# Performance Measurement and Evaluation Support Plan

## University of Washington ITS4US Deployment Project – Phase 1

www.its.dot.gov/index.htm

**Final Report – November 15, 2021**
**Publication Number FHWA-JPO-21-879**

COMPLETE TRIP
ITS4US

U.S. Department of Transportation

Produced by the University of Washington ITS4US Deployment Phase 1 Project
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-21-879** | N/A | N/A |

| 4. Title and Subtitle | 5. Report Date |
|---|---|
| Phase 1 Performance Measurement and Evaluation Support Plan—University of Washington ITS4US Deployment Project | November 15, 2021 |
| | **6. Performing Organization Code**<br>N/A |

| 7. Author(s) | 8. Performing Organization Report No. |
|---|---|
| Anat Caspi, Mark Hallenbeck, Erin Flanigan, Mark Jensen, Alice Marecek, Adam Danczyk, Guilherme Leao, | |

| 9. Performing Organization Name and Address | 10. Work Unit No. (TRAIS) |
|---|---|
| University of Washington<br>4333 Brooklyn Ave NE<br>Box 359472<br>Seattle, WA 98195-9472 | N/A |
| | **11. Contract or Grant No.**<br>693JJ321C000004 |

| 12. Sponsoring Agency Name and Address | 13. Type of Report and Period Covered |
|---|---|
| U.S. Department of Transportation<br>ITS Joint Program Office<br>1200 New Jersey Avenue, SE<br>Washington, DC 20590 | N/A |
| | **14. Sponsoring Agency Code**<br>HOIT-1 |

**15. Supplementary Notes**

Kate Hartman, COR

**16. Abstract**

This document discusses the Performance Measurement and Evaluation Support Plan (PMESP) for the University of Washington (UW) ITS4US Deployment Project, which is developing and deploying the Transportation Data Equity Initiative (TDEI). This project currently is in Phase 1, planning and systems engineering development, in which the preliminary idea is developed into a structured concept that is suitable for further design, building, testing, and operation. The structured concept will include identifying specific performance measures, targets, and capabilities associated with performance monitoring and performance management.

The PMESP is informed by the project's Concept of Operations (ConOps), the Safety Management Plan (SMP), and the Data Management Plan (DMP). The ConOps describes how the project will operate, as well as how those operations both meet the needs that motivated the project and the specific technical requirements from which the project is built. The SMP describes potential hazards that users of the deployed system might encounter and how the project will keep those users as safe as possible, while the DMP describe how data being collected for the project will be stored and shared, while maintaining the privacy of individuals participating in the project. These documents inform the PMESP, in that they describe the key topics that need to be evaluated, as well as the sources of the data that are needed to perform those evaluations,

This document is intended to be a living document. While performance measurements and evaluation recommendations are identified in this document, those plans are subject to refinement in Phase 2, as the project engineering is completed and deployed. It is anticipated that input from the USDOT's selected independent evaluation team, led by the Volpe Transportation Systems Center (Volpe) will be incorporated into future PMESP versions as those revisions and refinements are identified.

| 17. Keywords | 18. Distribution Statement |
|---|---|
| ITS4US; Complete Trip; Deployment; ITS; Intelligent Transportation Systems; Performance Measurement Plan; Evaluation Support Plan; Accessibility; Sidewalks; Navigation Software; Data Standards | N/A |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| N/A | N/A | 67 | N/A |

# Revision History

*Update the revision history table upon each update to the Performance Measurement and Evaluation Support Plan (PMESP).*

| Name | Date | Version | Summary of Changes | Approver |
|---|---|---|---|---|
| Mark Hallenbeck, University of Washington; Mark Jensen, Cambridge Systematics | 8 September 2021 | 1 | Initial Draft | |
| Mark Hallenbeck, University of Washington; Mark Jensen, Cambridge Systematics | 25 October, 2021 | 1 | Final Report | |
| Mark Hallenbeck, University of Washington; Mark Jensen, Cambridge Systematics | 15 November, 2021 | 1 | Refined Final Report | |

# Table of Contents

## List of Tables



## List of Figures

# 1. Introduction

This document presents the Performance Measurement and Evaluation Support Plan (PMESP) for the University of Washington's (UW) ITS4US Deployment Project, the Transportation Data Equity Initiative (TDEI), which is being performed as part of the U.S. Department of Transportation's (USDOT's) Complete Trip—ITS4US Deployment Program.

This PMESP is informed by the following project documents[1]: 1) Concept of Operations (ConOps) for the proposed system, which bridges the user needs that motivated the project and the specific technical requirements from which the project is built, 2) the Data Management Plan (DMP), which discusses how the data that is expected to be acquired or generated during the course of the ITS4US project will be managed, analyzed, protected, stored, and shared, and 3) the Safety Management Plan (SMP) which assesses the safety needs and risks in how travelers and others interact with the planned deployment.

This document is intended to be a living document. While preliminary recommendations are identified herein, many of these decisions regarding performance evaluation will be finalized in concert with the Independent Evaluation Team for the project, as well as the project's stakeholders as part of Phase 2 development. It is anticipated that many items will be adjusted, but the underlying intent of the proposed performance measurement plan will be preserved.

A glossary and acronym list are provided in the appendix.

## 1.1. Intended Audience

The PMESP is intended for use by the UW ITS4US Deployment Project team, its partners and peer agencies, the Independent Evaluation Team, and the United States Department of Transportation (USDOT). In addition, public agencies that wish to embark on a similar effort can refer to the PMESP for a better understanding of how one can demonstrate, quantify, and evaluate the impacts of deployment toward the identified project goals.

## 1.2. Project Background

The UW ITS4US Deployment Project, one of the Phase 1 Complete Trip – ITS4US Deployment Program projects selected to showcase innovative business partnerships, technologies, and practices that promote independent mobility for all travelers regardless of location, income, or disability, aims to create the foundational data tools necessary for both public and private entities to collect, share, manage, and use transportation data that provide equitable outcomes to all travelers. At its core, the project is about creating the foundational requirements for interoperable

[1] ITS4US project documents available at: https://www.its.dot.gov/its4us/htm/publications.htm

transportation data sharing that fulfills the informational needs of all travelers. This requires a specific focus on the unmet needs of people with mobility disabilities and other historically travel-disadvantaged communities that are the focus of this project. Without implementing this type of project, the needs of these communities will continue to remain unmet or underserved, limiting the ability of citizens in these communities to access destinations, explore opportunities, and be aware of all services available to them.

The project consists of five major parts. The performance of each of these parts will be the subject of project evaluation activities described in this document.

The first part of the project includes working with existing standards committees to extend and update three existing, early-stage international data standards: OpenSidewalks, GTFS-Flex, and GTFS-Pathways. These three data standards enable the consistent collection and reporting of data that provide the underlying information needed by the currently underserved target populations—people with disabilities, older adults, and individuals with low income—to efficiently travel.

The second part of the project is to develop a series of tools that help agencies, jurisdictions, and other stakeholders collect the data that can be stored with these refined data standards. These tools are needed to lower the cost and improve the quality and consistency of those data collection efforts to increase the availability of the data.

The third portion of the project is to develop tools, policies, and procedures that allow sharing and governance of the collected data. The tasks performed will enable effective and efficient vetting, aggregation, management, and fusion of the data that participating agencies, jurisdictions, and other stakeholders collect. This portion of the project also includes tasks required to enable and manage the sharing of those data with application developers that write software to deliver requested travel information.

The fourth portion of this project is the development of a data repository to contain the data to be shared within the six counties that represent the geographic boundaries for this ITS4US project. The data repository will be developed to illustrate how these data can be collected, stored, governed, updated, and maintained over time and then served upon request to application developers.

Finally, the fifth portion of this project is the development of three example applications that use the collected data. The three applications are intended to demonstrate three very different uses of the data that are collected, maintained, and made available to application developers as a result of the other four aspects of this project. Those data can be used to fulfill a variety of information needs, and those needs can be met through an almost infinite number of applications. The three applications deployed as part of this project are meant to show other application developers how the newly available data can be obtained and delivered.

**Figure 1** illustrates the overall “new mobility” ecosystem to which the UW’s ITS4US project is contributing. The outer circle consists of the variety of public transportation services that exist. Many of these services already generate data that can be readily obtained by applications via internet connections – the act which results in the discovery of “new mobility” options. These include fixed route transit services, micro-mobility services, and taxi services. The UW ITS4US Deployment project will help add the data sources that are particularly important to people with mobility disabilities, shown in purple at the bottom of the image. These are data that describe pedestrian pathways, transit station infrastructure, on-demand paratransit and community transit

services, and other on-demand shared ride modes. The UW ITS4US Deployment project is also building the interoperable integrated transportation data sharing layer and Application Programming Interfaces (APIs) shown in the green inner circle. This is the functionality needed to collect, fuse, and aggregate the data from disparate transportation services. Finally, the UWITS4US project will demonstrate a small number of applications used by the travelers shown in the center of the diagram. The applications take requests for information from the travelers, extract the required data from the data sharing layer (green circle), perform any required tasks—such as computing navigation directions—and deliver information to users in formats (audio, text, tactile displays) designed to meet their needs.

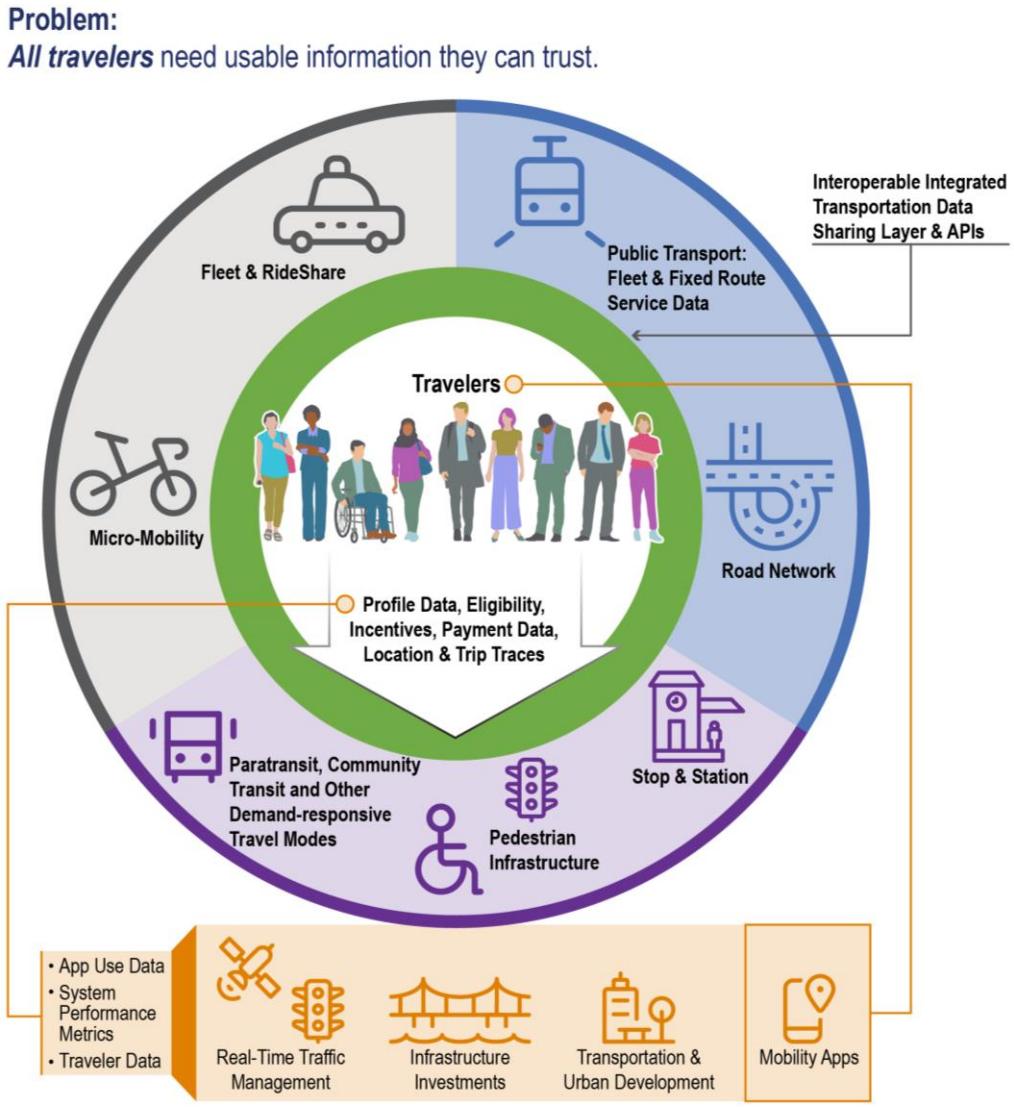


**Figure 1: Diagram. UW ITS4US Deployment Project's ecosystem.**

*Source: University of Washington.*

The project ConOps[2] describes a set of 62 user needs that drive the design of the system. The user needs statements were developed from extensive interaction with project stakeholders. Project stakeholders have been categorized based on the following five groups:

- Data generators (e.g., municipal infrastructure –owner/operators, private sector pedestrian-built-environment owner/operators, crowdsourced sidewalk reporters, elevation data providers),
- Transportation service providers (e.g., transit agencies and the companies that support the delivery of transit services operated by or for those transit agencies),
- Data service providers (e.g., mapping services, weather data providers),
- Application developers (e.g., AccessMap developers, Soundscape developers, Digital Twin developers, third-party application developers), and
- Digital device end users (e.g., travelers with sidewalk preferences, blind, vision disabled, or deafblind travelers, sighted older adults, multilingual or multicultural travelers, low-income transit users, rural transit users).

The needs described by these groups describe the basic functionality of a successful system deployment. The needs are presented in detail in Chapter 4 of the ConOps.

## 1.3. Scope

This project is currently in Phase 1 which focuses on the planning elements of the systems engineering process, in which the initial project idea is decomposed into a structured concept that serves as the foundation for more detailed design, building, testing, and operation. The structured concept includes identifying specific performance measures, targets, and capabilities associated with performance monitoring and performance measurement. The next phase, Phase 2, focuses on the design, testing, and deployment of the proposed system, while in Phase 3, the system will be operational and evaluated for its effectiveness. The PMESP is intended to inform and guide the evaluation activities to be performed in Phase 3, including setting up case studies as needed to examine changes from current conditions within the project study area.

This document is meant to plan and prepare for performance measurement and evaluation support prior to Phase 3 by 1) identifying and defining the performance measures to be used for evaluating the success of the system deployment, 2) describing the data to be collected and analysis methods to quantify those measures, as well as the sources from which those data will be obtained, and 3) describing the support being provided for that evaluation effort. For each of the proposed performance evaluation topics outlined in Chapter 2, this document covers potential performance measures and targets proposed for evaluation, confounding factors, and constraints to consider, recommended mitigation approaches to identified constraints, data sources and data

[2] Phase 1 Concept of Operations (ConOps)—University of Washington ITS4US Deployment Project, by the University of Washington and Cambridge Systematics, Inc., June 2021, Report Number FHWA-JPO-21-861. Available at: https://its.dot.gov/its4us/htm/publications.htm

collection plans, and proposed approaches to experimental design/analysis. In addition, this document also describes a data sharing framework to support the independent evaluation effort, performance reporting requirements, and high-level schedule for performance measurement activities.

The user needs documented in the ConOps serve as the basis for the identification of relevant performance measures. The data sharing framework included in this document indicates how and in what format the UW ITS4US team intends to release these performance measures as open data, which is consistent with the data sharing approach covered in the Data Management Plan. Requirements associated with each critical performance measurement and independent evaluation needs will be included in the System Requirements Specification document.

While the PMESP provides a complete plan based on the Phase 1 concept, future activities, such as Phase 2 Application Enhancement and Integration, may result in changes needed to the performance measurement and evaluation support activities. It is anticipated that many items will be adjusted, but the underlying intent of the proposed performance measurement plan will be preserved. The primary execution of this plan (e.g., data collection and analysis) will take place in Phase 3 (Operate and Maintain). Performance against baseline measurements and targets are anticipated to be routinely and publicly reported throughout Phase 3.

## 1.4. Performance Measurement and Evaluation Support Plan Purpose

The purpose of the PMESP is to develop a project evaluation approach that will determine whether the operational UW ITS4US system addresses the initial user needs documented in the ConOps. By conducting performance measurement and analysis activities, the project performance evaluation will be able to identify and quantify which deployed strategies, services, and/or components were of value in addressing the targeted Complete Trip challenges. Documenting the results with robust supporting data and analyses will enable others to understand and build upon the investments made in this project to progress toward Complete Trip goals more effectively in other deployments.

# 2. Goals and Objectives

A primary objective of the Complete Trip-ITS4US Deployment Program is to demonstrate, quantify, and evaluate the impact of advanced technologies, strategies, and applications toward addressing travelers' challenges to planning and executing a complete trip. The deployment goals and objectives introduced in the next section will be used to guide the design of the performance evaluation.

The evaluation goals and objectives for the UW ITS4US Deployment Project were refined from this initial USDOT starting point with user needs collected from data generators, transportation service providers, data service providers, application developers, and digital device end users. The main project intent is to generate data that are needed by travel-disadvantaged communities in order to plan and complete trips by creating the foundational data tools necessary for both public and private entities to collect, share, manage, and use transportation data that provide equitable outcomes to all travelers, and then demonstrate how the use of those tools can improve the mobility of travel disadvantaged populations by resolving current information gaps. Without implementing this type of project, the needs of these communities will continue to remain unmet or underserved, limiting the ability of citizens in these communities to access destinations, explore opportunities, and be aware of all services available to them.

In addition to generating data in six counties spread across three states, the project will develop tools which lower the cost of data collection throughout the country, as well as put in place procedures and processes that help make those data readily accessible to application developers, and continuously improve the quality and quantity of that data. Finally, the project will support three example applications which demonstrate three very different ways in which the data being generated and made available can be used to improve the mobility of disadvantaged populations

## 2.1. Deployment Goals and Objectives

To achieve this project's overall goal, as described above, the project's three stated goals, which are

- coordinate collaborative releases of data standards,
- publish and maintain interoperable data infrastructure
- deploy and sustain three accessible mobility applications

have been reorganized slightly into five specific activity goals, each of which is tied to multiple objectives. Essentially, the second of the original goals has been divided into three goals that are more readily evaluated. These goals and associated objectives are shown in Table 1. In the table, each project goal is followed by a short title, shown in underlined text, which is used to reference that goal throughout the remainder of this document.

**Table 1: UW ITS4US Deployment Project Goals and Objectives**

| Project Goals | Project Objectives |
|---|---|
| Extend three data standards, OpenSidewalks, GTFS-Pathways, and GTFS-Flex to better meet user needs (**Data Standards**) | 1. The extended data standards meet the informational needs of the underserved traveling public<br>2. The data standards are designed such that the data can be objectively collected by available technology<br>3. The data standards are widely accepted by the national transportation community |
| High quality data can be efficiently collected, quality assurance checked and shared (**Data Generation**) | 1. Data quality is high enough to satisfactorily meet user needs<br>2. Tools built as part of this project to collect and publish data work effectively<br>3. Data are widely available (more than half of all jurisdictions/agencies have data in the system by the start of Phase 3) across multiple counties and from multiple agencies<br>4. Data can be uploaded to a central data sharing platform<br>5. Feedback on data quality (or errors) is routinely used to improve and maintain the quality of available data |
| A bi-directional data feedback process can be implemented, is accepted, used, and maintained (**Data Vetting**) | 1. Agencies that own infrastructure or supply transportation services actively participate in the feedback process that vets the accuracy of the data which describe their infrastructure or services<br>2. Community and advocacy groups actively participate in the data vetting/feedback process |
| Collected data are readily available to application developers through a robust data sharing system (**Data Service Provision**) | 1. API performance meets system requirements goals.<br>2. API availability meets the system requirements performance targets.<br>3. 3rd party application developers routinely build applications that take advantage of the data available through the project APIs<br>4. The data sharing system passes all data security assessments, which are routinely performed |
| Demonstration applications successfully demonstrate the use of, and benefits from the deployed data sharing system (**Demonstration Applications**) | 1. Demonstration applications reliably obtain data from the data service and deliver it to end users<br>2. Demonstration applications are routinely used by a growing number of users<br>3. User satisfaction with the demonstration applications results in their continued use of the applications after the conclusion of testing<br>4. Users demonstrate improved travel outcomes from use of the demonstration applications |

Chapter 3 of this report describes how these goals and objectives will be evaluated. It includes a discussion of what performance measures will be used, what sources are available to the project team to supply data for those performance measures, what confounding factors exist and how those factors will be mitigated, and how the data collected will be shared with the USDOT's independent evaluator.

## 2.2. Use Cases/Scenarios and Performance Evaluation Topics

The UW ITS4US ConOps presented 13 operational scenarios that provide a wide range of examples of the ways in which different stakeholders will interact with the TDEI. It is important to note that several of the 13 scenarios illustrate how the TDEI will interact with third-party applications – which are not directly part of this project – and how the services provided by those third-party applications provide travel benefits to individuals with travel disabilities. Providing the data that facilitates the development of a wide range of third-party applications – and not those third-party applications themselves – is the primary intention of this project. The relationship between each operational scenario and the five primary project goals is shown in Table 2.

It can be seen in Table 2 that the majority of operational scenarios have attributes that fit within the goals and objectives that will serve as the basis for the project evaluation. The evaluation will therefore be oriented around how well these outcomes are achieved. This means that the evaluation is driven not by the achievement of specific travel activities impacted by a new service – as is common in many ITS technology deployment pilots – but instead will be driven by how effectively data can be collected and delivered that will effect a very large number of missing informational needs across multiple travel activities – thus achieving the overall ITS4US goal of making travel options easier to discover and use by individuals with a variety of mobility disadvantages.

Specific evaluation hypotheses, the performance measures that will be used to evaluate those hypotheses and the data sources for those measures are discussed in the next chapter of this document.

**Table 2: Relationship Between UW ITS4US Operational Scenarios and Project Goals and Objectives**

| Scenarios | Standards | | | Data Generation | | | | | Vetting | | System Performance | | | | Demo Application Performance | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Meets user needs | Uses objective data | Widely accepted | Quality is high | Tools work well | Are widely available | Uploads to central point efficiently | Feed-back routinely applied, improving quality | Agencies partici-pate in vetting | Vetting by local groups | API execution is good | API availability is good | 3rd party developers participate | Data security is high | Demon-stration application function is good | Use of demon-stration applica-tions is high | User satis-faction is good | Travel outcomes are safe and efficient |
| 1) Sidewalk data generation, collection, and vetting. | n | Y | Y | Y | Y | Y | Y | Y | Y | n | n | n | n | n | n | n | n | n |
| 2) Vetting of sidewalk data and street crossing identification. | Y | Y | Y | Y | Y | Y | Y | Y | n | Y | n | n | n | n | n | n | n | n |
| 3) Generation and vetting of GTFS-Pathways data. | Y | Y | Y | Y | Y | Y | Y | Y | Y | n | n | n | n | n | n | n | n | n |
| 4) Generation and vetting of GTFS-Flex data. | Y | Y | Y | Y | Y | Y | Y | Y | Y | n | n | n | n | n | n | n | n | n |
| 5) Individual with mobility disability uses verified sidewalk and transit data to navigate through several cities. | Y | Y | Y | Y | n | Y | n | n | n | n | Y | Y | n | Y | Y | Y | Y | Y |
| 6) Veteran with mobility disability traveling from a rural home to the Veterans Affairs (VA) hospital for a medical appointment. | Y | Y | Y | Y | Y | Y | n | n | n | n | Y | Y | n | Y | Y | Y | Y | Y |
| 7) Blind, vision disabled, or deafblind individual uses verified sidewalk and transit data. | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | n | Y | Y | Y | Y | Y |
| 8) Multilingual tourist tries to conduct pre-trip planning for a multilevel transit station. | Y | Y | Y | Y | Y | Y | Y | Y | n | n | Y | Y | n | Y | Y | Y | Y | n |

| Scenarios | Standards | | | Data Generation | | | | | Vetting | | System Performance | | | | Demo Application Performance | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Meets user needs | Uses objective data | Widely accepted | Quality is high | Tools work well | Are widely available | Uploads to central point efficiently | Feedback routinely applied, improving quality | Agencies participate in vetting | Vetting by local groups | API execution is good | API availability is good | 3rd party developers participate | Data security is high | Demonstration application function is good | Use of demonstration applications is high | User satisfaction is good | Travel outcomes are safe and efficient |
| 9) Low-income traveler utilizes a third-party application (One-Call/One-Click Service) to reach a destination. | Y | Y | Y | Y | Y | Y | Y | n | n | n | Y | Y | Y | Y | n | n | n | n |
| 10) Travelers with sidewalk preferences utilize data generated by a city government. | Y | Y | Y | Y | Y | Y | Y | Y | n | n | n | n | n | n | Y | Y | Y | Y |
| 11) Travelers with sidewalk preferences utilize data generated by civic organization. | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 12) Travelers with sidewalk preferences utilize data generated by an aerial mapping company's analytics engine for aerial images. | Y | Y | Y | Y | Y | Y | Y | Y | n | n | Y | Y | Y | Y | Y | Y | Y | Y |
| 13) Transit users utilize GTFS-Flex and GTFS-Pathway extensions through a navigation application. | Y | Y | Y | Y | Y | Y | Y | Y | n | n | Y | Y | Y | Y | Y | Y | Y | Y |

# 3. Design of the Performance Evaluation

This chapter presents the design of the planned performance evaluation of the UW ITS4US project, the Transportation Data Equity Initiative. It presents the hypotheses to be tested to determine the degree to which the project's goals have been met. It also presents the performance measures to be used to test those hypotheses, the data sources from which those measures will be obtained, and confounding factors which can limit the ability of the evaluators to accurately draw conclusions. In short, this chapter describes the plan for evaluating the TDEI.

## 3.1. Background

The USDOT's ITS4US program and Accessible Transportation Technologies Research Initiative (ATTRI) recognize that underserved individuals have different and unique travel needs, even among individuals within a specific stereotyped user group. In addition, the ITS4US Program recognizes that there is often overlap between these populations, so opportunities exist to implement a solution that serves individuals in a customizable manner, rather than categorizing travelers into user groups. For example, slope steepness is not a concern of only users with certain disabilities (e.g., wheelchair users) or older adults, and it would be inappropriate to exclude such concerns from mobility applications that offer nonvisual directions simply because the stereotype for travelers with visual disabilities tends to overlook such concerns.

The ultimate 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. It is specifically designed to provide data that are missing, but badly needed, so that they can be converted into personalized information that supports inclusive access to mobility.

To provide these new information sources in ways that can be scaled to meet the needs of individuals throughout the country, this project is performing several tasks. These tasks are:

- Developing enhanced data standards for sidewalks, transit center pathway and feature descriptions, and on-demand transit services, so that when populated with data, the information required by users about these services can be delivered to them via a variety of Internet enabled applications.
- Building tools and processes that dramatically improve the availability of sidewalk, transit center pathway, and on-demand transit service data that use those new standards.
- Building tools and processes that provide for effective data quality control associated with these data, including the ability for community and advocacy groups to provide feedback on where published data are missing or in correct.
- Developing systems and services that allow for sharing of the data being collected with a wide variety of 3rd party developers.

- Demonstrating the use of these data through a variety of internet enabled applications, including an improved ability to discover and perform complete trips using a combination of sidewalks, fixed, and on-demand transit services.

The UW ITS4US project will perform these tasks within six counties in three states. These are King and Snohomish Counties in Washington, Multnomah and Columbia counties in Oregon, and Baltimore and Harford Counties in Maryland.

The performance evaluation for this project is designed to use the data being generated in these six counties to determine the degree to which these tasks are being successfully accomplished.

## 3.2. System Deployment Impact Analysis

While many ITS evaluations focus on the measurement of travel behavior of a specific target population before and after a specific pilot deployment has occurred, this traditional approach to evaluation is not really appropriate for evaluating the TDEI, where the primary desired outcome is to generate a large amount of new, high-quality data that can be delivered to a wide variety of travelers, with very different travel disabilities, and therefore very different unmet information needs. These travelers are expected to gain access to the information through applications which will be primarily provided by 3rd party application developers. The applications built by these developers are not part of this funded project, and therefore the travel behavior changes that result from these as-yet-to-be defined applications will not be evaluated as part of this project. These are downstream benefits that can only occur if the data being generated and made available through this project become widely available to those 3rd party applications.

Consequently, the UW ITS4US team views the success of the project as being determined more by the availability of data and the ability to continue to scale the collection and sharing of that data than it is about the changes in travel behavior from the Multi-Modal AccessMap application or use of the Soundscape or Digital Twin applications. While we expect to see positive travel outcomes in terms of improved mobility and increased user satisfaction with their travel experiences from our demonstration applications, the UW team's expectation is that far more public benefit will come from widespread deployment of 3rd party applications that use the data the project generates or causes to be generated. This is because an unlimited number of 3rd party applications can be delivered that solve an almost unlimited number of transportation information problems for people with mobility disabilities, but only if the data are present for those applications to deliver. This project is designed to help generate and deliver those data. Our applications, while beneficial in their own right, are primarily intended to illustrate how the data and data services can be used.

Thus, this project's evaluation will focus on the collection and delivery of the data that make these future mobility applications possible. Other than a modest number of pilot datasets, there is no current collection and delivery of the data the project is seeking to generate and publish. Consequently, the project's evaluation focus is on the collection and delivery of that data, and key evaluation metrics will measure "change" in terms of "what has now been collected and delivered" as the baseline condition is "no data."

In support of illustrating the benefits from the collection and delivery of these data via future use of that data by unlimited numbers of 3rd party applications, this project does include the deployment of three applications, which will demonstrate how the data can be used in very different types of applications that improve the mobility and quality of life of travelers with disabilities. One of these applications, Multi-Modal AccessMap is expected to demonstrate some

direct travel behavior changes, in that it is designed to provide new travel routing capabilities that should positively change travel outcomes for its users. Travel behavior changes for that application will be examined as part of this evaluation but are not the primary focus of the evaluation.

The design of the evaluation can be divided into five major reporting efforts.

- The effectiveness and acceptance level of the three data standards being expanded to store and deliver data about sidewalks, pathways, transit center layouts, and on-demand transit services that are not currently available to travelers.
- The amount and quality of the data developed as part of the project, including the performance of tools developed to generate those data.
- The performance of data vetting systems put in place to further increase the quality and reliability of data shared with the public, through the design and application of multiple agency and crowd sourced error identification and reporting systems.
- The performance of the data sharing system which makes the data accessible to third party applications via Internet accessible application programming interfaces (APIs).
- The performance of the three demonstration applications, both in terms of their ability to access, download, and deliver the data being collected in this project, and in the case of Multi-Modal AccessMap, the travel outcomes achieved by delivering that data to users.

In addition to the technical evaluation of these five project components, the project will also examine the level of satisfaction stakeholders have with the data, systems, tools, and applications they interact with, to determine if the needs they have expressed as project stakeholders are being met.

### *3.2.1. Approach/Strategies for Focused Performance Analysis*

The TDEI is designed to develop, collect, and publish three datasets which currently do not exist across the vast majority of the U.S. (OpenSidewalks, GTFS-Pathways, and GTFS-Flex), and that are badly needed to improve the ability of people with mobility disabilities to travel successfully. Thus, the before condition is "little or no data are available." For this 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?" Since the project is essentially 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 only describe the degree to which the data collection and data delivery has been accomplished.

However, the UW ITS4US effort is not just about collecting and publishing data. The project will be delivering tools and procedures designed to significantly lower the cost and effort required to generate these data and put in place procedures and processes which help ensure the quality of those data. By treating these tools and systems as business processes, and the TDEI as the "business" that wishes to constantly improve on its performance (lowering costs of delivery, improving market size, increasing the customer base), the UW team will be adopting good business practices that generate management data from the work activities being performed. The evaluation will therefore use the management data associated with these tasks for the evaluation of those activities.

The data being collected for delivery to travelers as well as the data collected to actively manage the TDEI will serve as the core data sources for the project's evaluation. This is because the data needed to manage the TDEI are the same as that needed to evaluate its performance. That is, the TDEI management activities include the following:

- Understanding the quantity of data being collected and where geographically those data are and are not available,
- Understanding which data generation and collection tools are working well, and which are not, so that tools that are under performing can be refined or improved,
- Monitoring the success of the data vetting process, which is required to ensure that data accuracy remains high and that errors discovered in those data are quickly fixed,
- Monitoring the performance of the data sharing systems, to ensure that data are being uploaded efficiently and can be accessed efficiently by 3rd party applications.

To support the management process, the UW ITS4US team will be including in the development of the software tools which help generate, collect, store, and publish these data, a variety of performance reporting tools. For example, when artificial intelligence/machine learning software is used to identify routable sidewalk paths and path attributes from imagery, confidence intervals associated with those data will be generated. These statistics will indicate whether the imagery is robust enough to allow the AI software to be confident of the sidewalk's presence, location, and other attributes. From the "business management" perspective, these statistics will be used to determine the level of vetting required before these data can be published.

Where high levels of confidence in the data are present, the data can be uploaded to the OpenSidewalks dataset. Where lower levels of confidence are present, the data will need to be confirmed – or refined - by some other data source before being published. For example, a community group performing vetting activities might be assigned to manually review and confirm/refine those initial data points. The outcomes from these vetting activities (e.g., the number of changes made to the initial sidewalk attributes and the specific attributes that must be changed) can then be used to indicate where additional refinements in the data generation software is required. Similar evaluation activities will be performed for the vetting of on-demand transit service delivery and GTFS-Pathways data development.

When summarized, these data sets (both the initial data quality scores and the results of the vetting process), which are designed to inform the management of the system, are also excellent data sources for evaluating the degree to which the project's goals have been achieved. They will describe the amount of data that are available at the end of the project, the quality of that data, the degree to which agencies, cities, and community groups are participating in the project and maintaining those datasets, and the overall performance of the data sharing system put in place to make that data widely available to 3rd party applications.

Only two aspects of the system need "evaluation data" to be collected. The first is associated with obtaining the opinions and perspectives of the project stakeholders. The second are the outcomes of the demonstration applications.

### *3.2.2. Experimental Design*

The primary experimental design adopted by the UW ITS4US team is to report the data that will be routinely collected as part of the operation of the software systems, tools, and applications being developed and deployed. With a few exceptions, these same data are expected to be

routinely used to manage the system once it is deployed, as well as being used to guide further improvements to the tools and systems being developed.

In addition to the system's operational and management data, the evaluation effort will require a modest amount of additional data collection. These additional data are needed to:

- Understand how stakeholders actively participating in the project perceive the performance of the project, where these stakeholders include
  - agencies that supply data to the system (e.g., on-demand transit service providers),
  - agencies that own infrastructure that has been described in the published data stream, but may or may not have collected those data,
  - independent community groups and advocacy groups that collect or vet data that are being published about their community or services that they wish to see used,
  - 3rd party application developers that wish to use the TDEI in order to build applications that can deliver travel information to end users, and
  - travelers that use the demonstration applications.
- Measure the observed travel outcomes of a set of test subjects that use the Multi-Modal AccessMap application, which is the only demonstration application that is part of this project that is designed to 1) help travelers with mobility disabilities discover rotes and modal options that were not previously readily accessible to them, and 2) direct travelers with mobility disabilities through a complete trip.

The evaluation will primarily be performed by analyzing, summarizing, and reporting the statistics that describe the amount and quality of the data made available through the project, the quality of the data, the level of interaction the participants have with that data to ensure high ongoing data quality, and the performance of the central data system. These outcomes will then be supplemented by results of the stakeholder surveys which describe the degree of satisfaction the stakeholders have with the data system components and their use, and finally, by the travel outcomes associated with the use of the Multi-modal AccessMap application, which will illustrate the benefits to be expected to be delivered by 3rd party travel applications.

### *3.2.3. Selected Evaluation Scenarios*

The design of the UW ITS4US performance evaluation does not directly follow the 13 Use Cases described in the ConOps report for this project. Instead, the "evaluation scenarios" are oriented to the five goals being used to direct the evaluation effort for the project which are listed on in Section 2.1.

- Refinement of data standards for sidewalks, transit center pathways, and on-demand transit service to that those standards meet the needs of stakeholders.
- Generation of the large amounts of high-quality data that can be used to meet a vast array of information needs for travelers with mobility disabilities.
- Development and delivery of data vetting procedures that are actively used to ensure the continued high quality of the data being published.
- A data sharing software system which operates with a very high degree of availability and fast response times

- Demonstration applications that effectively illustrate a variety of uses of the data being collected, that have a high degree of customer satisfaction, and allow for beneficial travel outcomes that were not previously available.

The approaches to, and data to be used to perform each of these five evaluation topics, are described in the next five sections of this chapter. A crosswalk between the 13 ConOps Use Cases and the five primary goals was presented above in Table 2.

## 3.3. Data Standards

### *3.3.1. Introduction*

The goal of the data standards effort is to refine and update the three existing data standards (OpenSidewalks, GTFS-Pathways, and GTFS-Flex) so that they more effectively describe these three sets of transportation infrastructure and services which are vital for meeting the complete trip needs of many travelers with mobility disabilities. Objective and functional data standards are required if the data collection process is to generate the data needed by travelers.

The project team has adopted three basic objectives to be evaluated for this project goal. These objectives are:

- The extended data standards meet the informational needs of the underserved traveling public.
- The data standards are designed such that the data can be objectively collected by available technology.
- The data standards are widely accepted by the national transportation community.

The first two of these objectives will be evaluated by conducting reviews of stakeholder input to the data standards refinement process and by conducting survey of stakeholders participating in the project. The stakeholders to be included in this effort fall into two categories. The first are the agencies, organizations, and firms that are actively participating in the international standards committees for these three data standards. The second group to be included are the individuals with lived-experience that need to be able to use the data collected, stored, and published using these standards. This second group of individuals will be identified from our ongoing co-Design process and project stakeholder groups.

The UW team readily acknowledges that not all user needs will be met by any set of standards produced in conjunction with this project. For example, it is well documented that real-time status of sidewalks (e.g., is the sidewalk blocked by a sandwich board, or a badly parked bicycle?), transit centers (e.g., is the elevator between the street and platform level operating today or is it broken?), and on-demand transit service (e.g., when is my vehicle arriving?) is a need. Unfortunately, those data are not readily obtained, and are therefore outside of the scope for this project.

Therefore, in addition to the evaluation of the standards that are adopted, the evaluation will identify those data items (such as real-time status) for which data are not available, and for which further enhancement of the data standards will need to be made, as technology advances and data sources that can produce those missing data become available.

The third objective, the need to have the standards widely adopted, will be answered directly based on the degree to which these data are available within the six counties where this project is performing deployment. A second aspect to this evaluation topic is the degree to which other agencies and jurisdictions around the country have adopted these standards. For example, within the ITS4US project, the Calact team is potentially adopting both the GTFS-Flex and OpenSidewalks standards. The adoption of these standards outside of the project demonstrates the national scope of the benefits gained in part as a result of the work done in this project.

To measure the achievement of these objectives, the project team will report the level of success in achieving the basic objective. This is not a “before and after” analysis, because in all cases, the “before” condition is that the standard is not fully formed. Thus, the evaluation is simply designed to report on whether stakeholders agree that these standards meet their needs, can be collected in objective ways, and are being widely accepted.

### *3.3.2. Potential Performance Measures and Targets*

While the extension of the data standards is key to this project, directly measuring the effectiveness the standards is difficult, especially given the fact that many needs (such as real-time infrastructure conditions) cannot be collected. Thus, the UW team plans to use survey-based measures of user satisfaction from our stakeholders combined with summaries of the review comments submitted as part of the data standards adoption process to identify the degree of satisfaction the data standards have achieved with those working on those standards.

The UW team will also conduct surveys of

- digital device users (e.g., individuals with lived-experience),
- agencies that own or provide services (e.g., transit agencies and cities), and
- application developers,

to evaluate the degree of satisfaction the project stakeholders have with the standards being deployed. The UW team will use a 5-point Likert scale to judge overall satisfaction, with additional survey questions designed to identify specific areas of concern. Different surveys will be designed for each of the stakeholder groups, because the groups will have different areas of concern.

The performance measures to be used to evaluate these three objectives are shown below in Table 3.

**Table 3: Summary of Data Standards Performance Measures**

| OBJECTIVE | TOPIC | INFORMATION |
|---|---|---|
| Data standards meet user needs, including being objective | 1. Performance Measure | 1a. Likert scale survey response to stakeholder survey questions about the level of satisfaction with the ability to determine and use routes given the data available with the various data standards. Questions will be asked about the users' ability to identify routes that meet their specific travel needs, given specific types of available data. |
| | 2. Data Needs | 2a. Separate surveys with a) users with live-experience, b) OSW transit data generators, c) GTFS-Pathways data generators, d) on-demand transit service data generators, e) application developers |
| | | 2b. Review comments from data standards reviews |
| | 3. Experimental Design | 3a. Reporting of participant perceptions using surveys with Likert scale style questions. Multiple surveys will be performed, with each aimed at a different type of stakeholder group (e.g., OSW, versus GTFS-Pathways, vs. GTFS-Flex.) One survey per type of data is intended for project evaluation purposes. |
| | 4. Hypothesis | 4a. The OSW standard meets the travel needs of the project stakeholders |
| | | 4b. The GTFS-Pathways standard meets the travel needs of the project stakeholders |
| | | 4c. The GTFS-Flex standard meets the travel needs of the project stakeholders |
| | 5. Targets | 5a. Initially set at 4.0 out of 5 on five-point Likert Scale |
| | 6. Risks | 6a. Assumes stakeholder panel is sufficiently representative |
| | 7. Other information | 7a. Review comments from international standards process provides insight for future extensions that need to be implemented as data collection technology improves. |

| OBJECTIVE | TOPIC | INFORMATION |
|---|---|---|
| Data standards are widely accepted | 1) Performance Measures | 1a) The number and percentage of cities producing OSW data within our six counties<br>1b) The number and percentage of transit centers for which GTFS-Pathways data has been produced within our six counties<br>1c) The number and percentage of on-demand transit service providers that are producing GTFS-Flex data within our six counties<br>1d) The number of cities producing OSW data in the U.S.<br>1e) The number of transit agencies producing GTFS-Pathways data in the U.S.<br>1f) The number of on-demand transit service providers that are producing GTFS-Flex data in the U.S. |
| | 2) Data Needs | 2a) OSW database<br>2b) GTFS-Pathways database<br>2c) GTFS-Flex database<br>2d) Count of cities within the six county study area<br>2e) Count of transit agencies with transit centers in six county study area<br>2f) Count of on-demand service providers in six county study area (from State DOT partners) |
| | 3) Experimental Design | 3a) Reporting of trends in performance measures over time to measure the growth in acceptance over time. |
| | 4) Hypothesis | 4a) The number and percentage of cities/agencies producing data in these standards within our regions is high |
| | 5) Targets | 5a) 80 percent of cities within the six counties are participating in OSW by the end of Phase 3<br>5b) 100 percent of transit centers in the six counties have GTFS-Pathways data by the end of Phase 3<br>5c) 80 percent of on-demand service providers within the six counties are producing GTFS-Flex data by the end of Phase 3 |
| | 6) Risks | 6a) Assumes that state departments of transportation are aware of all on-demand service providers in the six counties, or that those providers can be identified by other means. |
| | 7) Other information | 7a) National expansion is outside of the scope of this project but is still a valuable performance measure to report. |

This project is not concerned specifically with the change from baseline to post-deployment conditions since the baseline condition is overwhelmingly a case of lack of use of the existing data standards. Thus, the focus of the evaluation effort will be on the level of deployment of the three standards achieved by the end of Phase 3 and the degree to which the wide range of stakeholders see value and benefit in that deployment.

### *3.3.3. Confounding Factors and Constraints*

The primary confounding factors to the evaluation of the data standards themselves is that other data standards, such as the General On-Demand Transit Feed Specification (GOFS) may take precedence over the standards being used in this deployment effort. A secondary constraint is that – like real-time data on elevator operations – it may prove difficult to collect required data at a cost that allows those data to be widely obtained and published. Such an outcome might hamper the overall adoption of the data standard.

### 3.3.4. Mitigation Approaches

The UW team does not view the adoption of competing standards such as GOFS as a negative to the project. If GOFS is embraced more than GTFS-Flex, and GOFS still allows the routing functionality that is desired from GTFS-Flex, the project team is willing to adopt that technology in place of GTFS-Flex. As a result, the UW team is participating in both the GTFS-Flex and GOFS standards development efforts to ensure that the "best" standard is adopted. In addition, the UW team, as well as others working on the international data standards committees for both GTFS-Flex and GOFS are working to ensure that these standards are interoperable. This allows data collected according to one standard to be easily converted into the other standard, thus mitigating the cost of shifting to an alternative standard.

To mitigate the potential failure to reduce the cost of the data collection tasks, the UW team is developing multiple approaches to data collection for all three standards. This approach lessens the impact of any one data collection approach proving to be overly expensive or inefficient.

### 3.3.5. Data Sources and Data Collection Plans

The data sources required for measuring the effectiveness and acceptance of the three data standards are twofold.

- A set of surveys and interview results with stakeholders that describe their acceptance of and perceptions of the degree to which the data standards meet their needs
- Direct measurement of the amount of data available within the project data system for the project's six county deployment region.

For additional supporting information that can describe the overall acceptance of the data standards, the evaluation will also examine the number of other agencies and cities in the U.S. that are publishing data using these standards.

#### 3.3.5.1. Data Needed

The data required for these analyses are summarized in Table 3. The data required for the evaluation of the data standards include the results from a series of survey questions asked of the project stakeholders, with different surveys given to different stakeholder groups. For example, the survey given to individuals with lived-experience will focus on their ability to use the data being generated and delivered to them as a result of the new data standards in order to resolve travel barriers. The survey questions given to on-demand transit service providers will ask questions about the clarity and objectivity of the variable definitions for which they are generating data about their services, as well as the ability of those data to accurately represent the transit services they are providing. The survey questions sent to application developers will examine the degree to which the data standards are meeting the needs of the developers in the development and delivery of new transportation services.
The result of these surveys will be the breadth of information needed to evaluate whether the data standards are meeting the needs of the various stakeholders participating in this project.

To examine the degree to which the standards are being widely accepted within the project study area, the data needed are already part of the central data sharing system. The evaluation simply needs to query and summarize those data. There is no baseline condition. (The baseline is that no data exist.) The project team will summarize and report statistics on the amount of data being published at least quarterly during Phase 3 of the project. This will allow the reporting of

deployment trends during Phase 3. To examine the more national level of deployment of these data, it is possible to directly determine the availability of data via examination of available Internet connections and published APIs.

#### 3.3.5.2. Data Collection and Quality Checks

The primary “data collection effort” for this portion of the evaluation will be surveys conducted during Phase 3 of this project. Data quality is not expected to be a significant problem with the survey responses, since the survey distribution will go to individuals already participating in the project as stakeholders.

Data quality controls for the data stored in the project data bases are discussed in other project documentation, such as the Systems Requirements. No additional data quality activities are planned for the data already stored in the project databases that will be summarized for this evaluation task. See section 3.4 for the evaluation of the quality of the data being collected in the three data standards.

#### 3.3.5.3. Cost of Data Standards Efforts

While costs for the data collection effort have not been estimated at this time, these costs are not expected to be substantial, as the only data collection activities required specifically for the evaluation effort are the surveys of the stakeholders. Since the stakeholders are already known, the cost of the surveys is simply the development, execution, and analysis of the surveys themselves.

### *3.3.6. Experimental Design*

The experimental design for these evaluation topics is based on straightforward reporting of summary outcomes. The surveys are intended to be performed during Phase 3 to learn how successful the Phase 1 and 2 co-design process has been at identifying the data needed to satisfy the information needs for route identification and navigation of the target population. They provide an evaluation of the data standards from the perspective of the end users. In contrast, the measurement of the trends associated with the amount of data published for use by 3rd parties that describe the geographic scope of the system describe, in part, the acceptance of the standard by the owners of the infrastructure and services that need to be described. In both cases, the experimental design is a straight-forward reporting of acceptance levels of the new standards at the time of the data collection.

## 3.4. Data Generation

### *3.4.1. Introduction*

The goal of the data generation evaluation effort is to determine if the data being collected, generated, and made available via the data service APIs are of high quality and are widely available. This evaluation task is the heart of the evaluation of this project, since the primary goal of this project is to collect and publish data that can be used to provide a wide range of mobility

improvements for people with mobility disabilities, understanding the quality and quantity of the data is the key to evaluating the project's outcomes.

The project team has adopted four objectives to be evaluated for the overall data generation goal. These objectives are:

- Data quality is high enough to satisfactorily meet user needs, meaning that the tools provided to generate that data work well.
- Data are widely available within the project study area.
- Data are being efficiently uploaded to a central data center
- Feedback about the quality of the data being uploaded is routinely being applied to those initially reported data to continuously improve data quality.

The first of these objectives will be evaluated by reviewing and summarizing the outcomes from the project's quality control procedures, both as the data are initially generated, and as a result of the data vetting process being developed to both improve the quality of data and provide mechanisms by which interested advocacy groups can aid in keeping the data system up to date and accurate over time.

The tools being developed for use by project stakeholders are designed to provide a first line of quality control. For example, when initially estimating the location and characteristics of sidewalk infrastructure, the machine learning algorithms report levels of confidence associated with model outcomes. These will be routinely summarized by the project team as part of the operation of those tools, which use those results for both determining the acceptability of those data outcomes prior to the publication of those data, and for the continued improvement of those data generation tools. For example, low confidence scores can be used to both identify where human data vetting is required, but also indicates the need for additional data collection for use in recalibration of the machine learning models used to generate sidewalk data in that specific sidewalk environment.

Similarly, the tools being developed for codifying on-demand transit services often require staff to enter service attributes into web forms, which convert those data entries into the required data standards. Quality assurance subroutines review those form entries prior to publication of that data to ensure that the data being entered match the data formats required and do not include numbers that exceed the boundaries of expected values (e.g., suggesting that the data were entered using the wrong units – meters instead of kilometers.) The outcomes of these data checks will both be reported to the agency performing the data generation, so that the flagged values can be check and fixed as needed and will be sent to the TDEI so that they can be summarized at the programmatic level and used to understand where improvements in the data generation software need to be made. The UW team will both use these data for its own evaluation purposes and share these data generation outcomes with the independent evaluation team.

The next objective examines the performance of the data upload process. This evaluation task simply examines the performance of the tasks and systems required to allow data generators to share the data they have created with the TDEI. The evaluation of this objective focuses on the performance of the data transfer process. If that process works smoothly and efficiently, it lowers the time and cost for obtaining data to both the data collectors and the central data repository. If this process does not work easily and efficiently, it creates barriers to participation in the data sharing process, reducing the likelihood that agencies/cities will continue to participate in the data

sharing process, to the detriment of travelers that need access to those data to travel more freely and safely.

The final evaluation topic examines the degree to which the “owners” of the infrastructure and transit services being described in the published data use the feedback they are being provided about the accuracy of the data that describe their services and infrastructure. All data systems contain errors. The key is finding those errors quickly and efficiently updating the data system with the correct information once it has been identified. Given the limited resources most agencies/cities have for data collection and reporting, the UW team has designed both a significant QA/QC process as the data are first generated, and a robust data vetting process which provides for continuing review of data quality both within an “ownership” agency and by enteral volunteers that have a stake in the quality of the data being reported.

This data vetting process can identify potential errors in the database, but to successfully result in an improvement in the data, the feedback which comes from these vetting activities need to be acted on. That is, a report of invalid data must be confirmed and the update to the database must be made. This evaluation task will examine the degree to which these feedback tasks are being performed by agencies that participate in this project by supplying data to the system.

### *3.4.2. Potential Performance Measures and Targets*

The performance measures to be used for the evaluation of the overall goals of data quality and availability fall into four basic categories.

- The quantity of data being generated
- The quality of that data
- The performance of the data upload process
- The degree to which the data vetting feedback process is being used.

Quantity will be measured in terms of the number of agencies and cities for which data are available within the study area and the fraction of those agencies and cities for which data are available. These statistics can also be reported as a fraction of the available network for which data are available. For sidewalk data, the “population” of possible sidewalks will be computed in terms centerline street miles, and it will be compared against the number of center-line street miles which have been examined for the existence of sidewalks. In this manner, the reporting statistic is not incorrectly punished for those instances where a street has been examined for sidewalks, only to determine that no sidewalk exists. Neither will the statistic over represent the fact that two sidewalks have been generated for one street segment, when sidewalks are present on both sides of that street. For transit service provision, the project team will rely on our state Department of Transportation partners for an accurate measure of the total number of service providers within the six county service area.

For data quality, the evaluation approach relies on the output from the data generation and data vetting procedures being developed and put in place for each of the three different types of data (OSW, GTFS-Pathways, and GTFS-Flex). The data generation tools are designed to provide immediate feedback to the staff performing the data generation tasks. These results will be stored for use in the evaluation. Similarly, the data vetting and feedback/improvement procedures also

are designed to provide an audit trail of what errors (or potential errors) have been identified and what action was taken as a result of the identification of those errors/potential errors.

By summarizing these quality assurance / quality control feedback reports, it is possible to gain a reliable understanding of the quality of the data contained in the TDEI that is being shared with travelers through either the demonstration applications or third-party application developers. Data quality summaries will be reported separately for each type of data and for each agency/city, and then summarized for the project as a whole.

The remaining two objectives are of lesser importance to the overall project goals but are still important evaluation outcomes. The first of these simply examines the performance of the data upload process. This can be accomplished by summarizing the outputs of the upload API logs. These indicate the number of uploads which occur, the time required, and if errors occur.

The final area of evaluation examines the number of agencies/cities that are actively responding to data vetting feedback. When agencies actively engage with the vetting process, they improve the accuracy of the published data, either by confirming that the currently published data need to be updated, or by showing that the current data are correct, and the change requested by the vetting process is itself in error. In both cases, data accuracy is improved. The goal of this objective is to determine if the growing use of the data by end users, combined with easy to use tools which lower the cost of collecting and maintaining data, encourages transportation agencies and cities to maintain these datasets more actively once they are built.

A summary of the performance measures to be used to evaluate these objectives is shown below in Table 4.

**Table 4: Summary of Data Generation Performance Measures**

| OBJECTIVE | TOPIC | INFORMATION |
|---|---|---|
| Data quality is high / Data generation tools work well | 1) Performance Measures | 1a) Number/fraction of errors reported in QA/QC/vetting of sidewalk / transit center / service process (e.g., how many errors in the sidewalk data are reported via the community vetting process, or how many errors are reported in the automated QA checks of the GTFS-Flex data files.) |
| | | 1b) Data quality score from AI/ML sidewalk system |
| | | 1c) Likert scale survey question about the level of satisfaction with the data quality asked of stakeholder agencies |
| | 2) Data Needs | 2a) Log of vetting updates from: OSW Task Manager software, GTFS-Pathways vetting software, and GTFS-Flex vetting software |
| | | 2b) Output of AI/ML sidewalk data generation software |
| | | 2c) Stakeholder survey of data generator stakeholders producing data |
| | 3) Experimental Design | 3a) Tracking the number of errors being reported, the fraction of reported data that contain errors, and the quality scores from the AI/ML software to report trends over time, along with the perceived level of quality from stakeholders |
| | 4) Hypothesis | 4a) Data quality is high for data being generated by tools built for this project: |
| | | 4a1) sidewalk and street crossing path data (OSW data |
| | | 4a2) transit center pathways and features (GTFS-Pathways) |
| | | 4a3) on-demand transit service data (GTFS-Flex) |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) Requires the vetting process reporting to be effectively captured – even when partner agencies are performing that vetting. |
| | 7) Other information | 7a) N/A |

| OBJECTIVE | TOPIC | INFORMATION |
|---|---|---|
| Data are widely available | 1) Performance Measures | 1a) The number of jurisdictions (e.g., cities, counties, etc.) producing OSW data within our six counties<br>1b) Percentage of centerline road miles within the six counties for which routable sidewalk paths have been developed<br>1c) The number and percentage of transit centers for which GTFS-Pathways data has been produced within our six counties<br>1d) The number and percentage of on-demand transit service providers that are producing GTFS-Flex data within our six counties<br>1e) The number of jurisdictions (e.g., cities, counties, etc.) producing OSW data in the U.S.<br>1f) The number of transit agencies producing GTFS-Pathways data in the U.S.<br>1g) The number of on-demand transit service providers that are producing GTFS-Flex data in the U.S. |
| | 2) Data Needs | 2a) OSW database<br>2b) OSM database (centerline miles)<br>2c) OSW centerline miles analyzed – from ML/AI software<br>2d) GTFS-Pathways database<br>2e) GTFS-Flex database<br>2f) Count of transit agencies with transit centers in six county study area<br>2g) Count of on-demand service providers in six county study area (from State DOT partners) |
| | 3) Experimental Design | 3a) Measurement of the amount of data delivered and available for use. (Including reporting of trends over time.) |
| | 4) Hypothesis | 4a) The number of agencies producing data in these standards within our regions is high<br>4b) The number of agencies producing data in these standards nationally is growing steadily |
| | 5) Targets | 5a) 80 percent of centerline road miles within the six counties are examined for sidewalks by the end of Phase 3<br>5b) 100 percent of transit centers in the six counties have GTFS-Pathways data by the end of Phase 3<br>5c) 80 percent of on-demand service providers within the six counties are producing GTFS-Flex data by the end of Phase 3 |
| | 6) Risks | 6a) Assumes that state departments of transportation are aware of all on-demand service providers in the six counties, or that those providers can be identified by other means. |
| | 7) Other information | 7a) National expansion is outside of the scope of this project but is still a valuable performance measure to report. |
| Data are uploaded to a central data center efficiently | 1) Performance Measures | 1a) Number/fraction of uploads without errors |
| | 2) Data Needs | 2a) Upload API |
| | 3) Experimental Design | 3a) Tracking the number of errors being reported, the fraction of reported uploads that contain errors, allowing for reporting trends over time. |
| | 4) Hypothesis | 4a) Data upload process works efficiently |
| | 5) Targets | 5a) 99 percent |
| | 6) Risks | 6a) None |
| | 7) Other information | 7a) N/A |

| OBJECTIVE | TOPIC | INFORMATION |
|---|---|---|
| Feedback is routinely applied to the initial data to continuously improve data quality | 1) Performance Measures | 1a) Number/fraction of vetting checks by agency/organization by outcome of vetting response<br>1b) Likert scale survey response from transportation infrastructure/service providers owners to vetting outcomes |
| | 2) Data Needs | 2a) Log of vetting API changes |
| | 3) Experimental Design | 3a) Tracking the number of vetting checks being performed, the number of vetting checks submitted to agencies that are then responded to, and the fraction of those being submitted that are responded to in order to report trends over time. |
| | 4) Hypothesis | 4a) Transportation infrastructure owners and service providers actively use feedback to improve their datasets |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) Once data have been vetted once, agencies may reduce the effort they place on maintaining those data, allowing feedback response to decline over time |
| | 7) Other information | 7a) N/A |

### 3.4.3. Confounding Factors and Constraints

These evaluation analyses are fairly straight forward. The primary confounding factor in this portion of the evaluation is if agencies/cities that are generating data are unwilling or unsuccessful at sharing the quality control reports that are generated as part of their data generation activities.

### 3.4.4. Mitigation Approaches

If data on the quality of the data generation tasks are not readily available, and if there is insufficient agency/community/advocacy group vetting of those data to make accurate assessments of data quality, the UW team can perform a limited number of independent data vetting tests as part of the evaluation effort. Contingency funds will be set aside for this activity in Task 3, however, the intent is to make the quality assurance and data vetting tasks a routine part of the data generation and acceptance testing activities associated with these data. Thus, limiting the need for the expenditure of these contingency funds.

### 3.4.5. Data Sources and Data Collection Plans

The data sources required for measuring the quantity and quality of the data being generated and made available through this project come from three sources.

- Data stored in the central data sharing repository.
- Output from the data quality assurance and data vetting tools.
- System logs from the data vetting tasks, and
- System logs from the data upload process.

All of these data sources exist either because they are intended outputs of the project or because they are needed for routine management and refinement of the system.

### 3.4.5.1. Data Needed

The data required for these analyses are summarized in Table 4. The data do not need to be specifically generated for the evaluation effort, as they are already being generated either because they are being published for use within the TDEI, or they are an output of software being developed for this project and will be used as part of the system management effort that is part of the project. Minor exceptions to this are 1) the need to measure centerline miles of roadway to be used as the basis for understanding the fraction of the road network within the six counties which have been examined for the presence of sidewalks, and 2) the development of the population size of cities which own/manage sidewalk infrastructure within the study area and on-demand transit service providers within the study area.

Centerline road mileage will be obtained by querying the OpenStreetMap database. The number of cities and transit agencies and transit service providers that make up the population with both counties within each of the three participating states will be obtained from the state departments of transportation that are participating in this project.

All other data sources required for the evaluation of the data generation goals for this project are already collected either by summarizing the data collected, vetted, stored in the central data repository, and published for use by 3rd-party application developers, or by summarized the results of the quality control and data vetting efforts performed as part of the project.

Data quality statistics will be reported for each of the three types of data (OSW, GTFS-Pathways, and GTFS-Flex.) and will include the amount of missing data within each type of record. For example, the project will report on the number of miles of sidewalk data, but it will also report how many of the miles for which routable sidewalks have been identified are missing data on specific data items such as surface type, side-slope, or other important variables.

A second set of data quality metrics will come from the data vetting process. In the data vetting process, tools are provided to cooperating agencies or groups. Those agencies/groups examine specific geographic areas and/or datasets and log specific errors that they find. The vetting software reports both what data (geographic area or dataset) the group has vetted, what errors they have found, and what the corrections should be. The TDEI staff then shares these reports with the “owners” of that infrastructure or transportation service in order to determine what the correct data record should be for each reported error. The appropriate changes are then made to the database. Records are kept of all of the changes made.

These change records are traceable to the agencies/organizations that perform the vetting and that respond to the vetting reports.

Analyzing these vetting reports provides the evaluation insight into the number of errors discovered and repaired by type of data. These records also indicate which agencies / organizations are actively participating in maintaining the database, as well as what fraction of the database records have been vetted, and when that vetting took place.

This information is useful not only for understanding the quality of the data being published for evaluation purposes, but for overall management of the data system. They indicate where vetting has and has not taken place, as well as where errors are being discovered as well as what types of data errors are occurring. These outcomes will then be used to direct project resources for both improving the data generation software and for directing data vetting resources.

#### 3.4.5.2. Data Collection and Quality Checks

The primary “data collection effort” for this portion of the evaluation will be to capture, organize and collate the results of the data vetting procedures done by data generators participating in the project. Additional data quality control is not expected to be a significant problem with this activity since these data are being used to manage the system being deployed.

Data quality controls for the data stored in the project data bases are discussed in other project documentation, such as the Systems Requirements. No additional data quality activities are planned for the data already stored in the project databases that will be summarized for this evaluation task.

#### 3.4.5.3. Cost of Data Generation Evaluation Effort

While cost of the data collection effort for the evaluation of the data generation goal has not been estimated at this time, these costs are not expected to be substantial, as the only data collection activities required specifically for the evaluation effort are collecting and collating the results of data vetting activities performed by stakeholder agencies and organizations. Data collection costs for the evaluation effort will be determined during Phase 2, once software being used for the data quality and vetting procedures has been developed, tested, and put into use.

### *3.4.6. Experimental Design*

The experimental design for these evaluation topics is based on straightforward reporting of the outcome summaries. Dates associated with different data collection activities (e.g., when sidewalk data are generated for different cities) will be reported along with the data quality control outcomes. This will allow a trend comparison over time for these procedures. But, in many respects, the trends over time are less important than the overall data quality outcomes, simply because the accuracy of any sidewalk data collection exercise may be significantly impacted by the quality of the imagery for that particular dataset, more than changes in the performance of the machine learning algorithm sed to generate sidewalk data. For example, one city might have good aerial imagery available from the past year, while another might have such imagery only from ten years ago, and those ten year old images have a great deal of tree cover, thus reducing the amount of sidewalk infrastructure that can be identified. Thus, while trends can be extracted from the data, trends the primary experimental design is simply reporting on the ability of the project team to work with its partner organizations on collecting and reporting data.

## 3.5. Data Vetting

### *3.5.1. Introduction*

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.

The data vetting process consists of two major activities, the first consists of automated reports at the time data are initial generated, if those data are generated with software developed as part of this project. That software will indicate the degree of confidence associated with the data being generated. (For example, if tree cover in an image obscures a sidewalk, the degree of confidence in the data associated with that sidewalk will be lower than for a sidewalk that is clearly visible in a different image.) In addition, the data vetting software will perform range and value checks for data being generated. Where low confidence in data occurs, or invalid data is identified, the data generating agency/organization is prompted to either use additional data sources to confirm the data or fix the invalid data. Where revisions are made, those changes are recorded and reported so that traceability to those changes exists.

The second vetting process is where community groups (e.g., active transportation advocacy groups, or community groups such as the boy and girl scouts) are trained in data mapping and they provide feedback on errors in the database using software provided by the project team. These data vetting reports are passed to the infrastructure owners who review those reports and accept or decline those data revisions.

### *3.5.2. Potential Performance Measures and Targets*

The results of the data vetting process are used as one of the inputs to the evaluation of the data generation process as described in the previous section. If the vetting process identifies a consistent type of error occurring in the data being generated, then revisions can be made in the process used to generate that data to avoid those errors. Alternatively, if the data vetting process falsely identifies a large number of data errors (i.e., the individual performing the vetting states that the data is in error, but a review shows that the data are correct as originally reported), the software mechanism responsible for identifying those vetting reports will need to be refined, or the training provided to organizations and agencies participating in the data vetting effort will need to be improved.

The evaluation of the data vetting effort is focused not on the data generation aspect of the vetting process (that is considered part of data generation), but on the participation and use of the vetting capabilities by participating agencies and organizations. Given the traditional lack of resources for the collection and review of the data that is at the heart of this project, one of the key outcomes is to increase the willingness and ability of agencies and organizations to collect, maintain, and publish these data. The vetting process is a key part of reducing the cost of performing these tasks, while also improving the quality of the data being published. By getting community and advocacy groups involved in data vetting, it is possible to both help those agencies improve the travel outcomes of their constituencies, while lowering the resource burden of the cities and transit agencies that supply and maintain the infrastructure and services. At the same time, the owners of that infrastructure and of those services need to have final say in what the truth is.

Therefore, the data vetting goal is concerned with the level of participation of these groups. Are they actively participating? For both agencies/cities and advocacy/community groups, this involves simply determining their level of participation. Performance measurement objectives track the level of participation of agencies and jurisdictions (e.g., infrastructure owners and service providers) separately from community organizations which provide “free” vetting services but that are often key to ensuring good levels of data quality and that are key advocates for the use of these data. For data owners, we are also concerned with whether they are using those

vetting reports to make changes in the data, but that specific evaluation outcome is address under the data generation subsection.

Consequently, the performance measures for data vetting, summarized in Table 5, are focused on the number of agencies/organizations that are actively participating in the project.

**Table 5: Summary of Data Vetting Performance Measures**

| OBJECTIVE | TOPIC | INFORMATION |
|---|---|---|
| Agencies participate in the data vetting process | 1) Performance Measures | 1a) Number/fraction of agencies actively vetting data |
| | 2) Data Needs | 2a) Vetting API log |
| | 3) Experimental Design | 3a) Measurement and reporting of the number of agencies actively participating in the vetting process, allowing for reporting of trends over time. |
| | 4) Hypothesis | 4a) Agencies that have infrastructure or services described in the TDEI actively participate in the data vetting process to ensure data published accurately represent that infrastructure or services. |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) Data with very high quality require few changes so that limited vetting interaction could be interpreted as lack of interest or resources to participate in the vetting activity. |
| | 7) Other information | 7a) Information gained during the co-Design process, as well as through project interviews and surveys with agency/city stakeholders will be used to describe the reasons behind the success or failure of specific approaches to vetting. |
| Community groups participate in the data vetting process | 1) Performance Measures | 1a) Number/fraction of organizations actively vetting data |
| | 2) Data Needs | 2a) Vetting API log |
| | 3) Experimental Design | 3a) Measurement and reporting of the number of community organizations actively participating in the vetting process, allowing for reporting of trends over time. |
| | 4) Hypothesis | 4a) Community organizations and advocacy groups can be successfully recruited to perform data vetting to improve data quality |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) Community and advocacy group support may not be evenly distributed, resulting in uneven data vetting support.<br>6b) It may be difficult to determine the total possible number of community and advocacy organizations that could participate in a given geographic region. |
| | 7) Other information | 7a) N/A |

## *Confounding Factors and Constraints*

The primary confounding factors with the analysis of the data vetting process are that the effectiveness of the use of outside organizations (e.g., community or advocacy groups) requires 1) there to be organizations present in that geographic location that are interested in performing the data vetting tasks, and 2) the project team needs to be able to both find and recruit those organizations. As an example, one individual in Mt. Vernon, Washington was primarily responsible for vetting the sidewalk data for that entire small city. Thus identifying that one individual resulted in a very large success, while not finding that individual would have resulted in far less data being available in that city.

This evaluation can determine when the project team is able to find and activate these types of organizations. It can also determine when such an organization is either not available or participating actively, but it is not clear that the results of this project are presentative of the success of this approach in other cities and counties across the U.S., as many community and advocacy groups are locally focused and may not be easily identified.

Similarly, the level of engagement of the city and county staff is expected to be a direct result of the importance placed on active transportation and equity of transportation within those cities and counties. While this evaluation can judge the degree to which this occurs in this project, it is not clear if the results of this project will be representative of the rest of the U.S.

This is true as well for transit agency participation, where the level of participation may well hinge on the availability of staff resources to perform these tasks, which will vary considerably from agency to agency.

### *3.5.3. Mitigation Approaches*

No direct mitigation is required for the evaluation of the data vetting approach for this project, other than working with the state Departments of Transportation to identify local groups that might be interested in participating in this project. That would allow the project team to indicate not just the amount of vetting activity taking place (from the data sources described in this subsection), but also the number and types of agencies which were not interested in – or able to – participate in the vetting process.

The confounding factors apply to the extrapolation of those findings to the rest of the country. To support that extrapolation, the UW team will work with U.S. DOT during Phase 3 to explore expansion of these concepts, and any findings from those efforts will be incorporated into the evaluation report for this project.

### *3.5.4. Data Sources and Data Collection Plans*

The data sources required for measuring the success of the project's approach to vetting of the collected data are available through the API logs associated with the vetting process. These logs will indicate the number of vetting interactions associated with each agency or organization. The vetting logs can also track how the size of these efforts change over time.

No specific data quality checks are planned for these data at this time.

#### 3.5.4.1. Cost of Data Vetting Evaluation Effort

The cost of the evaluation of the data vetting goal has not been estimated at this time. These costs are not expected to be substantial, as the data collection activities required specifically for this evaluation effort are minor, being mostly organizing and summarizing the data vetting API logs.

### *3.5.5. Experimental Design*

The experimental design for the data vetting evaluation is based on straightforward reporting of the outcome summaries. Data vetting will take place during the later stages of Phase 2 and

throughout Phase 3. API logs will be routinely downloaded and examined to determine the outcome of those vetting activities, as those activities help drive the management of the TDEI and are key indicators of the need to refine data generation software. The summarization of these reports can be performed as part of that activity and both the actual API logs and the summaries of those logs will be shared with the independent evaluation team as requested by that team.

## 3.6. Data Service Provision

### *3.6.1. Introduction*

The next major goal of the UW ITS4US project is to develop and deploy software and hardware systems that allow for the centralized storage and distribution of the data generated and submitted using the OpenSidewalks, GTFS-Pathways, and GTFS-Flex data standards. This portion of the project is essential for delivery of the data to 3rd party application developers, and therefore, is essential to achieving the overarching project goal of demonstrating systems which can deliver data widely across the nation.

Application developers need consistent data, available from a limited number of sources, in order to cost effectively build, deploy, and operate software with large enough markets to make their development efforts cost effective. Having a single source of data that provides access to data for multiple geographic areas allows application developers to build a single application that can be deployed widely, allowing one application to serve many cities, counties, and transit service providers.

An excellent example of this is the original deployment of the GTFS for fixed route transit systems. Portland TriMet, Google, and others to develop the GTFS standard. Google then announced that they would collect and publish these data and demonstrated that their navigation application could ingest that data and provide transit routing and navigation at no cost (other than access to the standardized data) to the transit agencies. Transit agencies quickly adopted the GTFS standard, and Google and other navigation service providers immediately gained the ability to allow potential transit users to discover transit routing options anywhere GTFS data were available.

This combination of events resulted in a major change in the general public's ability to discover and consequently use fixed route transit services. Travel behavior changed because Google worked with transit agencies to publish their schedules via GTFS feeds. Google, and other data aggregators, obtain and aggregate those data feeds, republishing them for use by application developers. Application developers access Google's central repository to obtain GTFS data for any transit agency they need that data from. This greatly reduces the cost of application development and maintenance because developers only need access to one transit data source. They do not need to discover, build, use, and maintain data service connections to transit service providers all over the country. Google (or another data aggregator) performs that task for them.

In addition to lowering their development cost, accessing a single data aggregator's repository means that the geographic area covered by their application is greatly increased over what would otherwise be possible if the developer must discover, connect to, and interact with multiple transit agency data feeds. The result is a far larger business opportunity for the application developer, at far lower cost, and this increases the number and success of applications available to travelers. It

is this basic business model that the TDEI is designed to emulate, only with OpenSidewalks, GTFS-Pathways, and GTFS-Flex data instead of GTFS for fixed route transit.

To get this process started, for this project, the TDEI will be building the initial central data service repository. This repository needs to function efficiently and effectively for 3rd party developers to both encourage the development of applications and to demonstrate the value of the data to other data aggregators, so that they too build and operate their own data repositories containing these key data items.

### *3.6.2. Potential Performance Measures and Targets*

The evaluation of the project's data service provisioning is focused on the performance of the software that performs the data service provisioning. The proposed measures are shown in Table 6. The UW team has selected four objectives to determine whether the deployed system meets the overall goal of having an effective and efficient data publication and provisioning system. These objectives are listed below.

- The API performance for the system is good.
- The API availability (uptime) is good.
- 3rd party developers are actively using the system.
- Data security for the system is strong.

For the first two objectives, the initially selected measures are the response times and availability (percentage uptime) of the APIs which are used by 3rd party developers to access the data. To supplement these data, it is also important to survey the developers using the service to determine their opinion on the performance of the data service, to ensure that its performance meets their expectations, and if it does not, why not.

To examine the other two objectives, two additional performance measurement areas are recommended. One is simply the number of 3rd party developers participating by the end of Phase 3 of the project. The number of independent developers accessing the data is an excellent measure of the potential impact of the data system being developed, with a large number of developers meaning that a large number of useful applications can be expected to be delivered across the nation as the base data needed by those applications are collected in regions outside of this project's boundaries.

The second set of measures to be reported are the outcomes of the routine data security assessments that will be performed to maintain the integrity of the system. This portion of the evaluation will show whether the data system is effectively securing the data that has been generated and submitted for publication. Among these performance measures are the need to track runtime errors and memory use, as these types of errors can be exploited as part of security breaches.

**Table 6: Summary of Data Publication/Provisions Performance Measures**

| OBJECTIVE | TOPIC | INFORMATION |
|---|---|---|
| API performance is good | 1) Performance Measures | 1a) API response time<br>1b) Network latency<br>1c) Stakeholder rating of API performance (5-point Likert scale) |
| | 2) Data Needs | 2a) API performance log<br>2b) Survey of application developer satisfaction with data provision. |
| | 3) Experimental Design | 3a) Measurement and reporting of API response time, network latency, allowing for reporting of trends over time., with stakeholder rating occurring after the system is in stable operations. |
| | 4) Hypothesis | 4a) API performance is fast, resulting in 3rd party application providers having a high level of satisfaction with the data service |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) API performance is likely to vary based on the complexity of the database query, making the API response time statistics biased towards common queries, but perhaps, away from other key queries.<br>6b) API performance may also change by time of day and day of week, so the analysis needs to be sensitive to temporal changes in API performance. |
| | 7) Other information | 7a) The evaluation's API performance reporting needs to be sensitive to temporal differences in performance as well as being sensitive to the types of queries being made (e.g., OSW versus GTFS-Pathways.) |
| API availability is good | 1) Performance Measures | 1a) API up-time/down-time percentage<br>1b) Stakeholder rating of API performance (5-point Likert scale) |
| | 2) Data Needs | 2a) API and server performance logs<br>2b) Survey of application developer satisfaction with data provision. |
| | 3) Experimental Design | 3a) Measurement and reporting of API availability for reporting of trends over time. |
| | 4) Hypothesis | 4a) API availability is high, resulting in 3rd party application providers having a high level of satisfaction with the data service |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) API performance is expected to change over time – both over the long term as the number of uses of the database grows, and between peak times of day and off-peak times of day. These differences need to be incorporated into the analysis. |
| | 7) Other information | 7a) N/A |
| 3rd party developers are participating | 1) Performance Measures | 1a) Number of API keys requested<br>1b) Number of active 3rd party application |
| | 2) Data Needs | 2a) System administrative data<br>2b) API logs |
| | 3) Experimental Design | 3a) Measurement and reporting of the number of 3rd party developers participating, and their level of activity, allowing for reporting of trends over time. |
| | 4) Hypothesis | 4a) Uptake of the database will grow quickly once data are available in Phase 3 of the project. |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) N/A |
| | 7) Other information | 7a) N/A |

| OBJECTIVE | TOPIC | INFORMATION |
|---|---|---|
| Data security is strong | 1) Performance Measures | 1a) Number and cause of runtime errors,<br>1b) Memory use<br>1c) Vulnerability assessments<br>1d) Number of entry point security flaws discovered, and number of flaws addressed as a result of routine assessment of electronic security entry points |
| | 2) Data Needs | 2a) Results of routine security audits and assessments |
| | 3) Experimental Design | 3a) Performance of, and reporting of, the outcomes from security audits performed routinely. |
| | 4) Hypothesis | 4a) The data system does not suffer from security breaches. |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) The more successful the system, the more likely it will be the target for cyber-attacks. |
| | 7) Other information | 7a) N/A |

### 3.6.3. Confounding Factors and Constraints

The major confounding factor to these analyses is that performance changes over time as well as for different types of data requests. For example, requests for data within a small geographic area might be retrieved and transmitted quickly, while requests for data covering a very large geographic area may take longer than expected or require longer than is acceptable to the developer. In addition, response time may change over time as the data service grows and as the hardware used to support the data service changes in response to the growth of the database. These changes can be both positive (e.g., new hardware or software is deployed in response to slowing response rates and as a result, response times improve), or negatively (e.g., growth in the amount of data being stored and thus the size of the database, or an increase in the number and timing of data queries being made, slow the response time.)

Similarly, attitudes may differ markedly from one application developer to another, based on how each of those developers is trying to access and use the data being collected. Therefore, the evaluation may need to be able to parse differing opinions and relate those opinions to different features and aspects of system performance. It may also be difficult to separate data availability from data service performance if developers are interested in data in locations for which those data are simply not available.

### 3.6.4. Mitigation Approaches

The primary mitigation for these issues is to ensure that the performance data being collected allow the tracking of activities requesting the data versus system performance those queries receive. Similarly, survey responses need to be tracked against the nature of system use being requested by those survey respondents. That is, the data collected need to be able to facilitate the level of detail required to describe the cause of differences in performance and stakeholder attitudes. This will need to be accounted for in the design of the API logging system as well as the stakeholder surveys.

### *3.6.5. Data Sources and Data Collection Plans*

The data sources required for measuring the success of the project's data provisioning services will be collected through the system logs output as part of the API services. In addition to responding to data requests made through the API, at a minimum, the logs will record the source of the request, the request itself, the time when the request was received, the time when the requested response was completed, and the size of the data package transmitted. Data will be collected once the system testing starts.

In addition to the API logs, a stakeholder survey will be performed of all 3rd party developers that have requested credentials for the APIs. The survey will be scheduled to occur roughly 18 months after the start of Phase 3 of the project in order to provide the stakeholders with sufficient time to develop firm opinion on the system's performance.

Finally, the UW team will schedule routine security assessments of the system. These will be performed quarterly starting just prior to the completion of final system testing.

#### 3.6.5.1. Approach to Data Quality Checks

The quality of the system performance checks is paramount to the management of the system, not just the evaluation of the system. Consequently, the robust testing of the API logging system will be conducted using the Digital Twin and Multi-Modal AccessMap applications as test cases. These applications, controlled by the UW project team, will periodically collect the same data recorded by the API logs, but from the application software, not the API service. For example, we will record the queries made for data, the time required to receive the data, etc. These results will then be compared to the data in the API logs to ensure that all data requests are being logged and that the data being logged are correct.

#### 3.6.5.2. Cost of Database Operations and Data Provisioning

The cost of the evaluation of the database operation and data provisioning has not been estimated at this time. The cost of data collection for this task is not expected to be substantial, as the data collection activities required specifically for this evaluation effort are minor, being mostly conducting and summarizing the stakeholder surveys, or extracting the logs from the API service. However, the cost of converting the API system logs into a useful analytical dataset that can be effectively summarized may not be trivial. This evaluation requirement will need to be considered in the design of the API logging system, which will occur in Phase 2 of the project.

## 3.7. Demonstration Applications

### *3.7.1. Introduction*

The final goal of the UW ITS4US project is to demonstrate the use of the data being generated and published. Three demonstration applications are being included within the scope of this project, Multi-Modal AccessMap, Microsoft's Soundscape, and a Digital Twin application for

previewing paths through transit centers. Each of these applications is intended to demonstrate a very different use case for the data being generated as part of this project. Combined, the three applications are intended to provide functional examples of how data can be retrieved from the data service and used to meet very different needs for different target populations. For example, Microsoft's Soundscape is an already fully functioning application that can be downloaded and used. However, it lacks the ability to provide users with details about sidewalk paths, because those data are not available to the application. As part of this project, Microsoft will add the available OpenSidewalks data to the data being ingested by Soundscape, thus making it available to users of the application.

Similarly, the AccessMap application currently exists and is in use in several cities. However, that application only describes walking paths. It does not have access to data that describe on-demand transit options or paths through transit centers. Thus, for this project, this application will be extended to incorporate the GTFS-Pathway and GTFS-Flex data streams, and the routing engine will be extended so that Multi-Modal Access Map will enable users to identify travel paths which include both on-demand and fixed route transit services, and the paths that lead between those services and their trip origins and destinations.

Finally, the UW team will build a digital twin application which allows users to visualize transit centers, allowing them to preview trips through those centers from their point of entry to a transit station (e.g., a sidewalk or station platform) and their exit point from that station (sidewalk or station platform.) The digital twin application is not intended as a routing guide, but instead is an exploratory tool, allowing a traveler to become comfortable with the layout of a station, and paths they can take through that station, prior to making a trip that involves the use of that station.

### *3.7.2. Potential Performance Measures and Targets*

The evaluation of the project's demonstration applications has three major focus areas. These demonstration applications are good examples of a variety of ways the data can be used, and how new services can be delivered. They should also be representative of the types of data queries many other applications will make to the data repositories.

The first evaluation focus area is the performance of the software that performs the data service provisioning. That is, are the three demonstration applications able to successfully request and obtain the data they are attempting to use from the data service. This evaluation task will be addressed as part of the evaluation of the data service's performance, which was described in Section 3.6. The proposed measures for this were previously shown in Table 6. Because the UW team can report on the performance of the system as observed specific by these three applications by examining the API logs for the API keys associated with these applications. The second focus area for evaluation is the overall performance of the applications themselves, and the third focus area is the travel outcomes that result from use of the demonstration applications. A summary of the evaluation measures for these two focus areas is shown in Table 7. Note that these portions of the evaluation are useful in that the applications are useful examples of how the data can be used, and it is important that the demonstration applications work well. However, the major benefit from the data will come from the use of the collected and published data by a wide variety of third-party applications, not from the three demonstration applications. Thus, the evaluation of these applications is a small part of the overall project evaluation, and the travel benefits from the demonstration applications is not the focus of the project.

**Table 7: Summary of Demonstration Application Performance Measures**

| OBJECTIVE | TOPIC | INFORMATION |
|---|---|---|
| Demonstration application performance is good | 1) Performance Measures | 1a) Number/fraction of routing errors occurring in lab tests of Multi-Modal AccessMap<br>1b) Number/fraction of errors occurring when using the Digital Twin application during lab tests and during field use (e.g., how often does the system crash when in use after deployment?) |
| | 2) Data Needs | 2a) TDEI data services, pre-selected O/D pairs and known routing alternatives. |
| | 3) Experimental Design | 3a) Lab tests |
| | 4) Hypothesis | 4a) The Multi-Modal AccessMap application is able to accurately and routinely identify efficient routing alternatives, given the user's personal profile, and provide the selected routing instructions. |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) The application is subject to the limitations in the available data, so routing errors can occur as a result of poor data quality, even while the application is performing correctly. |
| | 7) Other information | 7a) Lab tests are required in order to test for, and control for, both multiple types of multi-modal trips, and to ensure that the "correct" routes are known in advance, in order to determine the accuracy of the application's routing choices. |
| Use of demonstration applications is high | 1) Performance Measures | 1a) Number of application users and uses<br>1b) Usage patterns of users |
| | 2) Data Needs | 2a) System user profile (hashed IDs to determine the number of unique users, and their frequency of use)<br>2b) Log of the types of routing request results – e.g., what modes are selected) |
| | 3) Experimental Design | 3a) Measurement and reporting of the number of API calls from different applications, allowing for trend reporting over time. |
| | 4) Hypothesis | 4a) Usage grows over time during Phase 3, both in terms of the number of uses and the number of active users. |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) The is a trade-off of collecting data that could become PII (we would not collect actual trip data, but only the requests for routes.) |
| | 7) Other information | 7a) This analysis can only be performed for Muti-Modal AccessMap and Digital Twin, as these are the only applications where the required usage data can be collected. |
| User satisfaction associated with the demonstration applications is high | 1) Performance Measures | 1a) User rating on Likert scale for each of the three demonstration applications |
| | 2) Data Needs | 2a) Survey of user experiences with the three applications |
| | 3) Experimental Design | 3a) Reporting of participant perceptions using surveys with Likert scale style questions. |
| | 4) Hypothesis | 4a) Users have a high level of satisfaction with each of the three demonstration applications. |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) It will be necessary to recruit survey participants, and it may not be possible to recruit a sufficient sample to obtain a statistically significant result. The sample may also be biased in attitude. |
| | 7) Other information | 7a) N/A |

| OBJECTIVE | TOPIC | INFORMATION |
|---|---|---|
| Travel outcomes are safe when using the demonstration applications | 1) Performance Measures | 1a) Number of safety hazards occurring during field tests of the Multi-Modal AccessMap application<br>1b) User rating of their travel safety using the Multi-Modal AccessMap application based on user survey |
| | 2) Data Needs | 2a) Data from field test, modelled after the AbleLink field test, including<br>2a1) number of trips attempted<br>2a2) number of trips completed,<br>2a3) number of off-route notifications,<br>2a4) number of help requests,<br>2a5) summary outcomes of those help requests,<br>2a6) reporting of errors in the database (by type of data error) observed during the field tests<br>2a7) number and type of hazard occurring when data errors occur.<br>2a8) Survey of user experiences with the application |
| | 3) Experimental Design | 3a) Controlled field test, with secondary data collection in support of data collection for the application itself, using recruited test subjects |
| | 4) Hypothesis | 4a) Users of the Multi-Modal application will experience few, if any, hazardous travel outcomes, and if they do experience a hazardous outcome, the built in help system will limit the significance of that hazard. |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) The statistical validity of these tests can be difficult to achieve given the number of test subjects that can be recruited and the wide variety of different mobility disabilities that could be tested. |
| | 7) Other information | 7a) The UW team has existing data collection software that can be used during the field experiment to track the activities occurring, and outcomes of, trips taken as part of a field experiment using recruited test subjects. |
| Travel outcomes are efficient when using the demonstration applications | 1) Performance Measures | 1a) Overall trip completion rate for Multi-Modal AccessMap<br>1b) Number of navigation errors (user error vs. application error)<br>1c) Number of other safety hazards occurring during field tests of Multi-Modal AccessMap<br>1d) Number & success of help requests<br>1e) Localization is accurate and is used effectively<br>1f) User satisfaction ratings on Likert scale. |
| | 2) Data Needs | 2a) Survey of user experiences with the MultiModal AccessMap application |
| | 3) Experimental Design | 3a) Measurement and reporting of the number of API calls from different applications, allowing for trend reporting over time. |
| | 4) Hypothesis | 4a) Users' perception of the efficiency of the routing directions provided by the application are high, as a result of their comparison of the mobility and trip outcomes after using the application compared to those same outcomes prior to having access to the application. |
| | 5) Targets | 5a) TBD |
| | 6) Risks | 6a) It will be necessary to recruit survey participants, and it may not be possible to recruit a sufficient sample to obtain a statistically significant result. The sample may also be biased in attitude. |
| | 7) Other information | 7a) The results of this analysis are designed to complement the results of the analysis performed under the "Demonstration application performance is good" objective above. |

The second focus area for the performance of the demonstration applications is divided into three objectives, these are

- The demonstration application's functionality is good.
- The use of the demonstration applications is high, and
- User satisfaction with the demonstration applications is high.

The first of these objectives will be evaluated by examining the performance of the two applications being constructed by the UW as part of controlled laboratory testing of those applications, prior to their release. For the Multi-Modal AccessMap tests, routing options will be requested for selected origin/destination patterns for which alternative routes have been identified. The applications selected routes will then be compared against those known routes to determine if the algorithm is correctly selecting optimum routes. The success rate of the algorithm will be tracked, and the reasons for any failures identified (e.g., the algorithm works correctly, but errors found in the data being stored in the system result in incorrect routing outcomes versus the data are correct, but the algorithm does not identify the best path options.) This same testing regime will also be applied to the Digital Twin application. Many of these tests will be performed in controlled laboratory situations, where ground truth is known. This allows detailed testing of the performance of the routing algorithms by controlling for data quality, and where detailed independent analysis is possible of all routing outcomes for the trips being used for testing.

For Digital Twin, the application will operate on a central server and be access via a browser window. As a result, it will also be possible to track server errors (e.g., crashes) which occur while the system is in use, not just in laboratory-controlled experiments. The Digital Twin server application will track the number of uses of the system, and the number of errors which occur during that use. These statistics will be used for reporting continued performance.

To test Multi-Modal AccessMap, the UW team will recruit individuals that are willing to have their use of the system tracked over time. These individuals will have additional software installed on their devices (likely smartphones), and those devices will track the use of the Multi-Modal AccessMap application over time. This field evaluation effort is primarily intended to provide data on travel outcomes (see below) but will also provide data about the application's performance (e.g., the number of application errors) for this subset of users.

The third objective for the second focus area is the satisfaction levels users have of the three test applications. To obtain this information, users of the application will be invited to take a survey about their experiences with the application, with Likert scale questions being asked, along with free form text responses to explain any specific concerns those users have about the applications. In terms of the overall importance of the evaluation outcomes, this specific outcome is less important than many of the evaluation topics presented above, because the majority of physical travel benefits ar expected to come from 3rd party applications made possible by the data newly collected and made available. However, the demonstration applications are designed to be beneficial to travelers with mobility disabilities, and it is useful to report on whether participants liked or disliked the demonstration applications. Details about why stakeholders liked or disliked the demonstration applications provides useful feedback to 3rd party application developers making "better" or similar applications.

The third focus area for the evaluation of the demonstration applications concerns the travel outcomes experienced by users of the applications. The primary objective for this focus area is that travel being taken using the application have outcomes that are safe and efficient.

Microsoft's Soundscape is designed to provide improved exploration of the user's surroundings. Thus, this application is not a good choice for measuring changes in travel efficiency, since Soundscape is not intended to necessarily improve trip efficiency. It is better suited for giving the user better travel experiences while exploring geographic areas. These outcomes are bets examined through the survey mechanism mentioned above.

The Digital Twin application provides a trip preview function, not real-time virtual reality assistance. It is not intended to directly change user travel behavior, but instead is designed to simply provide users with a lower stress trip due to their having gained insight into the travel environments they can expect to walk or roll through in the future. Consequently, the evaluation will not attempt to measure changes in travel behavior or safety due to the use of this applications, other than as part of the survey questions about its use. While these answers will be useful for the evaluation, they are not expected to result in statistically valid measures of travel behavior change.

Consequently, the project activities for measuring actual travel efficiency and safety outcomes will focus on the use of the Multi-Modal AccessMap application. This application is designed to improve a traveler's ability to discover, select, and then follow multi-modal routes. The application will both provide pre-trip navigation directions, and mid-trip, step-by-step navigation instructions. To measure the safety and efficiency outcomes, the evaluation will focus on four specific travel activity outcomes for users of this application.

- The overall trip completion rate.
- The number and frequency of user errors (e.g., navigation errors) made during trips when the application is being used.
- The accuracy of user localization (i.e., does the application correctly identify locations where navigation directions need to be provided, and correctly provide those directions).
- The appropriateness and effectiveness of user alert notifications made by the application.

These statistics will be obtained from a subset of application users that have been specifically recruited for the evaluation effort. This recruitment is required in order to obtain permission from those users to capture, store, and analyze the data needed specifically for the purpose of the evaluation, since the Multi-Modal AccessMap application does not store or share the data required for the evaluation.

The UW team will recruit individuals from individuals that have downloaded the smartphone application or that have otherwise interacted with the project team for this project. These individuals must then allow the UW team to download additional software onto their phones. This software is designed to capture, store, and share data that describes the location of the phone, the requests being made to the application, and the instructions the application provides to the user. This allows the evaluation software to obtain and share with the UW team the detailed data needed to compute the performance measures listed above, for the recruited test population.

### 3.7.3. Confounding Factors and Constraints

The primary confounding problem with the evaluation of the demonstration applications is that these applications are not designed to collect data from their users. In fact, the opposite is true, they are specifically designed to protect the privacy of their users by storing as little information about those users as possible. Consequently, unlike the majority of the data needed for evaluating the TDEI which were described earlier in this chapter, the evaluation of the demonstration applications requires additional data collection activity intended specifically to support the evaluation effort.

The UW team is not expecting these demonstration applications to cause significant changes in travel behavior themselves, although users of the Multi-Modal AccessMap application may indeed change user travel behavior due to the increased discoverability of transportation services that were previously not discoverable. The demonstration applications are primarily intended to illustrate the wide variety of uses the new data sources make possible. The project team is not actively promoting behavioral changes as part of this project. As a result, the degree to which changes in travel behavior can be observed in the evaluation may be limited. Still, the evaluation can determine whether the Multi-Modal AccessMap application works as intended, whether users actively use that new functionality, and the users' level of satisfaction with both the application and the overall concept of a multi-modal travel planner which includes on-demand transportation services, fixed route transit services, and the "first/last 50 feet" pathway/sidewalk routing instructions that given them confidence that they have reliable access paths to and from the locations where they get on and off those transit services.

### 3.7.4. Mitigation Approaches

In order to determine the number of different users and track their use of the different applications, data collection functionality will need to be built into the Digital Twin and Multi-Modal AccessMap applications. Those data collection functions can collect identifiers such as Mac addresses or IP addresses, hash those IDs in a consistent manner, and store the minimum amount of information about those users and queries. However, storing these data will require a modification to the initially planned IRB.

To limit potential loss of privacy, at the end of the evaluation period, the hash key will be discarded, and the data collection stopped. Using hashed IDs and minimizing data collection limits the risk to privacy while allowing for analysis of usage patterns. The UW team will work with the IE team to determine the length of time these data will be collected.

### 3.7.5. Data Sources and Data Collection Plans

Data for the evaluation of the three demonstration applications will come from several sources which are noted below.

- Surveys sent to users of the applications (the general user population)
- Surveys of test subjects specifically recruited to measure the performance Multi-Modal AccessMap application in the field
- A database specifically created for the Multi-Modal AccessMap field evaluation which contains detailed data on the number of trips taken by test subjects using the application, their

trip completion statistics, the travel path actually taken versus the navigation path selected including deviations from the intended path, and any required requests for assistance made through the application

- A database of (controlled) laboratory tests of application performance
- Data on the Digital Twin application use and performance collected by the application server.

In addition, as noted in Section 3.6, data on total data service requests by application, and the performance of the data service will be tracked via API logs.

#### 3.7.5.1. Approach to Data Quality Checks

The quality of the system performance checks is paramount to the management of the system, not just the evaluation of the system. Consequently, the robust testing of the API logging system will be conducted using the Digital Twin and Multi-Modal AccessMap applications as test cases.

The UW team has quality control steps built into their existing data collection software, these identify significant discrepancies or potential errors in reported locations. While specific care will be taken with both the collection and storage of the survey and laboratory test data, no specific tests for quality control have been developed at this time.

#### 3.7.5.2. Cost of Data Collection for Demonstration Application Evaluation

The cost of the evaluation of the data vetting goal has not been estimated at this time. The cost of the evaluation of the demonstration applications can be divided into three tasks, two of which are modest efforts, and one which is more substantial. The two modest tasks are 1) the collection, analysis, and sharing of data collected from the API logs, and 2) the writing, conduct, and analysis of stakeholder surveys given to users of the applications.

The third task is more substantial, as it requires the development and conduct of both laboratory and field experiments. These both require the recruitment of individuals with lived experience, and this will require compensation of these individuals for the time they spend assisting with the evaluation. The sample size of these efforts (i.e., both field and laboratory tests) has yet to be determined. The sample size calculation involves both the number of test subjects needed and the number of trips each test subject is expected to perform. (Note that the lab and field tests are separate and require different sample size calculations, although individual test subjects may participate in both the field and laboratory tests.) Similarly, the UW team has not designed the actual laboratory and field tests, which will define the time required from each test subject, as well as the staff time required for interacting with each test subject. These time requirements will play a significant role in the cost associated with each test subject, the staffing resources required to oversee/perform the tests, as well as the time required to recruit the required number of test subjects.

The UW already has built and previously used software for collecting travel behavior ground truth data. The tool exists for both iOS and Android smartphones, and while minor modifications may be needed to this software, these modifications are not expected to be difficult or costly. This software tool produces trace and event data which can be used as ground truth for determining travel outcomes associated with the Multi-Modal AccessMap application.

## 3.8. Summary of Performance Evaluation Activities

Given the large amount of performance data being collected and described above, this section of Chapter 3 is present to summarize the basics of the performance evaluation effort. Table 8 describes the overarching project goal and objective being evaluated, along with a summary of the performance measure being used and a simplified description of the source for the data for that measure.

In most cases, to limit its size, Table 8 ignores the fact that there are multiple sources for similar types of performance measures when those measures cover different modal systems. For example, statistics that describe the outcome of data vetting activities will come from multiple sources, transit agency staff, staff working for on-demand transit service providers, agencies which own sidewalk infrastructure, and community or advocacy organizations. Each of these agencies/groups will use software provided by the UW ITS4US team to perform that vetting activity. The vetting activities will be tracked as will the actions taken as a result of those vetting reports. The project evaluation will report on the details of these outcomes. Similarly, the system operators will use these vetting reports to manage and improve the data system being developed, deployed, and operated as part of this project. Table 8 simply summarizes these activities.

**Table 8: Summary of Performance Monitoring Activities**

| Goals | Objective | Performance Measures | Data Sources |
|---|---|---|---|
| **Data Standards** | 1) Meet user needs<br>2) Use objective data<br>3) Widely accepted | 1.1) Likert scale survey question<br>2.1) Likert scale survey question<br>3.1) Number of participating agencies and cities | 1.2) Stakeholder review / survey<br>2.2) Stakeholder review / survey<br>3.2) Data repository |
| **Data Generation** | 1) Quality is high enough to meet user needs<br>2) Tools work effectively to collect and publish data<br>3) Are widely available (e.g., more than half of all agencies/jurisdictions have data in the system by Phase 3)<br>4) Uploads to central point efficiently<br>5) Feedback routinely applied improving quality | 1.1a) Number/fraction of errors in QA/QC/vetting of sidewalk / transit center / service data<br>1.1b) Data quality score from AI/ML sidewalk system<br>1.1c) Likert scale survey question to users<br>2.1a) (See “quality is high” measures above)<br>3.1a) Number/fraction of cities / agencies participating<br>3.1b) Percentage of centerline miles that have 95% of critical attributes<br>3.1c) Percent of services / centers in the repository<br>4.1) Number/fraction of uploads w/wo errors<br>5.1) Number/fraction of vetting checks by agency/organization | 1.2a) Log of vetting updates (Tasking Manager for OSW))<br>1.2b) Output from the AI/ML software<br>1.2c) Log of automated QA/QC checks of GTFS-Flex and GTFS-Pathways data<br>1.2d) Stakeholder survey<br>2.2a) (See “quality is high” measures above)<br>2.2b) Stakeholder survey<br>3.2a) Census data<br>3.2b) OpenStreetMap<br>3.2c) State DOT lists of transit providers<br>3.2d) Transit agencies<br>3.2e) Data repository<br>4.2) Upload API log<br>5.2a) Vetting API log<br>5.2b) Survey of stakeholders |
| **Vetting** | 1) Agencies participate in vetting<br>2) Vetting by community groups occurs | 1.1) Number/fraction of agencies actively vetting data<br>2.1) Number/fraction of organizations actively vetting data | 1.2) Vetting API log<br>2.2) Vetting API log |

| Goals | Objective | Performance Measures | Data Sources |
|---|---|---|---|
| **Data Publication/Provisioning Performance** | 1) API performance meets system requirements goals<br>2) API availability meets systems requirements targets<br>3) 3rd party developers are participating<br>4) Data security passes all security audit assessments | 1.1a) Response time<br>1.1b) Stakeholder rating on Likert scale<br>2.1a) Uptime/downtime percentage<br>2.1b) Stakeholder rating on Likert scale<br>3.1a) Number of API keys requested<br>3.1b) Number of active 3rd party applications<br>4.1) Security assessment results | 1.2a) API log<br>1.2b) Stakeholder survey<br>2.2) API & server logs<br>3.2a) System administrative data<br>3.2b) API logs<br>4.2) Routine security assessment |
| **Demonstration application performance** | 1) Demonstration application functionality meets user expectations<br>2) Use of demonstration are routinely used by a growing number of users<br>3) User satisfaction with the demonstration applications results in their continued use of the applications<br>4) Travel outcomes are safe and efficient | 1.1) Number/fraction of routing errors<br>2.1a) Number of application users and uses<br>2.1b) Usage patterns of users<br>3.1) User rating on Likert scale<br>4.1a) Number of safety hazards occurring during field tests of the Multi-Modal AccessMap application<br>4.1b) User rating of their travel safety using the Multi-Modal AccessMap application based on user survey Overall trip completion rate for Multi-Modal AccessMap<br>4.1c) Number of navigation errors (user error vs. application error)<br>4.1d) Localization is accurate and is used effectively<br>4.1e) Number & success of help requests<br>4.1f) User satisfaction ratings on Likert scale | 1.2) Lab tests of Multi-Modal AccessMap<br>2.2a) System profile<br>2.2b) Log of routing requests<br>3.2) User survey<br>4.2a) Lab & Field tests<br>4.2b) User survey |

# 4. Performance Reporting

As part of this project, the UW team will develop a series of reporting systems that produce summary statistics that will be used for ongoing project management and evaluation reporting purposes. These reports are intended primarily for the UW team to support management and operation of the system being developed. However, these reports are to inform USDOT and the independent evaluation team of the ongoing performance and use of the system being developed.

In general, summary reports will be produced for quarterly delivery to USDOT starting in Phase 2 of the project, but only after data are being generated and stored in the project's central data repository. Because data collection is dependent in most cases upon deployment of software being developed in Phase 2, many of the reporting measures will be blank for much of Phase 2 until that deployment occurs. By Phase 3, data should be widely available, and thus the size, completeness, and utility of these reports will be more substantial for evaluation of the operational phase of the project.

These reports include the following types of information:

- Data generation
  - Total miles of sidewalk data in the OSW database
    - Fraction of sidewalk links missing 1 key attribute (e.g., sidewalk width, or sidewalk surface type is not available)
    - Fraction of sidewalk links missing more than 1 key attribute
  - Total centerline miles of roadway network analyzed for the presence of sidewalks (note that in some cases, there will be two miles of sidewalk for a given centerline mile of roadway, while in other cases, there will be no sidewalk data for a given mile of roadway, as no sidewalks are present.)
  - Number of on-demand transit services available via GTFS-Flex
    - Number of GTFS-Flex descriptions that are "out-of-date"
    - Fraction of GTFS-Flex service descriptions that are missing key attributes (e.g., number of wheelchair tie-downs on the vehicle performing that service)
  - Number of transit centers for which GTFS-Pathways data are available
    - Number of centers for which GTFS-Pathways data are present, but where one or more "key" attributes is not available (e.g., fare vending machines or emergency equipment locations are not labeled at a center)
    - Fraction of GTFS-Pathways centers for which data are present but that are missing key attributes (e.g., number of wheelchair tie-downs on the vehicle performing that service)
  - Number of cities participating in the data generation or vetting process
  - Number of transit service providers participating in the generation and vetting of GTFS-Flex data

  - Number of transit service providers participating in the generation and vetting of GTFS--Pathways data
  - Number of community groups actively participating in the data vetting process
  - Number of vetting submissions provided, and number and percentage of changes required as a result of those vetting requests
  - Data quality summary statistics for sidewalk datasets generated and uploaded since the last report and total for the dataset.
- Data services
  - Number of active API keys / users
  - Number of API calls by user
  - Number of data uploads by type of data upload
  - Number of API calls for data by type of data
  - Number and fraction of API requests where errors occur when responding to the request (by type of error)
  - Response time statistics associated with those API calls (Mean response time, median response time, 90th percentile response time, mean data package size, median data package size, 90th percentile data package size)
  - Uptime/downtime fractions for the data service
  - Uptime/downtime fractions during peak travel periods
- Data security
  - Number of detected cyber attacks
  - Number of security breaches
- 3rd party developers
  - Number of active 3rd party developers (have and are using an API key)
  - Number of active 3rd party applications (have and are using an API key)
  - Number of new 3rd party application developers that have contacted the UW team during the reporting period.

The statistics included in these reports include data on the current performance of data generation tools (e.g., the data quality scores from sidewalk generation software), agency interactions (e.g., the number of cities that have supplied sidewalk and street crossing data, and the number of on-demand transit agencies which have delivered GTFS-Flex service descriptions, and demonstration applications (e.g., the number of active users of the Multi-Modal AccessMap application.) These summary reports will be provided to Volpe on a quarterly basis to allow Volpe staff to track trends in the key project activities, such as the number of agencies/cities participating in the project, the amount of data available as a result of that participation, and the amount of use demonstration applications are experiencing. The first of these quarterly reports will be delivered at the start of the project's Phase 3 effort. This will serve as the baseline for the project.

A limited number of these topics will be summarized and reported on the TDEI web site, available at https://transitequity.cs.washington.edu/.

During Phase 3 of the project, the UW team will work with the Independent Evaluation team to write and perform the surveys of project stakeholders described in previous chapters of this report. Stakeholder surveys will not be performed until after the various stakeholders have had an opportunity to work with the data, APIs, systems, and/or applications, as the surveys are intended to obtain their opinions, perceptions and experiences from working with or using these data and systems.

# 5. Support to the Independent Evaluation Effort

This section provides a summary of the support being given to the Independent Evaluation (IE) team for this project.

## 5.1. Data Sharing Framework

The UW team will support the USDOT's independent evaluation team, being led by the Volpe National Transportation systems Center, throughout all three phases of the project. This support includes providing Volpe staff with a variety of data sets, programmatic information, and access to UW team staff and some stakeholders for interviews. The UW team is aware of four categories of items to be provided to Volpe staff, including

- all draft and final reports,
- summary performance statistics,
- raw data,
- additional information as requested by Volpe staff.

It is expected that the project's COR (Kate Hartman) will supply the UW team with email addresses for Volpe staff. The Volpe staff will then be included on the email submittal of all formal draft documents to USDOT. Comments returned to the UW team by Volpe staff will be responded to at the same time that the team responds to USDOT comments on those reports. Volpe will also be included on the transmittal email for all Final reports.

As described in Chapter 4 above, the UW team will develop a series of reporting systems that produce summary reporting statistics that will be used for ongoing project management and reporting purposes. The statistics included in these reports include data on the current performance of data generation tools (e.g., the data quality scores from sidewalk generation software), agency interactions (e.g., the number of cities that have supplied sidewalk and street crossing data, and the number of on-demand transit agencies which have delivered GTFS-Flex service descriptions, and demonstration applications (e.g., the number of active users of the Multi-Modal AccessMap application.) These summary reports will be provided to Volpe on a quarterly basis to allow Volpe staff to track trends in the key project activities, such as the number of agencies/cities participating in the project, the amount of data available as a result of that participation, and the amount of use demonstration applications are experiencing. The first of these quarterly reports will be delivered at the start of the project's Phase 3 effort, which is expected to start in February 2024. This will serve as the baseline for the project.

The UW Team also expects the Volpe team to request some raw data, which will allow the Volpe staff to

- undertake additional analyses that are not included in the local evaluation,

- use different techniques than the UW team to perform performance analyses, and
- confirm that the UW team has not made computational errors in the technical performance of the system that the UW team is reporting.

During phases 1 and 2, the UW team will discuss with Volpe staff what raw data the IE team requires for their evaluation efforts. For example, the UW team expects that Volpe staff will be interested in the raw trace data collected as part of the field test of the Multi-Modal AccessMap application, in order to review the technical analysis the UW team performs to determine the number and type of navigation errors occurring when individuals are using the application for turn-by-turn navigation directions. The UW team is willing to share these data but will need to confirm that Volpe and USDOT follow the data security and release requirements that will be part of the UW team's IRB. This is because trace data, especially when that trace data includes a recurrent identifier of any kind (hashed or not hashed) is inherently identifiable given modern data science techniques, and the privacy of test subjects participating in this project need to be protected.

Similarly, Volpe may be interested in reviewing the output of the machine learning software that produces the original routable sidewalk network for cities, as opposed to the data uploaded to the data repository. If this is a valid assumption, the UW team will arrange to save specific examples of these data generation runs in order to transfer them to Volpe for secondary analysis.

Finally, the UW team is planning a number of surveys of stakeholders during Phase 3. Different surveys will be administered to different stakeholder groups, with each survey written specifically for that group. For example, the survey provided to on-demand transit service providers that are supplying GTFS-Flex data will ask about the agency's views on the performance of the software tool used to help generate those GTFS-Flex datasets, the cost to the agency of generating those data, as well as if there are improvements that should be made to the software tool. The cities that own sidewalks for which data are included in the OpenSidewalks data repository will receive a very different survey oriented around the generation, vetting, and use of sidewalk data within their jurisdiction. The UW team will include Volpe staff in the development of those surveys and will share the data from those surveys with Volpe.

In addition to the surveys the UW team will perform during Phase 3, Volpe has already announced that they will perform at least two sets of interviews independent from the surveys described above and in Chapter 3 of this report. One of these sets of interviews will occur during the pre-deployment portion of Phase 2 and one will during post deployment. Both sets of interviews will be with UW team members and (potentially) cooperating agencies and organizations. These interviews are being performed with the knowledge and agreement of the project COR.

For these interviews, the UW team will identify 2-3 deployment managers and provide Volpe with their contact information and electronic introductions. The Volpe team will coordinate the logistics of these interviews. The interviews will be used to better understand the goals, experiences, and results for each of the deployment sites, both before and after the deployment of the TDEI.

The UW team will also identify key deployment partners (e.g., deployment agency staff, technology partners, universities, government or policymakers, or others) to be interviewed by Volpe staff. For this second set of planned IE interviews, the current IE plan is to send to a questionnaire to 3-5 project stakeholders of the UW ITS4US project. The UW team leads (Dr. Caspi or Mr. Hallenbeck) will work with the IE team and the COR to select the types of stakeholders the IE team should interview. Given the large number of stakeholders associated

with the UW ITS4US project, it is distinctly possible that the IE team will need to interview – or send questionnaires to - more than the expected 3 to 5 stakeholders.

# 6. Performance Measurement and Evaluation Support Schedule

Chapter 5 presents the overall plan for the UW team's support for the independent evaluation team. This chapter summarizes the expected schedule for those support activities as well as for the UW team's own evaluation activities. Table 9 shows the schedule the IE team has provided for performing interviews with stakeholders in support of that evaluation effort. The UW team will provide all requested assistance for these interactions as noted in the previous chapter.

**Table 9. Timing of Planned Independent Evaluation Led Surveys and Questionnaires**

| Participant Role | Pre- Deployment Interviews | Post-Deployment Interviews | Questionnaire(s) / Stakeholder Surveys- TBD |
|---|---|---|---|
| **Time Frame** | **Phase 2 Jan-Feb '22** | **Phase 3 Jan-Feb '24** | **Phase 3 Dates: Feb '25 – Aug '25** |
| Federal Program Managers (ITS JPO, FTA, other) | **X** | **X** | no |
| Deployment Managers (e.g., lead agency and other decision makers) | **X** | **X** | no |
| Deployment Partners (deployment agency staff, technology partners, universities, government or policy makers, and others) | **X** | **X** | no |
| Project Stakeholders (transit agencies, healthcare organization, social service agencies, community organizations, etc.) | no | no | **X** |
| Government Entities (FTA Regional Offices, FHWA Resources Center, FHWA Division Offices, State DOT, County, City, etc.) | no | no | **X** |

The UW team will be conducting routine interviews and surveys as part of the co-design process. These surveys are short and specifically designed to gather user input to specific design decisions, either to help design the initial software implementation or to comment on the outcomes of those efforts and provide design insight into further refinements when those are needed. The co-design process is performed separately for different software development tasks. The timing of these interactions will be a function of the software development process. The surveys performed in 2025 that are listed in Table 9 are to be performed after software refinement has been completed and are designed to provide insight into stakeholder attitudes towards the

"finished product" – for this project, while the earlier project surveys are actually part of the software and process development process, not the formal evaluation process. However, the information gained from those survey efforts does inform the project team about the perspectives of the stakeholders.

In addition to these stakeholder interviews, The UW team is aware of four categories of data and reports that it will provide to Volpe staff, including

- all draft and final reports,
- summary performance statistics,
- raw data, and
- additional information as requested by Volpe staff.

Draft and final reports will be provided to the IE team at the same time they are delivered to USDOT. The remaining reports for Phase 1 of the ITS4US project are due on the dates shown in Table 10. Reports to be submitted as part of Phase 2 and Phase 3 of the ITS4US project have not yet been defined by USDOT.

**Table 10. Remaining Phase 1 Report Due Dates**

| Report | Due Date |
|---|---|
| Performance Measurement Plan (Draft) | 9/8//21 |
| Performance Measurement Plan (Final) | 10/25/21 |
| SyRS Document (Draft) | 9/20/21 |
| SyRS Document (Final) | 10/25/21 |
| Enabling Technology Readiness Assessment (Draft) | 8/30/21 |
| Enabling Technology Readiness Assessment (Final) | 11/15/21 |
| Human Use Approval Summary (Draft) | 11/29/21 |
| Human Use Approval Summary (Final) | 12/27/21 |
| Participant Training and Stakeholder Education Plan (Draft) | 12/06/21 |
| Participant Training and Stakeholder Education Plan (Final) | 1/03/22 |
| Institutional, Partnership, and Financial Plan (Draft) | 12/13/21 |
| Institutional, Partnership, and Financial Plan (Final) | 1/10/22 |
| Outreach Plan (Draft) | 11/29/21 |
| Outreach Plan (Final) | 12/27/21 |
| Systems Engineering Management Plan (Draft) | 11/29/21 |
| Systems Engineering Management Plan (Final) | 12/27/21 |
| Integrated Complete Trip Deployment Plan (Draft) | 12/13/21 |
| Integrated Complete Trip Deployment Plan (Final) | 1/10/22 |
| Deployment Readiness Summary (Draft) | 1/31/22 |
| Deployment Readiness Summary (Final) | 2/21/22 |

Given a February 2022 start to Phase 2, and a February 2024 start to Phase 3, the UW Team would suggest the preliminary schedule shown in Table 11 (not including the interview and questionnaire schedule shown in Table 9 above) for engagement with the independent evaluation team, and delivery of the datasets identified in Chapter 4. Table 11 also shows the proposed timing of the UW team's proposed evaluation reporting.

**Table 11. Initial recommended schedule for performance measurement activities and data sharing**

| IE Data Sharing Activity | UW Activities | Date |
|---|---|---|
| x | Stakeholder co-design questionnaires and surveys | Nov 2021 – Feb 2024 |
| x | Preparation of data quality and availability reporting and summarization software | June 2022 – Nov 2023 |
| x | Creation of monthly data status reports for internal management (performed monthly – actual status reports will be a function of which data are being generated and submitted.) | June 2022 – Feb 2026 |
| Initial discussion of data sharing procedures and datasets | Initial discussion of data sharing procedures and datasets | Nov. 2022 |
| x | Development of data sharing software | March 2023 - April 2023 |
| Initial trial of data sharing procedures | Initial trial of data sharing procedures | May 2023 |
| Finalization of data sharing procedures | Finalization of data sharing procedures | Dec. 2023 |
| x | Development of Internal performance reporting | Dec 2023 - Feb 2024 |
| Transfer of baseline data | Transfer of baseline data | Feb 2024 |
| Delivery of a quarterly summary performance update | Creation of quarterly summary performance updates | May 2024<br>Aug. 2024<br>Nov. 2024<br>Feb 2025<br>May 2025<br>Aug. 2025<br>Nov. 2025<br>Feb 2026 |
| Semi-annual delivery of raw data required by Volpe for the IE | Development of semi-annual raw data delivery package | Aug. 2024<br>Feb 2025<br>Aug 2025 |

| IE Data Sharing Activity | UW Activities | Date |
|---|---|---|
| | | Feb 2026 (end of Phase 3) |
| x | Development of detailed field test datasets for performance testing of demonstration applications | Feb 2025 - Aug 2025 |
| Transfer of detailed field test datasets for demonstration application performance testing | Transfer of detailed field test datasets | Aug 2025 |
| Design discussions for the stakeholder performance evaluation surveys | Design discussions for the stakeholder performance evaluation surveys | Feb 2025 |
| x | Performance of surveys and collection of survey data | March 2025 – July 2025 |
| Transfer of survey data from stakeholder evaluation surveys | Transfer of survey data | Aug 2025 |

The above dates will be adjusted as necessary in future revisions of this document. For example, given the large number of datasets which will be used for the evaluation, the initial discussions held between Volpe staff and UW team staff may result in spreading the trial of dataset transfer activities over several months. Similarly, changes in the delivery of technical aspects of the TDEI during Phase 2 could also require revisions to the schedule shown in Table 11.

This chapter of this document will be updated whenever revisions are made to the data sharing plan between the Volpe Center and the UW team, whether those involve adding or changing details about the content of the datasets being shared, the timing of those data sharing activities, or the procedures used to share those data. The COR and ITS JPO lead for this project will participate in the development of, and agree to, all agreements and decisions made with respect to data sharing between the UW team and Volpe team.

The performance reporting statistics listed in Chapter 4, will typically be tracked for internal use by the UW team on an ongoing basis. For example, the team will keep track of the number of organizations that have supplied data or are in the process of supplying that data for internal purposes. These statistics, will be compared against the list of organizations from which data can be obtained on a quarterly basis for performance reporting purposes. However, for management purposes, the UW team is more interested in the current status of where data have been obtained, and which jurisdictions, agencies, and groups still need to be contacted, or re-contacted, or followed up with.

Similarly, statistics like the fraction of centerline miles of roadway examined for the presence of sidewalks is a useful evaluation statistic, but the UW team is more interested in tracking what map sections (aerial images) have been analyzed, and which remain to be analyzed. Thus, much of the data collection and initial evaluation work will occur on an ongoing basis, while the “evaluation work” will be performed as part of summarization process performed to prepare the summary data for delivery to the IE team on a quarterly basis.

# 7. Cost of Participating in and Operating the Data Sharing Framework

As part of our data development activities, the UW team will work with participating agencies to gather information on the resources required to participate in the data collection activities. This includes such costs as the collection of transit center 3-D point cloud data, the staffing resources needed to enter and quality assurance check on-demand transit service data, and the costs associated with both transferring available sidewalk and road crossing inventory data (e.g., ADA curb ramp locations and characteristics, traffic signal and stop sign locations, etc.) and the resources required to work with community and advocacy groups on data vetting. These data will be used to provide estimates of the costs required by agencies to participate in the national expansion of the system being developed in this project.

Similarly, the UW team will track the cost of its own data development activities (e.g., cost of imagery licenses, computer costs of the machine learning process used to convert imagery to routable sidewalk networks), as well as the time and effort required to merge jurisdictional data on sidewalk and road crossing infrastructure with the routable sidewalk network. This will allow the team to estimate the data collection costs of further expansion of the system to the entity that maintains and expands the system once this project ends.

Finally, the UW team will track the computer resource costs (staff and computer resources) required to operate and maintain the system once it has been built. These costs will serve to estimate future operations and maintenance costs given various assumptions about the speed with which expansion of the system occurs.

# 8. References


- Accessible Transportation Technologies Research Initiative (ATTRI) Performance Metrics and Evaluation, Final Evaluation Framework Report, FHWA-JPO-20-784, https://rosap.ntl.bts.gov/view/dot/50748/.
- AccessMap GitHub site, https://github.com/AccessMap/accessmap.
- Bolten, Nicholas, Allie Deford, Reagan Middlebrook, Veronika Sipeeva, Alan Borning and Anat Caspi. AccessMap.io. Seattle, WA: N.p., 2015. Software. First Place Award at “Hack the Commute.” Sponsored by Seattle Department of Transportation.
- Bolten, Nicholas, Amirhossein Amini, Yun Hao, Vaishnavi Ravichandran, Andre Stephens, and Anat Caspi. “Urban sidewalks: visualization and routing for individuals with limited mobility.” First International Workshop on Smart Cities and Urban Analytics (UrbanGIS). Seattle, WA: 2015.
- Bolten, Nicholas, Sumit Mukherjee, and Anat Caspi. “Learning sidewalk path connectivity for accessible trip planning using crowdsourcing and open data.” 2016. ArXiv
- Bolten, Nicholas, Veronika Sipeeva, Sumit Mukherjee, Anissa Tanweer and Anat Caspi. A pedestrian-centered routing approach for equitable access to the built environment. 2017. IBM J. RES. & DEV. VOL. 61 NO. 6:10 [November/December 2017] 10.1147/JRD.2017.2736279.
- Bolten, Nicholas and Anat Caspi. 2019. AccessMap Website Demonstration: Individualized, Accessible Pedestrian Trip Planning at Scale. In The 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19). Association for Computing Machinery, New York, NY, USA, 676–678. DOI: https://doi.org/10.1145/3308561.3354598.
- Bolten, Nicholas and Anat Caspi. "Towards routine, city-scale accessibility metrics: Graph theoretic interpretations of pedestrian access using personalized pedestrian network analysis." PLoS one 16.3 (2021): e0248399.
- Digital Twins website, https://create.unity3d.com/real-time-3d-and-digital-twins.
- FHWA. University of Washington ITS4US Deployment Project—Phase 1 Needs Summary. Final Report—May 3, 2021.
- FHWA. Accessible Transportation Technologies Research Initiative (ATTRI)—User Needs Assessment: Stakeholder Engagement Report. Final Report—May 2016. FHWA-JPO-16-354.
- FHWA. Accessible Transportation Technologies Research Initiative (ATTRI) Policy and Impacts Assessment—Policy Assessment, Gaps & Needs. Final Report—July 2019. FHWA-JPO-17-506.
- (Fishman et al., 2020) Fishman, Tiffany, Kelkar, Mahesh, and Schwartz, Avi (2020). Transportation Trends 2020, https://www2.deloitte.com/us/en/insights/industry/public-sector/transportation-trends.html. Deloitte Services (4/13/2020).

- GTFS-Flex document (ongoing), http://bit.ly/gtfs-flex-v2.
- GTFS-Flex GitHub site, https://github.com/MobilityData/gtfs-flex.
- GTFS-Pathways document (ongoing), http://bit.ly/gtfs-pathways.
- GTFS-Pathways GitHub site, https://github.com/google/transit/pulls?q=is%3Apr+pathways.
- (GTiO) Data Interoperability: A Practitioner's Guide to Joining Up Data in the Development Sector. https://www.data4sdgs.org/sites/default/files/services_files/Interoperability%20-%20A%20practitioner%E2%80%99s%20guide%20to%20joining-up%20data%20in%20the%20development%20sector.pdf, accessed 4/13/2021.
- (ISO) https://www.iso.org/home.html, accessed 3/15/21
- (JUDS, 2016) Joined-Up Data Standards project (2016). The frontiers of data interoperability for sustainable development. Available at: http://devinit.org/wp-content/uploads/2018/02/The-frontiers-of-data-interoperability-for-sustainable-development.pdf.
- (Michel, 2018) Michel, John E. (2018). Mobility-as-a-Service: Enabling the Transformation of Transportation through Digitalization. Mass Transit Magazine (2/19/2018).
- National Academies of Sciences, Engineering, and Medicine 2020. *Development of Transactional Data Specifications for Demand-Responsive Transportation*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25800.
- Needs expressed by MVTransit: How data science is driving digital transformation at MV Transportation. https://www.dxc.technology/workplace_and_mobility/insights/148131-how_data_science_is_driving_digital_transformation_at_mv_transportation.
- (ODW, 2018) Open Data Watch, 2018.
- (OGC) Open Geospatial Consortium.
- OpenSidewalks website, https://tcat.cs.washington.edu/opensidewalks-2/.
- OpenSidewalks GitHub site, https://github.com/OpenSidewalks/OpenSidewalks-Schema.
- (SDF, 2021) San Diego Forward, (2021) "A bold new transportation vision in 5 big moves," (https://www.sdforward.com/mobility-planning/5-big-moves) Accessed on 4/15/2021.
- Soundscape website, https://www.microsoft.com/en-us/research/product/soundscape/.
- Tanweer, Anissa, Margaret Drouhard, Brittany Fiore-Gartland, Nicholas Bolten, Jess Hamilton, Kaicheng Tan, and Anat Caspi. Mapping for Accessibility: A case study of ethics in data science for social good. Bloomberg Data for Good Exchange Conference. 24-Sep-2017, New York City, NY, USA.
- Transportation Data Equity Initiative website, https://transitequity.cs.washington.edu/.
- (Trapeze, 2021) Trapeze Group (2021). Esri and Trapeze collaborating on integrated data platform, https://www.masstransitmag.com/technology/press-

release/21216834/trapeze-group-esri-and-trapeze-collaborating-on-integrated-data-platform. Mass Transit Magazine (4/1/2021).

- (W3C) https://www.w3.org, accessed 3/15/2021.
- Zhang, Yuxian and Anat Caspi. 2019. Stereo Imagery Based Depth Sensing in Diverse Outdoor Environments: Practical Considerations. In Proceedings of the 2nd ACM/EIGSCC Symposium on Smart Cities and Communities (SCC '19). Association for Computing Machinery, New York, NY, USA, Article 4, 1-9. DOI: https://doi.org/10.1145/3357492.3358627.

# Appendix A. Acronyms and Glossary

This appendix includes a list of acronyms and a glossary of key terms used in the document.

| Acronym | Definition |
|---|---|
| AD | Application developer |
| ADA | Americans with Disabilities Act |
| AI | Artificial intelligence |
| API | Application program interface |
| ATTRI | Accessible Transportation Technologies Research Initiative |
| BAA | Broad Area announcement |
| ConOps | Concept of Operations |
| COVID | Coronavirus disease |
| DG | Data generator |
| DMP | Data Management Plan |
| DOT | Department of transportation |
| DRSB | Deployment Readiness Summary Briefing |
| DS | Data service provider |
| DU | Digital device end user experiencing travel barriers |
| ETRA | Enabling Technology Readiness Assessment |
| FHWA | Federal Highway Administration |
| FTA | Federal Transit Administration |
| GIS | Geographic information systems |
| GOFS | General On-Demand Transit Feed Specification |
| GTFS | General Transit Feed Specification |
| GTFS-Flex | The Flex route extension to the General Transit Feed Specification, designed to describe demand-responsive or paratransit service |
| GTFS-Pathways | The Pathways extension to the General Transit Feed Specification which defines pathways linking together locations within stations |
| HUA | Human Use Approval |
| ICTDP | Integrated Complete Trip Deployment Plan |
| IE | Independent Evaluation |
| IEC | International Electrotechnical Commission |
| IEEE | Institute of Electrical and Electronics Engineers |
| IRB | Internal Review Board |
| ISO | International Organization for Standardization |
| IT | Information technology |
| ITS | Intelligent transportation system |
| ITS JPO | Intelligent Transportation Systems Joint Programs Office |
| ITS4US | The name of a USDOT program to enable communities to showcase innovative business partnerships, technologies, and practices that |

| Acronym | Definition |
|---|---|
| | promote independent mobility for all that is led by the Intelligent Transportation Systems Joint Program Office with support from the Office of the Secretary of Transportation, Federal Transit Administration, and Federal Highway Administration. |
| LEP | Limited English proficiency |
| LiDAR | Light detection and ranging |
| MARC | Mid-Atlantic Regional Council |
| MOOVEL | A software services provider to transit agencies |
| MVP | Minimum viable product |
| OGC | Open Geospatial Consortium |
| OSM | OpenStreetMap |
| OST | Office of the Secretary |
| OSW | OpenSidewalks |
| PII | Personally Identifiable Information |
| PMESP | Performance Measurement and Evaluation Support Plan |
| PMP | Project Management Plan |
| PPNA | Personalized pedestrian network analysis |
| PTSEP | Participant Training and Stakeholder Education Plan |
| REST API | Representational State Transfer Application Program Interface |
| ROI | Return on investment |
| SEMP | Systems Engineering Management Plan |
| SMP | Safety Management Plan |
| SyRS | System Requirements Plan |
| Taskar Center or TCAT | Taskar Center for Accessible Technology at the University of Washington |
| TCRP | Transportation Cooperative Research Program |
| TDEI | Transportation Data Equity Initiative |
| TRAC | Washington State Transportation Center at the University of Washington |
| TSP | Transportation service provider |
| U.S. | United States |
| U.S. DOT | United State Department of Transportation |
| USGS | United States Geological Survey |
| UW | University of Washington |
| VA | Veterans Affairs |
| W3C | World Wide Web Consortium |

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

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

FHWA-JPO-21-879

U.S. Department of Transportation