Phase 2 Data Management Plan (DMP)

Buffalo, NY ITS4US Deployment Project

www.its.dot.gov/index.htm

Final– January 27, 2023

FHWA-JPO-22-973

ITS4US

IT'S TRANSPORTATION FOR ALL OF US

U.S. Department of Transportation

Produced by NFTA
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-22-973		

4. Title and Subtitle	5. Report Date
Phase 2 Data Management Plan (DMP) Buffalo, NY ITS4US Deployment Project	January 27, 2023
	6. Performing Organization Code (Delete and insert information here or leave blank)

7. Author(s)	8. Performing Organization Report No.
Adel W. Sadek (UB), Polly Okunieff (ICF), Nayel Urena Serulle (ICF), Deepak Gopalakrishna (ICF), Robert Jones (NFTA), Kelly Dixon (GBNRTC), Jamie Hamann-Burney (BNMC), Chunming Qiao (UB), Stephen Still (UB), Victor Paquet (UB), and Jordana Maisel (UB)	(Delete and insert information here or leave blank)

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
NFTA, 181 Ellicott Street, Buffalo, NY 14203 BNMC, 640 Ellicott Street, Buffalo, NY 14203 ICF International, 9300 Lee Highway, Fairfax, VA 22031 University at Buffalo, Amherst, NY 14228 RSG, 55 Railroad Row, Suite 101, White River Junction, VT 05001 ETCH, 4696 Smothers Road, Westerville, OH 43081	**11. Contract or Grant No.** 693JJ32250011

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
U.S. Department of Transportation ITS Joint Program Office 1200 New Jersey Avenue, SE Washington, DC 20590	Final
	14. Sponsoring Agency Code HOIT-1

15. Supplementary Notes

Elina Zlotchenko (USDOT ITS-JPO) is the Award Officer Representative (AOR) and Sarah Tarpgaard (USDOT) is the Award Officer (AO).

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 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 (Allentown, 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 datasets and processes of the Buffalo, NY ITS4US deployment—such as data collection, analysis, protecting user data privacy, storage and sharing. The document also includes information on the datasets and their relationship to the performance metrics.

17. Keywords	18. Distribution Statement
ITS4US; Deployment; ITS; Intelligent Transportation Systems; Data Management Plan, Performance Measures, Data Quality	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	**unclassified**	76	

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

Revision History

Name	Date	Version	Summary of Changes	Approver
Buffalo ITS4US Phase 2 Data Management Plan	Nov 1, 2022	0.1	Initial Draft	Kelly Dixon
Buffalo ITS4US Phase 2 Data Management Plan	Dec 5, 2022	0.2	Responded to USDOT Comments	Kelly Dixon
Buffalo ITS4US Phase 2 Data Management Plan	Jan 27, 2023	1.0	Updated SOI and checked for 508 compliance	Kelly Dixon

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Document Purpose

The Phase 2 Data Management Plan (DMP) details the data, data collection, analysis, quality control, protection, storage and sharing of data acquired, used, derived, and processed as part of the Buffalo NY ITS4US Deployment Project. This Phase 2 document builds on the conceptual framework developed in Phase 1, specifically, the Phase 1 Data Management Plan (DMP)(FHWA-JPO-21-868)[1], and other Phase 1 documents including Concept of Operations (ConOps) (FHWA-JPO-21-860)[2], Performance Measurement and Evaluation Support Plan (PMESP) (FHWA-JPO-21-878)[3], System Requirements Specification (SyRS) (FHWA-JPO-21-883)[4], and privacy considerations in the Human Use Approval Summary (HUAS) (FHWA-JPO-21-898)[5]. As these documents are updated and the deployment progresses, this document will be updated and reissued as an “as-built” version.

1.1.1 Organization of this Document

This document is organized following the guidance provided by FHWA, which is based on the USDOT’s National Transportation Library’s guidance for Creating Data Management Plans for Extramural Research.

The sections are as follows:

- **Section 1** provides a brief overview of the ITS4US project including the document purpose, deployment concept, data management schedule, and data used in this deployment (e.g., needs and overview).
- **Section 2** describes the data stewardship and management approach.
- **Section 3** details the data standards, including the metadata.
- **Appendix A** lists the acronyms used in this document.
- **Appendix B** lists the references included in the Plan.
- **Appendix C** provides the Data Management Change Log.

1.1.2 Change Control

The DMP is a living document with revised versions generated during each of the three phases of this deployment project – concept (phase 1), deployment (phase 2), evaluation (phase 3). Additional specificity will be added as the data needs, requirements, architecture, and design mature throughout the system lifecycle. To ensure that the content is traceable from the proposed Phase 1 needs and requirements to the built system, the following rules will be established to ensure the integrity and accuracy of the data. These include the following:

- Identifiers related to datasets will be persistent throughout the three phases; each dataset will be assigned a unique identifier. Identifiers associated with deprecated datasets will not be reused.

- Changes to the dataset profile (e.g., type, stewardship, privacy provision, standard format) and datasets that are eliminated will be reported in a Data Management Change Log (included at the end of this document – starting in Phase 2). The Data Management Change Log is a log that describes the changes to the datasets from phase to phase and is included in Appendix C.

- Due to potential organization changes in format from phase to phase, only interim versions delivered by phase will be version controlled through the project's team site. These changes will be subject to approval by the USDOT.

1.2 Deployment Concept

Buffalo is moving toward a sustainable future at all levels of society, incorporating actions in the community, government, and private entities in the area. Providing access to the City's underserved populations to jobs and healthcare is the primary motivation for all the regional partners involved in this deployment. A lack of public transportation that adequately addresses "first/last mile" challenges is a major problem for community mobility, especially for people with disabilities. This often leads to compromised healthcare (e.g., rescheduled or missed appointments, delayed care) and/or dependence on paratransit service, which is much costlier for transit agencies and can be burdensome for riders. The ITS4US concept proposed here directly addresses these concerns by:

1. **Focusing on providing transit access to healthcare and jobs** to underserved residents or persons and allowing them to share in the economic development in downtown Buffalo.
2. **Putting technology to work in support of accessible transportation**, bringing leading edge researchers in accessible transportation, 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 deployment location is targeted around the downtown Buffalo area with a focus on travel to and from the Buffalo Niagara Medical Campus (BNMC). The deployment includes the 120-acre Medical Campus and surrounding neighborhoods with a focus on three nearby neighborhoods (Fruit Belt, Masten Park, and Allentown)—see Figure 1.

More than 16,000 people work or study at the BNMC and more than 1.5 million visit each year for health care and other services, generating significant transportation demand for the area, its visitors, and its employees. The demographics of the surrounding neighborhoods (see Table 1) are emblematic of a broader socioeconomic and racial divide in Buffalo along Main Street, which this deployment seeks to bridge. In Allentown (west of Main Street), the percentage of traditionally underserved populations is significantly less than other neighborhood east of Main Street, namely Fruit Belt and Masten Park. Table 1 indicates percentages for Allentown that are far below average of the Metropolitan Statistical Area (MSA) in many categories, and percentages for Fruit Belt and Masten Park that are above average for the MSA.

Figure 1. Buffalo Niagara Medical Campus Relative to the Neighborhoods of Focus

Source: Buffalo, NY ITS4US

While the Allentown neighborhood is not characterized by underserved populations, it contains a high concentration of transit service and commercial activity, including health care offices. Allentown hosts several significant bus lines (including the #20-Elmwood, the #25-Delaware, the #11-Colvin, and the #8-Main) that connect the BNMC and Downtown Buffalo with neighborhoods to the north, carrying over 10,500 riders on an average weekday.

Table 1. Demographics of Targeted Neighborhoods

Geography (ACS 2018 tracts)	Percent 0-veh. house-holds	Percent population 65+	Percent poverty	Percent Black	Percent Hispanic / Latino	Percent limited English ability	Percent income <$25k	Percent with a disability (18 to 65 yrs old)	Percent veteran	Percent commute by transit	Total house-holds	Total pop.
Fruit Belt	47.0%	21.9%	28.0%	77.0%	8.9%	4.2%	39.5%	20.0%	6.7%	16.1%	976	2,435
Allentown	18.4%	6.2%	28.8%	7.2%	6.6%	0.0%	17.4%	8.0%	7.8%	4.8%	1978	3,143
Masten Park	35.0%	18.5%	34.7%	89.7%	3.1%	2.9%	38.9%	15.2%	6.6%	11.7%	1496	3,208
Buffalo MSA	**36.6%**	**12.0%**	**31.1%**	**36.6%**	**11.6%**	**4.8%**	**30.7%**	**9.7%**	**5.7%**	**11.5%**	**11,0701**	**255,423**

BNMC sits adjacent to the Fruit Belt neighborhood, which has a poverty rate of 28%, and 47% zero-car households. Several community and social services are found within the neighborhood, which is relatively close to the wider array of services and jobs offered in downtown Buffalo. Several bus lines serve the area, although headways are relatively infrequent, ranging between ½ hour and one hour. Access to dispersed jobs in the suburbs via public transportation tends to be difficult. Although accessible to the Fruit Belt residents, the Niagara Frontier Transportation Authority (NFTA) Metro Rail station is 0.25 – 0.75 miles away, a distance that becomes amplified during the winter and for travelers with physical difficulties. While BNMC continues to improve

pedestrian accessibility, sidewalk quality and intersection crossings still are a challenge for wheelchair users and users with audible or visual impairments. The Fruit Belt struggles with aging infrastructure and infrastructure management issues, issues that have been consistently noted in community forums over the years.

This project seeks to improve transportation access for this population and utilize an innovative approach to support effective trip-making. BNMC's user population makes it an ideal location to test accessibility features for safety and usability. The ITS4US Buffalo project focuses on two primary trip purposes: employee-related travel and patient/visitor travel to the campus from the three adjacent neighborhoods.

1.2.1 System Overview

The Greater Buffalo-Niagara Regional Transportation Council (GBNRTC) established its vision for 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 communities which will 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 (CTP) 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 situational monitoring
- Performance metrics
- Trip history/ledger
- User Interface (UI): Mobile application
- UI: Web

- **Community Shuttle Subsystem** – The Community Shuttle (CS) subsystem provides demand-responsive transit services within a specified zone of operations, using a mix of vehicles, including both human-driven (HDS) 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 HDS 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** – The performance measurement dashboard (PMD) subsystem measures and presents the performance of the system to the agency operating the system.

While not directly part of the project, the CS will be complemented by NFTA Paratransit Access Line (PAL) spontaneous (i.e., same day trip booking and execution) and regular services (i.e., trip reservation done by at least 8pm ET the day before the trip). Other NFTA services, such as bus and rail, will also provide complement to the CS.

The envisioned service area for the proposed CS fleet is shown in Figure 2. The services to be provided within this area are detailed in Figure 3.

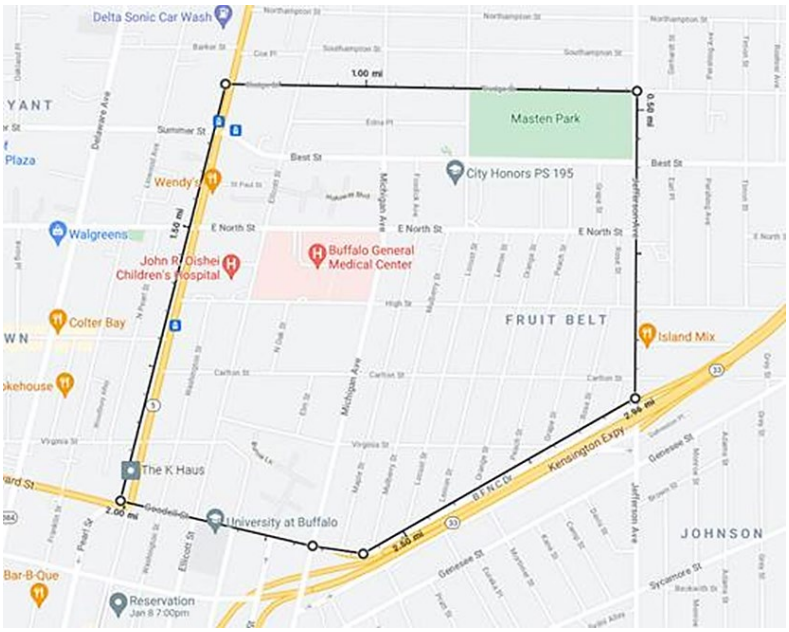

Figure 2. Proposed Service Area for the Community Shuttle
Source: Buffalo, NY ITS4US

	Services	Inside Area	Outside Area
Community Shuttle	HDS		
	SDS		
	PAL (spontaneous)*		
	PAL (regular)**		
	Other NFTA Serv.		

CTP- PAL eligible

CTP - General Traveler

* PAL (spontaneous) refers to same-day service
**PAL (regular) reservation must be done by 8pm night before the trip

Figure 3. Community Shuttle Services
Source: Buffalo, NY ITS4US

1.2.2 System of Interest

This section describes a high-level description of the four subsystems of the Buffalo, NY ITS4US system. The following SOI diagrams are updated versions of the Context Diagrams described in the Concept of Operations (ConOps) (FHWA-JPO-21-860) and System Requirements Specification (SyRS) (FHWA-JPO-21-883).

The high-level context diagram which includes the four subsystems (Smart Infrastructure, CTP, CS, and PMD) are shown in Figure 4.

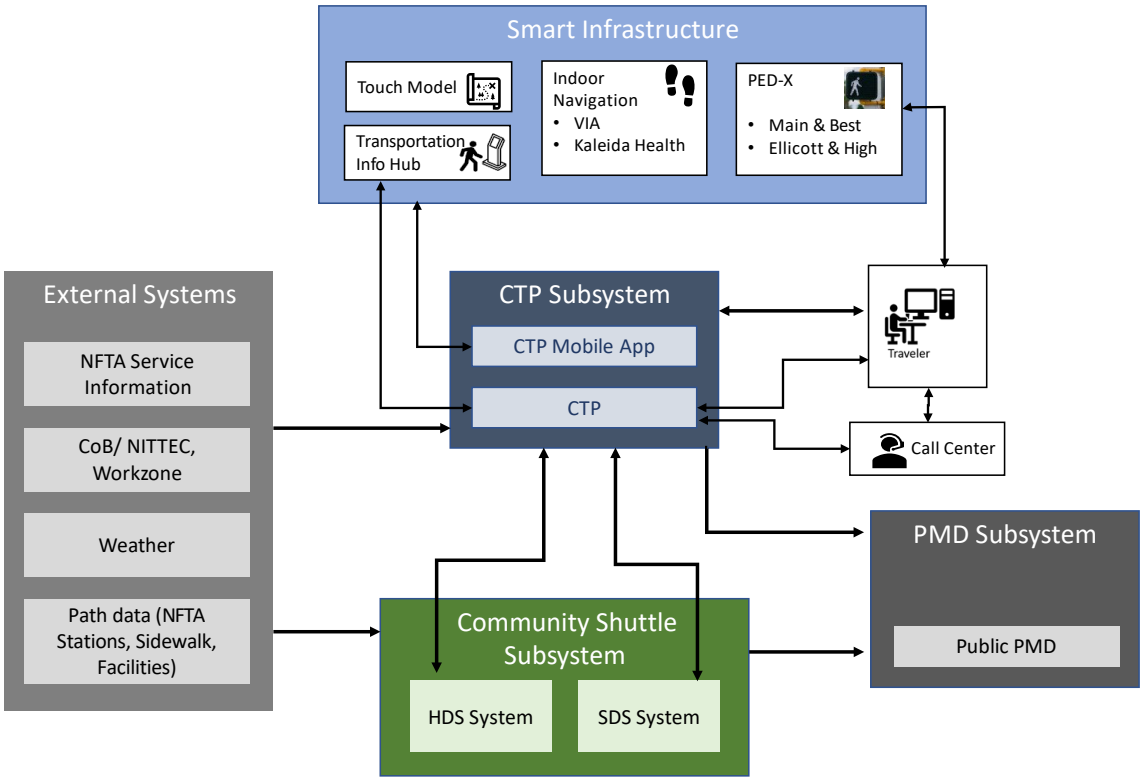

Figure 4. System of Interest: High Level Context Diagram
Source: Buffalo, NY ITS4US

Some of the details in the high-level context diagram were refined from the Phase 1 system of interest to Phase 2. Justification for these changes will be documented in the Phase 2 Concept of Operations, which will be available prior to June 14, 2024. The detailed context diagrams for each subsystem and associated Smart Infrastructure modules are described in the following four sections (each corresponding to one of the four subsystems). Where detail is needed, for example, the CTP, the subsystem functions and flows are well documented. Where subsystems are deployed as a service, pre-existing systems, or turn-key, fewer details are provided (e.g., SDS and HDS).

The following subsections describe the key components of the SOI. The detailed diagrams include information flows within and between subsystems and functions. A complete set of these information flows (tagged in the diagrams as [I-n] where n is a number between 1 and 23), is listed in Table 2.

Table 2. Information Flows in System of Interest

Information Flow #	Information Flow Name	Information Flow Description
I-1	UI Mobile App APIs	APIs and web services provisioned between the CTP mobile app and server
I-2	UI Web	APIs and web services provisioned to a thin client such as the web or TIH

Information Flow #	Information Flow Name	Information Flow Description
I-3	Service information	NFTA fixed route GTFS, GTFS-realtime
I-4	BNMC Facility Map Update	BNMC indoor facility pathways
I-5	NFTA Map Update	NFTA station pathways (GTFS-pathways)
I-6	(1) SDS Booking transactions (2) SDS Service Information	(1) Reservations, scheduling and status information to book a SDS service (2) SDS Service information (GTFS-realtime)
I-7	(1) HDS Booking transactions (2) HDS Service Information	(1) Reservations, scheduling and status information to book a HDS service (2) HDS Service information (GTFS-realtime)
I-8	PED-X request transactions	(1) Information exchange between the CTP mobile app and PED-X gateway to request actuation of the pedestrian crossing (2) Location of PED-X enabled crossing
I-9	BNMC Facility waypoint sensor broadcast	Transactions between mobile app native comm and indoor navigation waypoint sensors (e.g., beacons)
I-10	CTP operational log	Operational and performance data monitored and collected by the CTP
I-11	PED-X operational log	A summary of the requests and their results.
I-12	SDS operational log	Operational and performance information from SDS
I-13	HDS operational log	Operational and performance information from HDS
I-14	Performance measures	Performance Measurement results
I-15	Map services	APIs and web services that present performance measures
I-16	Direct access data files	APIs / links to access public data
I-17	API data (requires authentication)	APIs and web services that require authentication
I-18	External Data (NFTA Performance Data and OSM / path / map updates)	OSM / path / map updates: Map, sidewalk, indoor facility and asset data and updates from external sources NFTA Performance Data: Summary data for fixed route and PAL on time performance and other performance metrics
I-19	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

Information Flow #	Information Flow Name	Information Flow Description
		-- situational awareness information (incidents, work zones, planned events)
I-21	Communications between support personnel	Communications to support travelers on the CS SDS
I-22	Signal Control Exchange	Message from the PED-X gateway to a local traffic signal controller. The information flow forwards a request made by a traveler to request signal actuation
I-23	CoB PROW WZ	Update of work zone information associated with the public right of way developed and disseminated by the City of Buffalo

1.2.2.1 Complete Trip Platform

The 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 (e.g., book a trip reservation, check estimated time of arrival, etc.) 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 wayfinding assets using components specified in the smart infrastructure subsystem. Non-registered travelers will be able to use the trip planning 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.

The context diagram for the CTP is shown in Figure 5, where:

1) the subsystems are shown in blue boxes and boxes with icons;

2) functions are depicted in white boxes contained in the blue boxes;

3) terminators which are source or destinations (internal or external subsystems/systems) of the data are shown in various colored boxes and ovals as designated in the legend; and

4) information flows are shown as either green lines (designated as internal interfaces) or orange lines (designated as external interfaces) tagged with information flow indices [e.g., I-1]).

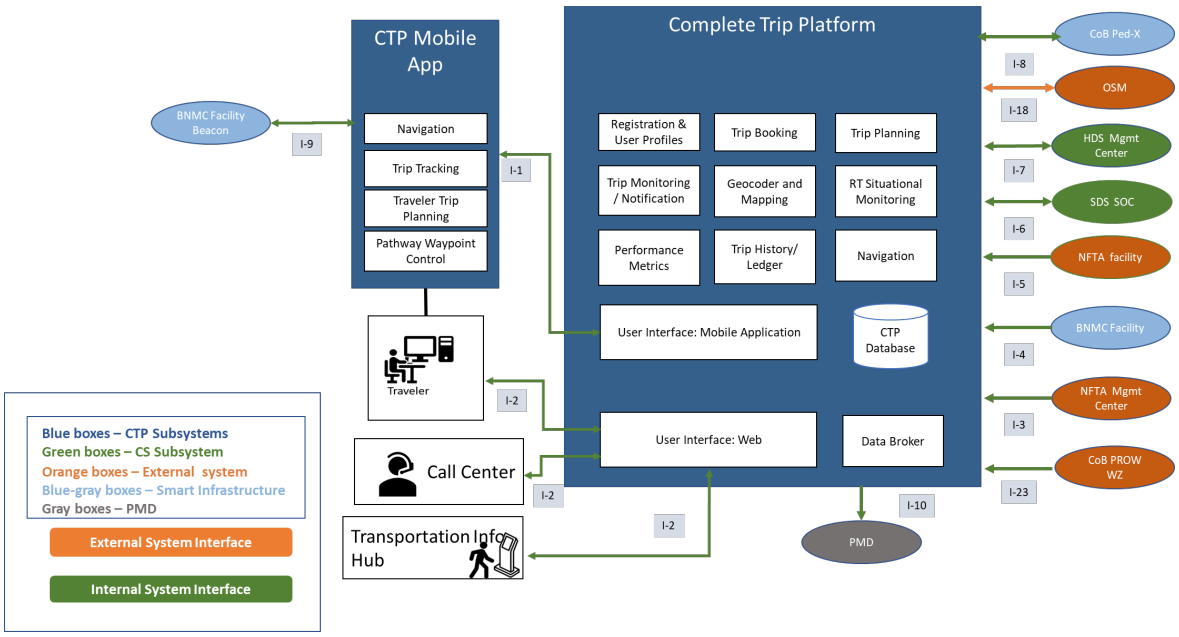

Figure 5. Complete Trip Platform Subsystem
Source: Buffalo, NY ITS4US

1.2.2.2 Community Shuttle Subsystem

The Community Shuttle (CS) subsystem will provide on-demand mobility services within the Fruit Belt, Masten Park and Allentown neighborhoods and BNMC. Although they share similar interfaces to external components/subsystems, the CS consists of two service types – Self Driving and Human Driven services, each designated as a subsystem within this architecture description. The two subsystems include:

- **SDS System** is composed of two autonomous vehicles and the shuttle operations center (SOC). The SDS will provide microtransit service on a pre-defined set of road segments that satisfy the Operational Design Domain (ODD) of the SDS.

- **HDS System** is composed of wheelchair accessible vehicles, dispatch, and scheduling software (from the NFTA paratransit fleet) and reservations services from the NFTA PAL Direct software system. The HDS will provide door-to-door, on-demand service similar to current PAL paratransit service, but available for all registered CTP participants.

1.2.2.2.1 The Self-Driving Shuttle (SDS) System

The SDS system will be procured as a turn-key service exchanging prescribed information flows between the SDS services and Buffalo project subsystems (e.g., PMD, CTP) and external systems (e.g., NITTEC). The SDS is composed of the Autonomous Vehicle and Shuttle Operations Center (SOC) components. The shuttle will operate on an on-demand schedule constrained to travel over a pre-defined route (i.e., a set of streets that satisfy the SDS ODD) and pre-designated pick-up/drop-off locations. The SDS Operations Center (SOC) will receive all calls for services and will track the status of each vehicle in the SDS fleet. The SDS system will be procured as a turn-key service exchanging prescribed information flows between the SDS services and SOI subsystems (e.g., PMD, CTP) and external systems (e.g., NITTEC).

Human actors include the SDS's Shuttle Operations personnel who will manage the SOC, manage incidents, provide Call Center Customer Support (who provide direct support to travelers) and provide SDS operators / stewards (who will be trained by the SDS vendor).

Figure 6 provides a representation of the SDS information flows with other systems and subsystems. Internal system interfaces are indicated by green lines and oval identification numbers and external system interfaces are represented with orange lines and oval identification numbers. Black lines represent voice calls. The CS/SDS System is enclosed within a dark blue box.

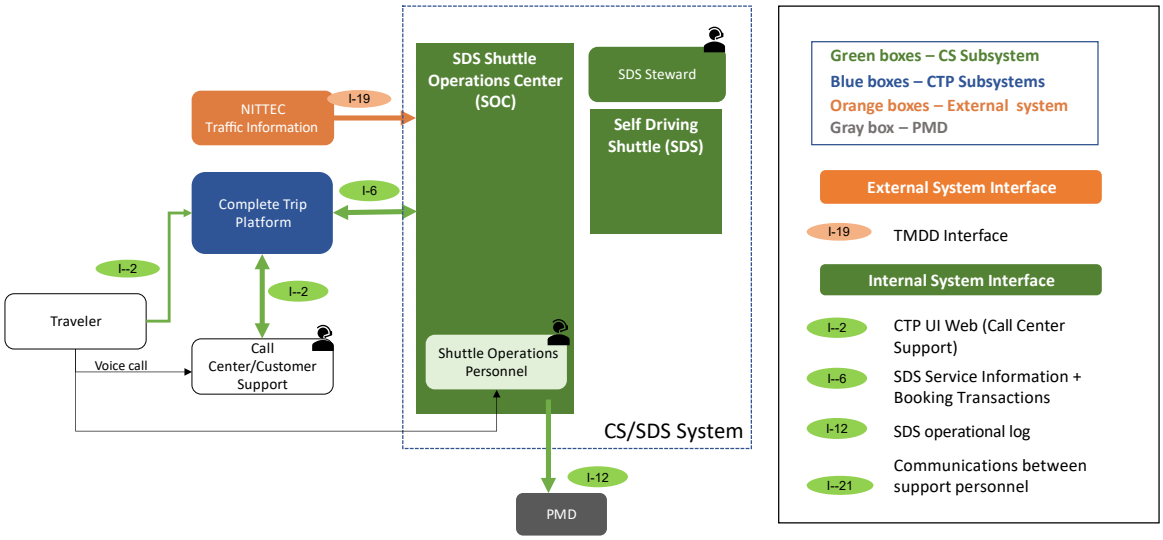

Figure 6. Community Shuttle SDS System
Source: Buffalo, NY ITS4US

1.2.2.2.2 The Human-Driven Shuttle (HDS) System

The HDS System (Figure 7) is composed of the NFTA Dispatch and Reservations System, HDS vehicle, and HDS operator. The HDS will use the NFTA PAL dispatch and software services and operate from the NFTA PAL dispatch and operations center. To that end, the major components are interfaces to and from the NFTA PAL Direct system.

The HDS will use an existing NFTA system that already provides the appropriate services (APIs) needed to transact customer booking and mobility services.

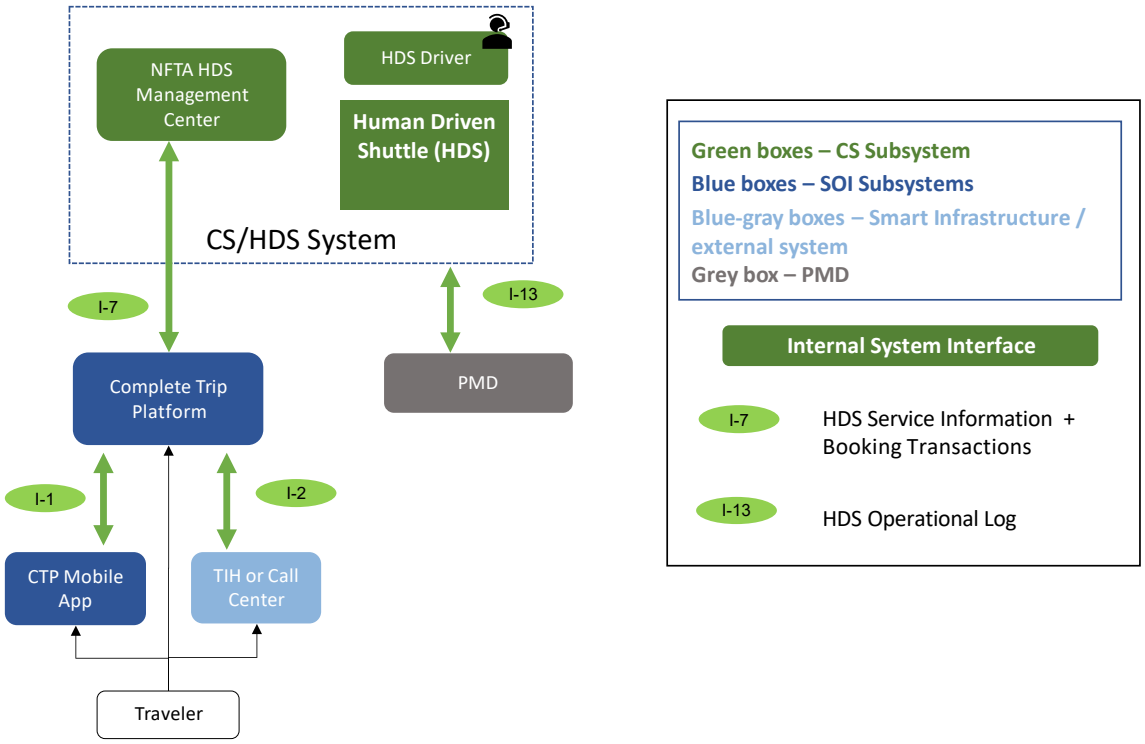

Figure 7. Community Shuttle HDS System Context Diagram
Source: Buffalo, NY ITS4US

1.2.2.3 Smart Infrastructure Subsystem

Smart Infrastructure (SI) supports personalized wayfinding capabilities for travelers. The technologies (modules) include:

- Transportation Information Hub (TIH) modules
- Indoor Navigation modules
- Pedestrian Intersection Crossing (PED-X) modules

SI provides support technologies for trip planning and wayfinding. The SI will be used for public access to trip planning activities, supplementary sensors to support indoor navigation at building / facilities and Metro rail stations, and broker / gateway services for pedestrians to request pedestrian actuation at selected signalized intersections—Main St. & Best St. and Ellicott St. & High St.

The context diagrams for the indoor navigation component and the pedestrian signal crossing component are provided below.

1.2.2.3.1 Indoor Navigation Component

Using commonly available communications technologies already deployed in mobile handsets, low-cost beacons will be deployed at two indoor spaces to provide indoor navigation (Figure 8). These beacons will provide waypoint (location) information for digital wayfinding integrated with the CTP mobile app navigation function.

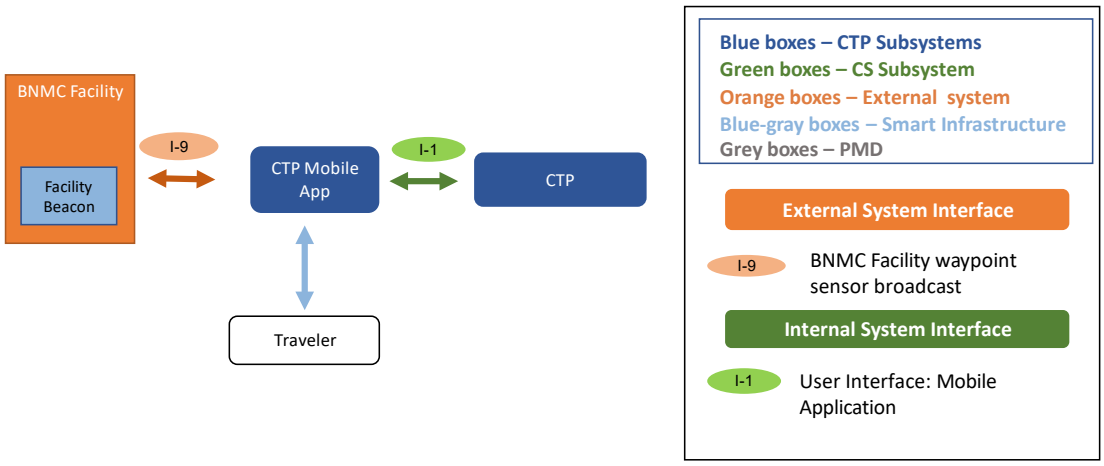

Figure 8. Indoor Navigation Module
Source: Buffalo, NY ITS4US

1.2.2.3.2 Pedestrian Signal Crossing Module

The PED-X module will be implemented as a PED-X Gateway. This gateway receives a CTP generated message triggered by their trip plan. The gateway serves as a conduit to authenticate and secure information exchange between the CTP and traffic signal system to actuate the pedestrian request. Information channeled from the pedestrian to the signal system includes the request information.

Note that the Traffic Signal System, using an audible pedestrian signal will display and annunciate *walk* and *don't walk* signals at the intersection. This link is not shown in the figure because it is not a physical communications message.

The Pedestrian Signal Crossing module is shown in Figure 9.

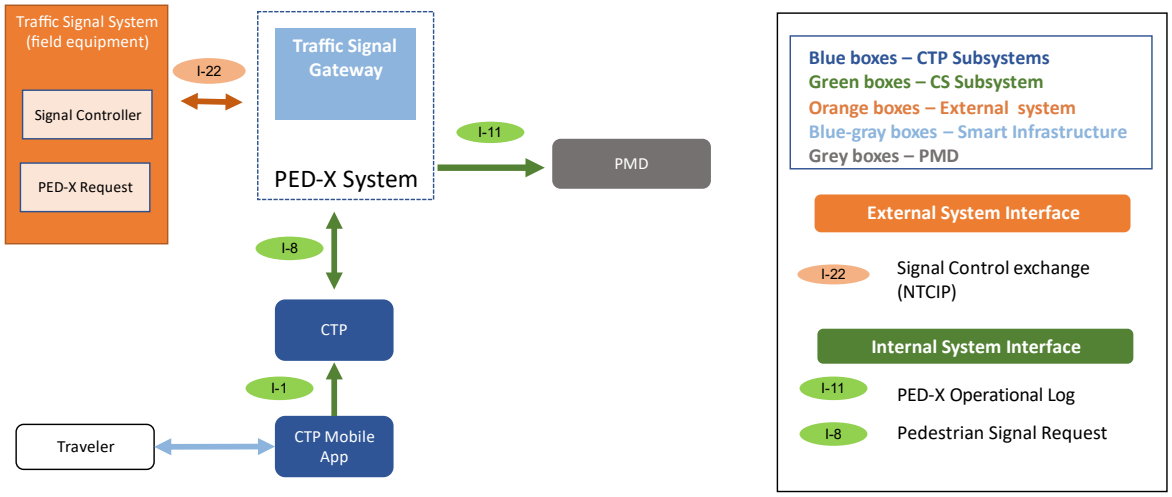

Figure 9. Pedestrian Signal Request (PED-X) Module
Source: Buffalo, NY ITS4US

1.2.2.4 Performance Measure Dashboard Subsystem

The Performance Measure Dashboard subsystem monitors, integrates, analyzes, and displays performance measures from other subsystems and external sources. The subsystem includes 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 PMD subsystem will be implemented using a three-tier architecture – data (data tier), analytic processes (processing tier), and access / visualization (presentation tier), see Figure 10.

Data Tier. The data tier will ingest and store data from the other subsystems and external systems as needed (defined in the Phase 1 PMESP (FHWA-JPO-21-878)[3], Phase 1 Data Management Plan (DMP) (FHWA-JPO-21-868) [1], and subsequent design documents). In addition, metadata management will be included to ensure data integrity as data is ingested and transformed for distribution. The data tier will ingest operational, maintenance, and performance summary data from each subsystem, as well as non-PII data from the performance measurement reports. The data will be ingested by the PMD and all PII will be removed.

Processing Tier. The processing tier will provide services to curate, transform, parse and query data stores to generate performance and aggregated measures. A user authentication function will provide access to users of different security levels. Security and privacy provisions will be implemented to protect, store and archive information.

Presentation Tier. The access and visualization channels will include a web-based dashboard showing key system performance measures as well as a data portal that will provide access via data feeds and APIs for authorized users.

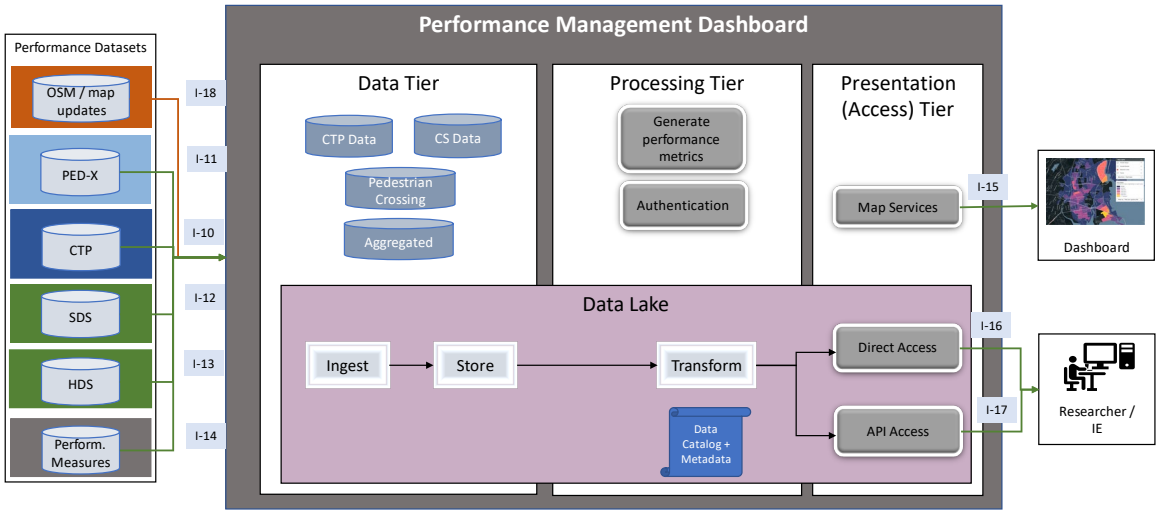

Figure 10. Performance Measure Dashboard
Source: Buffalo, NY ITS4US

1.3 Data Schedule

Limited data collection, restricted to qualitative interviews, was completed during Phase 1 of this project. The expected timing of data milestones for Phases 2 and 3 are listed in Table 3. Consistent with the schedule of the ITS4US program, Phase 2 will extend for a total of 24 months (8 quarters), and Phase 3 for 18 months (6 quarters) following the completion of Phase 2. Phase 2 started in June of this year (i.e., June 2022), and hence the yearly quarters mentioned in Table 3 are referenced to that start date (i.e., June 2022). In other words, Phase 2 Q1, for example, refers to the period extending from July 1, 2022 and ending on September 30, 2022.

Table 3. Expected Data Milestones

ID	Event Title	Description	Date
1.	Institutional Review Board (IRB) Approval	The ITS4US IRB submission was approved by UB's IRB, where it was determined that the project meets the "Exempt" determination.	Dec 8, 2021
2.	Phase 2 - Draft DMP	Initial Draft Phase 2 DMP submitted to USDOT for review.	Nov 1, 2022
3.	Phase 2 - Final DMP	DMP is updated with USDOT comments addressed.	Dec 20, 2022
4.	Initial Data Samples	Initial Data samples (based on the provided metadata) are created, validated and submitted to USDOT for review. Submission to USDOT will take place through a beta-version of the performance measurement dashboard (PMD) which will be developed. The beta version will only be accessed by those with log-in credentials, will show sample data with charts and tables, and include a function to download data in csv or shapefile format. No PII subject to IRB rule will be shared.	Phase 2, Q5
5.	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, Q5
6.	Baseline data collection starts	Initial collection of data on current conditions starts. May leverage existing surveys from local agencies and partners.	Phase 2, Q3
7.	Baseline data provided to USDOT	Complete Baseline data sets are uploaded to USDOT and the IE.	Phase 2, Q6/Q7

ID	Event Title	Description	Date
8.	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.	Periodically throughout Phase 2
9.	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.	Periodically throughout Phase 2
10.	DMP Update #2	DMP updated with any changes coming from the Agile process, testing and sample data schema.	Phase 2, Q7
11.	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.	Periodically throughout Phase 2/3
12.	Draft Final Analysis Report submitted	Draft Final Analysis Report submitted to USDOT. Content to be determined based on USDOT guidance for Phase 2/3.	Phase 3, Q4/Q5
13.	Final Analysis Report submitted	Draft Final Analysis Report submitted to USDOT.	Phase 3, Q6
14.	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.	End of Phase 2, Q8

With respect to item 6 above (i.e., the baseline data collection), it is to be noted that there are two broad classes of baseline data which the project plans to collect. First, the project plans to collect data regarding the condition of the sidewalks at the deployment site (from an accessibility standpoint), as well as information on pathways. These data will be needed to support the navigation and wayfinding functions of the CTP. The other class of baseline data will include traveler surveys documenting the pre- system deployment case, and these would need to be collected at the same time as when the recruitment of the system users starts. Collection of the first class of data could potentially start before the initial samples are made available.

1.4 Data Needs Summary

The data needs are derived from the SOI subsystems, functions and interfaces as described in Section 1.2.2, and more specifically from the context diagrams of the ITS4US subsystems, shown in Figure 5, Figure 6, Figure 7, Figure 8, Figure 9, and Figure 10, which depict the major subsystems, functions, information flows, 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 public datasets from each subsystem will be stored and archived in the PMD subsystem, and the output will feed the USDOT Public Data Store. Personally Identifiable Information (PII) collected by subsystems will remain in the subsystem secure storage and shared only as aggregated, anonymized datasets with the PMD.

Information flows between subsystems provide insight into input and output datasets. The information flows identified in the afore-mentioned figures represent the flows between subsystems, ingested and disseminated to external data sources. The information flows are listed in Table 2. Information flow indices are related to the datasets described in this document.

1.5 Data Overview

The DMP incorporates a diverse group of datasets that are needed to implement and evaluate the system of interest. The datasets are listed in Table 4 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 Context Diagram (see Table 2).
- **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 included 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 and their relationship described in Section 2.6.

    - 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.

    - 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.

    - GIS file format such as Esri shapefile, geodatabase, 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 were identified in the System Requirements document, but detailed data formats will not be determined until the design and development phase.

Table 4. Dataset Overview

Dataset ID	SOI Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
1	I-11	Pedestrian Signal Request Summary	Summary of transactions (message exchanges) between the CTP and PED-X gateway, and PED-X gateway and signal controller The summary data will be contained in a PED-X gateway log file.	Structured; Static	Data collection and forwarding tool	Flat
2	I-12, I-13	Community Shuttle (SDS / HDS) GTFS	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	I-3	NFTA Fixed Route	Schedule and stop information for NTFA light rail and bus services.	Structured; Static	IOO input	Flat
4	I-3	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	SOI Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
5	I-18	Public right of way (PROW) information in the project region	Sidewalk and facility entrances in project region integrated into OpenStreetMap: • 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	Structured; Static	IOO input	Flat
6	I-23	PROW Work Zone Data	Work zone information that describes current work zones on roads, public right of way (PROW) and crosswalks in the project region.	Structured; static	GIS	Shapefile or other GIS format
7	I-6, I-7	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.	Structures, dynamic	Data collection and forwarding tool	gtfs-realtime.proto
8	I-3	NFTA GTFS-realtime	Real time and event data for NFTA fixed route (light rail and bus).	Structured; dynamic	Data collection and forwarding tool	gtfs-realtime.proto

Dataset ID	SOI Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
9	Detailed: I-6, I-7, Summary Data: I-12, I-13	Shuttle Booking Summary	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. Note: detailed data will not be shared and stored in the PMD.	Structured; static	Data collection and forwarding tool	Flat
10	Detailed: I-7, Summary: I-13	PAL Direct Reservations Summary	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. Note: detailed data will not be shared and stored in the PMD.	Structured; static	Data collection and forwarding tool	Flat

Dataset ID	SOI Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
11	I-10	CTP Usage log files	Performance data shared from the CTP subsystem. 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 Agile process that will be developed during Phase 2. In addition, PII will be stripped from the data provided to the PMD 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
12	I-12	SOC Dispatch log summary files	Performance data shared from the SOC subsystems. Types of data include summary trip booking and tracking information. The data is dependent on the performance measurements and design documents that will be developed upon selection of the SDS vendor. In addition, PII will be stripped from the data provided to the PMD subsystem. (Note: trace data will not be included in this dataset).	Semi-structured; static	Data collection and forwarding tool	Schema based text files

Dataset ID	SOI Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
13 deprecated	~~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.~~ ~~Note: this data contains PII that cannot be shared. Summary data is contained in the summary booking data (dataset ID #9).~~	~~Semi-structured; dynamic~~	~~Data collection and forwarding tool~~	~~TBD~~
14	I-12, I-13	CS Vehicle Performance Data	This includes CS (SDS and HDS) performance data such as ridership, on time performance, travel times, etc. To be determined and refined in the Agile process.	Structured; static	Derived	Flat

Dataset ID	SOI Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
15	I-10	Trip Planning Summary Usage Data	Trip Planning summary usage data, includes scheduled and dynamic event data and interactive forms to support generating or reporting on trips/user satisfaction exchanged between the CTP 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 summary 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

Dataset ID	SOI Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
16	I-10	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. This data contains PII and should be secured and / or anonymized before it is shared.	Semi-structured; static	Data collection and forwarding tool	Schema based text files
17	I-9	Beacon Messages	Broadcast message (e.g., MAC Address) by beacon to support orientation and wayfinding signals. The specific format, content and communication methods will be determined during the acquisition stage. Note: other than a public broadcast, this information may be proprietary.	Unknown	IOO input	Flat
18	n/a	Beacon message log file	Log and summary data of smart sign (beacon) operations. Note: there may not be any summary data if beacon is not centrally managed.	Unknown	IOO input	Flat
19 deprecated		~~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~~

Dataset ID	SOI Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
20 deprecated		~~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	I-5	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.	Structured; static	IOO input	GTFS-Pathways
22	I-5	NFTA Conveyance Status	Real time data feed that provides information on elevator/escalator operational status.	Unknown	Data collection and forwarding tool	TBD

Dataset ID	SOI Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
23	I-19	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) Note: this data will not be saved by the NFTA Project. Data archiving is provided by NITTEC.	Structured; dynamic	Data collection and forwarding tool	Schema based text files
24	I-8	Intersection Crossing Assets	Static information on intersections and timing plans for crossing.	Structured; static	IOO input	TBD (will be a GIS file such as: shapefile, GeoJSON or KML)
25	I-18	Beacon asset locations	Location of beacon assets, interfacing requirements and the types of information they provide.	Structured; static	IOO input	TBD (will be a GIS file such as: shapefile, GeoJSON or KML)
~~26~~ Duplicated-removed	~~I-18~~	~~Beacon asset location~~	~~Location of beacon 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)~~

Dataset ID	SOI Flow #	Dataset Title	Description	Type	Collection Method	Data File Format(s)
~~27~~ Deprecate		~~Fare App APIs~~	~~The method and specification for mobile app deep linking.~~	~~Structured; dynamic~~	~~IOO input~~	~~Schema based text files~~
~~28~~ Deprecate		~~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	I-18	OpenStreetMap	Street and PROW network information downloaded from the Open Street Map foundation.	Structured; static	Download	Protocol buffer Binary Format (PBF)
30	I-14	Performance Measures	The performance metrics described in Section 2.6.	Structured; static	Download	Schema based text files
31	I-16, I-17	Researcher-accessible ITS4US Dataset	The data which will be available for access by researchers either directly or through API access. Note: I-15 applies direct access / API accessed datasets to present visualizations.	Structured; static	Download	Schema based text files
32	I-4	BNMC Facility Map	BNMC Facility Map describes the geometry of the BNMC facility, specifically walking pathway information and other waypoints of interest (ex: restrooms, elevators, etc.).	Structured; dynamic	TBD	TBD (will be a GIS such as: shapefile, GeoJSON or KML)

2 Data Stewardship

2.1 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 USDOT Program for any data collected for a USDOT-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 role 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 USDOT-funded research projects or one of the project’s partner organizations.
- **Federal Sponsor:** The USDOT sponsor for the dataset. The federal sponsor assumes the role of Data Owner once the dataset is provided to them per Award requirements later in the project. Note: Any data collected/created for a USDOT-funded research project is required to be provided to USDOT, unless otherwise noted in the award/contract document.

Table 5 lists the datasets and details the owner and steward of each set. Note that deprecated dataset identifiers are not included in subsequent tables. Active datasets include the following: 1-12, 14-18, 21-25 and 29-31. The last column of Table 5 indicates whether the dataset is sponsored by the current ITS4US project or provided from an external source.

Table 5. Data Owner and Steward Information

Dataset ID	Dataset Title	Data Owner	Data Steward	Sponsored by ITS4US Project
1	Pedestrian Signal Request Summary	COB	CTP	Yes
2	Community Shuttle (SDS/HDS) GTFS	NFTA	CS Operator	Yes
3	NFTA Fixed Route	NFTA	CTP	No (External data source)
4	GTFS Flex for PAL	NFTA	CTP	No (External data source)
5	Public right of way (PROW) information in the project region	OSM	CTP	No (External data source)
6	PROW Work Zone Data	COB	CTP	No (External data source)
7	Community Shuttle real time information	CS Operator	CS Operator	Yes
8	NFTA GTFS-realtime	NFTA	NFTA	No (External data source)
9	Shuttle Booking Summary	CS Operator, Traveler	CS Operator, CTP	Yes
10	PAL Direct Reservations Summary	PAL Direct	PAL Direct	No (External data source)
11	CTP Usage log files	CTP	CTP	Yes
12	SOC Dispatch log summary files	CS Operator	CS Operator	Yes
13	DEPRECATED			
14	CS Vehicle Performance Data	CS Operator	CS Operator	Yes

Dataset ID	Dataset Title	Data Owner	Data Steward	Sponsored by ITS4US Project
15	Trip Planning Summary Usage Data	COB, traveler	CTP	Yes
16	Customer Comment Forms	COB, traveler	CTP	Yes
17	Beacon Messages	IOO of asset	Project data steward	Yes
18	Beacon message log file	IOO of asset	IOO of asset	Yes
19	DEPRECATED			
20	DEPRECATED			
21	NFTA Conveyance Data Feed	NFTA	NFTA	Yes
22	NFTA Conveyance Status	NFTA	NFTA	Yes
23	NITTEC Traffic Information	NITTEC	NITTEC	No (External data source)
24	Intersection Crossing Assets	COB	COB	No (External data source)
25	Beacon asset location	IOO of asset	IOO of asset	Yes
26	DEPRECATED			
27	DEPRECATED			
28	DEPRECATED			
29	OpenStreetMap	OSM	CTP	No (External data source)
30	Performance Measures	USDOT	Project Team	Yes
31	Researcher-accessible ITS4US Dataset	USDOT	Project Team	Yes
32	BNMC Facility Map	IOO of asset	IOO of asset	Yes

2.2 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 duration of the project.

2.2.1 Storage Systems

Table 6 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 system is designated by the system that provides the storage. In the case of an external system, the stakeholder who provides the storage is listed. In the case of a project system, the subsystem name is designated as the data storage system name.
- **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. "External" indicates that the quality control, storage and archiving, though available to the project, are governed by external stakeholders.
- **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 6. 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	PMD	External	Based on Agile product roadmap	Each occurrence	Project POP
2	Community Shuttle (SDS/HDS) GTFS	PMD	External	Based on Agile product roadmap and SDS procurement	Quarterly	Project POP

Dataset ID	Dataset Title(s)	Data Storage	System Type	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
3	NFTA Fixed Route	NFTA	External	NFTA is currently producing this data (available from NYSDOT Data Manager)	Quarterly	Project POP
4	GTFS Flex for PAL	NFTA	External	NFTA is currently producing this data	Quarterly	Project POP
5	PROW information in the project region	OSM	External	OSM	As needed	Project POP
6	PROW Work Zone Data	CoB / GIS data storage	External	During construction period (est. Spring ’23)	As needed	Based on organization retention policy
7	Community Shuttle real time information	PMD	Project	Based on Agile product roadmap and SDS procurement	Continuously	Project POP
8	NFTA GTFS realtime	NFTA	External	NFTA is currently producing this data	Continuously	Project POP
9	Shuttle Booking Summary	PMD	Project	TBD during design phase	TBD - will be determined during design phase	Project POP
10	PAL Direct Reservations Summary	NFTA PAL Server	External	Based on Agile product roadmap	TBD - will be determined during delivery	Based on organization retention policy
11	CTP Usage log files	PMD	Project	Based on Agile product roadmap	Daily	Project POP
12	SOC Dispatch log summary files	PMD	Project	Based on Agile product roadmap	Daily	Project POP
13	DEPRECATED					
14	CS Vehicle Performance Data	PMD	Project	Based on Agile product roadmap and SDS procurement	Daily	Project POP

Dataset ID	Dataset Title(s)	Data Storage	System Type	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
15	Trip Planning Summary Usage Data	CTP	Project	Based on Agile product roadmap	Continuously	Project POP
16	Customer Comment Forms	CTP	Project	Based on Agile product roadmap	As needed	Project POP
17	Beacon Messages	Facility owner	External	Based on Agile product roadmap	As needed	Project POP
18	Beacon message log file	Facility owner	External	Based on Agile product roadmap	As needed	Project POP
19	DEPRECATED					
20	DEPRECATED					
21	NFTA Conveyance Data Feed	NFTA	External	Based on Agile product roadmap	As needed	Based on organization retention policy
22	NFTA Conveyance Status	NFTA	External	Based on Agile product roadmap	Continuously	Based on organization retention policy
23	NITTEC Traffic Information	NITTEC	External	Dependent on SDS procurement	Continuously	Based on organization retention policy
24	Intersection Crossing Assets	COB Server	External	Based on PED-X Agile product roadmap	As needed	Project POP
25	Beacon asset locations	CTP	Project	Based on PED-X Agile product roadmap	As needed	Project POP
26	DEPRECATED					
27	DEPRECATED					
28	DEPRECATED					
29	OpenStreetMap	CTP	External	Based on Agile product roadmap	As needed	Project POP
30	Performance Measures	PMD	Project	Based on system rollout	On-Demand	Project POP
31	Researcher-accessible ITS4US Dataset	PMD	Project	Based on system rollout	On-demand	Project POP

Dataset ID	Dataset Title(s)	Data Storage	System Type	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
32	BNMC Facility Map	Facility owner	External	Based on Agile product roadmap	As needed	Project POP

2.2.2 Data Storage System Description

The storage systems will be designed, developed, and tested over the remainder of Phase 2 of this project, and will include: (1) the CTP Storage system; (2) the CS Storage system; (3) the PMD Storage system; and (4) the system for storing the survey data which will be collected for the evaluation of the Buffalo ITS4US system, (may contain PII, static, and configuration datasets). Datasets will be assigned to subsystem level servers, for example CTP Storage, CS Storage, etc. 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.

2.3 Data Sharing Framework

The data sharing framework will be implemented through the PMD subsystem (see section 1.2.2.4 above and Figure 10). As described above, the PMD will be comprised of three tiers: (1) the data tier, which will ingest and store the data from the Buffalo ITS4US subsystems and external sources; (2) the processing tier; and (3) the presentation tier. Documentation for each dataset will be provided through the processing tier which will provide for metadata management capabilities to ensure data integrity. The processing tier will include a user authentication function, which will provide access to users of different security levels. Security and privacy provisions will also be implemented in the processing tier to protect, store and archive information.

Four access levels will be provided as described in more detail in the accompanying Data Privacy Plan (DPP) for the Buffalo ITS4US project. These levels are: (1) open; (2) private - research; and (3) private – operational; and (4) private - protected.

Open datasets will be made available to the public, without need to request permission. They will also be provided to the USDOT-managed public system. Open datasets may be anonymized or aggregated versions of the private datasets to protect any PII. Private datasets, on the other hand, will require users to meet specific requirements and secure appropriate permissions, since they may have PII. There are three different types (and access

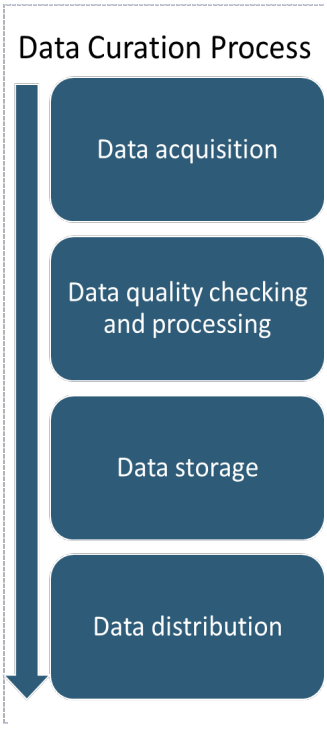

Figure 11. Data Curation Process
Source: Buffalo, NY ITS4US

levels) of private datasets. The first is research data, which will be made available, for research purposes, to users that meet Institutional Review Board (IRB) requirements. The second is operational data (e.g., from the SDS and SOC), which will be available for access by the data stewards for each dataset for operational purposes only. In that regard, it is to be noted that some data in this category may be proprietary licensed by a third party (e.g., the SDS vendor). Finally, protected data are datasets that have PII included. Access to this data will be restricted to protect the PII based on IRB-approved processes. More details about the specific datasets that are open, and those that are private (i.e., research, operational and protected) can be found in Section 2.2 of the accompanying DPP.

Access to the above-mentioned datasets will be provided via the presentation tier of the PMD, and specifically through password protected APIs.

2.4 Data Quality Control

The System Requirements developed for the Buffalo ITS4US project in Phase I include a specific data curation requirement, which covers the four steps of the data curation process shown in Figure 11, for each dataset that is internal to the system. Specifically, the following requirements were included:

Req-Sy-026 - Dataset Curation Plan which reads, **"**System developers shall develop a curation process plan for each dataset that is consistent with the DMP and PMESP. The dataset includes data collection, ingestion, validation/quality checking, ETL [extract, transform, load] processes, storage, distribution methods, metadata documentation. The frequency of the curation process will be determined in Phase 2."

Req-Sy-026.1 - Implement Dataset Curation, which reads, "Data stewards shall implement the dataset curation plan including quality control procedures."

Data curation processes will be designed to ensure the following:

- Accuracy of the data described.
- Relevancy to ensure that the data meets the requirements for the intended use.
- Completeness which refers to the fact that the data should not have missing values or records.
- Timeliness to ensure that the data are up to date.
- Consistency, designed to guarantee that the data are in the expected format and consistent across all subsystems.

While the quality of the datasets internal to the Buffalo ITS4US system will be controlled through the above-mentioned data curation system requirements, ensuring the quality of external datasets and external information flows will require specifying this within the procurement policy and/or contract. The responsibility of ensuring the data quality of a given dataset will be the prime responsibility of the data steward for that particular set. Table 7 lists the different datasets, along with their data steward, and the overall approach (or method) to the data curation process (e.g., System Requirements, Procurement Policy).

Table 7. Data Curation Approaches for the Buffalo ITS4US Datasets

Dataset ID	Dataset Title	Data Steward	Data Curation Approach
1	Pedestrian Signal Request Summary	PED-X Gateway	System Requirements implemented by PED-X Gateway
2	Community Shuttle (SDS/HDS) GTFS	CS Operator	Procurement Policy for SDS NFTA Data Curation Policies and Procedures for HDS
3	NFTA Fixed Route	CTP	NFTA Data Curation Policies and Procedures
4	GTFS Flex for PAL	CTP	NFTA Data Curation Policies and Procedures
5	PROW information in the project region	UB data collection	Processes to be jointly determined by UB data collection and CTP Vendor
6	PROW Work Zone Data	CoB	Processes to be jointly determined by CoB and CTP Vendor
7	Community Shuttle real time information	CS Operator	Procurement Policy for SDS NFTA Data Curation Policies and Procedures for HDS
8	NFTA GTFS realtime	NFTA	NFTA Data Curation Policies and Procedures
9	Shuttle Booking Summary	CS Operator	Procurement Policy for the SDS and SOC NFTA Data Curation Policies and Procedures
10	PAL Direct Reservations Summary	PAL Direct	NFTA Data Curation Policies and Procedures
11	CTP Usage log files	CTP	System Requirements implemented by CTP vendor for the CTP
12	SOC Dispatch log summary files	CS Operator	Procurement Policy
13	DEPRECATED		

Dataset ID	Dataset Title	Data Steward	Data Curation Approach
14	CS Vehicle Performance Data	CS Operator	Procurement Policy for SDS; NFTA Data Curation Policies and Procedures for HDS
15	Trip Planning Summary Usage Data	CTP	System Requirements implemented by CTP vendor for the CTP
16	Customer Comment Forms	CTP	System Requirements implemented by CTP vendor for the CTP in collaboration with Performance Management evaluators
17	Beacon Messages	Indoor Navigation System	System requirements implemented by the indoor navigation vendor
18	Beacon message log file	Indoor Navigation System	System requirements implemented by the indoor navigation vendor
19	DEPRECATED		
20	DEPRECATED		
21	NFTA Conveyance Data Feed	NFTA	System Requirements implemented by CTP vendor in collaboration with NFTA
22	NFTA Conveyance Status	NFTA	System Requirements implemented by CTP vendor in collaboration with NFTA
23	NITTEC Traffic Information	NITTEC	External data source – NITTEC's data curation policies and procedures
24	Intersection Crossing Asset Locations	CoB	External data source – CoB's data curation policies and procedures
25	Beacon asset locations	IOO of asset	Data curation policies and procedures of the IOO of the Asset
26	DEPRECATED		
27	DEPRECATED		
28	DEPRECATED		

Dataset ID	Dataset Title	Data Steward	Data Curation Approach
29	OpenStreetMap	OSM	External data source – Data curation policies and procedures implemented by the OpenStreetMap community
30	Performance Measures	Project Team	Data curation policies and policies implemented by the Buffalo ITS4US project team
31	Researcher-accessible ITS4US Dataset	Project Team	Data curation policies and policies implemented by the Buffalo ITS4US project team
32	BNMC Facility Map	IOO of asset	Data curation policies and procedures of the IOO of the Asset

2.5 Privacy

While Section 2.3 above has described the four different access levels which will be provided, more details are can be found in the project's DPP. In general, survey data that contains PII will be stored on secure servers at the University at Buffalo, in accordance with the IRB requirements and the approved, submitted protocol, and only when there are operational purposes for such data. Private datasets in the CTP will be encrypted, and only summarized / anonymized access will be provided to researchers via a secure interface to the PMD. Only researchers who have taken the proper training mandated by IRB will be able to access the summary datasets.

2.6 Relationship to Performance Measures

As described in more detail in the PMESP document of Phase I, a comprehensive performance measurement and evaluation framework has been developed for the Buffalo ITS4US project. The framework combines data collected from pre- and post-deployment surveys of the system users with data generated from the Buffalo ITS4US project itself and its different subsystems. Basically, six high-level performance measures were defined for the system. These are listed here:

1. Improved ability of the CTP users to make satisfactory Complete Trips in the study area
2. Usefulness of the CTP registration and trip preferences process
3. Usefulness of the CTP trip planning and booking process
4. Improved ability to find destinations efficiently using the CTP wayfinding functionality
5. Improved ability to cross specific intersections safely using CTP smart-signal functionality
6. Provision of an efficient, reliable and safe new on-demand transit shuttle system

Based on these high-level six performance measures, metrics were defined to allow for evaluating each measure. Table 8 lists the performance measures developed for evaluating the Buffalo ITS4US system, where the first column gives an ID for each performance metric, and the second column provides the title for the metric and the third column provides a brief description of the metric. Note that some metrics are divided into sub-metrics (i.e., A, B, C) depending on the information collected from the system.

Table 8. Performance Metrics Developed for the Buffalo ITS4US Project.

PM #	Metric	Data Description
1	**Ability to Make Complete Trips in the Study Area**	
1 (1)	System user ratings of how accessible door-to-door travel is for trips to, from and within the BNMC	Ratings from the Baseline and Periodic Post-Deployment Surveys, on 5-point or 5-star scale
1 (2)	System user ratings of how safe door-to-door travel paths are for trips to, from and within the BNMC, including level, slip-resistant paths	Ratings from the Baseline and Periodic Post-Deployment Surveys, on 5-point or 5-star scale
1 (3)	System user ratings of the adequacy and usefulness of information for making trips to, from and within the BNMC	Ratings from the Baseline and Periodic Post-Deployment Surveys, on 5-point or 5-star scale
1 (4)	System user ratings of the ability to make trips using integrated transit services. to, from and within the BNMC	Ratings from the Baseline and Periodic Post-Deployment Surveys, on 5-point or 5-star scale
1 (5)	System user ratings of the accessibility of the trip using the CTP	Aggregate daily responses across CTP users from a rating question in the CTP app at the end of each actual trip, on a 5-star scale
2	**Usefulness of the CTP Registration and Trip Preferences Process**	
2 (1)	System user ratings of the ease of the registration process	Aggregate daily responses across CTP users from a rating question in the CTP app at the end of the registration process about ease of use, on a 5-star scale
2 (2A)	System user ratings of the usefulness of providing preferences to get trip options that satisfy those preferences	Aggregate daily responses across CTP users from a rating question in the CTP app at the end of the registration process about perceived relevance of preference questions, on a 5-star scale
2 (2B)	System user ratings of the usefulness of providing preferences to get trip options that satisfy those preferences	Ratings from the Periodic Post-Deployment Surveys about getting trip options that satisfied preferences, on 5-point or 5-star scale
3	**Usefulness of the CTP Trip Planning and Booking Process**	
3 (1)	System user ratings of the ease of planning a door-to-door trip route/path	Ratings from the Periodic Post-Deployment Surveys about ease of planning trips, on 5-point or 5-star scale
3 (2)	System user ratings of the satisfaction with the specific route/path options provided by the CTP	Aggregate daily responses across CTP users from a rating question in the CTP app at the end of the planning a trip, on a 5-star scale
3 (3)	The fraction of trips that are planned using the CTP that are subsequently carried out using the CTP app	Aggregate daily data across CTP users of (a) number of new trips planned, (b) number of planned trips that were carried out

PM #	Metric	Data Description
3 (4)	The percent of CTP users who use the system to book on-demand transit trips (HDS or SDS)	Aggregate daily data across CTP users of number of new trips booked (a) on the HDS and (b) on the SDS
3 (5A)	System user ratings of the ease and convenience of booking on-demand transit trips via the system	Aggregate daily responses across CTP users from a rating question in the CTP app at the end of the CS booking process about satisfaction with the process, on a 5-star scale
3 (5B)	System user ratings of the ease and convenience of booking on-demand transit trips via the system	Ratings from the Periodic Post-Deployment Surveys about ease of booking CS trips, on 5-point or 5-star scale
3 (6A)	The percent of CTP users who use the CTP to report incidents or travel conditions during their trips	Aggregate daily data across CTP users of the number of incidents/conditions/resolutions reported, by issue code (TBD)
3 (6B)	The percent of CTP users who use the CTP to report incidents or travel conditions during their trips	Answers from the Periodic Post-Deployment Surveys about frequency of using reporting function
3 (7)	System user ratings of the ease of reporting incidents or conditions encountered during a trip in the CTP	Ratings from the Periodic Post-Deployment Surveys about ease of reporting incidents, etc. (if user has used the capability), on 5-point or 5-star scale
3 (8A)	System user ratings of the usefulness of reviewing past trip history in the CTP	Answers from the Periodic Post-Deployment Surveys about frequency of reviewing trip history
3 (8B)	System user ratings of the usefulness of reviewing past trip history in the CTP	Ratings from the Periodic Post-Deployment Surveys about usefulness of reviewing trip history (if user has used the capability), on 5-point or 5-star scale
4	**Improved ability to find destinations efficiently using the CTP wayfinding functionality**	
4 (1A)	The fraction of CTP users who elect to receive **outdoor** wayfinding notifications	Aggregate daily data across CTP users of the new users who prefer to receive outdoor wayfinding notifications (user-level data)
4 (1B)	The fraction of CTP users who elect to receive **outdoor** wayfinding notifications	Aggregate daily data across CTP users of the number of trips for which outdoor wayfinding notifications were sent (trip-level data) **and the total distance of the wayfinding paths**
4 (1C)	The fraction of CTP users who elect to receive **outdoor** wayfinding notifications	Aggregate daily data of the availability of **outdoor** wayfinding in terms of the percentage of time it was available and the percent of relevant trips for which it could be provided
4 (2)	System user self-reported frequency of using **outdoor** wayfinding notifications	Answers from the Periodic Post-Deployment Surveys about frequency of using **outdoor** navigation
4 (3A)	The fraction of CTP users who elect to receive **indoor** wayfinding notifications	Aggregate daily data across CTP users of the new users who prefer to receive indoor wayfinding notifications (user-level data)

PM #	Metric	Data Description
4 (3B)	The fraction of CTP users who elect to receive **indoor** wayfinding notifications	Aggregate daily data across CTP users of the number of trips for which indoor wayfinding notifications were sent (trip-level data) **and the total distance of the wayfinding paths**
4 (3C)	The fraction of CTP users who elect to receive **indoor** wayfinding notifications	Aggregate daily data of the availability of **indoor** wayfinding in terms of the percentage of time it was available and the percent of relevant trips for which it could be provided
4 (4)	System user self-reported frequency of using **indoor** wayfinding notifications	Answers from the Periodic Post-Deployment Surveys about frequency of using **indoor** navigation
4 (5A)	System user ratings of the how useful the **outdoor** wayfinding functionality is in reaching their trip destination on time	Aggregate daily responses across CTP users from a rating question in the CTP app at the end of making a trip for which notifications were sent, about satisfaction with the outdoor navigation, on a 5-star scale
4 (5B)	System user ratings of the how useful the **outdoor** wayfinding functionality is in reaching their trip destination on time	Ratings from the Periodic Post-Deployment Surveys (if user has used the capability), on 5-point or 5-star scale
4 (6A)	System user ratings of the how useful the **indoor** wayfinding functionality is in reaching their trip destination on time	Aggregate daily responses across CTP users from a rating question in the CTP app at the end of making a trip for which notifications were sent, about satisfaction with the indoor navigation, on a 5-star scale
4 (6B)	System user ratings of the how useful the **indoor** wayfinding functionality is in reaching their trip destination on time	Ratings from the Periodic Post-Deployment Surveys (if user has used the capability), on 5-point or 5-star scale
4 (7)	User ratings of various dimensions of using the CTP **outdoor** wayfinding functionality using the Rapid Assessment of Product Usability and Universal Design (RAPUUD) method	Ratings from the Periodic Post-Deployment Surveys (if user has used the capability), on 5-point or 5-star scale
4 (8)	User ratings of various dimensions of using the CTP **indoor** wayfinding functionality using the RAPUUD method	Ratings from the Periodic Post-Deployment Surveys (if user has used the capability), on 5-point or 5-star scale
5	**Improved ability to cross specific intersections safely using the CTP smart-signal functionality**	
5 (1)	The percent of CTP trips crossing at the relevant intersections who use the smart signal remote activation function.	Aggregate daily data across CTP users of (a) the number of trips that traverse each of the smart signal intersections and (b) the number of times the smart signals are activated from the CTP
5 (2)	Self-reported fraction of people who cross at the relevant intersections who use the CTP smart signal activation functionality	Answers from the Periodic Post-Deployment Surveys about frequency of crossing intersections and activating smart signals
5 (3)	Perceived ease of use and ratings of various aspects of using the smart signals, using the RAPUUD method	Ratings from the Periodic Post-Deployment Surveys (if user has used the capability), on 5-point or 5-star scale

PM #	Metric	Data Description
5 (4)	Perceived safety of crossing the intersections with smart signals	Ratings from the Baseline and Periodic Post-Deployment Surveys, on 5-point or 5-star scale
6	**The community shuttle is an efficient, reliable, safe mode of transportation**	
6 (1)	Percent of CS trips that arrive at the boarding stop within the targeted time allowance **of the scheduled arrival time**	Aggregate daily data across CTP users of the number of trips of trips that for which the differences between the scheduled and arrival times fall in different ranges (30+ min early, 20-30 min early,… etc., 30+ min late)
6 (2)	Percent of CS trips that arrive at the **alighting** stop within the targeted time allowance of the scheduled arrival time	Aggregate daily data across CTP users of the number of trips of trips that for which the differences between the scheduled and arrival times fall in different ranges (30+ min early, 20-30 min early,… etc., 30+ min late)
6 (3)	System user ratings of the how reliable the transit system is in reaching their BNMC trip destination on time	Ratings from the Baseline and Periodic Post-Deployment Surveys, on 5-point or 5-star scale
6 (4)	Self-reported frequency of using all transit services, including the CS as well as PAL services and NFTA bus and rail lines	Answers from the Baseline and Periodic Post-Deployment Surveys about frequency of using various transit services
6 (5)	Percent of transit trips made by PAL-eligible CTP users within the CS service area that are made via the HDS and SDS versus regular PAL trips	Aggregate daily data across CTP users of number of transit trips made (a) on the HDS and (b) on the SDS. Segmentation by PAL-eligibility status is necessary.
6 (6)	User ratings of CS service in terms of key service aspects – convenience, affordability and safety, as well as other aspects included in the modified RAPUUD questions	Ratings from the Periodic Post-Deployment Surveys separately for the HDS and SDS (if user has used each one), on 5-point or 5-star scale
6 (7)	Cost efficiency of the HDS and SDS shuttle services in terms of operating cost per passenger trip	Aggregate data from the CS Operations System on the actual number of boardings on the HDS and SDS during several periods of the day on each day.
6 (8)	Percent of CS bookings for which the earliest available pick-up time is within 45 minutes of the request time	Aggregate data from the CS Reservations System on the number of bookings within various ranges between the time request time and the earliest available pick up time (not always the same as the booked pick-up time)
6(9)	Objective measures of the reliability and safety of the Autonomous Driving component of the SDS	Aggregate SDS data from the CS Operations System such as number of disengagements, reasons behind disengagement, any hard or emergency braking events, etc.

With the performance metrics defined above, Table 9 below identifies the Buffalo ITS4US project datasets that will serve as a source of data from which the performance metrics, defined in Table 8, will be derived. Specifically, for each dataset, the specific performance measures which may utilize data within that specific dataset are listed. The performance metric ID is consistent with the metric ID listed in Table 8. No correlation is provided for those datasets that do not directly support performance metrics.

Table 9. The Buffalo ITS4US Project Datasets and the Performance Metrics they Support.

Dataset ID	Dataset Title	Performance Metrics Supported
1	Pedestrian Signal Request Summary	5(1)
2	Community Shuttle (SDS/HDS) GTFS	6(1), 6(2)
3	NFTA Fixed Route	6(1), 6(2), 6(7)
4	GTFS Flex for PAL	6(1), 6(2), 6(7)
5	PROW 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	
6	PROW Work Zone Data	
7	Community Shuttle real time information	6(1), 6(2), 6(8)
8	NFTA GTFS realtime	6(1), 6(2), 6(8)
9	Shuttle Booking Summary and Details	6(1), 6(2), 6(5), 6(7), 6(8)
10	PAL Direct Reservations Summary	6(5)
11	CTP Usage log files	3(3), 3(4), 3(5), 3(6), 3(7), 4(1), 4(3), 4(5), 4(6). 4(7), 4(8). 5(1)
12	SOC Dispatch log summary files	6(1), 6(2), 6(5), 6(7), 6(8)
13 deprecated	CS Service Vehicle Manifest	deprecated
14	CS Vehicle Performance Data	6(7), 6(9)
15	Trip Planning Summary Usage Data	1(1), 1(2), 1(3), 1(4), 5(4), 6(3), 6(4), 2(1), 2(2), 3(1), 3(2), 3(8), 4(2), 4(4), 5(2), 3(5), 3(7), 4(5), 4(6), 4(7), 4(8), 6(6), 3(3), 3(4), 4(1), 4(3), 5(1)
16	Customer Comment Forms	3(7)
17	Beacon Messages	
18	Beacon message log file	

Dataset ID	Dataset Title	Performance Metrics Supported
19 deprecated	HDS Canned and custom operations messages	deprecated
20 deprecated	SDS custodian Canned operations messages	deprecated
21	NFTA Conveyance Data Feed	
22	NFTA Conveyance Status	
23	NITTEC Traffic Information	
24	Intersection Crossing Assets	
25	Beacon asset locations	
26 deprecated	Facility smart sign (beacon) assets	deprecated
27 Deprecate	Fare App APIs	deprecated
28 Deprecate	Emergency Safety Messages	deprecated
29	OpenStreetMap	
30	Performance Measures	All Performance Measures, with some coming directly from performance-related surveys
31	Researcher-accessible ITS4US Dataset	All Performance Measures
32	BNMC Facility Map	

2.6.1 Baseline Data

The Buffalo ITS4US project will collect baseline data against which the post-deployment data will be compared to assess the effectiveness of the deployment. Table 10 describes the baseline data to be collected, as well the performance metric each baseline dataset will support. The Buffalo ITS4US project team sees that the best time to collect the baseline survey data would be at the same time as when the recruitment effort starts. For recruitment, the best time for that is when users can actually register to use the system, or a prototype of the system. At a minimum that would be the time when a demo version of the CTP app is available, and at least one of the system components is ready for testing (e.g., the HDS or the navigation functions of the CTP app).

Preferably, the baseline questions should be asked to the same groups of individuals who are potential system users (in fact, ideally baseline questions should be asked of exactly the same people who end up using the system and answering the post-deployment questions about the system). The Buffalo ITS4US project team plans to provide USDOT with an initial baseline data set in phase 2, and then to continue updating the baseline as more people are recruited early in phase 3.

So, the timing should be when it fits best in with recruitment and the rollout/testing of any system components in phase 2. We can give them an initial baseline data set in phase 2, and then we will continue updating the baseline as more people are recruited early in phase 3.

Table 10. Buffalo ITS4US Project Baseline Data to be Collected

#	Metric	Data Description
1 (1)	System user ratings of how accessible door-to-door travel is for trips to, from and within the BNMC.	Ratings from the Baseline on 5-point or 5-star scale
1 (2)	System user ratings of how safe door-to-door travel paths are for trips to, from and within the BNMC, including level, slip-resistant paths.	Ratings from the Baseline on 5-point or 5-star scale
1 (3)	System user ratings of the adequacy and usefulness of information for making trips to, from and within the BNMC.	Ratings from the Baseline on 5-point or 5-star scale
1 (4)	System user ratings of the ability to make trips using integrated transit services. to, from and within the BNMC.	Ratings from the Baseline on 5-point or 5-star scale
5 (4)	Perceived safety of crossing the intersections with smart signals.	Ratings from the Baseline on 5-point or 5-star scale
6 (3)	System user ratings of the how reliable the transit system is in reaching their BNMC trip destination on time.	Ratings from the Baseline on 5-point or 5-star scale
6 (4)	Self-reported frequency of using all transit services, including the CS as well as PAL services and NFTA bus and rail lines.	Answers from the Baseline about frequency of using various transit services

3 Data Standards

3.1 Data Standards

Table 11 lists the data standard(s) used for the datasets that will be ingested or disseminated using an existing standard. Many datasets will be delivered as flat, semi-structured or structured files with definitions included in their data curation plans. For datasets that will transferred using data standards, the table provides:

- **Dataset ID and Title:** Identifier and title of the applicable dataset.
- **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, a Uniform Resource Locator (URL) to the data standard(s) is provided. If one does not exist, “n/a” is documented in this field.
- **Open or Proprietary:** Indicates whether the data standard(s) is/are “Open” or “Proprietary.”
- **Data Standard(s) Rationale**: Explains the use of the chosen data standard(s).

Table 11. Data Standards

Data set ID	Dataset Title	Data Standard(s)	Data Standard DOI	Open / Proprietary ?	Data Standard(s) Rationale
1	Pedestrian Signal Request Summary	TBD	TBD	Open	MioVision Gateway data
2	Community Shuttle (SDS/HDS) GTFS	GTFS Flex	GTFS Flex	Open	Industry standard to implement flexible transit information.
3	NFTA Fixed Route	GTFS	GTFS	Open	Industry standard to implement static transit information.
4	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 information in the project region	Protocol buffer Binary Format	OSM	Open	OSM defines a standard set of features and attributes for the PROW networks. The network is used in the Open Trip Planner, which is the foundation of the CTP subsystem.
6	PROW Work Zone Data	Will adopt related standard for sidewalk WZ features	TBD	TBD	No known standard that supports work zones on sidewalks and other public right of ways.
7	Community Shuttle real time information	GTFS realtime	GTFS realtime	Open	Industry standard to provide real time transit location, event, and prediction information.
8	NFTA GTFS realtime	GTFS realtime	GTFS realtime	Open	Industry standard to provide real time transit location, event, and prediction information.
17	Beacon Messages	TBD	TBD	Open	May be link layer MAC address or other broadcast supported by Bluetooth.
21	NFTA Conveyance Data Feed	GTFS-Pathways	GTFS Pathways	Open	Open standard to describe pathways through transit stations.
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.

Data set ID	Dataset Title	Data Standard(s)	Data Standard DOI	Open / Proprietary ?	Data Standard(s) Rationale
29	OpenStreet Map	Protocol buffer 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.
30	Performanc e Measures	May develop standard set of performance measures	TBD	Open	No industry standard yet.
31	Researcher-accessible ITS4US Dataset	A diverse set of standards may apply here	TBD	Open	The datasets which will be made accessible will be diverse. While a subset of these datasets may have established standards, others may not.
32	BNMC Facility Map	TBD	TBD	TBD (may be proprietary)	TBD

3.2 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' file 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 2023."

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.

3.3 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 3.3.1 describes the types of information that are documented in the metadata files, Section 3.3.2 describes the file structure(s), and Section 3.3.3 provides information on the metadata update process.

3.3.1 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 the dataset including privacy and security levels associated with the 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 datasets.

- **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)

3.3.2 Metadata Structure

In this section, the 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

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

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.

- **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 *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.

3.3.3 Metadata Update Process

The metadata update process will occur throughout the dataset lifecycle. In Figure 12, 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.

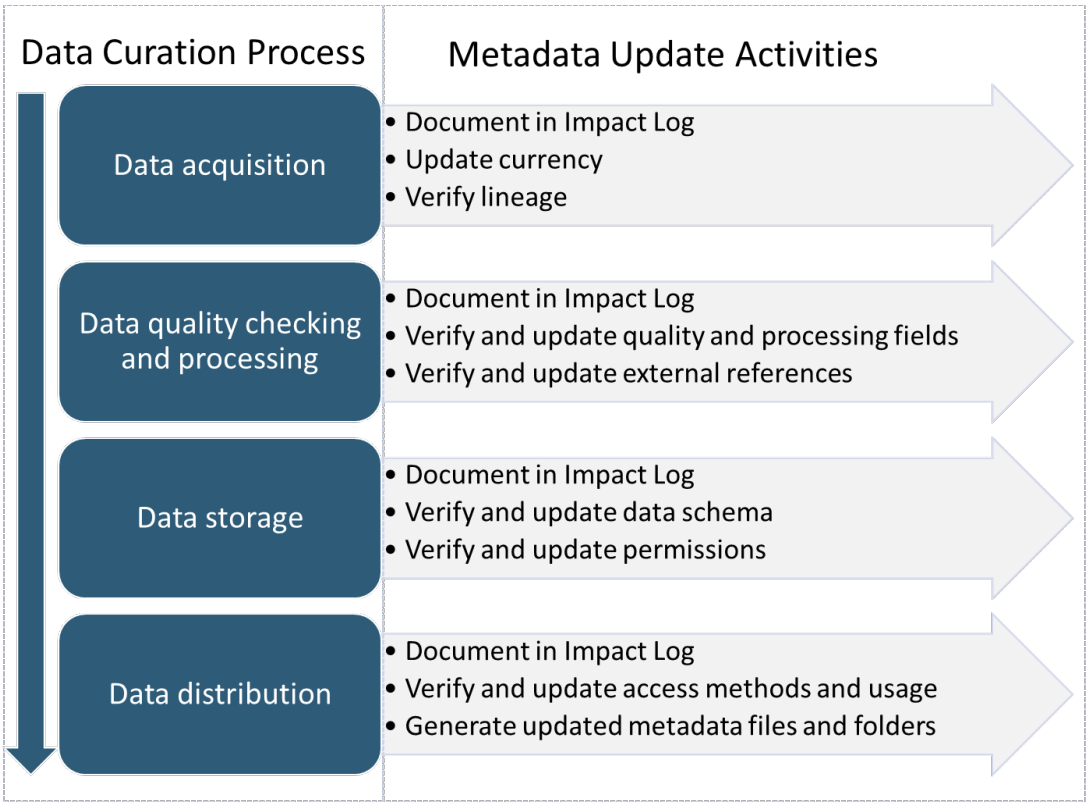

Figure 12. Metadata Update Process Associated with Data Curation Processes
Source: Buffalo, NY ITS4US

Appendix A. Acronyms

Table 12. Acronyms

Acronym	Description
ADA	American Disabilities Act
API	Application Programming Interface
BNMC	Buffalo Niagara Medical Campus
CoB	City of Buffalo
ConOps	Concept of Operations
CS	Community Shuttle
CSV	Comma Separated Value
DMP	Data Management Plan
DOI	Digital Object Identifier
DOT	Department of Transportation
DPP	Data Privacy Plan
ET	Eastern Time
FHWA	Federal Highway Administration
GBNRTC	Greater Buffalo-Niagara Regional Transportation Council
GIS	Geographic Information System
GTFS	General Transit Feed Specification
HDS	Human-Driven Shuttle

Acronym	Description
HUAS	Human Use Approval Summary
ID	Identification
IE	Independent Evaluator
IOO	Infrastructure Owner Operator
IRB	Institutional Review Board
ITS JPO	Intelligent Transportation Systems Joint Program Office
ITS4US	Intelligent Transportation Systems for Underserved Communities
JSON	JavaScript Object Notation
KML	Keyhole Markup Language
MSA	Metropolitan Statistical Area
MVP	Minimum Viable Product
NFTA	Niagara Frontier Transportation Authority
NITTEC	Niagara International Transportation Technology Coalition
NOFO	Notice of Funding Opportunity
NTCIP	National Transportation Communications for ITS Protocol
NYSDOT	New York State Department of Transportation
OSM	Open Street Map
PAL	Paratransit Access Line
PBF	Protocol buffer Binary Format
PED-X	Pedestrian Signal Request
PII	Personally Identifiable Information

Acronym	Description
PM	Performance Measure
PMD	Performance Management Dashboard
PMESP	Performance Measurement Evaluation Support Plan
POP	Period of Performance
PROW	Public Right of Way
RAPUUD	Rapid Assessment of Product Usability and Universal Design
SDS	Self-Driving Shuttle
SOC	Shuttle Operations Center
SOI	System of Interest
SUNY	State University of New York
SyRS	System Requirements Specification
TBD	To Be Determined
TIH	Transportation Information Hub
TMDD	Traffic Management Data Directory
UI	User Interface
URL	Uniform Resource Locator
USDOT	United States Department of Transportation
WZ	Work Zone
XML	Extensible Markup Language

Appendix B. References

[1] [DMP] Gopalakrishna, D., et al. (2021). *Phase 1 Data Management Plan (DMP) – Buffalo, NY ITS4US Deployment Project (FHWA-JPO-21-868)*. Federal Highway Administration.

[2] *[*ConOps*]* Gopalakrishna, D., et al. (2021). *Phase 1 Concept of Operations (ConOps) – Buffalo, NY ITS4US Deployment Project (FHWA-JPO-21-860)*. Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/57571.

[3] [PMESP] Gopalakrishna, D., et al. (2021). *Phase 1 Performance Measurement and Evaluation Support Plan (PMESP) – Buffalo, NY ITS4US Deployment Project (FHWA-JPO-21-878)*. Federal Highway Administration.

[4] [SyRS] Gopalakrishna, D., et al. (2021). *Phase 1 System Requirements Specification (SyRS) – Buffalo, NY ITS4US Deployment Project (FHWA-JPO-21-883)*. Federal Highway Administration.

[5] [HUAS] Gopalakrishna, D., et al. (2021). *Phase 1 Human use Approval Summary (HUAS) – Buffalo, NY ITS4US Deployment Project (FHWA-JPO-21-898)*. Federal Highway Administration.

[6] [DPP] Gopalakrishna, D., ET. al. (2022). *Phase 2 Data Privacy Plan (DPP) – Buffalo, NY ITS4US Deployment Project (FHWA-JPO-22-969)*. Federal Highway Administration.

Appendix C. Data Management Change Log

The Data Management Log contains metadata that records changes to the datasets described in the DMP. Impacts on the data are documented in the Change Log. Log entries include:

- Dataset ID
- Dataset Title
- Change from
- Change to
- Reason for Change
- Change Phase / version (from → to)

Table 13. Data Management Change Log

Dataset ID	Dataset Title	Change from	Change to	Reason for change	Change phase / version (from -to)
n/a	n/a	SOI data flows (Phase 1)	SOI data flows (Phase 2)	Refined SOI data flows due to architecture data flows. Only included external flows. SOI will be replaced in the next version of ConOps. Updated second column (SOI Flow #) in Table 4 to reflect changed SOI.	Phase 1 → 2
2	Community Shuttle (SDS/HDS) GTFS	Community Shuttle GTFS and GTFS Flex	Community Shuttle (SDS/HDS) GTFS	Refined name.	Phase 1 → 2
3	NFTA Fixed Route	NFTA Fixed Route GTFS	NFTA Fixed Route	Refined name.	Phase 1 → 2

Dataset ID	Dataset Title	Change from	Change to	Reason for change	Change phase / version (from -to)
4	GTFS Flex for PAL	NFTA GTFS Flex for PAL	GTFS Flex for PAL	Refined name.	Phase 1 → 2
5	Public right of way (PROW) information in the project region	PROW data	Public right of way (PROW) information in the project region	Refined name.	Phase 1 → 2
6	PROW Work Zone Data	Title from WZDx	PROW Work Zone Data	Refined name to specify specific type of work zone needed. Collection method and data file format is still an open issue.	Phase 1 → 2
8	NFTA GTFS realtime	NFTA GTFS real time	NFTA GTFS realtime	Refined name.	Phase 1 → 2
10	PAL Direct Reservations Summary	PAL Direct Reservations Summary and Details	PAL Direct Reservations Summary	Refined name.	Phase 1 → 2
12	SOC Dispatch log summary files	SOC Dispatch log files	SOC Dispatch log summary files	Refined name.	Phase 1 → 2
13	CS Service Vehicle Manifest			**Deprecated**. Dataset will not be stored by the system.	Phase 1 → 2
15	Trip Planning Summary Usage Data	Trip Planning Data	Trip Planning Summary Usage Data	Refined name.	Phase 1 → 2
17	Beacon Messages	Smart Sign Messages	Beacon Messages	Changed name of title and updated description to show that it is part of the indoor navigation system (beacon)	Phase 1 → 2
18	Beacon message log file	Smart sign message log file	Beacon message log file	Added clarification to description to show that it is part of the indoor navigation system (beacon). In addition, added note that this message log file may not exist.	Phase 1 → 2

Dataset ID	Dataset Title	Change from	Change to	Reason for change	Change phase / version (from -to)
19	HDS Canned and custom operations messages			**Deprecated**. Will not store canned messages. They are part of an existing system.	Phase 1 → 2
20	SDS Canned and custom operations messages			**Deprecated**. Will not store canned messages. They are part of a turn-key system.	Phase 1 → 2
25	Beacon asset locations	Smart Sign assets	Beacon asset locations	Changed title to clarify dataset content.	Phase 1 → 2
26	Facility smart sign (beacon) assets			**Deprecated.** Redundant with dataset #25.	Phase 1 → 2
27	Fare App APIs			**Deprecated**. NFTA is undergoing a change in fare management system and CS subsystems will not charge for trips.	Phase 1 → 2
28	Emergency Safety Messages			**Deprecated**. SDS SOC will contact emergency services through 911 and other existing means.	Phase 1 → 2
32	BNMC Facility Map		New 2022/10/28	**Added**	Phase 2

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

U.S. Department of Transportation