Phase 2 Data Management Plan (DMP)

University of Washington ITS4US Deployment Project

www.its.dot.gov/index.htms

Final Report — January 17, 2023
FHWA-JPO-22-976

ITS4US
IT'S TRANSPORTATION FOR ALL OF US

U.S. Department of Transportation

Produced by University of Washington
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-22-976		

4. Title and Subtitle	5. Report Date
Phase 2 Data Management Plan (DMP) University of Washington ITS4US Deployment Project	January 17, 2023
	6. Performing Organization Code N/A

7. Author(s)	8. Performing Organization Report No.
Adam Danczyk, Kristin Tufte, Anat Caspi, Mark Hallenbeck, Nick Bolten, Alice Marecek	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
The Taskar Center for Accessible Technology Department of Computer Science & Engineering University of Washington Box 352350 Seattle, WA 98195-2350	**11. Contract or Grant No.** 693JJ321C000004

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	**14. Sponsoring Agency Code** HOIT-1

15. Supplementary Notes

Kate Hartman, AOR

16. Abstract

This document is the Data Management Plan (DMP) for the University of Washington's (UW) Transportation Data Equity Initiative (TDEI) Project for the United States Department of Transportation's (USDOT) ITS4US Program. The DMP provides details about how the data will be managed for the project and is an update of the Phase 1 DMP that includes any new design developments that have been determined as part of Phase 2. The DMP discusses data management, analysis, storage, and other data-related topics that are relevant to the USDOT's data requirements, including management of Personally Identifiable Information (PII) that will exist in several instances among the full data set that is collected.

17. Keywords	18. Distribution Statement
ITS4US; Deployment; ITS; Intelligent Transportation Systems; Data Management Plan; Sensitive Personally Identifiable Information	N/A

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
N/A	Unclassified	50	

Revision History

Name	Date	Version	Summary of Changes	Approver
Kristin Tufte, University of Washington	22 November 2022	0.1	Initial Draft	
Kristin Tufte, University of Washington	22 December 2022	1.0	Final	
Mark Hallenbeck University of Washington	17 January 2023	1.0	Final 508 Compliant	

This page intentionally left blank.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Document Purpose

This document discusses the Data Management Plan (DMP) for the University of Washington (UW) ITS4US Deployment Project. This DMP—referred herein as the Phase 2 DMP—aims to serve the Phase 2 (Design and Test) and Phase 3 (Operate and Evaluate) portions of the UW Transportation Data Equity Initiative (TDEI) project and inform the data management processes that will occur during the five-year operations and maintenance period. This DMP serves as an update to the Phase 1 DMP [2] prepared in September 2021, and includes several updates made because of design developments made since Phase 1, as well as includes additional insights not previously known during Phase 1 that can be answered more clearly in Phase 2.

The purpose of this DMP is to describe the data that will be collected and how the data will be managed throughout the UW ITS4US project. Furthermore, the DMP will define a framework for appropriately preserving the necessary data for the Performance Measurement process, as well as for sharing the data with the USDOT ITS Data Hub [13]. This DMP is informed by the Phase 1 Concept of Operations (ConOps) [1] for the TDEI System, which bridges between the needs that motivated the project and the specific technical requirements from which the project is built. It is also informed by the Phase 1 Human Use Approval Summary (HUAS) [4] —which discusses the UW ITS4US team's approach to human subject participants and coordination with the UW Institutional Review Board (IRB)—and the Performance Measurement and Evaluation Support Plan (PMESP) [3] —which develops a project evaluation approach that will determine whether the operational UW ITS4US system addresses the initial user needs documented in the ConOps. Several privacy and security topics discussed in the Phase 2 Data Privacy Plan (DPP) [6] also inform this Phase 2 DMP.

A glossary and acronym list are provided in the appendix.

1.1.1 Organization of this Document

This document is organized as follows:

- Section 1 outlines the document purpose and deployment concept, as well as data schedule, needs summary, and overview.
- Section 2 discusses data stewardship in terms of ownership, storage/retention, sharing, quality control, privacy, and relationship to performance measures.
- Section 3 discusses data standards, versioning, and metadata.

1.2 Deployment Concept

The TDEI project will develop a national pipeline to create, disseminate, and share standardized data about pedestrian environments, transportation environments, and on-demand transportation services to enable better use, discoverability, and data analytics of these assets and services. The goal of the UW ITS4US Deployment project is to build a sustainable, inclusive data infrastructure to enable and accelerate the future of equitable mobility and access to transportation for the benefit of all travelers. Through community leadership, the TDEI System, the associated standards development, and the adoption by users (including both data generators and data consumers) will help provide a means to offer travel services, automate routing, and map out the transportation network in ways appropriate for every traveler. With this in place, previously underrepresented individuals will have tools available to make informed, customized travel decisions.

Systems developed in this project will enable users to have improved awareness of routes (specifically routes that align with their unique travel preferences) and transit services available to them. At a very high level, the TDEI System aims to achieve USDOT ITS4US Program goals by deploying the following key technology elements:

1. **Develop a Centralized Data Repository**. The UW Team will develop a centralized data repository that services many functions. It receives, validates, and quality assures incoming sidewalk and transit-related data that are provided by data generators and transit agencies. It then stores the latest versions of data in the data repository for use. It then accommodates data requests made through an Application Programming Interface (API) service from applications that request geographically focused data to satisfy trip making. This component represents the focal point of the project for moving data from those who produce it to those who want to use it.
2. **Create tools to support data collection**. The UW Team will develop tools for sidewalk infrastructure owner-operators and transit agencies to collect data, translate it into the preferred data standard, and submit data to the data repository. The goal of this component is to simplify the level-of-effort required to collect this data, thus, encouraging agencies to undertake this data collection initiative.
3. **Demonstrate use of the data by under-represented communities through three accessibility-focused mobility applications.** This project will utilize accessible mobility applications in the evaluation and testing of the usability and efficacy of the data standards and the supporting infrastructure.

The UW Team will publish collected data for the six U.S. counties that are part of this project. The six counties, as shown in Figure 1, are King and Snohomish Counties in Washington State, Multnomah and Columbia Counties in Oregon, and Harford and Baltimore Counties in Maryland. Our pilot area focuses on regions that combine urban centers with suburban sprawl and rural areas to demonstrate the use of this data in different conditions of travel networks.

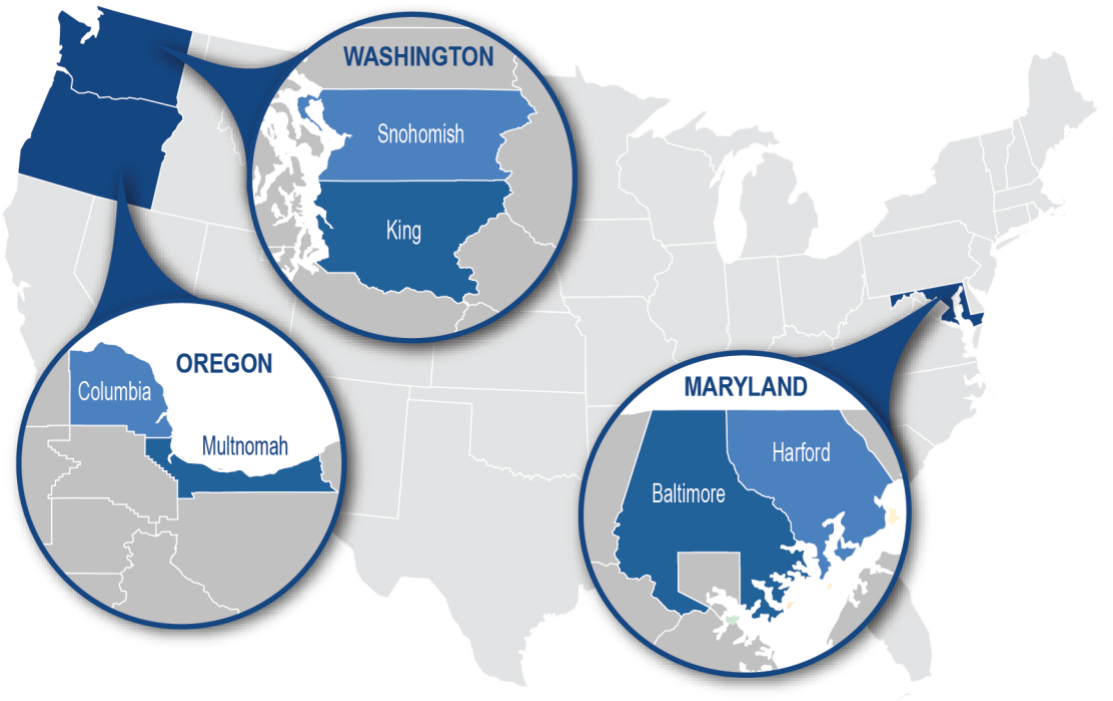

Figure 1. Map. Washington, Oregon, and Maryland Counties.

Source: United States Department of Transportation, University of Washington, and Cambridge Systematics.

As discussed in the Phase 1 DMP and the Phase 2 System Architecture Document (SAD) [5], this project involves movement of data between various external parties, most often using the public internet to facilitate data movement. Figure 2 shows an updated Phase 2 context diagram of the TDEI System. The efforts listed in that figure are identified as the following:

- Components that the UW Team will directly develop and test, which primarily include the data validation and data services technologies that are the focal point of this project. In the context of Figure 2, these components include the data processing pipelines, the data repository itself, and the service pipelines.
- Components that the UW Team will assist in developing to encourage data contributions, namely tool sets through which data providers will be encouraged to submit data. In the context of Figure 2, these tool sets will serve groups such as municipal governments, transit agencies, and other data providers.
- Components that represent software demonstrations whose development the UW Team will support to illustrate the success of the pipelines. These include the three applications that have been vetted to provide the services needed by underserved end users. In the context of Figure 2, these components include the UW Taskar Center for Accessible Technology's (TCAT) AccessMap Multimodal, Soundscape, and XR Navigation's Audiom.

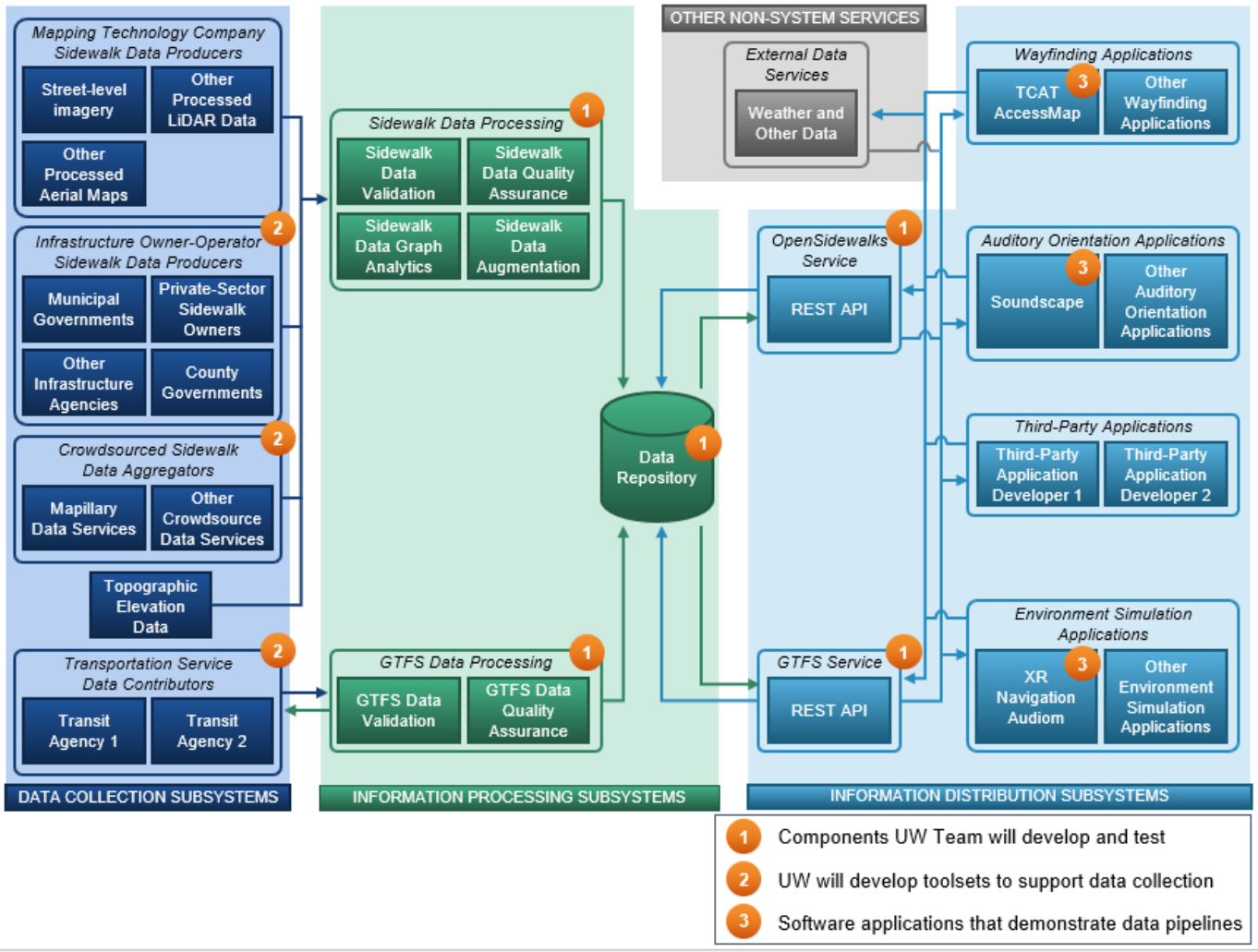

Figure 2. Diagram. Context diagram for the Transportation Data Equity Initiative system.

Source: University of Washington and Cambridge Systematics.

1.3 Key Phase 2/Phase 3 Adjustments

The UW ITS4US project team recognizes the importance of ensuring sufficient privacy controls to mitigate the risk of harm to individuals that would result in the improper handling or disclosure of Personally Identifiable Information (PII) or Sensitive Personally Identifiable Information (SPII) collected from individuals in connection with the project. The Phase 1 DMP extensively reviewed the data flows in the TDEI System and documented that most flows in the TDEI System deal with data on public infrastructure or transport services that reside in the public space and thus do not involve PII associated with an individual. A few data flows, however, were identified in the Phase 1 DMP as potentially including locational PII; these flows were data flows that included information on a trip origin, destination, and route-specific preferences that could be traced back to an individual's home address or to a specific person with a defining physical characteristic. This type of data requires limitations on reuse and redistribution of that data.

The updated understanding of PII in the TDEI System that is reflected in this Phase 2 DMP can be summarized as follows:

- The UW Team has determined that locational PII is not present in TDEI system operation data.
- The UW Team has determined that the user account data to be collected in Phases 2 and 3, which is part of the TDEI system administration data will contain limited PII. These user account data are exclusively for the UW Team's system administration activities – primarily communication with stakeholders and will not be provided to the USDOT or the USDOT Independent Evaluation (IE) Team.
- Locational PII is expected be present in the data collected in Phase 3 as part of the UW Team's local evaluation of the TDEI system. Specifically, the evaluation of the TDEI System using the AccessMap Multimodal demonstration application will involve the collection of PII. These data are termed UW Team local evaluation data in this report and are exclusively for the UW Team's local evaluation of the TDEI System. Data provided to the USDOT and the USDOT IE Team will not contain PII.

This Phase 2 DMP updates the data management needs with respect to these three updates, calling out throughout the document where deviations from the Phase 1 DMP occur. Additional details on securing privacy for sensitive data can be found in the Phase 2 DPP.

The three updates listed above are further described below.

1.3.1 No Locational PII in System Operation Data

Between Phase 1 and Phase 2, the team reviewed the concern about locational PII and identified that locational PII would, in fact, not be present within the TDEI system operation data. The locational PII does continue to exist between the application developers that are external to the TDEI System and the digital device end users (i.e., the individuals with privacy concerns) as described in the Phase 2 DPP.

The following are noted:

1. The application developers creating applications that use TDEI data and APIs act as a firewall between the digital device end user's requests that may contain PII data and the

TDEI System, as it is the applications themselves that choose the coverage area and extent of sidewalk and/or transit data that is necessary in order for them to provide routing or other service. The TDEI System is designed to provide general geospatial data for a geographic area of relevance, but the application developer's own algorithms are responsible for computing a route or providing any other service. The TDEI APIs [22] will be designed to receive requests for this geospatial data only. For example, the API will accept a message that conceptually says, "Provide me all sidewalk data and associated attributes for the Capitol Hill neighborhood in Seattle." While the TDEI System will know geographic areas of interest (i.e., neighborhoods), the TDEI System will not be able to ascertain the origins, destinations, and user physical characteristics (e.g., uses a wheelchair) based on requests for geospatial data alone.

2. The TDEI System is not aware if a query from an external application developer is for a user or for other software-serving purposes, such as the application developer building a data cache to allow quicker responses to digital device end users. In other words, an application may query for data (e.g., "Provide me all sidewalk data and associated attributes for the Capitol Hill neighborhood in Seattle.") either in response to a user seeking trip information or for the application to update its data for future purposes; the TDEI System will not know which motivation is prompting the request.

1.3.2 PII in Phase 2 and 3 – User Account Data (System Administration Data)

As described above, locational PII will not be collected by or stored in the TDEI System. However, an item not captured in the Phase 1 DMP is PII in TDEI user account data. The TDEI System design requires individuals and organizations to create user accounts in order to receive an API key and/or access token to be granted access to the TDEI System for consuming and/or producing data, respectively. At a minimum, an email address would be required to create a user account; the email address would reside in the TDEI System's authorization service to verify that a given user account is permitted to use the API to upload or download data. The full degree of PII present to support this authorization service is being worked out as part of the design, but it is anticipated that the data present there will meet the minimum requirements necessary to be considered PII, and thus will be treated as such for purposes of this Phase 2 DMP. This user account data will be stored in the TDEI authorization system with appropriate safeguards and will be used for purposes of managing communications with users, this data will not be provided to the USDOT IE Team.

1.3.3 PII – Phase 3 UW Local System Evaluation Data

PII data remains a concern for the UW ITS4US project with respect to the human participants that are being recruited in Phase 3 to help evaluate the TDEI System using demonstration applications. This evaluation is discussed in greater detail in the Phase 1 HUAS, the Phase 2 DPP, and the Phase 1 PMESP. The UW ITS4US project team will be using AccessMap Multimodal as one of the applications to demonstrate how the data are used in a real-world application. AccessMap Multimodal is the next generation of the UW's existing AccessMap software; the upgrades to AccessMap to create AccessMap Multimodal are not being funded by the USDOT.

To facilitate the demonstration of the TDEI in Phase 3, approximately forty volunteer participants will be recruited for a laboratory study to use AccessMap Multimodal and demonstrate how it performs (in terms of providing useful trip information) when the TDEI System is present to provide data. These volunteer participants will provide informed consent that they are willing to participate in this study, with the UW ITS4US project team providing information to participants in advance of the study which will include: the purpose of the study, the data that will be collected, how that data will be used, the tasks to be performed by participants and the risks associated with those tasks. These tasks involve performing their normal trip making, with the smartphone application making suggestions on paths that can be taken to perform those trips, with the participant choosing to accept or decline to use the suggested paths identified by the smartphone applications, and then recording the outcome of those travel decisions and reporting on their satisfaction levels concerning those outcomes. Participants can opt out of providing data that they do not wish to share, including the sensor or trace data collection or any of the surveys that are being generated. UW's IRB will review processes associated with these human subject experiments. The UW Team will ensure that approval from the UW IRB board is received prior to the data collection and experiments.

In the context of data privacy, the trace and survey data collected as part of the evaluation using AccessMap Multimodal and other demonstration applications will be stored in a separate system—independent of the TDEI System—to help inform evaluation outcomes. Necessary safeguards to protect user privacy will need to be applied to that independent system. Additional details are provided in the Phase 2 DPP. Handling and analysis of the data that describe the travel outcomes obtained by using AccessMap Multimodal will be described more fully in the Phase 2 Performance Measurement and Evaluation Support Plan (PMESP.) The data collected as part of this study, termed UW local system evaluation data, will not be provided to the USDOT or the USDOT IE Team; the USDOT IE Team will not receive data with PII. The USDOT IE Team will receive aggregated summaries of this study which are free of PII.

1.4 Data Schedule

Table 1 illustrates the anticipated schedule for data-related events, including those that took place as part of Phase 1. Parts of this schedule are subject to update, based on continued design developments that occur in Phase 2 and during the operations and evaluation activities of Phase 3. The UW ITS4US team will work with the USDOT to keep all parties informed of any deviations in data delivery.

Table 1. Transportation Data Equity Initiative data-related schedule.

Event Title	Description	Date
Phase 1 Draft DMP	Initial Draft DMP with basic information known at the time of writing.	July 2021
Phase 1 Final DMP	DMP is updated with USDOT comments addressed.	September 2021
Phase 2 Draft DPP	Initial Draft DPP with focus on privacy concerns.	September 2022

Event Title	Description	Date
Phase 2 Draft DMP	Phase 1 DMP is updated to align with Phase 2.	November 2022
Phase 2 Final DPP	DPP is updated with USDOT comments addressed.	November 2022
Phase 2 Final DMP	DMP is updated with USDOT comments addressed.	January 2023
System operation and system administration data samples	Initial data samples of system operation and system administration data are created, validated, and submitted to USDOT for review.	July 2023
Data review meeting with USDOT	Meeting to review data with USDOT and walk through the data schema and data samples.	July 2023
System operation and system administration data available to USDOT	The USDOT IE Team may retrieve public system operation data – General Transit Feed Specification (GTFS)-Flex [9,11], GTFS-Pathways [10] and Open Sidewalks (OSW) [17] data through the TDEI APIs. TDEI performance metric data based on system administration data will be available through a TDEI performance measures dashboard. The USDOT IE Team will have access to this dashboard.	Late Phase 2 (January 2024)
Baseline data collection	Use e-mission software, a mobility trace collection application used for understanding person movement [7,15] or similar, to collect baseline data for the AccessMap Multimodal demonstration application evaluation. Occurs just prior to start of demonstration application evaluation. These data are only for the UW local evaluation effort. The data for the TDEI performance measures dashboard for the first quarter of Phase 3 will be collected as baseline data for the IE Team evaluation as described in Section 2.6.1.	Early Phase 3
System evaluation data samples	Initial data samples of summarized UW local system evaluation data free of PII, are created, validated, and submitted to USDOT for review. These data are free of PII.	Early Phase 3

Event Title	Description	Date
System evaluation data available to USDOT	Summarized UW local system evaluation data free of PII will be provided to USDOT after the testing of demonstration applications is complete.	Phase 3
Performance metrics review	Ongoing Data review conducted with USDOT and the USDOT IE Team to ensure the performance metrics meet the IE Team's needs.	Phase 3
Draft Final Evaluation Report submitted	Draft Final Evaluation Report submitted to USDOT.	Phase 3
Final Evaluation Report submitted	Final Evaluation Report submitted to USDOT.	Phase 3

1.5 Data Needs Summary

Figure 2, earlier in Section 0, illustrates the context diagram for the TDEI System. The data flows shown in Figure 2 and detailed below and in Figure 3 are similar to those identified in the Phase 1 DMP, which focused mostly on data flows associated with the operation of the system. Considering design updates made since Phase 1, additional data sets of importance have been identified for the Phase 2 DMP. At a high level, the TDEI collects three types of data:

- System Administration – Data collected for the purpose of administering (managing and operating) the TDEI System and includes data collected for the purpose of monitoring the system and measuring system performance.
- System Operation – Data that is part of the transactional operation of the TDEI System. This data includes the GTFS-Pathways, GTFS-Flex and OSW data which together form the bulk of the data collected by the TDEI System.
- UW Local System Evaluation – Data collected for the purpose of the UW Team evaluating the TDEI System; primarily data collected during the use of the demonstration applications to demonstrate the TDEI's value, especially with regards to the use of Access Map Multimodal.

1.5.1 Context and Background

This section describes the data flows identified in the context diagram for the TDEI System. Figure 3 shows an updated version of the context diagram, with changes made to reflect design decisions made since the Phase 1 DMP was published and with callouts labelling the data flows.

Descriptions of the data flows identified in Figure 3 are provided in subsequent text. In addition, the data in these data flows was discussed at length as part of the Phase 1 DMP.

The Intent of Figure 3 and the descriptions of the data flows is to provide context and background for the descriptions of the System Administration, System Operation and System Evaluation data which follow in subsequent sections.

At a very high level, the following data flows, which correspond to the numbers shown in Figure 3, are present as the primary data movements for the TDEI System.

- **Sidewalk Data Contribution to the Data Repository (Data Flows #1 and #2):** The TDEI System needs data contributors—with first-hand knowledge of or data on sidewalk infrastructure—to provide information on sidewalk pathways and relevant accessibility attributes that can be added to a larger pool of sidewalk network data. Several types of entities may have information on sidewalk attributes or other characteristics of the pedestrian built environment. These entities will generally include infrastructure owner-operators (public and/or private), crowdsourced sidewalk data aggregators that report their observations on sidewalks, and mapping technology companies that utilize advanced analytics on aerial imagery or light detection and ranging (LiDAR) data to generate observations on sidewalk attributes. In the TDEI System, all data from these data contributors will be converted into the OpenSidewalks data format via the tools developed by UW and submitted to the TDEI System. Contributed sidewalk data will be screened and checked for accuracy by validators before being deposited in the data repository, primarily to validate that the data are compliant with the OpenSidewalks standard. A quality assurance check will be done for each data contribution to ensure that the contribution is reasonable; this process may be automated or manual (human verified) and is anticipated to evolve based on available technology. The quality assurance check will compare the contributed data with other reliable sources; for example, a contributed sidewalk segment may be verified by using aerial imagery to confirm that the path exists. After quality assurance is complete, the TDEI System will transform the data into a graph segment and determine how it augments the existing graphed sidewalk network. This data—now filtered by the previous checks—will then flow to the data repository for storage. That repository will comprise a fully comprehensive sidewalk network (based on data that have been submitted). Specifically:

    1. Data Flow #1 represents unvalidated sidewalk data submissions, sent from data producers to the TDEI System's vetting process. Unvalidated data are data on which the TDEI System's data checks (manual or automated) have not yet been conducted. This flow represents data flowing over the interface between external parties (i.e., data producers) and the TDEI System through the TDEI APIs.
    2. Data Flow #2 represents validated sidewalk data submissions, which include the data from Data Flow #1 that has gone through the TDEI System's vetting process and is stored in the data repository. This data flow is internal to the TDEI System.

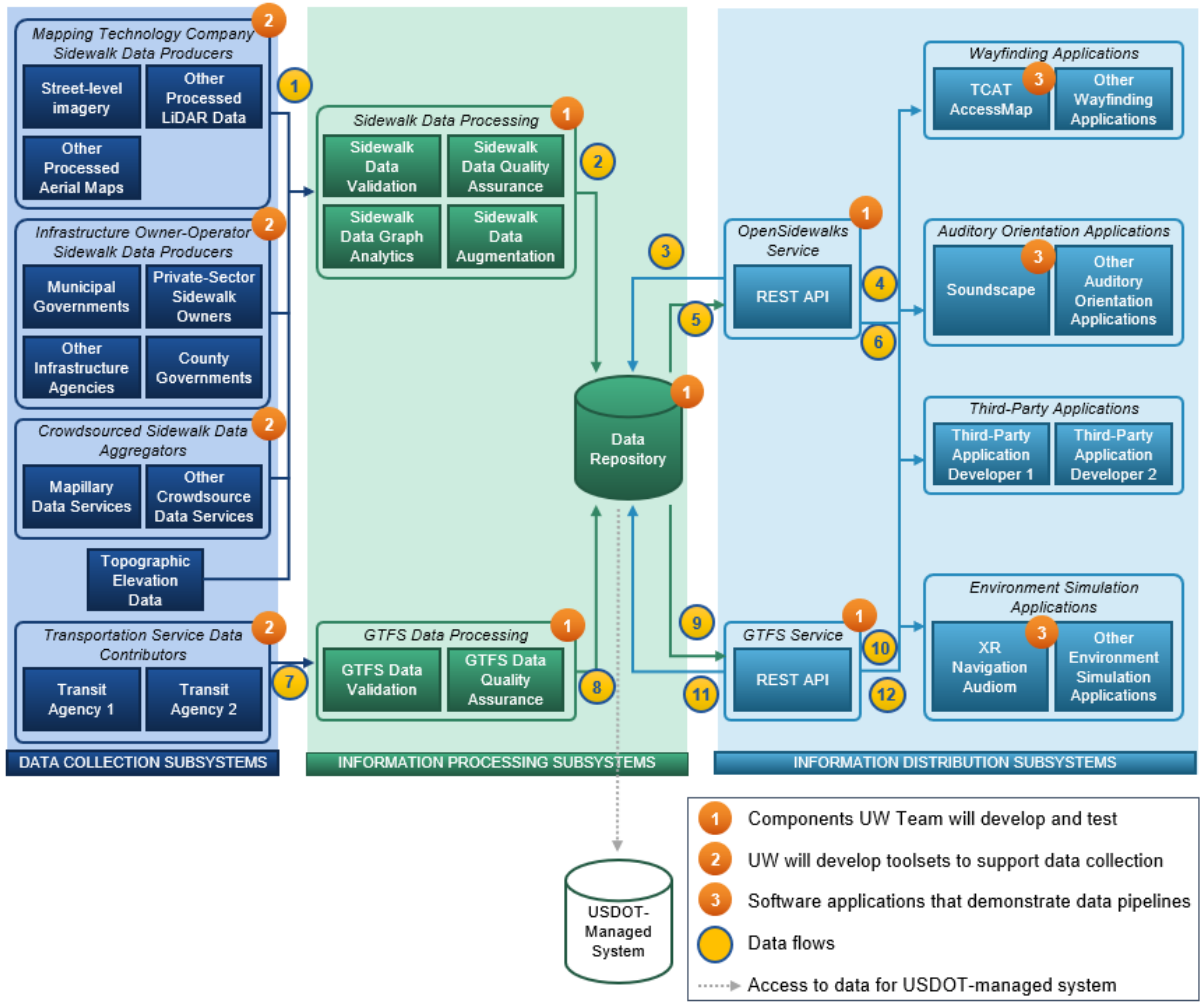

Figure 3. Diagram. Context Diagram with data flow callouts.

Source: University of Washington and Cambridge Systematics

- **Transit Service Data Contribution to the Data Repository (Data Flows #7 and #8):** The TDEI System needs data contributors—with first-hand knowledge of data on their offered on-demand transit services and station pathways—to provide information on these services and stations. Entities with this information will primarily include transit service providers that operate transit services or maintain station infrastructure to support transit services. In the TDEI System, these data contributors will transform their data into the relevant GTFS extension data format (GTFS-Pathways or GTFS-Flex v2). Contributed transit data will be screened for accuracy before being deposited in the data repository to validate that the data are compliant with the GTFS standard and its extensions. A quality assurance check will be done for each data contribution to ensure that the contribution is reasonable; this process may be automated or manual (human verified) and is anticipated to evolve based on the available technology. This data—now filtered by the previous checks—will then flow to the data repository for storage. Specifically:
    1. Data Flow #7 represents unvalidated transit GTFS-Pathways and GTFS-Flex data submissions, sent from data producers to the TDEI System's vetting process. Unvalidated data are data on which the TDEI System's data checks (manual or automated) have not yet been conducted. This flow represents data flowing over the interface between external parties (i.e., data producers) and the TDEI System through the TDEI APIs.
    2. Data Flow #8 represents validated transit GTFS-Pathways and GTFS-Flex data submissions, which include the data from Data Flow #7 that has gone through the TDEI System's vetting process and is stored in the data repository. This data flow is internal to the TDEI System.
- **Sidewalk Data Distribution to the Data Consumers (Data Flows #3, #4, #5, and #6):** The TDEI System needs to provide data upon request for a relevant geographic area of interest. At the application level, the end user may enter a preferred origin, destination, trip-specific travel preferences, or some other navigational feature. The application will retain these trip-specific travel preferences and query the TDEI System for sidewalk options deemed relevant by that application's algorithms for a geographic area of interest. The OpenSidewalks data service will receive and verify the request from the application through its secure API and pull all relevant data from the data repository. Specifically:
    1. Data Flow #4 represents unvalidated data queries sent from approved data consumers to the TDEI System's API for sidewalk data. This data flow represents data movement over an external interface.
    2. Data Flow #3 represents validated data queries for sidewalk data that have passed through the API and are being processed by the system, using requests made as part of Data Flow #4. This data flow is internal to the TDEI System.
    3. Data Flow #5 represents a package of geospatial sidewalk data that is pulled from the TDEI System and exposed through the API in response to a data query for sidewalk data. This data flow is internal to the TDEI System.
    4. Data Flow #6 represents a package of geospatial sidewalk data that moves from the TDEI System's API to the data consumers that made the original query. From a data content standpoint, it is the same as Data Flow #5 except that it is a data flow that represents data movement over an external interface.

- **Transit Service Data Distribution to the Applications (Data Flows #9, #10, #11, and #12):** The TDEI System needs to provide data upon request for a relevant geographic area of interest. At the application level, the end user may enter a preferred origin, destination, trip-specific travel preferences, or some other navigational feature. The application will retain these trip-specific travel preferences and query the TDEI System for transit pathways or flex options deemed relevant by that application's algorithms for a geographic area of interest. These data are only retained by the consumer application and are not stored in the TDEI. Each application is responsible for safely storing the PII. For Multi-Modal AccessMap, these data are retained on the user's personal device, significantly limiting the potential for loss of privacy for that individual through the TDEI. The GTFS-Flex or GTFS-Pathways data service will receive the request from the application through its API and pull all relevant data from the data repository. The relevant data will be sent in the GTFS-Flex or GTFS-Pathways data format to the application that made the original request for that application's use. Specifically:

    1. Data Flow #12 represents unvalidated data queries sent from approved data consumers to the TDEI System's API for transit pathways and/or flex data. This data flow represents data movement over an external interface.
    2. Data Flow #11 represents validated data queries for transit pathways and/or flex data that have passed through the API and are being processed by the system, using requests made as part of Data Flow #12. This data flow is internal to the TDEI System.
    3. Data Flow #9 represents a package of geospatial transit pathways and/or flex data that is pulled from the TDEI System and exposed through the API in response to a data query for transit pathways and/or flex data. This data flow is internal to the TDEI System.
    4. Data Flow #10 represents a package of geospatial transit pathways and/or flex data that moves from the TDEI System's API to the data consumers that made the original query. From a data content standpoint, it is the same as Data Flow #9 except that it is a data flow that represents data movement over an external interface.

These data movements allow the TDEI System to operate in a manner that addresses the user needs identified in the Phase 1 ConOps. Data that are retained for performance measures will be collected from these data movements and ultimately stored in the data repository for administrative use. The TDEI System itself does not collect any PII as part of these operations, which is an update from the Phase 1 DMP that noted a concern over Locational PII in some data flows. The PII associated with the user registration apply to the system administration data, discussed in Section 1.5.1.

1.5.2 System Administration Data

System administration data are data collected as part of the management, operation, monitoring and measurement of the TDEI System. System administration data may include information on the number of user accounts created, number of API keys generated, API response times, logging of API data ingest and data request calls, system uptime, system latency, and data security metrics, and other administrative task data such as data backup logs. The focus of system administration data is to evaluate how the system is performing – in terms of API usage, system health, system responsiveness, and system security. System Administration data excludes the GTFS-Flex, GTFS-Pathways, and OSW data that are collected by the system.

As discussed in the Phase 2 SAD and illustrated in the Functional Components diagram in the SAD, the TDEI System will have a monitoring service which will facilitate the collection of administrative data including data regarding API usage, system health, system security and system malfunctions or disruptions.

Much of the system administration data described in the prior two paragraphs will be used for TDEI performance measures and system evaluation and as described in the Phase 1 PMESP. These data do not include any sensitive information and will be provided to the USDOT IE Team through the TDEI performance measures dashboard.

In addition, for UW system management, system administration data will be collected by the authorization service in the context of establishing user accounts and distributing API keys and/or access tokens to approved parties. Users will need to voluntarily submit information – specifically an email address – to provide a point of contact for the users to receive system updates and other notifications relevant to ongoing use. While the exact data required for these user accounts is being determined as part of the design, it meets the minimum requirements to be considered PII, and thus will be a protected data resource. Once the user accounts are established and the API keys and/or access tokens are provided to eligible data producers and data consumers, the user's data interaction with the system will be considered part of the System Operation, discussed in Section 1.5.3. This user account data will be private to the TDEI and will not be provided to the USDOT IE Team.

1.5.3 System Operation Data

System operation data includes data that is part of the transactional operation of the TDEI System. System operation data includes the GTFS-Flex, GTFS-Pathways, and OSW data files collected by the system and information related to those files, such as the valid dates of the files, the entity that collected the file, and data confidence measures, that is data that flows over the data flows shown in Figure 3. Transactions over each of these data flows will be logged using the monitoring service; these logs are included in the system operation data. Thus, in summary, system operation data includes all data that flows over the flows number 1 through number 12 in Figure 3 and the logs of those data flows. In addition, data such as information about transit stations and transit service providers (e.g., transit agencies) will be stored in the system and accessible via API, these data are also considered system operation data. These data do not include any sensitive information and will be provided to the USDOT IE Team through the TDEI APIs.

1.5.4 UW Local System Evaluation Data

The UW Team will gather system evaluation data for the UW local evaluation of the TDEI system and its impacts on end user behavior, mobility, and overall satisfaction with travel opportunities and outcomes. These data are collected independent of TDEI System management or operations and are gathered as part of a separate effort that aims to evaluate the TDEI System during Phase 3 of the ITS4US Program. Data for this evaluation are divided into two approaches: 1.) the first is a stakeholder feedback surveys of users of the three demonstration applications and 2.) the second is user trace data collected to ascertain how successful the TDEI data—through AccessMap Multimodal—is for providing the user with the necessary information to make informed trip-making.

The trace data are collected from the forty informed voluntary participants who were described in Section 1.3. Data collected as part of the traces may include origin-destination pairs, routes offered, alternate routes offered, geospatial position of user on the route, and other metrics necessary to determine the accuracy of the application's routing choices when using the TDEI System data. The trace data and survey data will contain data that is considered PII, and thus will require safeguards for protection; therefore, this data will be stored in a system separate from the TDEI System with appropriate safeguards. Details of these safeguards are provided in the Phase 2 DPP for this project.

The raw trace and survey data will be used by the UW Team for their local evaluation and will not be provided to the USDOT IE Team. Summaries of these evaluation data and the results of the analyses performed using those data will be provided to the USDOT IE Team. Neither the data summaries nor the results of analyses will contain sensitive data.

1.6 Data Overview

This section provides an overview of the datasets collected by the TDEI organized by System Administration, System Operation, and System Evaluation data. Many of these datasets were discussed in the Phase 1 DMP. New to the Phase 2 DMP are the system administration dataset, which will help monitor system performance and the system evaluation-related datasets that will be collected as part of the Phase 3 evaluation process. Another key difference from the Phase 1 DMP is that fixed-route transit feeds, previously included in the Phase 1 DMP, are not included; collecting the fixed-route transit data is a future feature that is outside the scope of the ITS4US demonstration effort.

For each data set, the following information is provided:

- **Dataset Title:** The title of the dataset.
- **Description:** Description of the dataset.
- **Type:** The anticipated type of the data in the dataset.
- **Collection Method:** How the data will be collected, as well as whether collection is through a manual or automated process.
- **Data Formats:** The anticipated data file format. Note that this may be subject to change as the system is developed.

1.6.1 System Operation Datasets

Table 2 provides an overview of the system operation datasets in the TDEI System. These datasets primarily consist of the data flows in the TDEI System called out in Figure 3. The data flows referenced in the Dataset Title column refer to the data flows and numbers shown in Figure 3.

For each of the data flows, there are up to three types of data that are included in and related to the data flow: metadata, the data files themselves, and logs of the data flow, all of which are considered system operation data. The metadata data and the data files flow over the data flow as indicated in in Figure 3; the log files are collected by the monitoring system.

As described, a primary focus of the TDEI System is to collect transit station pathways (GTFS-Pathways), on-demand transit service (GTFS-Flex), and sidewalks (OSW) data. These data are collected through transportation service providers and other data generators using TDEI APIs and data collection tools. When data generators upload data files to TDEI, they are required to provide additional metadata about the files beyond what is captured in the files. For the purposes of this document, we call this type of metadata file-metadata to distinguish from data dictionary style metadata which will also be provided by the TDEI, and which is described in Section 3.3. The file-metadata for data uploads will include “valid_from” and “valid_to” dates, the method of data collection, the entity that collected the data, and other relevant data. The file-metadata are defined in the TDEI System and in the TDEI APIs and flows with the data files over the APIs and is part of the incoming data flow (data flows #1, #2, #7 and #8). The file-metadata will be provided by the data generators via the APIs in JavaScript Object Notation (JSON) [12] format.

Another key focus of the TDEI System is to allow users to download the transit station pathways, on-demand service, and sidewalk data files. When these data files are downloaded, file-metadata will be provided along with those files – the file-metadata for data downloads will include items like the upload metadata – “valid_from” and “valid_to” dates, method of data collection – and will also include TDEI-generated file-metadata, such as a confidence level. This file-metadata will be provided as part of the API responses in JSON format.

Note that this metadata, or information about the data files, is distinct from dictionary-type metadata which describes the metadata fields (otherwise thought of as metadata on metadata), and metadata that describes the files in the GTFS-Pathways, GTFS-Flex, and OSW files. This descriptive metadata will be discussed in Section 3.3.

As mentioned above, each of these flows will also be logged to understand system usage. Log files will be provided in Comma-Separated-Value (CSV) or JSON format.

A final type of system operation data is data about the transit stations and on-demand services themselves. This data includes items such as transit station name, on-demand service name, agency that manages the station or service, and other relevant information. These data will be uploaded into the TDEI System by TDEI staff using internal interfaces.

Table 2. Transportation Data Equity System Operation Datasets.

Dataset Title	Description	Type	Collection Method	Data Formats
Sidewalk Data (Data Flow #1)	Attributes of the pedestrian infrastructure.	Numerical data, text sequences, positional data (latitude, longitude)	Currently collected through manual entry by local entities, including infrastructure owner-operators and local community crowdsourcing groups. Automated collection through AI image processing is being explored.	File-Metadata[1]: JSON Data files: gzipped GeoJSON [8] file Log data: CSV or JSON
Validated Sidewalk Data (Data Flow #2)	Attributes of the pedestrian infrastructure transformed into a graph network. These data are validated and may be filtered.	Numerical data, text sequences, positional data (latitude, longitude)	Modified version of Data Flow #1. Validated, transformed version of sidewalk data submittal.	File-Metadata: JSON Data files: gzipped GeoJSON file Log data: CSV or JSON
Transit Flex Data (Data Flow #7)	Information on on-demand transit service, including operating hours, contact information, and other attributes.	Numerical data, text sequences, positional data (latitude, longitude)	Collected by transit agencies and on-demand service providers.	File-Metadata: JSON Data files: zip file consisting of a set of GTFS-Flex V2 CSV files Log data: CSV or JSON

[1] File-metadata is metadata describing the (OSW, GTFS-Pathways, and GTFS-Flex) data files as described above. The term file-metadata is used to distinguish this metadata from data dictionary style metadata.

Dataset Title	Description	Type	Collection Method	Data Formats
Transit Pathways Data (Data Flow #7)	Attribute data on pathways within a transit station and other attributes.	Numerical data, text sequences, positional data (latitude, longitude)	Collected manually by transit agencies.	File-Metadata: JSON Data files: zip file consisting of GTFS-Pathways CSV files Log data: CSV or JSON
Validated Transit Flex Data (Data Flow #8)	Information on on-demand transit service, including operating hours, contact information, and other attributes. These data are validated and may be filtered.	Numerical data, text sequences, positional data (latitude, longitude)	Modified version of Data Flow #7. Validated, transformed version of on-demand transit service flex data.	File-Metadata: JSON Data files: zip file consisting of GTFS-Flex V2 CSV files Log data: CSV or JSON
Validated Transit Pathways Data (Data Flow #8)	Information on pathways within a transit station and other attributes. These data are validated and may be filtered.	Numerical data, text sequences, positional data (latitude, longitude)	Modified version of Data Flow #7. Validated, transformed version of transit station pathways data.	File-Metadata: JSON Data files: zip file consisting of GTFS-Pathways CSV files Log data: CSV or JSON

Dataset Title	Description	Type	Collection Method	Data Formats
Request Geographic Data (Data Flow #4 for sidewalks and Data Flow #12 for transit service and/or stations)	End-user-initiated requests for trip information. Information sent to TDEI as a result of these requests will include a bounding box or other geographic area specification (e.g., a polygon) and other filtering attributes such as minimum confidence level. Query is made to both the OpenSidewalks Service (Data Flow #4) and/or the GTFS service (Data Flow #12) in accordance with user's preferences.	Numerical data, text sequences, positional data (latitude, longitude)	Requests generated by external parties. Data specified as parameters to the API.	File-Metadata: JSON with geographic data in GeoJSON format Log data: CSV or JSON
Filtered Request Geographic Data (Data Flow #3 for sidewalks and Data Flow #11 for transit service and/or stations)	End-user-initiated requests for trip information. Information sent to TDEI as a result of these requests will include a bounding box or other geographic area specification (e.g., a polygon) and other filtering attributes such as minimum confidence level. Data are filtered via the Application Program Interface (API). Data requests are made from the service to the data repository for data of relevance.	Numerical data, text sequences, positional data (latitude, longitude)	Modified version of Data Flow #4 or Data Flow #12, respectively.	File-Metadata: JSON with geographic data in GeoJSON format Log data: CSV or JSON

Dataset Title	Description	Type	Collection Method	Data Formats
Response Geographic Data (Data Flow #5 for sidewalks and Data Flow #9 for transit service and/or stations)	Geospatial sidewalk, transit stations pathways, and on-demand transit flex data, produced in response to a query and provided to the service.	Numerical data, text sequences, positional data (latitude, longitude)	Requests generated by external parties. Data specified as parameters to the API	File-Metadata: JSON Data files: gzipped GeoJSON file, Zip file consisting of GTFS-Flex V2 CSV files or GTFS-Pathways CSV files Log data: CSV or JSON
Filtered Response Geographic Data (Data Flow #6 for sidewalks and Data Flow #10 for transit service and/or stations)	Geospatial sidewalk, transit stations pathways, and on-demand transit flex data, produced in response to a query and provided to the service.	Numerical data, text sequences, positional data (latitude, longitude)	Modified version of Data Flow #5 or Data Flow #9, respectively. Captured as part of response log to a specific query.	File-Metadata: JSON Data Files: gzipped GeoJSON file, Zip file consisting of GTFS-Flex V2 CSV files or GTFS-Pathways CSV files Log data: CSV or JSON
Station and Service Information	Information about transit stations and transit service providers (e.g., transit agencies).	Numerical data, text sequences, positional data (latitude, longitude)	Manually collected by TDEI staff.	CSV

1.6.2 System Administration Datasets

Table 3 describes the system administration datasets. These datasets are new in the Phase 2 DMP and include data collected in the process of managing and monitoring the system. These data will be used for system administration as well as for the independent evaluation of the system as described in Section 2.6.

System administration data also includes user account data, which does contain PII and is managed by the authorization service. PII data will not be provided to the USDOT or the USDOT IE Team. The system administration data sets are further described in Section 1.5.3.

Table 3. Transportation Data Equity Initiative System Administration Datasets.

Dataset Title	Description	Type	Collection Method	Data Formats
System Administration Data	Metrics associated with measuring uptime performance, quantity of registered users, and other system health-related metrics.	Numerical data, text sequences	Collected through the monitoring service which logs system operation and monitors system health. This includes logging of user accounts as part of the authorization service.	JSON or CSV
User Account Data	Information about user accounts which are required for uploading and downloading TDEI data.	Text data – email address.	Collected from the user upon user registration through a registration form.	CSV

1.6.3 UW Local System Evaluation Datasets

Table 4 describes the system evaluation datasets. The UW local system evaluation datasets consist of trace data from users of AccessMap Multimodal and survey data from users of AccessMap Multimodal, Audiom, and Soundscape. These data contain PII and will be used by the UW Team for their local evaluation and will not be provided to the USDOT or the USDOT IE Team. Summaries of these data, free of PII, will be provided to the USDOT IE Team. The UW local system evaluation data sets are further described Section 1.5.4.

Table 4. Transportation Data Equity Initiative System Evaluation Datasets.

Dataset Title	Description	Type	Collection Method	Data Formats
Stakeholder Feedback Surveys	Stakeholder feedback to structured survey questions on the use of the demonstration applications.	Numerical data, text sequences	Stakeholder feedback surveys will be done with a tool called Qualtrics, which is an online survey tool that can output data in various common file types.	JSON or CSV
AccessMap Multimodal UW Local System Evaluation Data	User-specific routing and trip data that aims to show the ability of a user to make a successful trip with the TDEI System data present via AccessMap Multimodal. These data will be collected for the use of the UW team following UW IRB protocols and will be stored in the Evaluation Data Repository. No PII will be provided to the USDOT.	Numerical data, text sequences	Collected automatically as part of a data logging service in connection with testing use of AccessMap Multimodal, with a focus on trip performance. User interactions (i.e., what was the request, what route received) is output in JSON. Traces will use e-mission (or a similar application), which has a schema-less data store, with a user identifier to cross-correlate between tracking and traces. *Note: These data are considered private and will not be shared without IRB approval.*	JSON; e-mission data store format

2 Data Stewardship

This section provides details regarding data stewardship for the TDEI System. Data stewardship involves proper data management throughout the data's lifecycle, including maintenance of data quality and safeguards. This section discusses the data owner and stewardship; respective access levels; and policies regarding reuse, redistribution, and derivation.

This section focuses on data stewardship in the context of the TDEI System for the duration of the UW ITS4US Deployment Project. Data stewardship covered in this section applies only to data that are contributed to the data repository for use in the TDEI System.

2.1 Data Owner and Stewardship

This section defines the data owner and data steward for the datasets that will be part of the TDEI System. Three roles are defined for the ownership and stewardship of the TDEI data:

- **Data Owner:** The anticipated dataset owner, which is the person or organization with the authority, ability, and responsibility to access, create, modify, store, use, share, and protect the data.
- **Data Steward:** The anticipated dataset steward, which, at the direction of the data owner, is the person or organization that is delegated the privileges and responsibilities to manage, control, and maintain the quality of a data asset throughout the data lifecycle. The data steward also may apply appropriate protections, restrictions, and other safeguards, depending on the nature of the data, subject to the direction of the data owner.
- **Federal Sponsor:** The USDOT sponsor for the datasets.

For all TDEI data sets, which are listed in Table 2, Table 3, and Table 4 the following roles apply.

- **Data Owner**: The **University of Washington** will be the owner of all data collected, processed, and distributed as part of this USDOT-funded research project.
- **Data Steward**: The **Washington State Transportation Center (TRAC)** at the University of Washington and the **TCAT** will be the data stewards, composing the primary group developing the TDEI System and coordinating with associated third-party entities.
- **Federal Sponsor**: The **Intelligent Transportation Systems Joint Programs Office (ITS JPO)** is the USDOT sponsor for these datasets.

2.2 Data Storage and Retention

This section describes the data storage systems that will be used to store the project's data, the details of those data systems, and the duration for which data will be stored in each system. At a high level, three data storage systems will be required to support the needs of this initiative:

1. **TDEI System Data Repository:** This repository represents the location where data will be stored to facilitate primary administration and operation of the TDEI System. As the central focal point, it will ingest nearly all the identified datasets as part of its operation. Data stored in the TDEI System's data repository will include public datasets, but storage will still be controlled. This data repository will include user accounts for managing access to data and enabling communication with users. The data related to these user accounts is a restricted dataset and will have safeguards in place to protect user privacy. Further details on management of this restricted data are discussed in the Phase 2 DPP.

2. **Evaluation Data Repository:** This repository represents the location where the UW local system evaluation data being collected from forty voluntary participants using AccessMap Multimodal will be stored. This repository is separate from the TDEI System's data repository and will include sensitive datasets that qualify as PII; therefore, the evaluation data repository will be access controlled so that only staff named in the IRB documents will have access to it. Permissions will be controlled using the UW permission system that is regularly used to safeguard access to other critical research data sets that involve human use subjects. This data also stores stakeholder survey data. Further details on access control are discussed in the Phase 2 DPP.

3. **USDOT Data Repository:** This system is the USDOT ITS Data Hub and represents the USDOT-managed data storage for purposes of providing allowable data to the public or to approved users. It may include all non-restricted datasets that are provided as part of this project by the TDEI System's data repository or the evaluation data repository. Any data stored in the USDOT-managed storage systems will include only data sets that are considered public datasets. Restricted datasets with PII will not be submitted to this repository, but rather will be available on UW systems to approved parties that meet the IRB approval requirements. To minimize unnecessary data duplication, USDOT and UW are exploring opportunities to federate the data, where interested parties looking for this data in the ITS Data Hub will simply be pointed to the public datasets available from the TDEI System via APIs. This federation would significantly reduce the data stored on the USDOT Data Repository. These details are part of ongoing coordination between the TDEI team and USDOT.

Further details on these storage systems are discussed in the following sections.

2.2.1 Data Storage Systems Overview

Table 5 lists information about each of these three storage systems, including the datasets anticipated to be stored in each system, the initial storage date (based on the current TDEI approach envisioned in Phase 2 of this project), the estimated update frequency, and the archiving and preservation period.

The datasets to be shared with the USDOT include all validated data and all API requests and responses from and to external parties. The unvalidated data and the internal/filtered API requests and responses are not listed in the data sets to be provided to and federated with the USDOT. It is anticipated that the USDOT would be primarily interested in the validated data and the external API requests. In addition, this data will be available via TDEI dashboards and API requests. The unvalidated data are not expected to be available via API requests. If USDOT or the USDOT IE Team wishes to obtain these data, the data can be provided by the TDEI team.

Table 5. Data storage systems – Data, Initial Storage, Updates, Archiving.

Data Storage System Name	Dataset Titles	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
TDEI System Data Repository	• System Administration Data • User Account Data • Sidewalk Data • Validated Sidewalk Data • Transit Flex Data • Transit Pathways Data • Validated Transit Flex Data • Validated Transit Pathways Data • Request Geographic Data • Filtered Request Geographic Data • Response Geographic Data • Filtered Response Geographic Data • Station and Service Data	Mock data and test data to be stored starting November 2022. This data will be deleted from the system after testing is complete. Storage of actual data begins after system acceptance testing.	Continuously, as data are submitted.	5 years after completion of Phase 3.
Evaluation Data Repository	• Stakeholder Feedback Surveys • AccessMap Multimodal Evaluation Data	Storage begins after system acceptance testing.	Continuously, as data are submitted.	5 years after completion of Phase 3.

Data Storage System Name	Dataset Titles	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
USDOT Data Repository	• System Administration Data (excluding user account data) • Evaluation Data (non-restricted datasets only) Datasets accessible via federation: • Validated Sidewalk Data • Validated Transit Flex Data • Validated Transit Pathways Data • Request Geographic Data • Response Geographic Data Additional datasets provided to USDOT on request	Storage begins after system acceptance testing.	Data are available continuously through dashboards and APIs.	5 years after completion of Phase 3

2.2.2 Data Storage Systems Description

This section describes the two data storage systems that will be maintained by the TDEI team – the TDEI System Data Repository and the Evaluation Data Repository. The intent of this section is to provide information on the technical configuration of these storage systems.

2.2.2.1 TDEI System Data Repository

The first row of Table 5 lists the titles of the datasets that will be stored in the TDEI System Data Repository. Reviewing that list in the first row of Table 5 in conjunction with the tables that describe those datasets (Table 2, Table 3, and Table 4) leads to the observation that the data which will be stored in the TDEI System Data Repository is of varied types and structures. Table 6 provides information on data storage formats for the various types of data. Whereas Table 2, Table 3, and Table 4 describe the datasets themselves, Table 6 is organized by data type and structure. In addition, Table 6 provides a mapping from the data type and structure to the list of datasets in Table 5.

The types of data stored in the TDEI System repository organized by data type and structure can be summarized as follows:

- **Data files:** zip or gzipped files containing GTFS-Flex files, GTFS-Pathways, and/or OSW files.
- **File-metadata:** As described in Section 1.5.1 the term file-metadata are defined in the context of this document as metadata that provides auxiliary information about the data files, such as file valid dates and confidence level, is stored in the TDEI. These file-metadata are conceptually name-value pairs – for example: “valid_from”=”2022-01-01” and “valid_to” =”2022-01-31”.
- **Data flow logs:** The data flows, shown in Figure 3 and described in Table 2, are all logged. In practice, these data are formed of logs of external API calls and responses (Data Flows 1, 4, 5, 7, 9 and 12) and internal system communications (Data Flows 2, 3, 6, 8, 10 and 11), which are collected through the TDEI monitoring process.
- **System Administration Data:** This data consists of metrics associated with measuring system performance and is collected through the TDEI monitoring process.
- **User Account Data:** This data consists of information about user accounts which are required for uploading and downloading TDEI data. The data conceptually can be thought of as name-value pairs – for example: username=”someuser” and email=”someuser@email.com.”
- **Station & Service Data:** This data consists of metadata about transit stations and on-demand services with data stored in TDEI, which supplements the information in the file-metadata and the data files. This data conceptually can be thought of as name-value pairs.

Table 6 Storage in TDEI System Data Repository

Data Type	Structure	Associated Datasets	Storage Method
Data Files	zip and gzipped files consisting of CSV and GeoJSON files. (GTFS-Pathways, GTFS-Flex and OSW files)	• Sidewalk Data • Validated Sidewalk Data • Transit Flex Data • Transit Pathways Data • Validated Transit Flex Data • Validated Transit Pathways Data	Blob storage. Azure Blob storage in TDEI.
File-metadata	Name-value pairs providing supplemental information about data files.	• Sidewalk Data • Validated Sidewalk Data • Transit Flex Data • Transit Pathways Data • Validated Transit Flex Data • Validated Transit Pathways Data	Relational database. Managed PostgreSQL database in TDEI.

Data Type	Structure	Associated Datasets	Storage Method
Data flow logs	Audit logs of API calls from external sources and logs of internal communications between microservices. Audit logs have a tabular structure.	• Sidewalk Data • Validated Sidewalk Data • Transit Flex Data • Transit Pathways Data • Validated Transit Flex Data • Validated Transit Pathways Data • Request Geographic Data • Filtered Request Geographic Data • Response Geographic Data • Filtered Response Geographic Data	Centralized logging service. Microsoft AppInsights in TDEI.
System Administration Data	Audit logs of metrics for system performance monitoring. Audit logs have a tabular structure.	• System Administration Data	Centralized logging service. Microsoft AppInsights in TDEI. TDEI will have a single logging service which will be used for logging data flows and System Administration data.
User Account Data	Name-value pairs of user account information.	• User Account Data	Relational database. Managed PostgreSQL database in TDEI. For security purposes, this will be a separate instance of a relational database from the instance used to store the File-metadata.

Data Type	Structure	Associated Datasets	Storage Method
Station & Service Data	Name-value pairs of supplemental information about transit stations and on-demand services.	• Station and Service Data	Relational database. Managed PostgreSQL database in TDEI. This will be the same instance of a relational database from the instance used to store the File-metadata.

The TDEI System Data Repository is part of the TDEI System and is a cloud-based system using a microservices architecture, as further described in the Phase 2 SAD and the Phase 2 System Design Document (SDD). As the TDEI System is using a microservices architecture, several independent microservices will be used to store the data listed above. Table 6 provides for each of the types of data listed above: the structure of that data, the datasets associated with that data type, and the storage mechanism that will be used for that data. The intent of including the associated datasets in this table is to provide a mapping between these data types and the datasets. For storage method, the table provides a general name of the storage method and the specific storage solution that the TDEI is using in its Azure cloud environment.

The TDEI System Data Repository will house the user account data, which contains limited PII in the form of user information – primarily user emails. In terms of security, the user account data will be processed by a dedicated microservice – the authorization service – and will be in a dedicated database with strict access controls and encrypted storage. Details of security measures are provided in the Phase 2 DPP.

2.2.2.2 UW Local System Evaluation Data Repository

The UW local system evaluation data repository, shown in Table 4 sabove, contains data describing trips being made by forty volunteer participants that use the AccessMap Multimodal application. The data describe when, where, how, and how successfully trips are made by participants. The repository includes trace data from trips being made, as well as data about the trips which are being performed to make those traces. The repository will be created as part of the UW local evaluation, in order to demonstrate and evaluate the mobility benefits that can be gained from use of the TDEI data in applications such as AccessMap Multimodal. The repository will be a file storage system on a server housed in the University of Washington Computer Science and Engineering's Information Technology networks. The university routinely stores data that includes various levels of sensitivities for critical research studies, and the evaluation data repository will be configured in accordance with university policies for storage of PII data. Access will be controlled using OAuth-2.0 [16] authorization methods via University of Washington or equivalent accounts with access limited to TDEI staff and collaborators with IRB approval to access the data. The UW Team will ensure that this storage is created by the Information Technology services with levels of security consistent with sensitive research data.

2.3 Data Sharing Framework

The UW Team will support the USDOT's IE Team throughout the project. This support includes providing USDOT staff with access to a variety of data sets, programmatic information, and UW Team staff. The data available to the USDOT IE Team will include all pathways (GTFS-Pathways), on-demand transit (GTFS-Flex), and sidewalks (OSW) data stored in the TDEI System, the metadata for these files, summarized system evaluation data cleaned of PII, and system administration data, except for user account data. Additional data requests from USDOT staff will be accommodated as is possible.

The UW Team will create a dashboard that will make performance measures data (generated from system operation and administration data) available to the USDOT and the USDOT IE Team. Additionally, as mentioned earlier in Section 2.2, operational data from TDEI will be federated; USDOT IE Team members going to the USDOT Data Repository will be linked to the UW Team's data repository, where public data will be available for consumption. UW local system

evaluation data stored on the UW Team's data repository will not be available to the USDOT IE Team or USDOT staff. Researchers outside the UW Team may be granted access to raw data that includes sensitive PII data by going through the IRB approval process.

2.4 Data Quality Control

As a data-focused project, many steps to preserve data quality inherently exist within the TDEI System's scope and design. On the data producer side of the system, all contributions to the system go through a vetting process that, depending on the development of standards, may involve automated processes to determine data completeness and automated or manual processes to verify accuracy against real-world conditions. Part of the evaluation effort includes a focus on how successful a data producer can be with submitting data, which involves a careful vetting of contributions necessary to preserve data quality. The data vetting process is both a key task in the project's quality control and quality assurance effort, and it is a way for providing direct value to the agencies and organizations participating in this project. For agencies that own infrastructure or provide transportation services, the data vetting process helps ensure that data being maintained and published about their infrastructure or services is correct. In many cases, this will allow the agency to correct data errors in their own databases. Over time, some of these validation processes may become automated, potentially through a versioning schema, such as Semantic Versioning 2.0.0 [20].

On the data consumer side of the system, data quality control is important to ensure that data consumers receive data that is timely, appropriate, and reliable to aid in informed trip-making. The API's design will distribute a structured dataset that provides data in response to a query. This design is being vetted with various stakeholder groups as part of design, and its performance is directly measurable as part of the evaluation efforts to determine how useful the data are for providing travelers with information.

2.5 Privacy

The UW ITS4US project team will employ various technical and policy controls to protect user privacy. As noted earlier, the TDEI System itself will likely house PII data in the form of data collected for user accounts, which will include at a minimum email address. Another location of potential PII is within the system evaluation data, specifically among the datasets of evaluation data collected from the forty voluntary participants (discussed earlier in Section 1.3). Privacy controls focus on both data sets, which would be classified as Private (Research): Data that cannot be shared with external users. Access to these data are limited and only granted with IRB and Project Team approvals.

User account data are stored as part of the TDEI System's authorization service, which is its own microservice and hence is separate from other microservices. As such, user account data will reside in that authorization service and will be used, but not stored by, the TDEI System's notification service (i.e., in the event that a user account holder needs to be notified of API changes or other system updates), but the user account data will not be intertwined with other operational data that is passing through the TDEI System. User account holders will not have access to private data affiliated with other user account holders.

UW local system evaluation data collected from the forty participants will be stored in a system that is separate from the TDEI System, so as to not unnecessarily place additional PII or SPII data into the TDEI System and create additional restrictions of data use, particularly with respect to the operational data that is intended to be publicly available and is expected to be of interest for research efforts outside of the TDEI and additional third-party application development. Evaluation data will be placed in a directory that is only accessible to pre-approved individuals that are part of the TDEI staff, have a purpose and role in evaluating the system, and have IRB approval to access the data. Access permissions will be controlled using the UW's Information Technology (IT) permission system, which is consistent with practices adopted for other studies that involve human subjects. All data in the analysis dataset will fall under the same protocols until a decision to share information outside of the core research group is made, at which point access would be restricted to clean copies of the data which would not contain PII or SPII. IRB protocols will be followed—as discussed in the Phase 1 HUAS—to provide access to approved individuals.

Further information on data privacy can be found in the Phase 2 DPP.

2.6 Relationship to Performance Measures

As discussed in the Phase 1 PMESP, the evaluation can be divided into five major data reporting categories:

- Data Standards: The effectiveness and acceptance level of the three data standards being used by TDEI: GTFS-Flex v2, GTFS-Pathways, and OSW.
- Data Generation: The amount and quality of the data developed as part of the project.
- Data Vetting: The performance of the data vetting systems put in place.
- Data Sharing: The performance of the data sharing system, which makes the data accessible to third-party applications via application programming interfaces (APIs).
- Demonstration Applications: The performance of the demonstration applications, in terms of their ability to access and deliver the data being collected in this project, and for AccessMap Multimodal, the travel outcomes achieved by delivering that data to users.

Data to support these reporting efforts will be collected through stakeholder surveys, through AccessMap Multimodal trace evaluation data, through data logged as part of operational data collection, and through monitoring of access and usage of the TDEI System. The TDEI will have a dedicated microservice – the monitoring service – which will track user registrations and API keys issues, API calls, internal system communications and metrics related to system performance and health. The TDEI monitoring service will be able to track data errors which are identified by the TDEI's automated data validation service. In the cases where human review is required, for instance when there are conflicts between sidewalk data sets, the system used by the human reviewers will track data errors. The TDEI monitoring service provides both the data flow logs for the system operation data, described in Sections 1.5.2 and 1.6.1, and the system administration data described in Sections 1.5.3 and 1.6.2.

The specific metrics that will be collected by the TDEI monitoring service include:

System Administration Data – API Usage & Users:

- API keys issued: dates/times/entities/number/IP to track user and system actions.
- API calls to create users

System Administration Data – API data flow logs:

- API calls made (by date, time, user/key, data type/schema); every API request to the gateway will be logged
- API calls for data ingest (i.e., file submissions) will be tracked; information to be tracked includes data schema, acceptance or rejection, and reasons for rejection
- API calls for data requests will be tracked; information to be tracked includes data schema, parameters for requests, accept/reject/response, and reasons for rejection

System Administration Data – System Health:

- Uptime/downtime percentages for the TDEI data service
- Uptime/downtime percentages during peak travel periods

System Administration Data – Performance/Load Testing

- Latency and performance of API calls will be tested under increasing numbers of concurrent users and increasing data size

Table 7 shows the relationships between the reporting efforts and the primary performance measure objectives and descriptions, as well as the data sets that will be used to meet those reporting objectives. Note that Table 7 shows a summarized version of the performance measures table from the Phase 1 PMESP. The performance measure summary in Table 7 is intended to capture the key objectives from the PMESP.

Table 7: Relationship to Performance Measures

Reporting Effort Category	Objective	Description	Relevant Datasets
Data Standards	Data standards meet user needs, including being objective and are accepted by community	Surveys with a) users with lived-experience, b) OSW transit data generators, c) GTFS-Pathways data generators, d) on-demand transit service data generators, e) application developers.	Stakeholder Feedback Surveys
Data Generation	Data are widely available.	Volume and geographical coverage of data submissions for OSW data, GTFS-Pathways data, and GTFS-Flex data. Logs of API calls for data ingest of relevant data flows.	Validated Sidewalk Data (Data Flow #2) Validated Transit Flex Data and Validated Transit Pathways Data (Data Flow #8)
Data Generation	Data quality is high, data generation tools work well	Logs of API calls for unvalidated data and validated data.	Sidewalk Data and Validated Sidewalk Data (Data Flows #1 and #2) Transit Flex Data, Transit Pathways Data, Validated Transit Flex Data and Validated Transit Pathways Data (Data Flows #7 and #8)
Data Generation	Data are uploaded to a central data center efficiently	Latency and performance of API calls. Logs of API calls for data ingest of relevant data flows.	System Administration Data Validated Sidewalk Data (Data Flow #2) Validated Transit Flex Data and Validated Transit Pathways Data (Data Flow #8)
Data Vetting	Agencies participate in the data vetting process	Logs of API calls to create users and logs of API keys issued.	System Administration Data

Reporting Effort Category	Objective	Description	Relevant Datasets
Data Sharing	API performance and availability is good	Latency and performance of API calls. Logs of API calls for data requests of relevant data flows.	System Administration Data
Data Sharing	API availability is good	Uptime/downtime percentages for the TDEI data service.	System Administration Data
Data Sharing	3rd party developers are participating	Logs of API calls to create users and logs of API keys issued. Logs of API calls for data requests for relevant data flows.	System Administration Data Request Geographic Data (Data Flow #4 for sidewalks and Data Flow #12 for transit service and/or stations) Filtered Request Geographic Data (Data Flow #3 for sidewalks and Data Flow #11 for transit service and/or stations) Response Geographic Data (Data Flow #5 for sidewalks and Data Flow #9 for transit service and/or stations) Filtered Response Geographic Data (Data Flow #6 for sidewalks and Data Flow #10 for transit service and/or stations)
Data Sharing	Data security is strong	Number of attempted/detected cyber-attacks. Number of security breaches.	System Administration Data

Reporting Effort Category	Objective	Description	Relevant Datasets
Demonstration Applications	Demonstration application performance is good	TDEI data services, pre-selected O/D pairs and known routing alternatives.	AccessMap Multimodal Evaluation Data
Demonstration Applications	User satisfaction associated with the demonstration applications is high; travel outcomes are safe and efficient	Survey of user experiences with the three applications.	Stakeholder Feedback Surveys

2.6.1 Baseline Data

The TDEI System aims to fill a gap that exists in the new mobility ecosystem. As a result, establishing a comparative non-zero quantitative baseline for a before/after assessment is a challenge. For the UW local system evaluation data, the TDEI team is intending to use the e-mission software to obtain before travel data for the AccessMap Multimodal travel behavior outcome evaluation; otherwise, the before condition is "little or no data are available." For the UW local evaluation, the after condition is "the degree to which data are available" combined with "are the available data sufficient (good enough, detailed enough, widespread enough) to meet traveler needs?"

"The degree to which data are available" is key for after data collection since the project is starting from zero in terms of the availability of sidewalk, on-demand transit, and transit center pathway data, it is not necessary to design a before and after comparison, but rather to describe the degree to which the data collection and data delivery has been accomplished, as a key goal of this project is simply to make previously unavailable information available to travelers. This goal assumes that the delivery of that information to those travelers will come in a wide variety of forms, intended to achieve a wide variety of travel benefits, to a variety of different intended audiences, with the majority of that information delivery occurring after the completion of this project. Thus, many of the measurable mobility benefits will occur outside of the project itself, but only if the data exist and can be effectively delivered to future application developers. Similarly, since data that do not currently exist are not being delivered to travelers, understanding whether the data that are collected and delivered as part of this project actually meet user's travel needs is key to understanding whether those data improve mobility outcomes.

For the performance measures described in the PMESP that are beyond the UW local system evaluation and the demonstration application evaluation, and specifically the performance measures based on system administration data collected through the TDEI monitoring system, opportunities exist to evaluate the change in system performance over time. For these performance measures which will be provided to the USDOT IE team, the measurements taken in the first quarter of Phase 3 will serve as the baseline for the project.

3 Data Standards

3.1 Data Standards

Table 8 shows the data standards that are anticipated to be used as part of the TDEI System. This table provides descriptions of the following items:

- **Dataset Title:** The title of the dataset.
- **Data Standard(s):** The data standard anticipated to be used by the dataset.
- **Data Standard(s) Digital Object Identifiers (DOI):** The DOI(s) of the standard(s) for the data, if available (marked “N/A” if not available).
- **Open or Proprietary:** Whether the data standard is considered open or proprietary.
- **Data Standard(s) Rationale:** The rationale for using the chosen data standard for the dataset in question.

Applicable data standards apply to the data that is part of the TDEI System operation. System administration data and evaluation data do not follow a formal data standard, as these data are of a custom design that is unique to the TDEI System. Geographic data made in response to requests to the API made by data consumers (referred below as “response” data) will be done through REpresentational State Transfer (REST)ful [19] APIs, which is a best practice and will provide consistency between the APIs for each data format.

Table 8. Data standards.

Dataset Title	Data Standard(s)	Data Standard(s) DOI(s) or URL(s)	Open or Proprietary	Data Standard(s) Rationale
Sidewalk Data and Validated Sidewalk Data	OpenSidewalks	https://wiki.openstreetmap.org/wis/sidewalk_schema	Open	OpenSidewalks provides graph-network attributes that objectively describe the pedestrian built-environment.
Transit Flex Data and Validated Transit Flex	GTFS-Flex v2	https://github.com/MobilityData/gtfs-flex	Open	This proposed extension builds upon GTFS, GTFS is widely adopted among transit agencies, the Flex extension is gaining acceptance with transit agencies.
Transit Pathways Data and Validated Transit Pathways Data	GTFS-Pathways	https://developers.google.com/transit/gtfs/reference#pathways.txt	Open	These proposed extensions build upon GTFS. GTFS is widely adopted among transit agencies, the Pathways extension is gaining acceptance with transit agencies.
Response Geographic Data and Filtered Response Geographic Data	OpenSidewalks; GTFS; GTFS-Flex Extension; GTFS-Pathways Extension and other affiliated extensions	See above	Open	These data standards describe the sidewalk and transit service/station information available for a requested location.

3.2 Versioning

Data versioning is an important to the part of the TDEI System, as this system is expected to receive operational data – specifically GTFS-Flex, GTFS-Pathways, and OSW data files – on an ongoing basis and, as such, cataloguing which operational data are valid for which time periods is important to ensure that data consumers are receiving the most relevant versions of data. In addition, the TDEI System itself, particularly its APIs, may be updated to accommodate new features or an update to the various data standards. Finally, the data standards on which the TDEI relies may be updated. While these standards are managed by the various standards committees, updates to those standards will impact the TDEI System. Registered user accounts will be used to enable TDEI staff to communicate updates to the users, such as API updates or system updates.

3.2.1 Operational Data Updates

Versioning for new and updated GTFS-Flex, GTFS-Pathways, and OSW data files will be handled by a combination of valid dates for the files and a valid/invalid flag. The system will enforce the requirement that there is only one valid file for a given time period and a given transit station, on-demand service, or geographic sidewalk area. For example, there will be only one valid GTFS-Pathways file for the Northgate transit station for June 1, 2022 through June 15, 2022. This policy is consistent with current GTFS practices and avoids confusion for users that may occur if multiple files were considered to be valid for a single time period.

To manage this requirement and to allow file updates, the file-metadata for a particular file will contain a “valid_from” date, a “valid_to” date, and a valid/invalid flag. Note that there is one file-metadata record for each data file uploaded into the TDEI System and that the file-metadata record also contains a link to the location where the file is stored. As an example, when a data generator wants to upload a new GTFS-Pathways file for a transit station, they specify “valid_from” and “valid_to” dates for that file as part of the API call, and the file is loaded into the system with those dates. If the valid range of the file conflicts with the valid range of another valid file for that station, the user is given an error. If a data generator wants to replace a file in the TDEI System, they will use an API call to set the flag for the file to be replaced as invalid and then they can upload a new file.

Thus, the operational data files (GTFS-Flex, GTFS-Pathways, and OSW) are versioned by keeping track of valid time ranges for each file and by allowing data generators to invalidate files previously uploaded.

3.2.2 TDEI API Updates

TDEI APIs may be updated during the project period based on stakeholder and API user feedback. To facilitate updates, the TDEI APIs have been designed with a version number in anticipation of updated API versions so that users can be aware of which version of the API they are using. As mentioned earlier, TDEI users must be registered, which allows notifications of updates to be pushed to the respective contact personnel. TDEI APIs are specified using a JSON file in the Open API specification, which is maintained in a TDEI GitHub repository. API documentation is maintained on a Swagger [21] web site and on a GitHub Pages web site, both of which are fed from TDEI GitHub repositories. The version control in the GitHub repositories will

be used to track the version updates to the API and associated documentation. API releases will be tagged with Git tags for reference purposes.

3.2.3 TDEI System Software Updates

TDEI System software will be maintained in the TDEI GitHub repository, which is a public GitHub repository, and GitHub version control and tags will be used to keep track of software versions.

3.2.4 Standards Updates

During the course of the project, the GTFS-Flex, GTFS-Pathways and OSW standards may be updated. The updates to the standards will be managed by the relevant standards organization. GTFS-Flex and GTFS-Pathways are managed by Mobility Data [14], while OSW is managed by the TDEI team. The TDEI team manages OSW similar to other standards by using a GitHub repository for the standard. As with other items, the GitHub version control and tagging mechanisms will be used for versioning of the OSW standard.

3.3 Metadata

Metadata indicate where the data come from, how they should be translated or transformed, their form, and their functions. Metadata that are relevant to the datasets in this TDEI System are outlined in Table 9. This information is anticipated to be further worked out as part of subsequent Phase 2 development (i.e., some updates potentially in later releases), and the proposed metadata listed herein will be updated as further information is garnered.

3.3.1 Metadata Types

There are several different types of metadata that occur in and are related to the TDEI System. For the purposes of this report, we define the following types of metadata.

- **File-metadata:** Metadata that provides supplemental information about the GTFS-Flex, GTFS-Pathways and OSW data files in the TDEI System. This file-metadata may consist of attributes such as “valid_from” and “valid_to” dates, collection method, and confidence metric. Some of the file-metadata are collected from data generators through data ingest API calls, some file-metadata, such as confidence metric, is generated by the TDEI System. File-metadata are provided to data consumers when they download files. File-metadata are conceptually name-value pairs and are stored in a relational database.
- **Standards-metadata:** TDEI relies on three data standards (GTFS-Flex, GTFS-Pathways, and OSW). These data standards provide descriptions of the information and data fields that are contained in GTFS-Flex, GTFS-Pathways, and OSW data files that are ingested into and stored in the TDEI. Standards metadata are maintained on GitHub sites and web pages maintained by standards management organizations. The TDEI team is responsible for managing and updating the OSW standards-metadata, and Mobility Data are responsible for managing and updating the GTFS-Flex and GTFS-Pathways standards-metadata.
- **Descriptive-metadata:** All datasets provided by the TDEI will need data-dictionary type metadata that explains the meaning of fields in datasets. For example, descriptive-

metadata or data dictionaries will be required to describe fields in file-metadata and for system-administration data.

3.3.2 TDEI Datasets Metadata

OpenSidewalks, GTFS-Pathways, and GTFS-Flex files individually have their own metadata, based on design requirements associated with that data schema. In the TDEI System, the APIs also define metadata that is relevant for each file type as part of the data transaction. The University of Washington maintains a list[2] that defines the proposed metadata that is relevant to the various API requests. Metadata are subdivided between queries for API information, queries for each of the three data types that can be exchanged, and the relevant schemas for various API call attributes (e.g., polygon sizes, geometries, login, token response, etc.). Proposed data in the working list are shown in Table 9 in terms of how they relate to the key TDEI System's datasets.

Table 9. Anticipated metadata.

Dataset Title	Anticipated Metadata
• **System Administration Data**	Descriptive-metadata: System Administration Data will need descriptive-metadata to describe the fields in the system administration data sets. This metadata will be maintained on a public GitHub Pages web site.
• **Sidewalk Data** • **Graphed Sidewalk Data**	Standards-metadata: Data collected through the OpenSidewalks data schema will follow the OSM data schema. OSM utilizes element, change set, and user metadata to describe information, such as the username making edits to the network, the number of edits, which editor was used, the commit message, and other data that support analysis of user metrics and insight into contributions. File-metadata: Data producers who submit the files will be required to provide the metadata, which will include information about how and when the data was collected and any valid dates of the file that are applicable. Descriptive-metadata: Descriptive-metadata will be needed to describe the fields in File-metadata.

[2] https://tdei-gateway.azurewebsites.net/swagger-ui/index.html

Dataset Title	Anticipated Metadata
• **On-Demand Transit Service Data** • **Transit Station Pathway and Attribute Data** • **Validated On-Demand Transit Service Data** • **Validated Transit Station Pathway and Attribute Data**	Standards-metadata: GTFS and its associated extensions will utilize existing GTFS metadata resources. The GTFS dataset specification includes various files, including a conditionally required "feed_info.txt." This dataset file will include dataset metadata, such as publisher (name and webpage), version, language codes, and expiration information. Even though the extensions are still in the process of being adopted, it is likely that similar metadata will be adopted. File-metadata: Data producers who submit the files will be required to provide the metadata, which will include information about how and when the data was collected and any valid dates of the file that are applicable. Descriptive-metadata: Descriptive-metadata will be needed to describe the fields in File-metadata.
• **Request Geographic Data** • **Filtered Request Geographic Data**	Descriptive-metadata: The parameters for the data requests will be documented in the TDEI API OpenAPI document. These parameters include parameters like bounding box, minimum confidence level, schema version, time period, and other attributes. Beyond the limited information about the API parameters in the OpenAPI specification document, descriptive-metadata will be required to provide details on the parameters.
• **Response Geographic Data** • **Filtered Response Geographic Data**	File-metadata: The response for the data requests will be documented in the TDEI OpenAPI specification and will include file-metadata fields, including confidence level, valid dates, schema version, and other attributes. Descriptive-metadata: Beyond the limited information about the API parameters in the OpenAPI specification document, descriptive-metadata will be required to provide details on the parameters in the response.
• **Station and Service Data**	Descriptive-metadata: Descriptions of the meanings of the fields.
• **Stakeholder Feedback Survey Data**	Descriptive-metadata: Digitized survey data will include a brief description of the survey and a collection time period.
• **AccessMap Multimodal Evaluation Data**	Descriptive-metadata: Digitized evaluation data collected will include a brief description of the data, phone type, and a collection time period. Any metadata associated with user attributes will be kept in a separate file.

3.3.3 Metadata Structure

The metadata structure for the TDEI System will be posted on a public GitHub Pages web site. This web site will serve as the published resource for all metadata that is directly relevant to the files created within the TDEI System. This includes the latest metadata for administration and operation data. The GitHub Pages site will explain the context of the metadata, similar to a traditional “readme” file. Using a GitHub Pages web site drawn from a TDEI GitHub repository will facilitate metadata updates.

This GitHub Pages web site will include all descriptive-metadata listed in Table 9 and would be categorized by data set. For metadata associated with the data schema standards themselves (i.e., OSW, GTFS-Pathways, and GTFS-Flex), this GitHub Pages site will reference viewers to the associated web sites of each of the standards, maintained by the respective standards bodies. File-metadata are stored in the TDEI System, and the descriptive-metadata needed to interpret that data will be on the TDEI GitHub Pages web site.

The GitHub Pages site will include metadata references to evaluation data, where applicable. As noted earlier in Section 1.6, Qualtrics—an online survey tool used to gather product and customer experience feedback—will be used to gather feedback survey data, and the trace data (private data) will be collected through AccessMap Multimodal using e-mission to correlate user identifiers between tracking and traces. The UW Team will include a reference to published metadata made available for Qualtrics and AccessMap Multimodal, where not limited by private datasets. Metadata for private data sets will protected and require IRB approval for access.

3.3.4 Metadata Update Process

The goal for this project is to adopt a standard metadata set that will be used throughout the duration of the project to avoid any discrepancies in data that are reported. If a metadata update is necessary (adding, removing, or changing) after the system has entered a demonstration period and is logging relevant data, the following process will be followed:

1. A project log will note the proposed changes to the metadata record with supporting reasons for making the change. This log will be added to the project record.
2. The metadata record(s) will be adjusted in accordance with the proposed changes, with the noted date of change and any related documentation will be updated. Approved API users that might be affected by metadata changes would be notified of changes.
3. As applicable, updated metadata will be stored in the data repository as applicable.
4. Metadata changes will be reflected in an updated version of the DMP, where applicable.

Appendix A. Acronyms and Glossary

Table 10. Acronyms

Acronym	Definition
API	Application Programming Interface
ConOps	Concept of Operations
CSV	Comma Separated Value
DMP	Data Management Plan
DOI	Digital Object Identifier
DOT	Department of Transportation
DPP	Data Privacy Plan
GTFS	General Transit Feed Specification
HTTPS	Hypertext Transfer Protocol Secure
HUAS	Human Use Approval Summary
IE	Independent Evaluator
IRB	Institutional Review Board
IT	Information Technology
ITS	Intelligent Transportation System
ITS JPO	Intelligent Transportation System Joint Program Office
JSON	JavaScript Object Notation
LiDAR	light detection and ranging
OSW	Open Sidewalks
PII	Personally Identifiable Information
PMESP	Performance Measurement and Evaluation Support Plan
SAD	System Architecture Document
SDD	System Design Document
SPII	Sensitive Personally Identifiable Information
Taskar Center or TCAT	Taskar Center for Accessible Technology at the University of Washington
TDEI	Transportation Data Equity Initiative
TRAC	Washington State Transportation Center
U.S.	United States
USDOT	United States Department of Transportation
UW	University of Washington

Appendix B. References

1. Caspi, Anat, et al., Phase 1 ConOps, University of Washington ITS4US Deployment Project, Final Report—June 28, 2021, Report Number FHWA-JPO-21-861. https://rosap.ntl.bts.gov/view/dot/58675 (Accessed Nov 9, 2022)
2. Caspi, Anat, et al., Phase 1 Data Management Plan (DMP), University of Washington ITS4US Deployment Project, Final Report—August 23, 2021, Report Number FHWA-JPO-21-869. https://rosap.ntl.bts.gov/view/dot/60117 (Accessed Nov 9, 2022)
3. Caspi, Anat, et al., Phase 1 Performance Measurement and Evaluation Support Plan (PMESP), University of Washington ITS4US Deployment Project, Final Report—November 15, 2021, Report Number FHWA-JPO-21-879. https://rosap.ntl.bts.gov/view/dot/60781 (Accessed Nov 9, 2022)
4. Caspi, Anat, et al., Phase 1 Human Use Approval Summary (HUAS), University of Washington ITS4US Deployment Project, Final Report—February 16, 2022, Report Number FHWA-JPO-21-899. https://rosap.ntl.bts.gov/view/dot/62488 (Accessed Nov 9, 2022)
5. Caspi, Anat, et al., Phase 2 System Architecture Document (SAD), University of Washington ITS4US Deployment Project, Draft Report—October 24, 2022, Report Number FHWA-JPO-22-984.
6. Caspi, Anat, et al., Phase 2 Data Privacy Plan (DPP), University of Washington ITS4US Deployment Project, Draft Report—September 30, 2022, Report Number FHWA-JPO-22-972.
7. e-mission-server https://github.com/e-mission/e-mission-server (Accessed Dec 21, 2022)
8. GeoJSON. https://geojson.org (Accessed Nov 17, 2022).
9. “GTFS Flex V2”, Nov 9, 2021. https://github.com/MobilityData/gtfs-flex/blob/master/spec/reference.md (Accessed Sept 6, 2022)
10. GTFS Pathways https://developers.google.com/transit/gtfs/reference#pathwaystxt (accessed Nov 17, 2022)
11. GTFS Schedule Reference”, May 9, 2022, https://gtfs.org/schedule/reference/ (Accessed September 6, 2022)
12. Introducing JSON. https://www.json.org/json-en.html. (Accessed on November 11, 2022).
13. ITS Data Hub https://www.its.dot.gov/data/ (Accessed Nov 17, 2022)
14. Mobility data https://mobilitydata.org (Accessed Nov 17, 2022)
15. NREL OpenPATH: Open Platform for Agile Trip Heuristics https://www.nrel.gov/transportation/openpath.html (Accessed Dec 21, 2022)
16. OAuth 2.0 https://oauth.net (Accessed Nov 18, 2022)
17. OpenSidewalks. https://www.opensidewalks.com (Accessed Nov 9, 2022)
18. Qualtrics. https://www.qualtrics.com (Accessed Dec 15, 2022)

19. What is REST. https://restfulapi.net (Accessed Dec 15, 2022)
20. Semantic Versioning 2.0.0. https://semver.org/. (Accessed Nov 21, 2022)
21. Swagger. https://swagger.io (Accessed Nov 21, 2022)
22. TDEI API Definition. Working Copy. https://tdei-gateway.azurewebsites.net/swagger-ui/index.html#/ (Accessed Nov 9, 2022)

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

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

FHWA-JPO-22-976

U.S. Department of Transportation