Phase 2 Data Privacy Plan

University of Washington ITS4US Deployment Project

www.its.dot.gov/index.htm

Final Report — November 22, 2022 FHWA-JPO-22-972





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

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

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

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-JPO-22-972		
4. Title and Subtitle		5. Report Date
Phase 2 Data Privacy Plan		November 22, 2022
University of Washington ITS4US Depl	oyment Project	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 Add	dress	10. Work Unit No. (TRAIS)
The Taskar Center for Accessible Technology	0;	N/A
Department of Computer Science & Engineering University of Washington		11. Contract or Grant No.
Box 352350 Seattle, WA 98195-2350		693JJ321C000004
12. Sponsoring Agency Name and Address		13. Type of Report and Period Covered
U.S. Department of Transportation ITS Joint Program Office		N/A
1200 New Jersey Avenue, SE		14. Sponsoring Agency Code
		HOIT-1

15. Supplementary Notes

Kate Hartman, AOR

16. Abstract

This document is the Data Privacy Plan (DPP) 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 DPP provides details about how the privacy of participants in the UW's ITS4US project will be protected. The DPP describes actions that will be taken to protect the privacy of users, guard against potential breaches of the system, and prevent unauthorized use of sensitive participant data, specifically Sensitive Personally Identifiable Information (SPII) and other Personally Identifiable Information (PII). Much of this DPP is informed by the Phase 1 Data Management Plan (DMP), with updates made based on design developments that occurred between Phase 1 and Phase 2. The TDEI system can meet the requirements of protecting user privacy because the TDEI system itself does not collect PII and SPII as part of its regular data transactions; for those transactions, the PII is exchanged between the digital device end users and the application developer. This data flow exists outside of the TDEI system that is being developed as part of the USDOT ITS4US program. The primary instance of PII in the TDEI project is data collected from the 40 participants that are being recruited—through an informed consent process—to demonstrate use of the TDEI system through a separate mobile application for the purpose of evaluating the TDEI. Limited PII – primarily an email address – also exists in the TDEI system's authorization service. The TDEI requires users to register and create a user account to access data in the TDEI; there are no restrictions on who can receive a user account. The primary purpose of requiring users to create accounts is so that TDEI can communicate with users and notify them of system updates and manage accounts that violate TDEI usage policies. This limited PII data will reside within the TDEI system, but will be stored separately from the operational data, and will be managed with privacy in mi

17. Keywords		18. Distribution Statement		
ITS4US; Complete Trip; Deployment; ITS; Intelligent Transportation Systems; Data Privacy Plan; Sensitive Personally Identifiable Information		N/A		
19. Security Classif. (of this report) 20. Security Classification 20. Security 20.		ssif. (of this page)	21. No. of Pages	22. Price
N/A Unclassified			24	N/A

Form DOT F 1700.7 (8-72)

Reproduction of completed page authorized

Revision History

Name	Date	Version	Summary of Changes	Approver
Kristin Tufte, University of Washington	30 September 2022	0.1	Initial Draft	Mark Hallenbeck, University of Washington
Kristin Tufte, University of Washington	31 October 2022	0.2	Revised Draft	Mark Hallenbeck, University of Washington
Kristin Tufte, University of Washington	22 November 2022	0.3	Final	Mark Hallenbeck, University of Washington

This page intentionally left blank.

Table of Contents

Re	vision History	i
Tak	ole of Contents	iii
1	Introduction	1
1.1	Document Purpose	1
	1.1.1 Organization of this Document	1
1.2	Deployment Concept	1
1.3	Phase 2/Phase 3 Privacy Update	6
2	Privacy Approach	9
2.1	Access Requirements	9
	2.1.1 Data Generators	9
	2.1.2 Data Consumers	10
	2.1.3 Participants Using Multimodal AccessMap	11
2.2	Security Assessments	12
	2.2.1 Confidentiality, Integrity, and Availability Assessment	13
2.3	Data Security Requirements	16
2.4	Risk Assessment of Threats	18
2.5	Data Sharing and Provision	20
2.6	Specific System Hardware Security Analysis	20
	2.6.1 TDEI System	20
	2.6.2 Data Translation Tool	21
	2.6.3 Applications to Demonstrate How Data Are Used	21
3	Security Controls	23
3.1	Technical Controls	23
3.2	Policy Controls	23
	3.2.1 Breach Plan	23
Ap	pendix A. Acronyms and Glossary	25
۸n	nondix B. Poforonoos	27

List of Tables

Table 1. CIA Assessment of TDEI Data Flows	14
Table 2. CIA Assessment of Evaluation Data	16
Table 3. List of Additional Security Requirements for Data Privacy (TDEI System)	16
Table 4. List of Additional Security Requirements for Data Privacy (Independent System)	18
Table 5. Key threats in relation to PII or SPII	19
List of Figures	
Figure 1. Map. Washington, Oregon, and Maryland Counties	2
Figure 2. Diagram. Context diagram for the proposed Transportation Data Equity Initiative	
system	4
Figure 3. Diagram. Context Diagram with data flow callouts	5

1 Introduction

1.1 Document Purpose

The Data Privacy Plan (DPP) provides details about how the privacy of participants in the University of Washington's (UW) Transportation Data Equity Initiative (TDEI) Project for the United States Department of Transportation's (USDOT) ITS4US Program will be protected. The DPP describes actions that will be taken to protect the privacy of users, guard against potential breaches of the system, and prevent unauthorized use of sensitive participant data, specifically Sensitive Personally Identifiable Information (SPII) and other Personally Identifiable Information (PII).

1.1.1 Organization of this Document

This document is organized as follows:

- Section 1 outlines the document purpose and deployment concept.
- Section 2 outlines the privacy approach regarding access, security, data security, risk assessment of threats, data sharing and provision, and hardware security analysis.
- Section 3 outlines the technical and policy controls.

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, this proposed system, the associated standards development, and the adoption by users (including both data generators and data consumers) will help provide a means to offer appropriate 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 under any situation.

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:

 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.
- 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.



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 Data Management Plan (DMP) and the Phase 2 System Architecture Document (SAD), 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 above 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) Multimodal AccessMap, Microsoft's Soundscape, and XR
 Navigation's Audiom.

Various data flows in this context diagram will facilitate the overall system; and various components within subsystems will serve different roles to process, transform, and/or store these data. The Phase 1 DMP illustrated the data flows between entities, with callouts added to refer to specific data flows. **Figure 3** shows this diagram, updated for Phase 2. This figure shows the primary data flows associated with key operational features provided by the TDEI system that are associated with ITS4US program goals. Specific names of data flows are discussed in the DMP.

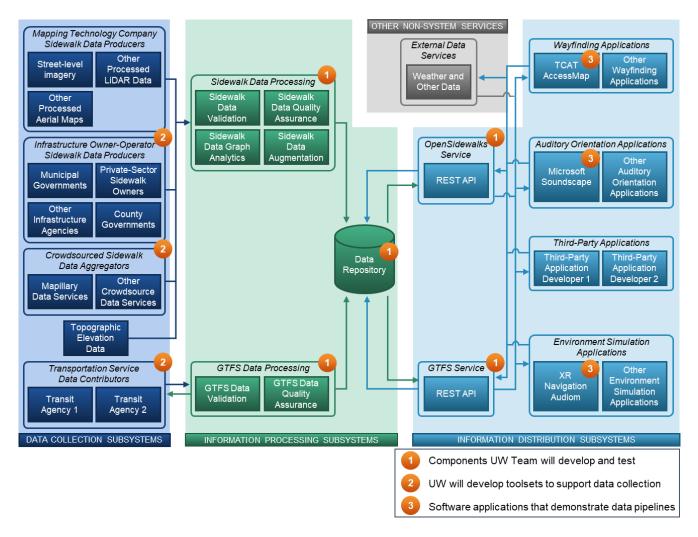


Figure 2. Diagram. Context diagram for the proposed Transportation Data Equity Initiative system.

Source: University of Washington and Cambridge Systematics.

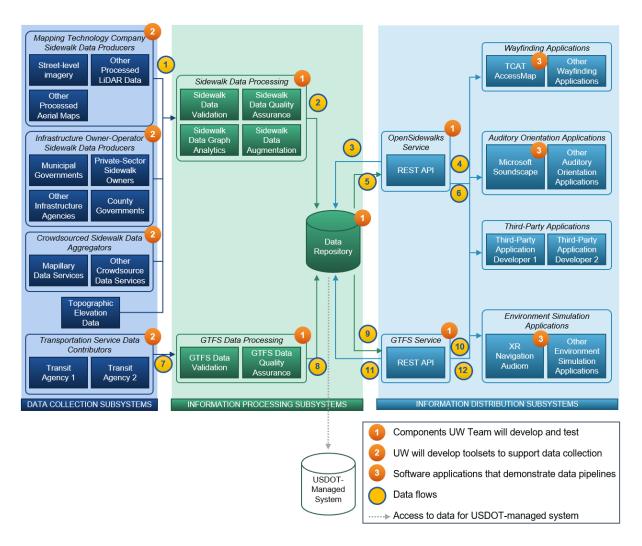


Figure 3. Diagram. Context Diagram with data flow callouts.

Source: University of Washington and Cambridge Systematics.

1.3 Phase 2/Phase 3 Privacy Update

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 PII or 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 primarily 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 as potentially including locational PII; these flows were primarily 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.

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 data sets collected as part of the TDEI system's operation. As discussed in the DMP, the locational PII exists between the application developers and the digital device end users (i.e., the individuals with privacy concerns). The following are noted:

- 1. The application developers 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 for them to provide routing or other service. The TDEI system is designed to provide 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 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 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. The TDEI will know that a particular API key is making the query (which could be tied to a specific application developer account), but the TDEI will not know the motivation for the query (i.e., the TDEI does not know if a request for data for "the Capitol Hill neighborhood of Seattle" is made by a user in that neighborhood, a user not in that neighborhood, or just as a general query to cache data for future use).

Where PII data remains a concern for the UW ITS4US project is with the human participants that are being recruited in Phase 3 to collect evaluation data of the TDEI system, which is discussed

in greater detail in the Phase 1 Human Use Approval Summary (HUAS). The UW ITS4US project team will be using Multimodal AccessMap as one of the applications to demonstrate how the data are used in a real-world application. Multimodal AccessMap is the next generation of the UW's existing AccessMap software, and development of the upgrade to AccessMap is not being funded by the USDOT. However, to facilitate the demonstration of the TDEI, in Phase 3, approximately 40 volunteer participants will be recruited for a laboratory study to use Multimodal AccessMap 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 Institutional Review Board (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 privacy data, the USDOT-funded portion of the TDEI system will receive filtered requests for data from Multimodal AccessMap, like the processes identified in the first two noted resolutions in this list. The TDEI system itself is not storing the routing and/or user data. Those data are being gathered as part of a UW study to evaluate the TDEI system for the ITS4US program and will be warehoused in a separate system—independent of the TDEI system—to help inform evaluation. As such, the necessary safeguards to protect user privacy will need to be applied to that independent system. The evaluation study is being paid for as part of the ITS4US project, and the separate UW system used to house the evaluation data will be part of the IRB review. Greater details on proposed security requirements are discussed in Section 2.3.

An additional item not captured in the Phase 1 DMP is data submitted for creation of user accounts. The TDEI requires that users create accounts in order to be granted access to the TDEI system as either a data generator or a data consumer. Thus, in order for a user to receive access credentials, including an API key, they must supply a contact email address. This is the only information required to be provided to create a user account; users may optionally provide a first and last name. This information will reside in the TDEI system's authorization service in order to facilitate communication with the user. Thus, it is anticipated that the data present in the authorization service will meet the minimum requirements necessary to be considered PII, and thus will be treated as such for purposes of this Phase 2 DMP.

The Phase 2 DMP reflects on the data management needs for these two systems. This DPP focuses on any remaining privacy considerations for the data collected by the system.

2 Privacy Approach

2.1 Access Requirements

The TDEI system aims to be a scalable system with national applicability and to serve a wide variety of users, so the design of this system needs to accommodate growth while also not having access controls that are overly restrictive and prohibitive. That said, the TDEI needs some access restrictions in place not only to protect any data considered sensitive or private, but also to keep a tally of user accounts to help support training, outreach, and updates, as well as to help minimize impacts to system service and performance from unapproved actors (whether intentional or unintentional).

As shown in **Figure 2**, two key types of groups exist, each with their own interface to the TDEI system. These interfaces are described in greater detail as part of the Phase 2 SAD and Interface Control Document (ICD). The first group is data generators—which include transportation service providers, and data service providers—who are the groups that submit sidewalk, transit pathway, and/or transit flex data to the TDEI system. The second group is data consumers—which include application developers and digital device end users—who are the groups that request and use the sidewalk, transit pathway, and/or transit flex data that the TDEI system can provide. As part of the TDEI system's design, both groups—data generators and data consumers—will have to be granted access prior to interfacing with the system. The process for providing access to the TDEI system and the data collected for registration is discussed in greater detail as part of the Phase 2 ICD. It is anticipated that this registration data, as described in the prior section, will meet the minimum requirements necessary to be considered PII and will need to be safeguarded as such. Data storage will be discussed part of the Phase 2 DMP. Further details on the data generator and data consumer groups are discussed in subsequent subsections.

In the Phase 1 Concept of Operations (ConOps), two user needs directly address user privacy. Both of these user needs are affiliated with application developers or digital device end users:

- UN-AD13: Application Developers (ADs) need to protect end user privacy by ensuring that interoperable transportation data sharing does not offer access to personal data, whether intentionally or unintentionally
- UN-DU1: Digital Device End Users Experiencing Travel Barriers (DUs) need to be able to set boundaries on the allowed release of their personal data in order to gain functionality, while being protected from unapproved data releases.

2.1.1 Data Generators

Data Generators possess sidewalk, transit pathways, and/or transit flex data that they can submit to the TDEI system. As described in the Phase 2 SAD and ICD, data from this group is submitted through an API to the TDEI system, utilizing public internet to facilitate the transmission (in **Figure 3**, this includes data flows #1 and #7). Access must be granted to a given data generator in advance of data submission, which will require user account information that meets the

minimum requirements to be considered PII. In the context of the Phase 2 and Phase 3 demonstration, access is granted as followed:

- 1. Data Generator registers as a TDEI user
- 2. Access credentials (username, password) and an API key for the new user will be generated
- 3. The API key and access credentials will be stored
- 4. The Data Generator provides their access credentials and requests the API key
- 5. The API Gateway validates the credentials using the database
- 6. The API Gateway retrieves the API key from the database
- 7. The API key is provided to the Data Generator
- 8. An (authorized) Transit Service Provider Point of Contact or Representative provides their access credentials and requests the (new) user be granted the Data Generator Role for that Transit Service Provider
- The system stores the Data Generator access for the user for that Transit Service
 Provider; the user is now authorized to act as a Data Generator for that Transit Service
 Provider.

It is anticipated that data generators will be manually approved (i.e., approved by a human reviewer). During the ITS4US demonstration, approval will be provided by a transportation service provider point of contact or approved representative as described in the Phase 2 ICD. This manual approval is not for data privacy concerns, but for other reasons that focus on data integrity (see **Section 2.2.1**), specifically for making sure proper training and/or resources are offered for data generators to properly structure their data. The sidewalk and/or transit-related data submitted by data generators does not contain any PII, and the only information offered from the TDEI system to data generators is information on whether their data submission has been accepted. That said, user account information will need to be separately stored and protected within the TDEI system. User account information is anticipated to be stored in the TDEI's authorization service, which is a separate service from other services offered by TDEI.

As part of the Phase 2 and Phase 3 effort, the UW ITS4US project team will develop a tool to help data generators translate data sets into the correct data schema. This tool will operate independent of the TDEI system's data repository, and only translate data associated with public infrastructure (i.e., not personal individual data).

2.1.2 Data Consumers

Data Consumers include individuals and/or application developers that consume data offered by the TDEI system. As described in the Phase 2 SAD and ICD, data from this group is submitted through an API to the TDEI system, utilizing public internet to facilitate the transmission (in **Figure 3**, the query follows data flows #4 and #10, the response follows data flow #6 and #12). Access must be granted to a given data consumer in advance of data submission, which will require user account information that meets the minimum requirements to be considered PII. In the context of the Phase 2 and Phase 3 demonstration, access is granted as followed:

- 1. Data Consumer registers as a TDEI user by completing a web form
- Access credentials (username, password) and an API key for the new user will be generated
- 3. The API key and access credentials will be stored
- 4. The Data Consumer provides their access credentials and requests the API key
- 5. The API Gateway validates the credentials using the database
- 6. The API Gateway retrieves the API key from the database
- 7. The API key is provided to the Data Consumer

Unlike data generators, it is anticipated that data consumers will be automatically approved (i.e., data consumer fills out a webform, then receives access without need for human approval). One key motivation for this is to facilitate making the data open and accessible to the public, but with some access controls in place to help coordinate with outreach and provide updates when parts of the API change. The API also does not receive any trip-purpose information, so the system will not know whether geospatial data are being provided in response to a trip request within a geographic area or if the application developer is simply pulling data for other purposes, such as to build a local cache.

The transactions across this interface include structured queries from application developers to the TDEI system, and geospatial data from the TDEI system to the application developer. Since the APIs restrict PII from being shared with the system and anonymize the purpose of a specific geographic request, any data stored on the query requests will not include data privacy concerns. As such, data sent back to application developers—made up of sidewalk, transit pathways, and/or transit flex data submitted by data generators—does not include any privacy-related data.

Similar to the case for Data Generators, user account information will need to be separately stored and protected within the TDEI system. User account information is anticipated to be stored in the TDEI's authorization service, which is a separate service from other services offered by TDEI.

To prevent system issues caused by excessive data requests, the TDEI system will rate limit the number of queries that an approved data consumer can make to the system in a given period of time.

2.1.3 Participants Using Multimodal AccessMap

The 40 participants that demonstrate the TDEI system in use through Multimodal AccessMap will supply three sets of data. The first set of data consists of the demographic and mobility profile of the user. The second set of data consists of the trip routing and navigation requests they make using the application. The third set of data consists of the user's "summary thoughts and perceptions" through a survey. None of these data will be stored with direct identifiers associated with the user, instead this data will be stored with indirect identifiers that only indicate the demographic and mobility disabilities associated with that traveler. This includes all interview and survey responses. No one outside the TDEI research team will see the raw data without IRB approval. Data from this effort that is shared publicly and without restriction will be aggregated

data summaries that are both aggregated and stripped of the indirect identifiers, which is a data management practice regularly used by the UW for other sensitive research studies to protect participant privacy. Aggregated data cannot be re-identified and lacks indirect identifiers. Participants will have access to the Multimodal AccessMap application for their use, but the evaluation data that is collected will be stored separately and not be accessible to them.

All data in the contact information dataset will be protected by best-practices protocols: access will be mediated by OAuth-2.0 authorization methods via University of Washington or equivalent accounts following the requirements outlined in Health and Safety Guidance (HSG). 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. As noted earlier, the evaluation data for these 40 participants will be stored on a directory in which only TDEI staff have access to it, with permissions controlled using the UW permission system that is regularly used to safeguard access to other critical research data sets that involve human use subjects.

2.2 Security Assessments

The Phase 1 DMP identified two levels of access:

- 1. Open: Data that can be used by the public with no or limited licensing restrictions. These data are available to the public without needing to request permissions and will be provided to the USDOT-managed Public System. These may be anonymized or aggregated version of private datasets to protect PII. Permission to access these data will be available across the entire program's scope. Any user may request access to the dataset in question, in accordance with the OPEN Government Data Act. This permission will include data that are non-PII and non-confidential business information (CBI) but will include no other information that threatens the privacy of an individual or group.
- 2. Private (Research): Data that cannot be shared with external users. Access to these data is limited and only granted with IRB and Project Team approvals. Permission to access these data will be limited to research team members with IRB approval. A specific concern will be the presence of PII, which includes potential PII, actual PII, locational PII, and sensitive PII, as well as other potential information that threatens the privacy of an individual or group. The proposed system will not capture CBI. These data are considered research private data, meaning it is available for research, but users of the data must meet IRB requirements before gaining access.

With the updates to Phase 2 design that remove the presence of locational PII in the TDEI, most datasets qualify as open data. The two exceptions include:

- Evaluation data being collected from the 40 consenting and informed participants through Multimodal AccessMap for evaluation of the TDEI system, which is being collected, processed and stored in a separate system than the TDEI.
- 2. User account data that is stored in the TDEI system's authorization service to issue API keys, provide notifications of system changes, and other system-related outreach.

Most of the security elements implemented for this effort for the TDEI system—specifically the access requirements mentioned in **Section 2.1**—are present to maintain system integrity and

data preservation. The APIs are structured in a manner that does not reveal any of the PII stored in the TDEI system regarding user accounts, which are stored separately as part of the authorization service. Security elements are required to protect user privacy, particularly in the authorization service of the TDEI system. In the event of a data breach with the TDEI system (discussed later), the risk of personal data being exposed does exist, but the UW team aims to keep the amount of PII to the lowest minimal value necessary to facilitate system operation and help reduce impact to users if data were exposed. The evaluation data collected from participants also requires security elements to protect user privacy in the separate system from the TDEI system.

2.2.1 Confidentiality, Integrity, and Availability Assessment

Confidentiality, integrity, and availability (CIA) is a model designed to guide policies for information security within an organization. In this context, confidentiality is a set of rules that limits access to information, integrity is the assurance that the information is trustworthy and accurate, and availability is a guarantee of reliable access to the information by authorized people.

Confidentiality: When PII is present in a given data set, a great degree of effort is needed as part of design to protect confidentiality. The TDEI system (as well as the data translation tool) is collecting and storing user account data that is considered PII or SPII in order to grant access, but this data will be stored in the authorization service of the TDEI system and will not be intermingled with other data resources. The APIs also provide only certain data sets back to data generators and consumers, which do not include PII data about another user. Data that is in transit will use Hypertext Transfer Protocol Secure or HTTPS to ensure security of the data transmission. Similarly, evaluation data, which also includes PII, will be stored separate from the TDEI system. Access to these sensitive data sets will require IRB approval.

Integrity: When data is critical for the operation, a great degree of effort is needed as part of the design to ensure data integrity is maintained. The TDEI system provides access control to track how parties interact with the data, which requires user account registration and approval prior to being able to interface with the TDEI's data storage. On the data generator side, any infrastructure or transit-related data contribution that is submitted by an individual or an agency undergoes a vetting and quality assurance process that makes sure that the data appears valid prior to adding it to the data repository. On the data consumer side, any data that are queried are returned as a structured set that is directly attributable to the request made to the API; these data queries are rate limited to reduce the chance of overloading the TDEI system with data requests and impacting the ability of other users to interface with the system. Similarly, evaluation data, which is critical to measure performance, requires design effort to maintain integrity.

Availability: When PII is present in a given data set, a great degree of effort is needed to make sure its availability is contained to permitted parties, or that sufficient effort is taken for the system to not distribute PII as part of another data set. Access to infrastructure and transit-related data in the TDEI system is envisioned to be consistently and readily available to authorized parties. For data consumers, the availability of data is dependent on the data consumers' ability to correctly query the TDEI system via the API. The APIs are designed in a manner that provide straightforward and well-documented request parameters to facilitate the application developers' data requests. Use of registered accounts will allow outreach from the UW ITS4US project team to inform of how the API gateway is structured and help troubleshoot when an error occurs, as opposed to a random participant attempting to connect to the TDEI with no insight as to why a

connection may not be made. The TDEI system is being operated in a cloud environment with the necessary service levels to be reachable through the public internet, with the necessary backups to make sure data are not lost. Additionally, in the event that the TDEI system is offline, registered data consumers will be able to receive email notifications to keep them aware that data are not available.

To assess all data sets for confidentiality, integrity, and availability, a low-medium-high scoring will be used for each category:

- High This requires a great degree of effort for the metric in question to meet its stated need listed above.
- Medium This requires a moderate degree for the metric in question to meet its stated need listed above.
- Low This requires none or a very limited degree of effort for the metric in question to meet its stated need listed above.
- Not Applicable The TDEI system does not have control over the data, the data are being provided by a third party, or the data flow is internal to the TDEI system and is represented under another data flow.

Table 1 provides a CIA assessment for each of the TDEI data flows, as represented in the Phase 1 DMP and updated based on Phase 2 design developments. Please refer to **Figure 3** for reference of where each data flow falls within the TDEI system (i.e., Data Flow numbers listed in **Table 1** can be found on **Figure 3**, for reference). The user account registration data, which is a one-time entry to receive an API key and thus is not a routine data flow to be shown on the concept diagram, is included with these TDEI data flows.

Table 1. CIA Assessment of TDEI Data Flows

Data Flow	Confidentiality	Integrity	Availability
User Account	High – PII is	High – Since data is	High – PII will be isolated
Registration Data	present. Data will	necessary to permit	in the authorization
	be stored separate	system access,	service. Other data flows
(one-time entry, part of	from other	maintaining integrity is	will not intermix with this
authorization service)	operational	crucial to avoid an	data. No other users can
	elements of the	approved user from	access other PII. Access
	TDEI to avoid	being disallowed	to data is limited to TDEI
	mixing and will be	access to the system.	staff or parties that have
	stored encrypted.		IRB approval.
Sidewalk Data	Low – No PII are	High – Data are vetted	Medium – Well-
	present in public	by human verifier	documented API for
(Data Flow #1)	infrastructure, data	and/or automated	submission is provided to
	vetted prior to	processes prior to	approved data generators
	system entry.	system entry to provide	to submit certain types of
		good quality.	data.
Validated Sidewalk	N/A – Internal to	N/A – Internal to TDEI	N/A – Internal to TDEI
Data	TDEI system, uses	system, uses data from	system, uses data from
	data from Data	Data Flow #1	Data Flow #1
(Data Flow #2)	Flow #1		

U.S. Department of Transportation

Data Flow	Confidentiality	Integrity	Availability
Transit Pathway or	Low – No PII are	High – Data are vetted	Medium – Well-
Flex Data	present in public	by human verifier	documented API for
	infrastructure, data	and/or automated	submission is provided to
(Data Flow #7)	vetted prior to	processes prior to	approved data generators
	system entry.	system entry to provide	to submit certain types of
		good quality.	data.
Validated Transit	N/A – Internal to	N/A – Internal to TDEI	N/A – Internal to TDEI
Pathway or Flex Data	TDEI system, uses	system, uses data from	1 -
	data from Data	Data Flow #7	Data Flow #7
(Data Flow #8)	Flow #7		
Request Geographic	Low – API is	Medium – Data must	Medium – Well-
Data	structured to	be structured by the	documented API for
,,	receive non-PII	data consumer to	submission is provided to
(Data Flow #4 for	geospatial requests	accommodate API, and	approved data consumers
sidewalks and Data	from approved data	only requests following	to query for certain types
Flow #10 for transit	consumers and	the correct structure	of data.
service and/or stations)	disregards any	are accepted.	
	extraneous data.		
Filtered Request	N/A – Internal to	N/A – Internal to TDEI	N/A – Internal to TDEI
Geographic Data	TDEI system, uses	system, uses data from	system, uses data from
(D. L. El //0 f	data from Data	Data Flow #4 or Data	Data Flow #4 or Data
(Data Flow #3 for sidewalks and Data	Flow #4 or Data Flow #10	Flow #10	Flow #10
	FIOW # IU		
Flow #11 for transit			
service and/or stations)	N/A – Internal to	N/A – Internal to TDEI	N/A – Internal to TDEI
Response			
Geographic Data	TDEI system, provides data to	system, provides data to Data Flow #6 or	system, provides data to Data Flow #6 or Data
(Data Flow #5 for	Data Flow #6 or	Data Flow #12	Flow #12
sidewalks and Data	Data Flow #12	Data Flow #12	FIOW #12
Flow #9 for transit	Data Flow #12		
service and/or stations)			
Filtered Response	Low – Data	Medium – Data	Medium – Well-
Geographic Data	includes public	includes vetted data	documented API for
Sograpino Data	infrastructure	from the data	submission is provided to
(Data Flow #6 for	and/or	repository.	approved data consumers
sidewalks and Data	transportation	Topository.	to query for certain types
Flow #12 for transit	service data only		of data.
service and/or stations)	for a geographic		
	area.		
	u. 3u.	l .	

For the data collected for evaluation, a similar assessment can be made, with a low-medium-high scoring will be used for each category:

High – This requires a great degree of effort for the metric in question to meet its stated need listed above.

- Medium This requires a moderate degree for the metric in question to meet its stated need listed above.
- Low This requires none or a very limited degree of effort for the metric in question to meet its stated need listed above.

Data from the 40 participants is collected by specific evaluation tools used in conjunction with the Multimodal AccessMap application, and thus does not show up as a data flow in the context of the TDEI system. Instead, **Table 2** provides a CIA assessment of that data set in the independent system. Since this is independently collected outside of the TDEI system, it does not have an affiliated callout on Figure 3.

Data Flow	Confidentiality	Integrity	Availability
Evaluation Data	High – PII will likely		High – Access to data is limited to TDEI staff and has oversight from an IRB. Individual users of Multimodal AccessMap will not be granted access
	system that is used to protect other similar sensitive data sets.	mamam data integrity.	to evaluation data.

Table 2. CIA Assessment of Evaluation Data

2.3 Data Security Requirements

Data security requirements in this DPP are focused on the necessary security requirements for addressing privacy concerns, and do not focus on the more general security needs of the TDEI system. Since the TDEI system (and its associated data translation tools) do not use any restricted data sets that include PII or SPII from a routine operations standpoint, no additional security requirements are necessary to secure the data for these data flows. The exception is the user account registration data that will reside in the TDEI system's authorization service, which necessitates permission controls, encryption, and cybersecurity systems to protect user data from unauthorized access. **Table 3** provides a list of additional security requirements for data privacy.

Table 3. List of Additional Security Requirements for Data Privacy (TDEI System)

Data Flow	Additional Security Requirements for Data Privacy
User Account Registration Data (one-time entry, part of authorization service)	Requires a permission-controlled system designed for storing sensitive data, with appropriate levels of encryption and cybersecurity systems. These data are isolated from other data flows to avoid association with other data sets. Must satisfy IRB approval.
Sidewalk Data (Data Flow #1)	None

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

Data Flow	Additional Security Requirements for Data Privacy
Validated Sidewalk Data (Data Flow #2)	None
Transit Pathway or Flex Data (Data Flow #7)	None
Validated Transit Pathway or Flex Data (Data Flow #8)	None
Request Geographic Data (Data Flow #4 for sidewalks and Data Flow #10 for transit service and/or stations)	None
Filtered Request Geographic Data (Data Flow #3 for sidewalks and Data Flow #11 for transit service and/or stations)	None
Response Geographic Data (Data Flow #5 for sidewalks and Data Flow #9 for transit service and/or stations)	None
Filtered Response Geographic Data (Data Flow #6 for sidewalks and Data Flow #12 for transit service and/or stations)	None

Since none of the data sets that are involved with data generator or data consumer transactions involve PII or SPII (i.e., all data sets except user account registration, which is not an operational transaction after the API key is issued), any combination of data sets does not create PII or SPII either that necessitate additional data security requirements. Any external systems—independent of the USDOT-funded TDEI system—may need to adopt additional data security requirements where they involve personal data, but that is on the application developers to decide the appropriate level necessary. For the data that is stored on an independent system for evaluation, additional requirements exist, as shown in Table 4.

Table 4. List of Additional Security Requirements for Data Privacy (Independent System)

Data Flow	Additional Security Requirements for Data Privacy
Evaluation Data	The PII data will be stored in a file system local to the University of Washington, permissions to access files containing PII data will be controlled with a system that uses the OATH 2.0 security protocol ¹ . Access to these files will be limited to the personnel named in the IRB approval. All files will be encrypted using a FIPS-approved encryption algorithm such as Advanced Encryption Standard (AES) ² as recommended by NIST ³ .

2.4 Risk Assessment of Threats

A risk assessment of threats in the context of data privacy examines the likelihood and impact of a potential threat in regard to exposure of privacy information. National Institute of Standards and Technology (NIST) Special Publication 800-30 Revision 1, Guide for Conducting Risk Assessments, identifies a sequence of events for assessing risks and threats:

- Identify threat sources that are relevant to organizations
- Identify threat events that could be produced by those sources
- Identify vulnerabilities within organizations that could be exploited by threat sources through specific threat events and the predisposing conditions that could affect successful exploitation
- Determine the likelihood that the identified threat sources would initiate specific threat events and the likelihood that the threat events would be successful
- Determine the adverse impacts to organizational operations and assets, individuals, other organizations, and the Nation resulting from the exploitation of vulnerabilities by threat sources (through specific threat events); and
- Determine information security risks as a combination of likelihood of threat exploitation of vulnerabilities and the impact of such exploitation, including any uncertainties associated with the risk determinations.

In the context of data privacy, several key threats exist, as outlined in Table 5.

¹ https://oauth.net/2/

² http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf

³ https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-111.pdf

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

Table 5. Key threats in relation to PII or SPII

Threat	Risk to PII or SPII
An unauthorized breach to the TDEI system that accesses PII or SPII data.	Some Risks – PII exists in the TDEI system's authorization service in order to register a user account and issue an API key. The UW team is structuring this user account to require a minimal amount of information, so such a breach would capture minimal information (like an email address, as opposed to something more sensitive).
An interception of PII or SPII data that Is sent over the public internet, either to or from the TDEI system.	None – The TDEI system does not send PII or SPII data. An external party that attempts to send PII or SPII data as part of the query will not meet the API requirements and be rejected. Any risks associated with interception of user account registration data during transmission would occur prior to the TDEI system over the public internet, which is encrypted with HTTPS.
Erroneous data is submitted to the TDEI by a data generator.	None – Erroneous data would be screened by the TDEI system, either through manual or automated processes. The user account associated with the API key would receive a notification of acceptance or rejection, likely via their provided email, but no PII data for other users is sent as part of that email.
An approved data generator or data consumer does not structure a request or submission in accordance with the API.	None – Incorrectly structured requests or submissions are rejected. The user account associated with the API key would receive a notification, likely via their provided email, but no PII data would be included as part of that.
A placement of a TDEI tool (i.e., the data translation tool) in a location where PII or SPII data are present, and the tool subsequently puts the PII or SPII data at risk.	None – The data translation tool will only translate certain infrastructure or transportation service parameters into the data schema used by the TDEI. The data translation tool does not have inputs for PII or SPII data.
Evaluation data collected through Multimodal AccessMap is compromised and exposed to unauthorized parties.	Some Risks – Even though data are stored with indirect identifiers between the three sets, some locational PII and other PII is present that could be stitched together to expose private information about a person's travel choices and/or preferences. While a risk exists, it also will take an extensive level of effort to re-identify an individual through locational PII.

Risk may exist outside of the work being done by the TDEI system, specifically among the application developers that are interfacing directly with the digital device end users (beyond the evaluation data in Multimodal AccessMap), as these two enterprise groups may be sharing

sensitive data that includes locational PII. It is not feasible to assess the risk of a given threat, as the architecture and security practices in place of these application developers are not known (in the context of the TDEI system) and are considered out of scope for this report.

2.5 Data Sharing and Provision

As discussed in **Section 2.2**, most data affiliated with the TDEI system will be open license, and only the user account registration data stored in the authorization service will be a restricted license. As such, all data except that data housed in the authorization service can be transferred to the USDOT ITS JPO Data Hub System without concerns about restriction. This will allow data to be provided and licensed in a way that allows free and open use for data transactions that align with the evaluation goals of the system, which focus on quantity and coverage of data collected and used. In other words, user account registration was not one of the key evaluation goals of TDEI, and this restricted data set does not include data of interest to the ITS4US effort.

Data affiliated with the evaluation will continue to be classified as private data due to the presence of human participants and the risks of PII being present in the dataset. The raw data itself will contain PII that qualifies under the private data definition defined earlier in **Section 2.2**, meaning it is available for research, but users of the data must meet IRB requirements before gaining access. It is anticipated that processed/aggregated evaluation data will be able to provide results in aggregate that strips away any PII, but whether that qualifies as being open license will need to be determined at the time of evaluation. The UW has frequently aggregated data sets to strip away any PII or indirect identifiers for public use as part of other research efforts, so expectations are that the same approach—which is reviewed by IRB—would be used here. As described in Section 2.1.3, the standard approach is to store demographic information about the participants separately from the locational trace and survey data sets using only indirect identifiers in the other data sets, thereby limiting the potential harm if those data sets were exposed. Further, as the trace data sets contain locational information which potentially could be used to re-identify persons, the trace data is clipped and aggregated using standard IRB-approved processes prior to any public distribution.

Details on data transfer to USDOT are discussed in the Phase 1 and Phase 2 DMP.

2.6 Specific System Hardware Security Analysis

The security analysis for key system hardware components in the context of data privacy are discussed in the following subsections. These sections outline the key features identified in **Section 1.2**, specifically the TDEI system that the UW ITS4US project team is building, the data translation tool that the UW ITS4US project team will provide to data generators who need to translate data, and the application developers that are external to the TDEI system's effort, but will be used to help demonstrate the use of the system in a real-world environment.

2.6.1 TDEI System

The TDEI system exchanges and stores data to and from approved parties. It will likely be hosted in a cloud environment. PII data in the form of user account data is present in the TDEI system as part of the authorization service, so the cloud environment will need to have security requirements to safeguard the data.

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

2.6.2 Data Translation Tool

The data translation software would reside with a data generator or a transportation service provider to help them convert existing data into the correct data schema. It will likely be a downloadable tool that would operate on a transportation agency's computer. Since this data translation only includes public infrastructure data, no individual privacy data are associated with it, and thus no security concern exists in the context of data privacy. If any PII or SPII is stored on the machine that also houses the data translation tool, the data translation tool will not accept it, as its translation will only work on infrastructure or transportation service data.

2.6.3 Applications to Demonstrate How Data Are Used

The applications to demonstrate the TDEI system in use are the one major element that is external to this effort, and thus may qualify as having potential security concerns depending on how their hardware is designed. Most application developers will house a central server to exchange and process data for their client application software that resides on the mobile devices of digital device end users. For third-party applications not including Multimodal AccessMap, it is unknown what hardware security requirements will be used to preserve any private information that is shared between those two external devices, but it is important to remember that the USDOT-funded portion of this effort is simply evaluating whether these demonstration applications are able to consume the data, not evaluate how the data is used. For data collected through Multimodal AccessMap (which is being used to evaluate how the data is used), the hardware security requirements for the independent server that stores the evaluation data will need to meet the UW's security requirements for storage of PII and other sensitive data and be subject to IRB review and approval.

3 **Security Controls**

3.1 Technical Controls

The TDEI system will have processes in place for users to interface with the system, discussed earlier in Section 2.1. As part of the performance reporting process, the TDEI system will log and monitor activity of each approved API key: for example, logs may track data submissions, data queries, associated access issues, and time stamps affiliated with activity. Data consumers will be rate limited in the number of queries that can be done during a period of time to reduce the chance of overwhelming the system's operation. These technical controls serve primarily system operation and performance.

3.2 Policy Controls

The TDEI system will use access permissions to provide access to various parts of the system. For example, administrator level access will allow TDEI developers to build, operate, and maintain the system; external users, on the other hand, will only have access to an API to exchange data. These policy controls exist for security of the system itself (i.e., to prevent unauthorized changes, removal, or addition of software code and/or data), as well as prevent unauthorized access to another user account's data.

Evaluation data collected from the 40 participants will be stored in a system that is separate from the TDEI system, so as to not place additional PII or SPII data into the TDEI system and create restrictions of data use on data sets that currently qualify as open to the public. Evaluation data will instead be stored on an independent server, placed in a directory that is only accessible to pre-approved individuals that are part of the TDEI staff and have a purpose and role in evaluating the system. 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. IRB protocols will be followed—as discussed in the HUAS—to provide access to approved individuals.

3.2.1 Breach Plan

A breach in data security for the TDEI system would be an instance where an unauthorized actor has accessed or modified data in the TDEI system. Examples of this may include a disruption to services or a change to the data environment, such as modifying the data that is stored. When a breach is detected, immediate near-term actions will be to re-establish a secure operating environment that removes the unauthorized actor, which may necessitate shutting down services and rolling back data versioning to a version that occurred prior to the breach. These actions would be necessary to restore system functionality and service. The UW Team will coordinate with the USDOT to inform of the breach and the necessary action taken.

Since no PII or SPII data are part of the TDEI system's regular transactional data (i.e., sidewalks, transit pathways, transit flex, or structured data requests), any breach of that data set will not

expose or compromise data privacy. For user account data that is either part of the TDEI system's authorization service or data evaluation collected and stored separate from the TDEI system, a breach will require an investigation on the extent of data that was accessed by unauthorized parties, whether that data included any sensitive data (directly or indirectly), and whether the data can be recovered (i.e. if the unauthorized party is not a bad actor and TDEI staff are able to contact the party, there is a possibility that data could be recovered). If a breach is detected or brought to the attention of TDEI staff, the TDEI staff will work with the UW IT administrator and/or the cloud service provider to remedy the breach and secure data from any further acquisition.

Appendix A. Acronyms and Glossary

Acronym	Definition
AD	Application Developer
AOR	Agreement Officer's Representative
API	Application Programming Interface
CBI	Confidential Business Information
CIA	Confidentiality, Integrity, and Availability
ConOps	Concept of Operations
DMP	Data Management Plan
DPP	Data Privacy Plan
DOT	Department of Transportation
DU	Digital Device End Users Experiencing Travel Barriers
HSG	Health and Safety Guidance
HTTPS	Hypertext Transfer Protocol Secure
HUAS	Human Use Approval Summary
ICD	Interface Control Document
IRB	Institutional Review Board
IT	Information Technology
ITS	Intelligent Transportation System
NIST	National Institute of Standards and Technology
PII	Personally Identifiable Information
SAD	System Architecture 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
U.S.	United States
USDOT	United States Department of Transportation
UW	University of Washington

Appendix B. References

- 1. Data Integrity: Identifying and Protecting Assets Against Ransomware and Other Destructive Events. National Institute of Standards and Technology (NIST) Special Publication 1800-25. December 2020. https://www.nccoe.nist.gov/publication/1800-25/VolA/index.html
- 2. Guide for Conducting Risk Assessments. National Institute of Standards and Technology (NIST), September 2012, https://csrc.nist.gov/publications/detail/sp/800-30/rev-1/final
- 3. Caspi, Anat, et al., Phase 2 Data Management Plan (DMP), University of Washington ITS4US Deployment Project, Draft Report—November 15, 2022, Report Number TBD.
- 4. Caspi, Anat, et al., Phase 2 System Architecture Document (SAD), University of Washington ITS4US Deployment Project, Draft Report—October 24, 2022, Report Number TBD.
- Caspi, Anat, et al., Phase 2 Interface Control Document (ICD), University of Washington ITS4US Deployment Project, Draft Report—October 24, 2022, Report Number TBD.
- Caspi, Anat, et al., Phase 1 ConOps, University of Washington ITS4US Deployment Project, Final Report—June 28, 2021, Report Number FHWA-JPO-21-861.
- 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.
- 8. Caspi, Anat, et al., Phase 1 Performance Measurement and Evaluation Support Plan (PMESP), University of Washington ITS4US Deployment Project, Draft Report—September 8, 2021, Report Number FHWA-JPO-21-874.
- 9. Caspi, Anat, et al., Phase 1 System Requirements Specification (SyRS), University of Washington ITS4US Deployment Project, Draft Report—October 25, 2021, Report Number FHWA-JPO-21-884.

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-22-972



U.S. Department of Transportation