Data types including in the dataset may include txt, csv, png, mpeg, and others. Dataset types include:
        - API – transaction set or data exchange services between two systems.
        - Geographic/attribute file – geographic data (features and attributes).
    - **Semi-Structured** – formatted data where the data entries may be ordered differently or content may change. Dataset types include:
        - Log file – semi structured data set that includes a line (row) for each entry. Typically, entries include a date and time entered.
        - Web forms – semi-structured data.
    - **Unstructured** – unformatted data where data is unknown when delivered. Crowdsource input and surveys are examples of unstructured data.
    - **Dynamic Velocity** – datasets that are streamed or require real time data acquisition methods. Dataset types include:
        - Real time data feed – real time data may be streamed, pushed, or pulled from source.
    - **Static Velocity** – datasets that are typically pulled from a source. They may change on a daily, monthly, or other frequency. Dataset types include:
        - Static data feed – static data that is typically pulled from source. This data does not change very often.

  - **Unknown** – may not be known at this time, to be determined by the end of the design phase.

- **Collection Method**: The method for collecting data, including:

  - Data collection and forwarding tools – tools used to generate, aggregate, store and forward datasets generated by SOI. These tools may generate datasets (e.g., tools to generate GTFS-Flex), aggregate transaction sets (for example, reservations, traveler trip histories), store (e.g., database or flat file), and forward (e.g., to an open data portal).

  - Derived – data derived from one or more sources. Many of these datasets will be identified in the PMESP.

  - Download – open data pulled from open data portal (e.g., OpenStreetMap).

  - User (traveler, infrastructure owner and operator (IOO)) input – data input / ingested from user (e.g., using web form or transaction set).

- **Data File Format(s)**: The data file format(s) the datasets are anticipated to be in when they are made accessible to the USDOT. The values include:

  - Flat file -- which may be in csv or plain text format (TBD later).

  - Schema based text files – including JSON (JavaScript Object Notation) and XML (Extensible markup language specified by an XML schema).

  - Specific formats – in several cases, specific encoding formats are defined for the standards that are applied to datasets, for example, GTFS real-time uses gtfs-realtime.proto.

  - Geographic feature / attribute dataset type formats for example ESRI Shape file, Keyhole Markup Language (KML), GeoJSON.

  - TBD – the format has not yet been determined because this is a new dataset that may be subject to a standard or emerging data dictionary. High level data requirements will be identified in the System Requirements document, but detailed data formats will not be determined until the design and development phase.

Table 5. Dataset Overview.

Dataset ID	Context Diagram Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
1	6	Pedestrian Signal Request Summary	Request-response message sets between the CTP mobile app requesting pedestrian signal crossing to the signal system	Structured; dynamic	Data collection and forwarding tool	TBD - National Transportation Communications for ITS Protocol (NTCIP)
2	8	Community Shuttle GTFS and GTFS Flex	Data feed that includes information on community shuttle operations including fixed/flex routes, demand responsive and on-demand service as well as fixed pick-up, drop-off shuttle stops.	Structured; Static	Data collection and forwarding tool	Flat
3	2	NFTA Fixed Route GTFS	Schedule and stop information for NFTA light rail and bus service	Static	IOO input	Flat
4	2	NFTA GTFS-Flex for PAL	Data feed that includes information on demand-responsive service for PAL and on-demand service for project region.	Structured; Static	IOO input	Flat

Dataset ID	Context Diagram Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
5	4a	PROW data	Public right of way information in the project region. • Parking garage entrances and pathways • Public right of way data includes features and conditions of sidewalk, crosswalk, curb cuts, and walking paths • Access to transit stops and community shuttle pickup/drop off locations • Access to accessible building entrances	Unknown	Data collection tools TBD	TBD
6	4a, 4b	Work Zone Data Specification	Work zone information that describes current work zones on PROW and crosswalks in the project region.	Structured; static	TBD	Schema based text files
7	8	Community Shuttle real time information	Real time and event data for shuttle operations. The specific format will be determined during the design stage to meet universal design for multiple user interface delivery mechanisms.	Unknown	Data collection and forwarding tool	TBD
8	2	NFTA GTFS-real time	Real time and event data for NFTA fixed route (light rail and bus). The datasets are currently under development by NFTA for a separate project—see Buffalo, NY ITS4US ConOps for more details.	Structured; dynamic	Data collection and forwarding tool	gtfs-realtime.proto
9	8	Shuttle Booking Summary and Details	Reservations exchanges between the CTP and Shuttle reservations module for both the SDS and HDS. The specific format will be determined during the design stage.	Structured; static	Data collection and forwarding tool	Flat

Dataset ID	Context Diagram Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
10	1	PAL Direct Reservations Summary and Details	A set of transaction APIs to book ADA paratransit trips that include messages that support CTP functionality for booking, booking confirmation and status update messages. Additional messages may include cancellation requests. The APIs are available and will be provisioned by the external paratransit vendor so it can be integrated into the CTP trip booking planning function.	Structured; static	Data collection and forwarding tool	Flat
11	9	CTP Usage log files	Performance data shared from the CTP subsystems. Types of data include trip plans and executions, user preferences and other log files that provide insight into traveler usage and utilization of Services. The data also includes utilization of user interface channels (web, mobile, call center). The data is dependent on the performance measurements and design documents that will be developed during the Design Phase. In addition, personally identifiable information (PII) will be stripped from the data provided to the PM Dashboard subsystem. (Note: trace data will be removed from the data sent from the CTP to the PMD).	Semi-structured; static	Data collection and forwarding tool	Schema based text files

Dataset ID	Context Diagram Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
12	9	SOC Dispatch log files	Performance data shared from the SOC subsystems. Types of data include trip booking and tracking information. The data is dependent on the performance measurements and design documents that will be developed during the Design Phase. In addition, PII will be stripped from the data provided to the PM Dashboard subsystem. (Note: trace data will be included in this dataset).	Semi-structured; static	Data collection and forwarding tool	TBD
13	Internal to CS	CS Service Vehicle Manifest	Ordered set of reservations information for passenger pickup including the customer(s), and pickup/drop off locations and times for each dispatched vehicle from the SOC. The format of the information will depend on whether the recipient is a SDS or human operator / steward.	Semi-structured; dynamic	Data collection and forwarding tool	TBD
14	Internal to PMD	CS Vehicle Performance Data	This includes performance data such as ridership, on time performance, travel times, etc. To be determined the performance plan and refined in the design phase.	Structured; static	Derived	Flat

Dataset ID	Context Diagram Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
15	12, 13, 14	Trip Planning Data	Trip Planning, scheduled and dynamic event data and interactive forms to support generating or reporting on trips/user satisfaction exchanged between the CTP backoffice and the user interfaces including TIH, call center, traveler web, web app or mobile app. The TIH, call center and web browser use a subset of the mobile app APIs. The APIs will include the following types of transactions: • Trip Plans • Traveler profile (stripped of contact information) including preferences • Trip histories (including timestamped notifications and reservations, requests for pedestrian signal requests, pickup/drop off times/locations) • Responses to survey questions (e.g., user satisfaction ratings) • User reviews and comments	Structured; static	Data collection and forwarding tool	Schema based text files
16	12, 13, 14	Customer Comment Forms	Forms that are completed by CTP access channels (web, mobile and call center) on trip conditions and obstacles to travel along indoor and outdoor pathways.	Semi-structured; static	Data collection and forwarding tool	Schema based text files

Dataset ID	Context Diagram Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
17	13	Smart Sign Messages	Messages broadcast by Smart Signs to support orientation and wayfinding signals. The specific format, content and communication methods will be determined during the design stage.	Unknown	IOO input	Flat
18	13	Smart sign message log file	Log and summary data of smart sign operations.	Unknown	IOO input	Flat
19	15, 16	HDS Canned and custom operations messages	A set of canned operations messages and format for custom messages that may be exchanged between the shuttle operations personnel and HDS operators. These will follow the types of messages that are exchanged between NFTA PAL Dispatch and PAL operators.	Unknown	IOO input	Flat
20	15, 17	SDS custodian Canned operations messages	A set of canned operations messages and format for custom messages that may be exchanged between the shuttle operations personnel and SDS steward. These will follow the types of messages that are exchanged between NFTA PAL Dispatch and PAL operators.	Unknown	IOO input	Flat
21	2	NFTA Conveyance Data Feed	Data feed that provides information on NFTA station pathways including vertical paths (elevators, escalators). The data will be restricted to the stations in the project area.	Unknown	IOO input	TBD

Dataset ID	Context Diagram Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
22	4b	NFTA Conveyance Status	Real time data feed that provides information on elevator/escalator operational status.	Unknown	Data collection and forwarding tool	TBD
23	4a, 4b	NITTEC Traffic Information	The static network data and dynamic information includes right of way (ROW) data feeds and situational awareness TRANSCOM data fusion engine SPATAL data feeds. The following data feeds are currently identified: • mobile maps • situational awareness information (incidents, work zones, planned events)	Structured; dynamic	Data collection and forwarding tool	Schema based text files
24	4a	Intersection Crossing Assets	Static information on intersections and timing plans for crossing.	Structured; static	IOO input	TBD (will be a geographic exchange file such as: shape file, GeoJSON or KML)
25	4a	Smart Sign assets	Location of smart sign assets, interfacing requirements and the types of information they provide.	Structured; static	IOO input	TBD (will be a geographic exchange file such as: shape file, GeoJSON or KML)

Dataset ID	Context Diagram Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
26	4a	Facility smart sign assets	Location of smart sign assets, interfacing requirements and the types of information they provide. May need to augment location referencing method to locate indoor spaces).	Structured; static	IOO input	TBD (will be a geographic exchange file such as: shape file, GeoJSON or KML)
27	5	Fare App APIs	The method and specification for mobile app deep linking.	Structured; dynamic	IOO input	Schema based text files
28	10	Emergency Safety Messages	Message set with information about the SDS vehicles to EMC including information on current location and event information.	Unknown	Data collection and forwarding tool	TBD
29	4a	OpenStreetMap	Street and public right of way network information downloaded from the Open Street Map foundation.	Structured; static	Download	Protocolbuffer Binary Format (PBF)

4 Data Stewardship

Governance of the data includes people, policies and rules associated with ensuring quality, curation (data acquisition through distribution or retirement), management and ownership. Section 4.1 describe the “people”, Sections 4.2 through 4.4 describe policies related to these datasets. Rules will be described in other documents such as the PMESP and Phase 2 design documents.

4.1 Data Owner and Stewardship

The “people” part of governance includes stakeholders and their roles and responsibilities. Each dataset is assigned a stakeholder who is described as the data owner and steward. In addition, USDOT ITS JPO, as the sponsor of this project, will have an oversight role on datasets that are generated or derived through this project. External datasets (those imported from external sources) are not overseen by the USDOT sponsors. Each column is defined as follows:

- **Dataset ID:** a unique number assigned to each dataset.
- **Dataset Title:** The title of the datasets that is assigned to the designated data owner and/or data steward.
- **Data Owner:** The owner of 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. The data owner will be the U.S. DOT Program for any data collected for a U.S. DOT-funded research project. Initially, the data owner is the team stakeholder creating the dataset.
- **Data Steward:** The data steward for the dataset. 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. 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. The data steward may be the recipient of U.S. DOT-funded research projects or one of the project’s partner organizations.
- **Federal Sponsor:** The U.S. DOT sponsor for the dataset. The federal sponsor assumes the role of Data Owner once the dataset is provided to them per BAA/NOFO requirements later in the project. Note: Any data collected/created for a U.S. DOT-funded research project is required to be provided to USDOT, unless otherwise noted in the BAA.

Table 6 lists the datasets and details the owner and steward of row.

Table 6. Data Owner and Steward Information.

Dataset ID	Dataset Title	Data Owner	Data Steward	Federal Sponsor
1	Pedestrian Signal Request Summary	COB, traveler	CTP	USDOT ITS JPO
2	Community Shuttle GTFS and GTFS Flex	NFTA	CS Operator	USDOT ITS JPO
3	NFTA Fixed Route GTFS	NFTA	CTP	USDOT ITS JPO
4	NFTA GTFS-Flex for PAL	NFTA	CTP	USDOT ITS JPO
5	PROW data	IOO of PROW	CTP	TBD (the acquisition approach of data has not yet been determined)
6	Work Zone Data Specification	COB	CTP	Not applicable. External data source
7	Community Shuttle real time information	CS Operator	CS Operator	USDOT ITS JPO
8	NFTA GTFS-real time	NFTA	NFTA	USDOT ITS JPO
9	Shuttle Booking Summary and Details	CS Operator, Traveler	CS Operator, CTP	USDOT ITS JPO
10	PAL Direct Reservations Summary and Details	PAL Direct	PAL Direct	USDOT ITS JPO
11	CTP Usage log files	CTP	CTP	USDOT ITS JPO
12	SOC Dispatch log files	CS Operator	CS Operator	USDOT ITS JPO
13	CS Service Vehicle Manifest	CS Operator	CS Operator	Not applicable because these may contain PII/ Health Insurance Portability and Accountability Act (HIPAA) data

Dataset ID	Dataset Title	Data Owner	Data Steward	Federal Sponsor
14	CS Vehicle Performance Data	CS Operator	CS Operator	USDOT ITS JPO
15	Trip Planning Data	COB, traveler	CTP	USDOT ITS JPO
16	Customer Comment Forms	COB, traveler	CTP	USDOT ITS JPO
17	Smart Sign Messages	IOO of asset	Project data steward	USDOT ITS JPO
18	Smart sign message log file	IOO of asset	IOO of asset	USDOT ITS JPO
19	HDS Canned and custom operations messages	CS Operator	CS Operator	USDOT ITS JPO
20	SDS custodian Canned operations messages	CS Operator	CS Operator	USDOT ITS JPO
21	NFTA Conveyance Data Feed	NFTA	NFTA	USDOT ITS JPO
22	NFTA Conveyance Status	NFTA	NFTA	USDOT ITS JPO
23	NITTEC Traffic Information	NITTEC	NITTEC	Not applicable. External data source
24	Intersection Crossing Assets	COB	COB	Not applicable. External data source
25	Smart Sign assets	IOO of asset	IOO of asset	USDOT ITS JPO
26	Facility smart sign assets	IOO of asset	IOO of asset	USDOT ITS JPO
27	Fare App APIs	NFTA	Fare vendor (operator)	Not applicable. External data source
28	Emergency Safety Messages	NFTA	NFTA	USDOT ITS JPO
29	OpenStreetMap	OSM	CTP	Not applicable. External data source

4.2 Access Level

Four access levels are defined relative to this project. The access levels combine levels of privacy and license restrictions (e.g., limited copying or redistribution). The four types of data access include:

- **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 may be anonymized or aggregated version of private datasets to protect PII.
- **Private**
    - **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.
    - **Operational (Proprietary)** – Data stewards and operational personnel can access data for operations purposes only. This includes both third party (licensed) datasets and protected data. A subset of the data may be proprietary data that is licensed by third party or commercial business interests. Access to CBI data is determined by usage agreements between the parties.
    - **Protected (PII)** – Data that has PII included in the dataset. Access to this data is restricted to protect the PII based on IRB-approved processes. Data in this category should have an operational purpose that justifies its storage.

Section 4.2.1 describes those datasets that are private and will not be released to the public that includes datasets that fall under the Research, Operational and Protected access levels. Some datasets will be available for research and summarization purposes once they are anonymized although they may still contain trace data. Other datasets will not be made available because they are protected and may contain PII and HIPAA information.

4.2.1 Private Datasets

Table 7 lists the datasets that will be private (i.e., not shared with the public). For each private dataset, the table provides the access level, reason why the data is private and the safeguarding methods and processes. It is important to note that the access level of that dataset may change as additional information becomes available.

Table 7. List of Private Datasets.

Dataset ID	Dataset Title	Access Level	Reason(s) the Data is Private	Safeguarding Methods and Processes
1	Pedestrian Signal Request Summary	Operational	may contain PII	Anonymize and stored on secure server

9	Shuttle Booking Summary and Details	Research	Contains PII	Encryption and store on secure server
10	PAL Direct Reservations Summary and Details	Research	May contain PII	Encryption and store on secure server
11	CTP Usage log files	Research	May contain PII / HIPAA and even anonymized may still contain trace data.	Encryption and store on secure server
12	SOC Dispatch log files	Research	May contain PII	Anonymize and stored on secure server
13	CS Service Vehicle Manifest	Protected	Contains PII and HIPAA related information	Encryption and store on secure server
15	Trip Planning Data	Research	Contains PII	Encryption and store on secure server
16	Customer Comment Forms	Research	May contain PII	Encryption and store on secure server
19	HDS Canned and custom operations messages	Open and Research	May contain PII	Anonymize and stored on secure server
20	SDS custodian Canned operations messages	Open and Research	May contain PII	Anonymize and stored on secure server
23	NITTEC Traffic Information	Operational	External data source	Will integrate the data and then discard
24	Intersection Crossing Assets	Operational	May be subject to security provisions	Security protocols (TBD during the design phase)
27	Fare App APIs	Operational	Proprietary and subject to license agreement	encrypted and secured

4.2.2 Access Request

Access to all datasets is expected to be managed by the data steward in close coordination with the data owner. However, management of access to research data will be determined after a full IRB process is completed in Phase 2. This section will be updated then, as stated in Table 1. Detailed procedures and application forms will be developed during Phase 2.

4.2.3 Related Tools, Software and/or Code

No special tools or software will be required to process or access the datasets.

4.2.4 Relevant Privacy and/or Security Agreements

During the project design phase, a formal set of data privacy and security agreements will be developed that incorporates all the datasets. The agreement will include existing privacy and security agreements issued from stakeholder data providers including NFTA PAL and City of Buffalo.

4.3 Re-Use, Redistribution, and Derivative Products Polices

All dataset identified so far in this project are expected to be open licensed, as detailed in Table 8. This section will be updated in Phase 2 once there are more details about the design. Note that some datasets, although not subject to a license restriction, are subject to access level provisions (e.g., no redistribution of the data). The access level drives some restrictions on the data.

Table 8. Re-Use, Redistribution, and Derivative Products Polices

ID	Dataset Title	Access Level	License Used	Reason(s) for Non-Open License
1	Pedestrian Signal Request Summary	Research	Open	n/a
2	Community Shuttle GTFS and GTFS Flex	Open	Open	n/a
3	NFTA Fixed Route GTFS	Open	Open	n/a
4	NFTA GTFS-Flex for PAL	Open	Open	n/a
5	PROW data	Open	Open	n/a
6	Work Zone Data Specification	Open	Open	n/a
7	Community Shuttle real time information	Operational	Open	n/a
8	NFTA GTFS-real time	Open	Open	n/a
9	Shuttle Trip Booking Summary and Details	Protected (PII)	Open	n/a
10	PAL Direct Reservations Summary and Details	Protected (PII)	Open	n/a
11	CTP Usage log files	Research	Open	n/a

ID	Dataset Title	Access Level	License Used	Reason(s) for Non-Open License
12	SOC Dispatch log files	Research	Open	n/a
13	CS Service Vehicle Manifest	Protected (PII)	Open	n/a
14	CS Vehicle Performance Data	Open	Open	n/a
15	Trip Planning Data	Research	Open	n/a
16	Customer Comment Forms	Protected (PII)	Open	n/a
17	Smart Sign Messages	Open	Open	n/a
18	Smart sign message log file	Research	Open	n/a
19	HDS Canned and custom operations messages	Research	Open	n/a
20	SDS custodian Canned operations messages	Open	Open	n/a
21	NFTA Conveyance Data Feed	Open	Open	n/a
22	NFTA Conveyance Status	Open and research	Open	n/a
23	NITTEC Traffic Information	Open and research	Open	n/a
24	Intersection Crossing Assets	Open	Open	n/a
25	Smart Sign assets	Open	Open	n/a
26	Facility smart sign assets	Operational	Open	n/a
27	Fare App APIs	Operational	TBD	Depends on vendor and agreement
28	Emergency Safety Messages	Open	Open	n/a

ID	Dataset Title	Access Level	License Used	Reason(s) for Non-Open License
29	OpenStreetMap	Open	Open	n/a

4.4 Data Storage and Retention

Data storage and retention policies will be developed as part of the design and implementation phase. This section identifies several types of storage systems and related policies and procedures that will be developed during Phase 2. It is important to note that, once available, all open data will be provided to the USDOT in near real time throughout the length of the project.

4.4.1 Storage Systems

Table 9 details the systems used to store the project's data. For each dataset, the following information is provided:

- **Data Storage System Name**: The name of the data storage system the dataset will be stored in. The storage systems are depicted in Figure 7 in Section 3)
- **Data Storage System Type**: The data storage system type. The system type describes where governance to the data store policies and agreements will be made. "Project" indicates that the storage systems will be developed, implemented, operated, managed, and maintained by the project team's assigned data steward.
- **Initial Storage Date**: The initial date that data will be available in the storage system.
- **Frequency of Update**: How frequently the data will be updated in the data storage system once ingestion begins.
- **Archiving and Preservation Period**: The expected duration for which the dataset will be maintained in each data storage system. The archiving will be from the beginning through the project period of performance (POP).

Table 9. Storage Systems.

Dataset ID	Dataset Title(s)	Data Storage	System Type	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
1	Pedestrian Signal Request Summary	CTP storage	Project	Will be determined during design phase	Each occurrence	Project POP

Dataset ID	Dataset Title(s)	Data Storage	System Type	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
2	Community Shuttle GTFS and GTFS Flex	CTP storage	Project	Will be determined during design phase	Quarterly	Project POP
3	NFTA Fixed Route GTFS	CTP storage	Project	Will be determined during design phase	Quarterly	Project POP
4	NFTA GTFS-Flex for PAL	CTP storage	Project	Will be determined during design phase	Quarterly	Project POP
5	PROW data	CTP storage	Project	Will be determined during design phase	As needed	Project POP
6	Work Zone Data Specification	TBD	TBD	Will be determined during design phase	As needed	Based on organization retention policy
7	Community Shuttle real time information	CS storage	Project	Will be determined during design phase	Continuously	Project POP
8	NFTA GTFS-real time	NFTA Transit server	NFTA Transit	TBD during design phase	Continuously	Project POP

Dataset ID	Dataset Title(s)	Data Storage	System Type	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
9	Shuttle Booking Summary and Details	CS storage	Project	TBD during design phase	TBD - will be determined during design phase	Project POP
10	PAL Direct Reservations Summary and Details	NFTA PAL Server	NFTA PAL	TBD during design phase	TBD - will be determined during design phase.	Based on organization retention policy
11	CTP Usage log files	CTP storage	Project	TBD during design phase	Daily	Project POP
12	SOC Dispatch log files	CS storage	Project	TBD during design phase	Daily	Project POP
13	CS Service Vehicle Manifest	CS storage (secure)	Project	TBD during design phase	As needed (for each dispatched vehicle)	Project POP
14	CS Vehicle Performance Data	Performance Management Dashboard (PMD) storage	Project	TBD during design phase	Daily	Project POP
15	Trip Planning Data	CTP storage	Project	TBD during design phase	Continuously	Project POP
16	Customer Comment Forms	CTP storage	Project	TBD during design phase	As needed	Project POP
17	Smart Sign Messages	Message file storage	Project	TBD during design phase	As needed	Project POP
18	Smart sign message log file	Message file storage	Project	TBD during design phase	As needed	Project POP

Dataset ID	Dataset Title(s)	Data Storage	System Type	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
19	HDS Canned and custom operations messages	Message file storage	Project	TBD during design phase	As needed	Project POP
20	SDS custodian Canned operations messages	Message file storage	Project	TBD during design phase	As needed	Project POP
21	NFTA Conveyance Data Feed	NFTA Conveyance storage	NFTA	TBD during design phase	As needed	Based on organization retention policy
22	NFTA Conveyance Status	NFTA Conveyance storage	NFTA	TBD during design phase	Continuously	Based on organization retention policy
23	NITTEC Traffic Information	NITTEC servers	External (NITTEC)	TBD during design phase	Continuously	Based on organization retention policy
24	Intersection Crossing Assets	COB Server	External	TBD during design phase	As needed	Project POP
25	Smart Sign assets	Message file storage	Project	TBD during design phase	As needed	Project POP
26	Facility smart sign assets	Message file storage	Project	TBD during design phase	As needed	Project POP
27	Fare App APIs	N/A	External	Unknown	As needed	Based on organization retention policy
28	Emergency Safety Messages	Message file storage	Project	TBD during design phase	As needed	Project POP
29	OpenStreetMap	CTP storage	Project	TBD during design phase	As needed	Project POP

4.4.2 Data Storage System Description

The storage systems will be designed, developed, and tested during Phase 2 and are unknown at this time. The datasets are assigned to subsystem level servers, for example CTP Storage, CS Storage. In addition, static message, and configuration dataset, (e.g., Emergency Safety Messages) will be stored in a file storage system. These files will not be subject to significant change over the course of the project. Even within the subsystem level, the systems will implement role based access. PII will be in separate servers and be encrypted to enforce security

and limit access. As noted above, the storage of protected PII data will be designed and developed in Phase 2.

4.4.3 Cybersecurity Polices

During the project design phase, cybersecurity polices will be developed that cover operations and networks of all the subsystem back-office systems, vehicles, field devices and their interactions. The policies will include existing cybersecurity policies issued from stakeholder data providers including NFTA PAL and City of Buffalo.

4.4.4 Data Security Policies and Procedures

The data security policies and procedures developed during Phase 2 will provide layered security policies and procedures to protect data storage while at rest and data exchange while in transit. The policies will address:

1. **Confidentiality** requirements to ensure that information is not made available or disclosed to unauthorized persons or systems.

2. **Availability** to ensure that data is accessible, functioning, and able to meet the needs of the system. The types of procedures include addressing Denial of Service attacks and other ransom attacks and corruption of data.

3. **Integrity** to preserve the quality of the data including accuracy and consistency. Integrity also includes protecting against unauthorized modification to prevent unauthorized modification.

4. **Authenticity and non-repudiation** to ensure that data and information is authentic, the user is confirmed, and the message exchange can be audited. The procedures include identity management and role-based access provisions.

4.4.5 Back-up and Recovery Policies and Procedures

During Phase 2, back-up and recovery policies and procedures will be developed that cover all datasets including operational data of all the subsystem back-office systems, vehicles, field devices and their interactions. The policies will incorporate existing back-up, recovery and retention policies issued from stakeholder data providers including NFTA PAL and City of Buffalo.

5 Data Standards

5.1 Data Standards

Table 10 lists the data standard(s) used for each dataset. For each dataset, the table provides

- **Data Standard(s)**: Name(s) of the data standard(s) in which the data are made available to the USDOT.
- **Data Standard Digital Object Identifier (DOI)**: List the DOI(s) of the standard(s) for the data. If possible, provide a Uniform Resource Locator (URL) to the data standard(s). If one does not exist, just document it with "n/a" in this field.
- **Open or Proprietary:** Indicating whether the data standard(s) is/are "Open" or "Proprietary."
- **Data Standard(s) Rationale**: Explaining the use of the chosen data standard(s).

Table 10. Data Standards.

Data set ID	Dataset Title	Data Standard(s)	Data Standard DOI	Open / Proprietary?	Data Standard(s) Rationale
1	Pedestrian Signal Request Summary	NTCIP 1211	NTCIP 1211 version v02	Open	Industry standard to implement Signal Control message exchange.
2	Community Shuttle GTFS and GTFS Flex	GTFS Flex	GTFS Flex	Open	Industry standard to implement flexible transit information.
3	NFTA Fixed Route GTFS	GTFS	GTFS	Open	Industry standard to implement static transit information.
4	NFTA GTFS-Flex for PAL	GTFS Flex	GTFS Flex	Open	Industry standard to implement flexible transit information.

Data set ID	Dataset Title	Data Standard(s)	Data Standard DOI	Open / Proprietary?	Data Standard(s) Rationale
5	PROW data	Several standards are under consideration including Open Geodata Standard, OpenStreets, emerging ISO standard.	TBD	Open	Will adopt a compatible standard to provide sidewalk and indoor facility information.
6	Work Zone Data Specification	WZDX v3.1 or most recent	WZDX	Open	Emerging industry standard to provide work zone information particularly for automated vehicles.
7	Community Shuttle real time information	May be GOFS or GTFS real-time	GTFS realtime	Open	Industry standard to provide real time transit location, event, and prediction information.
8	NFTA GTFS-real time	GTFS real-time	GTFS realtime	Open	Industry standard to provide real time transit location, event, and prediction information.
17	Smart Sign Messages	May develop standard set of messages	TBD	Open	No industry standard yet.
19	HDS Canned and custom operations messages	May use TCIP messages or adopt existing NFTA CAD/AVL message set	TBD	Open	No industry standard yet.
20	SDS custodian Canned operations messages	May use TCIP messages or adopt existing NFTA CAD/AVL message set	TBD	Open	No industry standard yet.
21	NFTA Conveyance Data Feed	GTFS-Pathways	GTFS Pathways	Open	Although there are no industry standards, the GTFS pathway specification includes information on vertical conveyances in transit stations.

Data set ID	Dataset Title	Data Standard(s)	Data Standard DOI	Open / Proprietary?	Data Standard(s) Rationale
22	NFTA Conveyance Status	May develop standard set of messages	TBD	Open	No industry standard yet.
23	NITTEC Traffic Information	Institute of Transportation Engineer Traffic Management Data Dictionary	TBD	Open	NITTEC implemented their system using the TMDD.
27	Fare App APIs	N/A	N/A	Proprietary	Must use NFTA Fare System, which is deployed by a commercial vendor. There are no open standards for fare payment systems.
28	Emergency Safety Messages	May develop standard set of messages	TBD	Open	No industry standard yet.
29	OpenStreet Map	Protocolbuffer Binary Format	OSM	Open	OSM defines a standard set of features and attributes for the road and PROW networks. The network is used in the Open Trip Planner, which is the foundation of the CTP subsystem.

5.2 Versioning

Many dataset specifications include attributes or files that identify the dataset version. These native files and attributes will be the primary reference for the dataset version. For example, GTFS includes a feed_info.txt that includes version information. Datasets stored in a database will include attributes for each data item. Several date types typically contribute to the version naming convention. These date types include:

- **Activation date** – date and time collection of data began.
- **Deactivation date** – date and time collection of data end.

- **Acquisition date** – date and time dataset were secured, obtained, downloaded, or verified.

- **Publication date** – date and time dataset were published.

For derived data, the effective set of versioning dates is typically based on the data source versions and the date the summary information was derived, for example “Ridership for July 2021.”

Each dataset will have a version naming convention that includes reference to these date types. The naming convention will be described in the Design Phase. The version information will also include a reference to the appropriate Metadata files, which contain summary information about the specific dataset or dataset collection version.

A list of datasets with their data types will be published so downstream users can identify the most active and planned changes to the datasets. In addition, a summary of any changes that impact the dataset format will be included in this summary information.

5.3 Metadata

For each dataset and dataset collection, metadata will be generated that describe “data about the data.” These metadata files will provide summary information that describes the dataset, purpose, lineage, applied methods, quality, organization, lifecycle changes, currency, and access. Section 5.3.1 describes the types of information that are documented in the metadata files, Section 5.3.2 describes the file structure(s), and Section 5.3.3 provides information on the metadata update process.

5.3.1 Metadata Types

The metadata files will contain information to facilitate discovery and to understand the purpose, usage, license provisions, lineage, currency, quality, and lifecycle processes used in curating the dataset. The metadata types are described as follows:

Discovery file – Metadata that describes high level information about the dataset. These include:

- **Dataset name** – file name of dataset associated with this metadata.

- **Abstract** – summary of the dataset content, purpose, and usage.

- **Data Tags** – tags that are associated with the dataset, for example “transit.”

- **Data owners and stewards** – contact information of data owners and / or steward.

- **Currency** – activation, deactivation, acquisition, and publication dates associated with the dataset. The currency also summarizes any gaps in the dataset.

- **Lineage** – summary of the data source(s) or collection methods used to collect information.

- **Usage** – who should use the data and how the data is intended to be used.
- **Permissions** – restrictions on dataset including privacy and security levels associated with dataset.
- **Quality** – summary of quality including location and attribute accuracy, (location scale, resolution, and precision), gaps in data, processes applied to verify data.
- **Access Method** – location of dataset and methods used to access dataset. This includes information on acquiring dataset.
- **External references (optional)** –The location (URL) of an external reference that describes or specifies the data or data format. For example, the URL for the GTFS real-time reference defines the schema, data definitions and encoding descriptions. Transcom DFE/SPATAL reference specify the specification for NITTEC event datasets.
- **Included files** – lists the names of additional files that are associated with this dataset (see below).

Additional files will also be included in the metadata collection. The files include the following:

1. **Licensing** – Metadata that provides licensing information or a reference to licensing information for the data. The licensing terms and conditions provide information on user rights to use, copy, publish, distribute, transmit, cite, or adapt the data.
2. **Data Schema** – Metadata that documents the fields in the data including field name, description, data type, and notes. This information describes the contents of the data or dataset generally, in accordance with the Project Open Data Metadata Schema[1] or another appropriate standard. The metadata file does not need to be included if the External Reference (in the Discovery file) includes a user specification—note that it is important that all metadata is included for all datasets, even if it is a link to where the details are explained.
3. **Data Processing** – Metadata that documents any curation processes that was performed on the data from the data acquisition to publication (that includes when it is delivered to the USDOT).
4. **Data Impact Log** – Metadata that logs information on impacts to data during the collection period. Impacts on the data are documented in an Impact Log. Log entries include:
    i. Data collection start and end date/times
    ii. Gaps in collection and reason for gap (e.g., weather, sensors down, construction, etc.)
    iii. Dataset State (e.g., baseline, testing, operations, etc.)
    iv. Changes to implementation that impacts data
    v. Data organization (schema) or data field change (e.g., added, changed, or eliminated data element)

[1] https://project-open-data.cio.gov/v1.1/schema/

5.3.2 Metadata Structure

In this section, structure used to store the metadata and associate it with the appropriate dataset is described.

Metadata provided to the USDOT will be stored in a flat file format such as plain text files, CSV, XML, JSON, or Structured Query Language. The specific format will be designed during the Development Phase in coordination with the USDOT. Metadata will be stored with the data to allow users of the data, including future deployers, researchers and the public all key information in a unique location. Files will be provided in a way that maintains the structure of the data, so users of the data can easily determine what is project metadata and what is specific dataset metadata. This will be accomplished using a folder format and naming convention. Here is an example of a possible complete metadata package may look like.

- **ITS4US Buffalo** - Project Folder
    - **ITS4US_Buffalo_DataInventory.txt** – Discovery information for all datasets. Includes a link to the current DMP and the location of all datasets and their version information. The version information will include a “read me” file which summarizes changes that impact data ingestion, integrity or quality.
    - **Dataset 1 *DatasetName* Folder**
        - ***Dataset*_Discovery.txt -** Discovery information for this dataset
        - **Dataschema.csv (optional)** – Table of all the fields in the dataset.
        - **Dataprocessing.txt** – List of dataset processes conducted on the data before delivery.
        - **DataImpactLog.csv** – Table of events that changed the dataset.
        - **License.txt** – License information for the dataset collected for the Project. Depending on the data this could have its own unique license from the rest of the Project.
    - **Dataset 2 Folder**
        - ***Dataset*_Discovery.txt -** Discovery information for this dataset
        - **Dataschema.csv (optional)** – Table of all the fields in the dataset.
        - **Dataprocessing.txt** – List of dataset processes conducted on the data before delivery.
        - **DataImpactLog.csv** – Table of events that changed the dataset.
        - **License.txt** – License information for the dataset collected for the Project. Depending on the data this could have its own unique license from the rest of the Project.
    - **Dataset N Folder…**

5.3.3 Metadata Update Process

The metadata update process will occur throughout the dataset lifecycle. In Figure 5, the metadata update activities are driven by each step in the data curation process. At each stage of the curation process, the impact log is updated and specific metadata fields are verified for correctness and updated if the process deviated from the “typical” process. Prior to dataset publication, the metadata files are updated and published to new folders. The metadata for public datasets will be published to location that is open and accessible to third party users. Metadata for private datasets will be restricted to authorized personal to internal servers that manage the datasets.

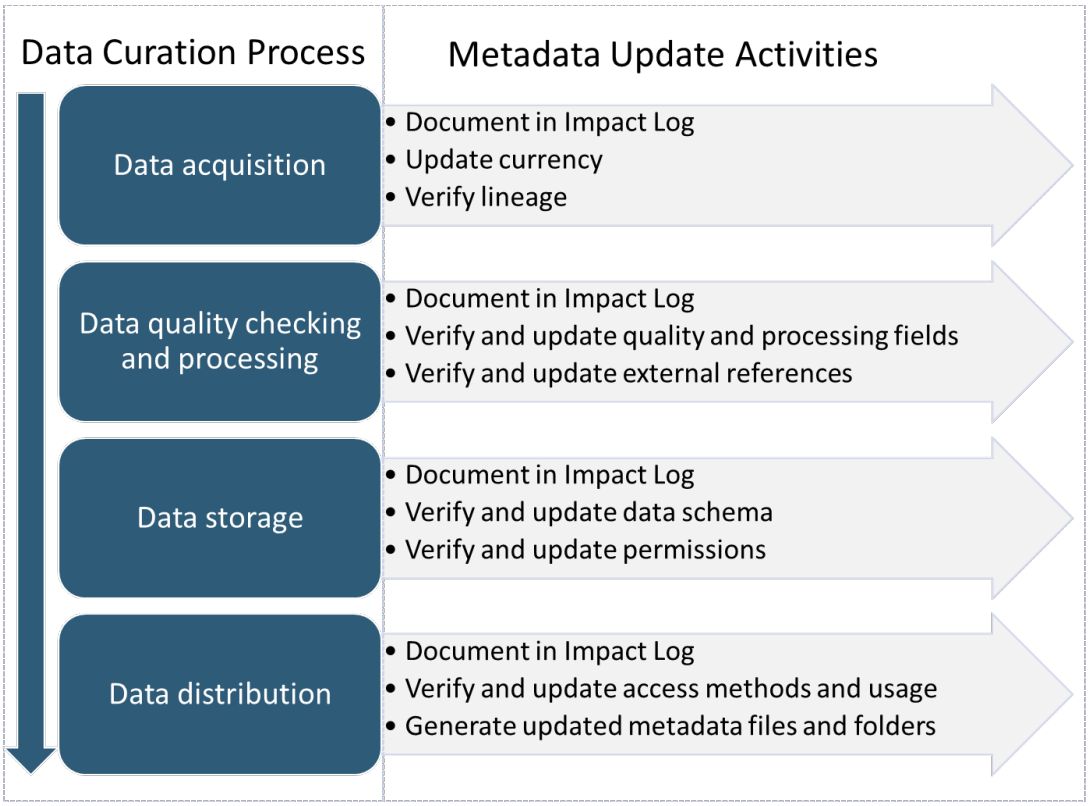

Figure 5. Metadata update process associated with Data Curation Processes.
Source: Buffalo, NY ITS4US

Appendix A. Acronyms

Table 11 lists the acronyms used in the document.

Table 11. List of acronyms

Acronym	Description
API	Application Programming Interfaces
BAA	Broad Agency Agreement
BNMC	Buffalo Niagara Medical Campus
ConOps	Concept of Operations
CS	Community Shuttle
CTP	Complete Trip Platform
EMC	Emergency Management Center
FHWA	Federal Highway Administration
FTA	Federal Transit Administration
GTFS	General Transit Feed Specification
HDS	Human-Driven Shuttle
HIPAA	Health Insurance Portability and Accountability Act
IE	Independent evaluator
IOO	Infrastructure owner and operator
ITS	Intelligent Transportation System
IVR	Interactive Voice Response
JPO	Joint Program Office
JSON	JavaScript Object Notation
KML	Keyhole Markup Language
NITTEC	Niagara International Transportation Technology Coalition
NFTA	Niagara Frontier Transportation Authority
NTCIP	National Transportation Communications for ITS Protocol
NY	New York
PAL	Paratransit Access Line
PROW	Public Right of Way
ROW	Right of way
SDS	Self-Driving Shuttle
SOC	Shuttle Operations Center
SOI	System of Interest
TBD	To be determined
TIH	Transportation Information Hub
TMDD	Traffic Management Data Dictionary
UB	University at Buffalo
UI	User Interface
URL	Uniform Resource Locator
U.S.	United States
USDOT	United States Department of Transportation
XML	Extensible Markup Language

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-21-868

U.S. Department of Transportation