Phase 1 Data Management Plan (DMP)

California Association for Coordinated Transportation ITS4US Deployment Project

www.its.dot.gov/index.htm

Final Report — August 23, 2021
FHWA-JPO-21-866

COMPLETE TRIP
ITS4US

U.S. Department of Transportation

Produced by California Association of Coordinated Transportation ITS4US Deployment Program, Phase 1
U.S. Department of Transportation
Intelligent Transportation Systems Joint Program Office
Federal Highway Administration
Office of the Assistant Secretary for Research and Technology
Federal Transit Administration

Notice

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-JPO-21-866		

4. Title and Subtitle	5. Report Date
Phase 1 Data Management Plan (DMP) **California Association for Coordinated Transportation ITS4US Deployment Project**	August 23, 2021
	6. Performing Organization Code (Delete and insert information here or leave blank)

7. Author(s)	8. Performing Organization Report No.
Natalie Ortiz (Compiler LA), Thomas Craig (CALACT), Taylor Bailey (Garnet Consulting LLC)	(Delete and insert information here or leave blank)

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
California Association for Coordinated Transportation – CALACT 4632 Duckhorn Drive, Sacramento, CA 94583	
	11. Contract or Grant No. 693JJ321C000007

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
U.S. Department of Transportation ITS Joint Program Office 1200 New Jersey Avenue, SE Washington, DC 20590	(Delete and insert information here or leave blank)
	14. Sponsoring Agency Code HOIT-1

15. Supplementary Notes

Robert Sheehan, COR

16. Abstract

The California Association for Coordinated Transportation (CALACT) ITS4US project aims to address the need for riders who use demand-responsive services, rural riders, and riders with disabilities to have equal access to real-time trip planning technologies already available for fixed-route transit and by Transportation Network Companies. The Data Management Plan (DMP) for the CALACT Phase 1 ITS4US project details the data needed to operationalize an improved, technology-enabled transit system within the three state region of California, Oregon, and Washington and details the project's data ownership and stewardship practices across various types of project partners, including state departments of transportation, local/regional transit systems and providers, and technology partners, all of which have preferences toward data management to balance and make work for the project more globally. The DMP also addresses the ways the datasets from the project will be structured. The DMP is a living document that reflects what is currently information known and anticipated as of its last update; changes and revisions will be documented.

17. Keywords	18. Distribution Statement
ITS4US; Complete Trip; Deployment; ITS; Intelligent Transportation Systems; California; Oregon; Washington; Data; Transit; GTFS	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
		(Delete and insert information here or leave blank)	

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

Revision History

Name	Date	Version	Summary of Changes	Approver
Natalie Ortiz, Compiler and Taylor Bailey, Garnet Consulting LLC	26 July 2021	1.0	Draft Final	Thomas Craig, CALACT
Natalie Ortiz, Compiler and Taylor Bailey, Garnet Consulting LLC	23 August 2021	1.1	Updates based on USDOT comments	Thomas Craig, CALACT
Natalie Ortiz, Compiler	14 September 2021	1.2	Updates based on USDOT comment verification	Thomas Craig, CALACT

Table of Contents

Revision History ..... **i**

Table of Contents ..... **iii**

1. Introduction **1**

1.1. Project Background 1

2. Project Overview **5**

2.1. Change Control 6

2.2. Relevant Sources 6

2.3. Data Schedule 6

3. Data Overview **9**

3.1. Data Needs Summary 9

3.2. Data Overview 15

4. Data Stewardship **25**

4.1. Data Owner and Stewardship 25

4.2. Access Level 28

4.2.1. Private Datasets 29

4.2.2. Access Request 32

4.2.3. Related Tools, Software and/or Code 33

4.2.4. Relevant Privacy and/or Security Agreements 33

4.3. Re-Use, Redistribution, and Derivative Products Polices 33

4.4. Data Storage and Retention 36

4.4.1. Data Storage System Description 39

4.4.2. Cybersecurity Polices 40

4.4.3. Data Security Policies and Procedures 40

4.4.4. Back-up and Recovery Policies and Procedures 40

5. Data Standards **41**

5.1. Data Standards 41

5.2. Versioning 47

5.3. Metadata 47

5.3.1. Metadata Types 47

5.3.2. Metadata Structure 47

5.3.3. Metadata Update Process 49

Appendix A. Acronyms and Glossary **50**

List of Tables

Table 1. Schedule and Data Related Milestones....6
Table 2. Data Overview Table....16
Table 3. Data Owner, Steward, and Custodian Information for CALACT ITS4US Data....25
Table 4. Data Access Tier Information....28
Table 5. Private CALACT ITS4US Datasets....30
Table 6. Re-Use, Redistribution, and Derivative Products Licensing for CALACT ITS4US Data....33
Table 7. CALACT ITS4US Storage Systems....37
Table 8. Data Standards for Datasets....41
Table 9. CALACT ITS4US Metadata Structure....47

List of Figures

Figure 1. CALACT ITS4US Context Diagram....9
Figure 2. CALACT ITS4US Proposed System Data Needs....10

1. Introduction

The purpose of this document is to describe the data that will be collected and generated during the course of the CALACT ITS4US project. This plan provides an overview of how data will be managed, described, analyzed, protected, and stored as well as how data will be presented and shared.

The intended audience of this document is the CALACT team, including its subcontractors and stakeholder subcommittee chairs and members, as well as the USDOT program management team. Academic and practitioner stakeholders as well as future technology deployers who may find this document useful are considered as well.

1.1. Project Background

The CALACT project addresses the clear need for riders who use demand-responsive services, including riders with disabilities, to have equal access to the real-time trip planning technology that is already available for urban fixed-route transit. Nearly 300 of the over 500 transit operators in California, Oregon, and Washington deliver a form of demand-responsive service.[1] Rider characteristics of these services likely differ substantially from those on fixed-route services as rural residents and people with disabilities are more likely to be low-income, unable to use fixed-route services due to disability, and/or are living in a physically isolated environment.

The demand-response systems themselves offer a lower quality of rider experience, where would-be passengers must find a transit provider that will serve their needs, call a dispatch system to plan and reserve their trip, requiring a long lead time (typically at least a day in advance), and allowing little room for flexibility. The trip planning experience of demand-response systems is further and uniquely burdened by a complex web of determining operator coverage area, for what qualifications that operator or specific service within that operator's service menu they qualify, if the operator has availability, if they need to pay and how. Unlike fixed route services, which have a well-established data standard and a stable industry of third-party trip planning services, and private Transportation Network Companies (TNCs), which produce their own seamless and instantaneous booking and payments flows, demand-responsive transit lacks the technical solutions which could ease these burdens for their riders. There's no comparable desktop or smartphone experience and no other innovations which exist to untangle these webs of availability, reservations, or payments.

Most fixed route users in the three-state region have access to real-time information about transit services through any mobile device. However, very few users have that information about public

[1] Numbers calculated based on internal lists of agencies and metadata provided by ODOT, WSDOT, and Caltrans.

demand-responsive transit, and none have that information except through custom proprietary systems implemented at a few local agencies. Further, users of fixed-route services who would like more access to details regarding the transit system accessibility features and other amenities often cannot easily find that information.

The particular underserved communities the project focuses on are people with mobility disabilities, people with vision disabilities, people with cognitive and developmental disabilities, people with hearing disabilities, older adults, low-income populations, rural residents, veterans, and people with limited English proficiency.

This project is one of five deployments of the Complete Trip - ITS4US Deployment Program, led by the ITS JPO and supported by Office of the Secretary (OST), Federal Highway Administration (FHWA), and Federal Transit Administration (FTA). These deployments were selected to showcase innovative business partnerships, technologies, and practices that promote independent mobility for all travelers regardless of location, income, or disability. The Complete Trip - ITS4US Deployment Program is carried out in three phases over five years: Concept Development (current phase), Design and Testing, and lastly Operations and Evaluation. There is a post-deployment operations and maintenance phase for an additional five years. The intended outcomes for the CALACT deployment are to improve the user experience and cost efficiency of demand responsive transit for riders at agencies throughout the Washington, Oregon, and California.

Project partner (subcontractor) organizations include:

- Oregon Department of Transportation (ODOT): Agency outreach in Oregon, member of Project Management Team (PMT), transit directory product manager
- Washington Department of Transportation (WSDOT): Agency outreach in Washington, member of PMT, transit analysis product manager
- California Department of Transportation (Caltrans): Agency outreach in California, member of PMT, payments product manager
- Washington State Transit Association (WSTA): Support agency outreach in WA and assist with event coordination
- Trillium, an Oregon small business: Concept design, report writing and product management support
- Compiler LA, a California small business: Software systems requirements and data management lead
- Tamika L. Butler Consulting, a California small business: Internal evaluation and stakeholder engagement
- Mark Wall Associates, a California small business: Agency outreach and support for reporting and project administration
- Estolano Advisors, a California small business: Agency and stakeholder outreach support

- California Partners for Advanced Transportation Technology at UC Berkeley: Project evaluation and stakeholder safety and human use leads
- MobilityData IO, a Canadian nonprofit: Data specification development and technology readiness assessment lead
- Transit, a Canadian private corporation registered for business in the US: Technical advise on customer interface needs and development
- Navilens, a Spanish private corporation registered for business in the US: Digital accessible signage and text to speech product leads
- Google, an American public corporation (unfunded): Participation in an advisory and user testing coordination role

2. Project Overview

The proposed system would put into place a new transit data governance approach to ensure that the General Transit Feed Specification (GTFS) published by transit agencies provides for the needs of riders with disabilities, older riders, low-income riders, rural riders, Limited English Proficiency (LEP) riders, and riders with other safety concerns. Further, the governance system would provide tools and resources to allow all users to access this data and look up information regarding transit systems, as well as to support the development of rider applications by private parties which put that data to use.

These high-level goals imply three different general approaches to measuring the performance of the system:

- Is there more GTFS data published, which complies with the accessibility-focused enhancements fostered by this project?
- Are users able to successfully answer their questions regarding transit services by using the tools and resources provided by the project?
- Do third-party application developers implement the accessibility features suggested by the project?

These general performance measurements will be broken down into a series of precise performance metrics, to be detailed in the Performance Measurement and Evaluation Support Plan (PMESP). The data sources to be used by the project are itemized in Section 3. Specifically the PMESP identifies 6 performance measures related to data sources identified within this DMP.

1. Increase in number of transit agencies for which demand responsive transit and/or paratransit services appear with a booking option in at least two open-data-based apps
    a. Data flow 1.a
2. Increase in number of transit agencies for which complete GTFS data including all project-sponsored extensions is published on an open directory
    a. Data flow 1.a
3. Increased rider satisfaction with regard to station and stop wayfinding as reported in rider surveys (among general public, riders with disabilities, riders with device language is not English)
    a. Data flow 6.b
4. % of riders reporting satisfaction with the trip planning process (among general public, riders with disabilities, riders with device language is not English)
    a. Data flow 6.b
5. More riders will use online or mobile booking tools for demand responsive transit and/or paratransit trips
    a. Data flow 5.a

6. Increase in origins and destinations of demand responsive trips near key fixed-route transfer locations
    a. Data flow 5.a

More information on the specific nature of these data flows is included in section 3.1.

2.1. Change Control

The DMP is a living document that will be updated regularly in alignment with the Schedule and Related Milestones events and dates described in Table 1. These events and related updates are expected to occur throughout Phases 1, 2, and 3. Updates are expected to include providing additional information and detail as decisions are made about project implementation and deployment specifics, and as data are collected, as well as accounting for any unexpected changes to the DMP and to the data and datasets themselves. As a living document, the CALACT ITS4US project intends to use the DMP to document the life of the project as it relates to the project's data needs and present a clear picture to the USDOT, the Independent Evaluator (IE), and other stakeholders of the progression of project-related decisions and activities that impact the data needed, produced, and used by the project.

2.2. Relevant Sources

Phase 1 Concept of Operations (ConOps) California Association for Coordinated Transportation ITS4US Deployment Project, USDOT (2021).

Phase 1 Performance Measurement and Evaluation Support Plan (PMESP) California Association for Coordinated Transportation ITS4US Deployment Project, USDOT (2021).

FIPS Pub 199: Standards for Security Categorization of Federal Information and Information Systems. National Institute of Standards and Technology (2004).

NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, Version 1.0. National Institute of Standards and Technology (2020).

2.3. Data Schedule

The following table outlines key events during the project that will affect the collection and use of project data and, thus, the information provided and details conveyed in the DMP. The table provides a description of the event and notes when subsequent changes to the DMP will result.

Table 1. Schedule and Data Related Milestones

ID	Event Title	Description	Date
1	Draft DMP Is delivered to USDOT	Initial Draft DMP with basic information known at the time of writing	July 2021
2	Finalized DMP Phase 1 Deliverable	DMP is updated with USDOT comments addressed	Aug 2021

ID	Event Title	Description	Date
3	DMP updated	DMP is updated as needed to reflect new information stemming from completion and reviews of Performance Management and Evaluation Support Plan (Task 5); Systems Requirements Plan (Task 6); Human Use Plan (Task 8) and related IRB comments or concerns; and selection and negotiation processes with deployment regions that meet criteria and secure commitments from necessary parties for participating in Phases 2 and 3 of the project	October/November 2021
4	Memoranda of Understanding (MOUs) with deployment sites and partners executed	An MOU that details the expectations of the partnership between the CALACT ITS4US project, project vendors (e.g., software; consultants), and/or the deployment site (including any role of a government agency), including data collection and sharing needs, is signed by all parties	TBD – Task 10 or Phase 2
5	Data sharing agreements (DSAs) between project partners executed (as needed)	A DSA that formalizes and specifies expectations of what, when, and how project-related data will be handled and shared between data owner(s) and steward(s) is signed by all parties	Phase 2
6	DMP updated (as needed)	DMP is updated to reflect new information stemming from execution of MOUs and DSAs	Phase 2
7	Initial data samples created/ collected and provided to USDOT	Initial data samples are created/collected, inspected/validated/analyzed, and submitted to USDOT for review	Phase 2

ID	Event Title	Description	Date
8	GTFS extensions and related data standards finalized and/or adopted	DMP is updated to reflect that the GTFS extensions and other related data standards promoted by the adoption have been officially released either in a 'final' form or officially adopted by a relevant governance body.	Phase 2
9	Project baseline data collection starts	Initial collection of data starts	Phase 2/3
10	DMP updated	DMP updated with any changes reflecting ins and outs of collecting real data	Phase 2/3
11	Baseline data provided to USDOT	Complete baseline data sets are uploaded to USDOT and the IE	Phase 3
12	Initial release of directory/analysis frontend website analytics	Upon the launch of the public directory/analysis frontend the website analytics suite used on the site will begin publishing data, allowing an update to the DMP.	Phase 3
13	Other release milestones as determined	From time-to-time, various systems and technologies used will go through subsequent releases as significant updates are published. The DMP will be updated to reflect these changes	Phase 3
14	Month(s) of user testing of applications begin	Initial upload to USDOT after testing datasets are collected and cleaned	Phase 3
15	Data transferred to USDOT	Datasets are provided to USDOT and IE	Phase 3
16	Data review	Data review conducted with USDOT and IE to ensure datasets are complete	Phase 3
17	Draft final analysis report submitted	Draft version of final analysis report submitted to USDOT	Phase 3
18	Final analysis report submitted	Final version of final analysis report submitted to USDOT	Phase 3

3. Data Overview

3.1. Data Needs Summary

The purpose of this summary is to describe the data that are needed by the CALACT ITS4US project to design, develop, operationalize, and assess the performance of the system proposed in the Concept of Operations (ConOps) Context Diagram as illustrated in Figure 1.

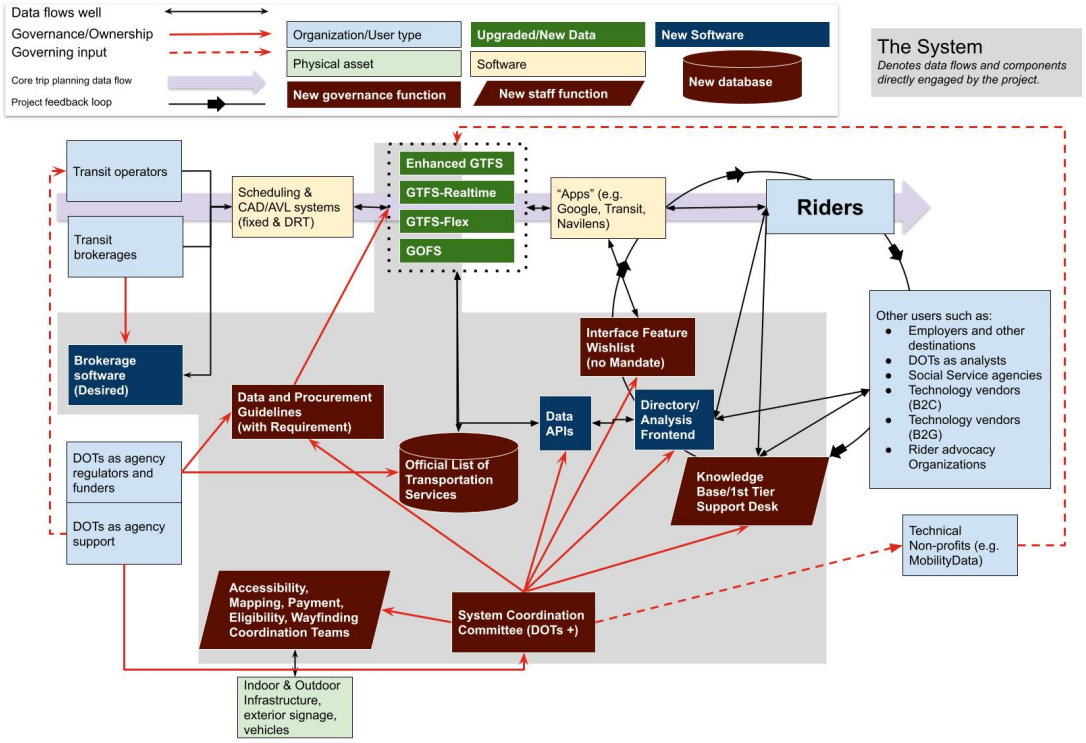

Figure 1. CALACT ITS4US Context Diagram

The data needed by the project are both quantitative and qualitative, supporting the creation of new digital tools and resources—such as websites and software—for state departments of transportation, riders, operators, and other stakeholders of transit systems. These data will also support the project's IEs in measuring and assessing the effectiveness of the system to meet the needs and operational scenarios identified in the ConOps. The data layer of the context diagram is illustrated in Figure 2. Figure 2 depicts the six subsystems that the project anticipates will have a specific role in either generating, sharing, or using data to operationalize and measure the performance of the proposed system. The yellow in the figure highlights a subsystem and the numbers in the figure identify a dataset that will technically move between components of the proposed system. In addition to the description in this section, other details about these datasets are discussed further in this plan.

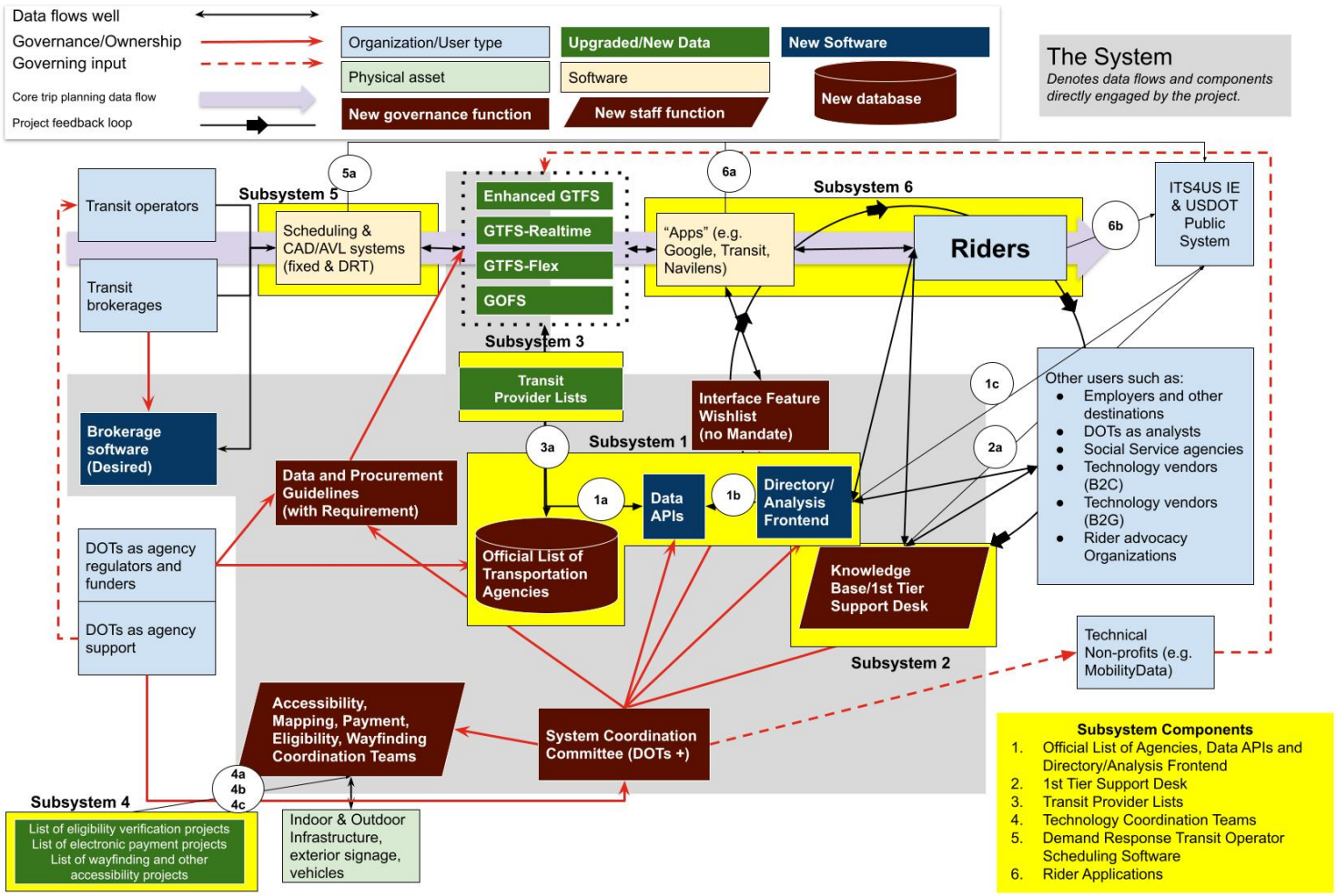

Figure 2. CALACT ITS4US Proposed System Data Needs

The core substantive capabilities of the proposed system depend on better, enhanced GTFS and General On-Demand Feed Specification (GOFS) data that enable an ecosystem around the planning and completing of complete trips for riders who use demand-responsive services, rural riders, and riders with disabilities (see Section 5 of ConOps, pp. 73 - 74). As with the current system, these data are generated by transit agencies (see Section 3 of ConOps, pp. 11-19). The scope of the DMP does not include the production and ongoing maintenance of these data at the agency-level. The project will develop the first draft of these datasets where needed and feasible (specifically pathways, text-to-speech, translations, and flex) on behalf of agencies during the deployment phase to import into their local systems, but the long-term connection between agency systems and rider applications through GTFS not part of the proposed system although it is affected by the system through the Data and Procurement Guidelines.

A brief description of each system component related to data flows is included below before the description of the related data flows, but additional details regarding the proposed system can be found in the CALACT ITS4US Concept of Operations Section 5.2.

Subsystem 1. Official List of Agencies, Data APIs and Directory/Analysis Frontend

The Official List of Agencies, Data APIs and Directory/Analysis Frontend are a database and software applications which ingest, store and distribute GTFS feeds from transit services in the region.

1.a. Official List of Agencies and Data APIs. These data are the GTFS and metadata components collected in the Official List, processed by an application that assigns that data to a database, then produces a series of APIs which can be used by applications like the Directory/Analysis Frontend.

- **Data flows to** the Directory/Analysis Frontend and other registered users
- **Data are stored on** servers rented by project

1.b. Directory/Analysis Frontend. These data describe what, when, and where transit is available across all geographies within a state and are intended for display in a graphical user interface. Data for this system will be largely derived from GTFS and GOFS feeds that transit operators produce and publish for consumption by software applications through application programming interfaces (APIs). The transit directory website data will include data extracted from GTFS and GOFS feeds, including transit service, geographic scope of services, routes, trips, and fares. The data infrastructure for the website will be provided by a party that will be determined later in Phase 1 and it will have responsibility for ensuring the website conforms to national best practices and standards for cybersecurity such as NIST 800-53 or ISO 27001 as well meets the needs of people with disabilities by conforming to the latest version of Web Content Accessibility Guidelines (WCAG) for websites. There may be distinct deployments in different states, or the deployments may be shared with parties outside the states.

Additional data for the transit directory may be sourced from the National Transit Directory (NTD) and state databases. It is possible that additional data may be needed for the website to meet users' needs, pending further user research that cannot be conducted in Phase 1 (i.e., user research requires Institutional Review Board [IRB] review and approval, which is scoped for Phase 2). The development of system requirements may also identify additional data requirements that cannot be foreseen at this exact time.

- **Data flows to** users of the transit directory website
- **Data are stored on** servers rented by project

1.c. Directory/Analysis Frontend website analytics. In addition to the directory itself, this subsystem will produce website usage data that provide information on how the website is navigated by visitors. These data may include geographic location of visitors' IP addresses, search terms (queries) submitted by visitors, click paths, and redirects/referrals from other websites, with specific elements depending on the analytics provider selected. The CALACT ITS4US project largely envisions these data to be an export from the analytics provider. In the long-term, these data—when analyzed for frequencies, patterns, and trends—will be used by the System Coordination Committee (SCC) to better understand visitors' behavior on the website and inform updates to display and navigation in later phases of the project and beyond.

- **Data flows to** later defined evaluation activities only
- **Data are stored on** servers rented by project

Subsystem 2. 1st Tier Support Desk

The 1st Tier Support Desk is a system component that will be piloted in only one deployment region, and includes both services and technology that answer conversational questions directly from riders through online, text, and voice media. The exact design of communication endpoints have not yet been determined and are subject to low-level design choices during Phase 2.

2.a. Transit directory website user support/help desk (knowledge base/first tier support). A data component of this subsystem is the exchange records of the help desk function of the website. The help desk connects visitors who submit questions in conversational language in real-time with assistance in getting answers to their questions, thus creating a chat-based exchange that is recorded. The CALACT ITS4US project largely envisions these data to be used as an export from the help desk software provider. The collection of the help desk data provides a direct feedback loop with the rest of the proposed system, with the responsibility being with the SCC for how the data is meaningfully used to influence new website features, improve operators' services, and/or to propose new software features to transit planning application developers (i.e., Interface Feature Wishlist).

- **Data flows to** later defined evaluation activities only
- **Data are stored on** servers rented by project

Subsystem 3. Transit Provider Lists

Transit provider lists are state-level inputs to the Official List of transportation agencies.

3.a. List of state transit providers and services. These data describe the scope of transit providers and services for each state that is a CALACT ITS4US project partner. Data for this product are generated through state departments of transportation (DOTs) as a function of their responsibility to regulate transportation services in their states. A list for each state will be curated using a standard to dictate which data elements are needed to ensure a comprehensive regional list of all agencies and that allow each agency's GTFS or GOFS feeds to be coordinated and in sync with all other transit services available within the region specifically and the state generally.

An example of the data to be collected by each state is Caltrans' Cal-ITP transit providers dataset. These lists will act, in part, as a data source for Subsystem 1. Lists may take one of several forms, depending on the state DOT that is producing the list. For example, the list could be a single csv, it may be several csvs that could be merged into a single larger file using unique identifiers, or it could be one or more application programming interfaces (APIs).

Lists will be the responsibility of the state to produce and subsequently maintain, and each state will have its rules and policies for how these data are stored, protected, accessed, and governed.

For the purposes of the CALACT ITS4US project, the development of the list will be coordinated through the SCC and each state will steward its own list content. In the long-term, this data resource may also be used for other purposes as determined by each state.

- Data flows to Official List of Agencies and Data APIs
- Data are stored by state departments of transportation

It is worth noting that at this moment in Phase 1 and drafting of the DMP, the project is still determining if these data will be needed to operationalize the system. A future version of the DMP informed by additional work on system requirements may determine these data are unnecessary; the DMP and Figure 2 would be updated accordingly.

Subsystem 4. Technology Coordination Teams

Technology Coordination Teams are staff and contractors of the project that provide training and support to DOTs, agencies, and project partners and identify and develop ways to coordinate technologies which cannot easily be standardized in the near-term.

4.a. List of eligibility verification projects. These data describe the scope, technology, and other design and implementation processes and features of eligibility verification projects that have been completed or are in development within each state that is a CALACT ITS4US project partner. The intent of having these data and making them available is to provide a reference tool for local transit operators to understand 1) the contours of modernizing and digitizing the verification process, 2) how to use verified eligibility data to facilitate the planning, booking, and paying of trips within the transit ecosystem, and 3) who to contact for additional information on the project and/or how to link/coordinate services within a region.

Data collected under this subsystem will be coordinated through the SCC. The project management organization (PMO) identified by the SCC will be responsible for producing and subsequently maintaining the list, and will have rules and policies for how these data are stored, protected, accessed, and governed.

4.b. List of electronic payment projects. These data describe the scope, technology, and other design and implementation processes and features of payment projects that have been completed or are in development within each state that is a CALACT ITS4US project partner. The intent of collecting these data and making them available is to provide a reference tool for local transit operators to understand 1) the contours of modernizing and digitizing payment methods and systems, 2) how to use new payment methods to book and pay for trips within the transit ecosystem, and 3) who to contact for additional information on the project and/or how to link/coordinate services within a region.

Generating the data for this product is coordinated through the SCC. The PMO identified by the SCC will be responsible for producing and subsequently maintaining the list, and will have rules and policies for how these data are stored, protected, accessed, and governed.

4.c. List of wayfinding and other accessibility projects. These data describe the scope, technology, and other design and implementation processes and features of projects that have been completed or are working on improving wayfinding and accessibility for transit users, including non-English speakers and people using assistive devices. The intent of collecting these data and maintaining them within the system team is to provide a reference tool for local transit operators to understand 1) the ways, methods, and systems used to improve wayfinding and accessibility, 2) how wayfinding and accessibility improvements are integrated into the rest of the transit ecosystem, and 3) who to contact for additional information on the project and/or how to link/coordinate services within a region.

Data to be collected will be coordinated through the SCC. The PMO identified by the SCC will be responsible for producing and subsequently maintaining the list, and will have rules and policies for how these data are stored, protected, accessed, and governed.

It is worth noting that at this moment in Phase 1 and drafting of the DMP, the project is still determining if there is a need to conceptualize these lists within a subsystem and thus as datasets needed to operationalize the system. A future version of the DMP informed by additional work on system requirements may determine these data are unnecessary; the DMP and Figure 2 would be updated accordingly.

Subsystem 5. Demand Response Transit Operator Scheduling Software

Demand Response Transit Operator Scheduling software is outside of the CALACT system, but affected by it. These systems are purchased by agencies and made subject to the Data and Procurement Guidelines through agreements with those agencies.

5.a. Usage Statistics. These data describe various generalized, non-user specific aspects of demand response transit operations, such as when and how scheduling and booking take place or whether origins or destinations are near any transit stop but not specific OD pairs or other potential PII. These data would be derived from pre-populated reports, static and/or dynamic dashboards, and/or custom queries of backend software databases. The intent of collecting these data is to understand high-level trends and patterns in a local transit system—i.e., a deployment site—that occur within operations over one or more time intervals. These data would be aggregated, meaning single, individual, or row level observations would be calculated and analyzed along one or more attributes. These statistics enable and fully support tracking changes in transit system operations that stem directly from CALACT ITS4US project investments and are needed to enable the project's Performance Measurement and Evaluation Support Plan (PMESP).

- **Data flows to** later defined evaluation activities only
- **Data are stored on** servers maintained by a project partner

Subsystem 6. Rider Applications

Rider Application software is outside of the CALACT system, but affected by it. These applications have the opportunity to use the GTFS data published by agencies according to the

Data and Procurement Guidelines and to incorporate suggestions from the Interface Feature Wishlist.

6.a. Usage Statistics. These data describe various aspects of rider application usage and are derived from pre-populated reports and/or custom queries of backend software databases. The intent of collecting these data is to understand trends and patterns in rider behavior that correspond with software application use within one or more deployments site that occur over one or more time intervals during implementation and monitoring of the CALACT ITS4US project. These data would be aggregated, meaning single, individual, or row level observations would be calculated and analyzed along one or more attributes. These statistics enable and support tracking changes in rider behavior that can be correlated with the project's investment in transit system changes and are needed for the project's PMESP.

- **Data flows to** later defined evaluation activities only
- **Data are stored on** servers maintained by a project partner

6.b. Rider Surveys. These data describe various aspects of subjective rider attitudes toward various aspects of their transit experience, including digital experiences (e.g., application-based planning) and in-person experiences (e.g., wayfinding). The intent of collecting these data is to understand patterns within riders' experience that are related to transit system changes affected by the CALACT ITS4US project. These data would be aggregated, meaning single, individual, or row level observations would be calculated and analyzed along one or more attributes. These statistics enable one or more performance metrics detailed in the project's PMESP and are therefore useful for understanding the impact of the project's investment in transit system changes.

- **Data flows to** later defined evaluation activities only
- **Data are stored on** servers maintained by a project partner

More information on the approaches to protecting rider privacy is provided in later sections of the DMP.

3.2. Data Overview

Table 2 provides a description and additional detail on the nature, scope, and detail of each unique dataset that is anticipated by the CALACT ITS4US project:

Table 2. Data Overview Table

Dataset Title	Description	Type / Scale	Collection Method	Data File Format(s)
State of California Transit Services Directory (3a)	This dataset includes known transit providers in California, as defined by the State of California DOT and to which it can regulate, and includes public transit operators as well as services that may not fall within commonplace definitions of transit. It is currently limited to providers who operate their own services.	String; hundreds of entries	Manual; derived from GTFS and GOFS feeds; extracted from national datasets	.csv
State of Oregon Transit Services Directory (3a)	This dataset includes known transit providers in Oregon, as defined by the State of Oregon DOT and to which it can regulate, and includes public transit operators as well as services that may not fall within commonplace definitions of transit. It is currently limited to providers who operate their own services.	String; dozens of entries	Manual; derived from GTFS and GOFS feeds; extracted from national datasets	.csv

Dataset Title	Description	Type / Scale	Collection Method	Data File Format(s)
State of Washington Transit Services Directory (3a)	This dataset includes known transit providers in Washington, as defined by the State of Washington DOT and to which it can regulate, and includes public transit operators as well as services that may not fall within commonplace definitions of transit. It is currently limited to providers who operate their own services.	String; dozens of entries	Manual; derived from GTFS and GOFS feeds; extracted from national datasets	.csv
State of California Eligibility Verification Projects List (4a)	This dataset describes the projects in progress or completed in the State of California at any level of operation (e.g., provider, county, statewide) and by any funding source that are creating new automated ways for transit riders to verify their eligibility for discounts for or access to transit services.	String; a dozen or so entries	Manual	.csv

Dataset Title	Description	Type / Scale	Collection Method	Data File Format(s)
State of Oregon Eligibility Verification Projects List (4a)	This dataset describes the projects in progress or completed in the State of Oregon at any level of operation (e.g., provider, county, statewide) by any funding source that are creating new automated ways for transit riders to verify their eligibility for discounts for or access to transit services.	String; a dozen or fewer entries	Manual	.csv
State of Washington Eligibility Verification Projects List (4a)	This dataset describes the projects in progress or completed in the State of Washington at any level of operation (e.g., provider, county, statewide) by any funding source that are creating new automated ways for transit riders to verify their eligibility for discounts for or access to transit services.	String; a dozen or fewer entries	Manual	.csv

Dataset Title	Description	Type / Scale	Collection Method	Data File Format(s)
State of California Electronic Payment Projects List (4b)	This dataset describes the projects in progress or completed in the State of California at any level of operation (e.g., provider(s), county(ies), statewide) by any funding source that are creating new ways for transit riders to pay for transit fares.	String; a dozen or so entries	Manual	.csv
State of Oregon Electronic Payment Projects List (4b)	This dataset describes the projects in progress or completed in the State of Oregon at any level of operation (e.g., provider(s), county(ies), statewide) by any funding source that are creating new ways for transit riders to pay for transit fares.	String; a dozen or fewer entries	Manual	.csv

Dataset Title	Description	Type / Scale	Collection Method	Data File Format(s)
State of Washington Electronic Payment Projects List (4b)	This dataset describes the projects in progress or completed in the State of Washington at any level of operation (e.g., provider(s), county(ies), statewide) by any funding source that are creating new ways for transit riders to pay for transit fares.	String; a dozen or fewer entries	Manual	.csv
State of California Wayfinding and Accessibility Projects List (4c)	This dataset describes projects in progress or completed in the State of California at any level of operation (e.g., provider(s), county(ies), statewide) by any funding source that are improving wayfinding and accessibility for riders.	String; a dozen or so entries	Manual	.csv

Dataset Title	Description	Type / Scale	Collection Method	Data File Format(s)
State of Oregon Wayfinding and Accessibility Projects List (4c)	This dataset describes projects in progress or completed in the State of Oregon at any level of operation (e.g., provider(s), county(ies), statewide) by any funding source that are improving wayfinding and accessibility for riders.	String; a dozen or fewer entries	Manual	.csv
State of Washington Wayfinding and Accessibility Projects List (4c)	This dataset describes projects in progress or completed in the State of Washington at any level of operation (e.g., provider(s), county(ies), statewide) by any funding source that are improving wayfinding and accessibility for riders.	String; a dozen or fewer entries	Manual	.csv

Dataset Title	Description	Type / Scale	Collection Method	Data File Format(s)
Data APIs and related database (1a)	This dataset covers the full scope of transit services known to the CALACT ITS4US project across the states of California, Oregon, and Washington.	JSON or GeoJSON; dozens of endpoints each with differing scales ranging from size of statewide lists to millions of entries (e.g. for stop_times)	Derived from state transit service directories; derived from GTFS and GOFS feeds	.csv
Directory/Analysis Frontend Analytics (1b)	This dataset represents how visitors of the website directory interact and navigate the website itself.	Unknown at this time; analytics provider has not been selected at this point in Phase I; scale depends on usage but would presumably be millions of rows over time.	Unknown at this time; analytics provider has not been selected at this point in Phase I.	Unknown at this time; depends on analytics provider, which has not been selected at this point in Phase I.
Website Help Desk Exchange Records (2a)	This dataset consists of the question-answer response exchanges between visitors of the transit directory website and the answer from the help desk.	String or TBD; scale depends on usage but would presumably be thousands of rows over time.	Unknown at this time; help desk software provider has not been selected at this point in Phase I.	Unknown at this time; depends on help desk analytics provider, which has not been selected at this point in Phase I.

Dataset Title	Description	Type / Scale	Collection Method	Data File Format(s)
Rider Survey(s) (6b)	This dataset represents the responses provided by riders using services in one or more of the deployment regions.	String or TBD; scale depends on usage but would presumably be thousands of rows over time.	Possible methods include app-based (i.e., in partnership with a vendor that is also a project partner, such as Google, Transit, or Navilens) or via a transit agency involved in deployment. Likely to be collected via digital/web-based design versus in-person design.	.csv; there may be multiple files taking the same format.
Scheduling Application Booking Statistics (5a)	This dataset tracks high-level statistics regarding the usage of a demand-response scheduling system, such as the number of rides scheduled.	String or TBD; scale depends on usage but would presumably be thousands of rows over time.	Unknown at this time; analytics provider has not been selected at this point in Phase I.	Unknown at this time; depends on analytics provider, which has not been selected at this point in Phase I.
Rider Application User Statistics (6a)	This dataset tracks user behavior within rider applications which adopt project technologies and partner through a deployment site.	Unknown at this time; analytics provider has not been selected at this point in Phase I; scale depends on usage but would presumably be thousands of rows over time.	Unknown at this time; analytics provider has not been selected at this point in Phase I.	Unknown at this time; depends on application provider, which has not been selected at this point in Phase I.

4. Data Stewardship

4.1. Data Owner and Stewardship

Table 3 identifies the data owners and stewards of each dataset the CALACT ITS4US project anticipates creating and managing (see Table 2). The USDOT ITS JPO is the federal sponsor of each dataset. To facilitate accurate interpretation of roles for Phase 1 of the project, the table and any subsequent references to such roles uses the following definitions of data owner and steward. This delineation ensures that the meaning of USDOT ITS JPO's role in managing the datasets that the CALACT ITS4US project produces is provided. The very nature of the ITS4US project implies a lineage across datasets and there is a decision-making structure around the development and production of all of the project's datasets. The definitions used in the CALACT ITS4US project and in this section are:

- **Data owner:** The data owner is the person or organization with the authority, ability, and responsibility to access, create, modify, store, use, share, and protect the data. Data owners have the right to delegate these privileges and responsibilities to other parties.
- **Data steward:** The data steward, at the direction of the data owner, is the person or organization that is delegated the privileges and responsibilities to manage, control, and maintain the quality of a data asset throughout the data lifecycle. The data steward may also apply appropriate protections, restrictions, and other safeguards depending on the nature of the data, subject to the direction of the data owner.
- **Federal sponsor:** The federal sponsor will assume the role of Data Owner once the dataset(s) are provided to them per Broad Agency Announcement requirements later in the deployment program.

The roles various project partners play with regard to data will be officiated within the agreements with those partners. Note that software vendors providing data related to data flow subsystems 5a and 6a are external to the proposed system and agreements governing that data will only reference specific data sets provided to the system for the purpose of performance evaluation pursuant to the CALACT PMESP. Additionally, these data will generally be covered through agreements between the deployment site agencies and vendor with which those agencies have sub agreements, rather than directly between the ITS4US project contractor and the vendor.

Table 3. Data Owner, Steward, and Custodian Information for CALACT ITS4US Data

Dataset Title	Data Owner	Data Steward	Federal Sponsor
State of California Transit Services Directory	Caltrans	SCC	ITS JPO

Dataset Title	Data Owner	Data Steward	Federal Sponsor
State of Oregon Transit Services Directory	ODOT	SCC	ITS JPO
State of Washington Transit Services Directory	WSDOT	SCC	ITS JPO
State of California Eligibility Verification Projects List	Caltrans	SCC	ITS JPO
State of Oregon Eligibility Verification Projects List	ODOT	SCC	ITS JPO
State of Washington Eligibility Verification Projects List	WSDOT	SCC	ITS JPO
State of California Electronic Payment Projects List	Caltrans	SCC	ITS JPO
State of Oregon Electronic Payment Projects List	ODOT	SCC	ITS JPO
State of Washington Electronic Payment Projects List	WSDOT	SCC	ITS JPO
State of California Wayfinding and Accessibility Projects List	Caltrans	SCC	ITS JPO
State of Oregon Wayfinding and Accessibility Projects List	ODOT	SCC	ITS JPO

Dataset Title	Data Owner	Data Steward	Federal Sponsor
State of Washington Wayfinding and Accessibility Projects List	WSDOT	SCC	ITS JPO
Website Directory of Transit Services	The vendor/partner that will develop the website has not yet been determined	SCC	ITS JPO
Transit Directory Website Analytics	The vendor/partner that will provide website analytics has not yet been determined	SCC	ITS JPO
Website Help Desk Exchange Records	The vendor/partner that will provide the help desk has not yet been determined	SCC	ITS JPO
Rider Survey(s)	To be determined by collection method, which has not yet been decided	SCC	ITS JPO
Scheduling application booking statistics	The vendor(s) that will provide the scheduling application has not yet been determined	SCC	ITS JPO
Rider application user statistics	The vendor(s) that will provide the rider application has not yet been determined	SCC	ITS JPO

4.2. Access Level

The CALACT ITS4US project will use a category-based approach to control access to datasets the project will produce and use. There are two major categories that datasets can be assigned to: 1) Open and 2) Private. Within the Private category, there are three mutually exclusive subcategories to which a dataset can be assigned: 1) PII Data, 2) Proprietary Data, and 3) Research Data. Furthermore, these categories are aligned with the Federal Information Processing Standard 199 to control the risks and impacts stemming from the loss of confidentiality, integrity, and/or availability of the data to the organizations and individuals who are represented in the datasets and take into account the NIST Privacy Framework.

Table 4. Data Access Tier Information

Access Category	Definition	Impact
Open	Data that can be used by the public with no or limited licensing restrictions. This data is available to the public without needing to request permissions and will be provided to the USDOT-managed Public System. These data may be anonymized or aggregated version of private datasets to protect personally identifiable information (PII).	No known impacts, risks, jeopardy, or adversity to individuals and/or organizations if disclosed or breached. No public laws, regulations, or contracts restrict access.
Private	Data that cannot be shared with external users. Access to these data is limited and only granted with IRB and Project Team approvals. Private data are further categorized into PII Data, Proprietary Data, and Research Data.	Impacts, risks, jeopardy, or adversity to individuals and/or organizations could be reasonably anticipated if data were released. These consequences might be mitigated through transformations (e.g., redaction, sanitization) or other processes.
Private – PII Data	Data that has PII included in the dataset. Access to this data should be as restrictive as possible to protect the PII based on IRB-approved processes. Data in this category should have an operational purpose that justifies its storage.	Impacts, risks, jeopardy, or adversity to individuals and/or organizations are reasonably anticipated if data were released. Public laws, regulations, contracts, including consent procedures, restrict use.
Private – Proprietary Data	Data that is either licensed from third parties or represents confidential business information (CBI). This data may be used for operational purposes. Any access to the	Impacts, risks, jeopardy, or adversity to individuals and/or organizations could be reasonably anticipated if data were released without consideration of

Access Category	Definition	Impact
	data is determined by usage agreements between the parties.	consequences, which might be mitigated through conditions described in a data use agreement between parties.
Private – Research Data	Data that is available for research, but users of the data must meet IRB requirements before gaining access to the data. These datasets may have PII.	Impacts, risks, jeopardy, or adversity to individuals and/or organizations could be reasonably anticipated if data were released without consideration of consequences, which might be mitigated through conditions prescribed by IRB.

The project may also develop column and/or row level access criteria, but at this point in developing the project relative to the timeline for drafting the DMP, it is too early to define with any sort of specificity. The project anticipates having additional information on whether it will implement granular data access features in Phase 3. Future revisions to the DMP will reflect updates to the decisions and rationale for restricting and permitting access to certain observations within a dataset and/or certain fields within a dataset.

4.2.1. Private Datasets

At this point in the development of the CALACT ITS4US project, there is no intention around keeping one or more datasets private in their entirety and not shared with public users, including researchers. Subsets of particular datasets could contain Private Data (see Table 4 for definition), but such datasets could still be categorized as Open and thus shared if anonymized, aggregated, or through other approved processes. As noted above, some determinations about vendors have not been made, including vendors that will partner on the project on providing rider applications. Some of these vendors have systems (e.g., databases) that will provide usage statistics and the vendors may desire stricter access controls to the data, going so far as to desire the data be categorized as Private – Proprietary Data and thus may not be available for public use. Still, datasets that represent rider-level usage statistics are not indented to contain any individual-level records. Furthermore, two of the subcategories for Private Data reference IRB approvals and requirements; the project will not engage with IRB until Phase 2. As such, the information contained in this section is subject to change.

PII includes anything that can distinctly identify a specific person directly or indirectly. It includes information like full name or legal name, phone number, mailing or physical address, and/or email address that can be directly associated with a person. To operate one or more parts of the proposed system, there is some anticipation around the collection and use of geographic indicators, such as IP address or nearest transit center or bus stop, demographic information, and health status, such as having health-related conditions like physical or cognitive disabilities, and information on languages spoken (e.g., signaling LEP riders, one of the project's target populations), which could be used by data users to falsely assume characteristics about race/ethnicity or citizenship. While these geographic and demographic data elements would not be collected along with elements that directly identify a specific person, the combination of IP address, primary language spoken, and other identifiers such as gender identity and disability could infer someone's personal identity, especially in rural areas with relatively small populations of demand-response transit users (e.g., paratransit users). For these reasons, datasets with this

type of user-level information will be classified as Private (see Table 4) and be subject to additional rules and processes for accessing. Table 5 provides more information on the datasets that will be Private. Additional information about the privacy of datasets may be determined by an IRB.

Table 5. Private CALACT ITS4US Datasets

Dataset Title	Reason(s) the Data is Private	Safeguarding Methods and Processes
Website Help Desk Exchange Records	The use of natural language in a website-based chat exchange may reveal identifiable information, such as home address, that could jeopardize the privacy of website help desk users.	Data revealing identifiable information such as address will be redacted before making the data accessible because it risks identifying people and personal identity is not needed to accomplish the project's purpose or its performance management and analytic strategy. While additional details still need to be determined, the project will work with the help desk vendor to ensure that PII is redacted, while also balancing the project's interests in discovering potential patterns associated with help desk users providing identifying information.

Dataset Title	Reason(s) the Data is Private	Safeguarding Methods and Processes
Rider Survey(s)	Demographic information, when combined with geographic information, could narrow potential identities because of small cell sizes within the deployment area(s), and jeopardize the privacy of riders.	Data elements and combinations of data elements that contain demographic and geographic information may be redacted before making the data accessible because of the risk associated with identifying respondents of the survey. While responses from certain categories of survey respondents (riders) are fundamental to CALACT's ITS4US performance measurement and evaluation strategy, unrestricted access to the data could jeopardize the privacy of riders in some cases and in some areas. Data owners that have a role in data collection are expected to store the complete, unredacted dataset on a secured system and follow industry control standards for access management. Project personnel involved in tracking system performance and analyzing these specific data are expected to have secure systems that are independently monitored and audited and password protect files as necessary.

Dataset Title	Reason(s) the Data is Private	Safeguarding Methods and Processes
Rider Application Usage Statistics	Rider information, such as the geographic location where a trip starts or ends, combined with other application setting, application data, or other non-project datasets could jeopardize the privacy of riders. Application providers that use average use statistics to generate business could be impacted if a dataset suggests usage is different than marketed.	Data elements and combinations of data elements that contain demographic and/or geographic information may be redacted before making the data accessible because of the risk associated with identifying riders. While usage data from certain categories of riders are fundamental to CALACT's ITS4US performance measurement and evaluation strategy, unrestricted access to the data could jeopardize the privacy of riders in some cases and in some areas. Data owners that have a role in data collection are expected to store the complete, unredacted dataset on a secured system and follow industry control standards for access management and perform necessary redactions, aggregations, or apply differential privacy techniques before releasing to project teams. As described above in section 3.2, the planned form of these data sets would be aggregated and not include specific geographical locations. However, this section identifies that depending on the vendor system (which is not yet known) unrestricted access to the source of the data (i.e. a data dashboard within the proprietary system) could potentially expose some information which qualifies as PII. For that reason these data sets will be considered at risk of including PII, and controls will be determined during Phase 2 to ensure that only aggregated statistics that do not qualify as PII are published.

4.2.2. Access Request

To the extent that is permissible by an IRB, datasets that are classified as Private (see Table 4 for definition) will be made available to the public with the necessary redactions of cells, elements (columns), and/or observations (rows).

If approved by the IRB, the full, unredacted datasets may be requested by transit and mobility researchers to fulfill a compelling research purpose that is determined by both the data owner and the SCC in conjunction and consensus with each other. A data use request should explain the purpose, the significance of the unredacted CALACT ITS4US project data to the purpose, the methods for protecting the confidentiality of the dataset from unauthorized users, and other information on rational and use as determined by the SCC and/or at the suggestion of the IRB. Potential data users would be required to sign a data use agreement with the SCC to formalize their acceptance and agreement with the terms and conditions allowing access to and use of the data.

4.2.3. Related Tools, Software and/or Code

At this point in what is known about the CALACT ITS4US project datasets, there is no anticipated need for specialized tools, software, and/or code for reuse or replication. To the extent the project is able, software code developed for operationalizing the proposed system will be open source and accessible through a repository service (e.g., GitHub).

4.2.4. Relevant Privacy and/or Security Agreements

The CALACT ITS4US project anticipates executing agreements with data owners, including state DOTs, as well as deployment partners like transit agencies and software vendors to implement and test the proposed system and its subsystems. Whether these agreements are two-party or multi-party is still to be determined and may be based on preferences within partnerships that are still under development. The project anticipates executing agreements within the timeframe noted in Table 1.

The project intends to execute data sharing agreements that base terms and conditions on current and emerging standards for privacy and security, such as NIST 800-53, and in the spirit of California Consumer Privacy Act and similar privacy protection laws. To the extent necessary for government partners involved in the project, the project may also ensure software vendors have recently completed a standards-based security audit, such as SOC 2 Type 1 or 2.

4.3. Re-Use, Redistribution, and Derivative Products Polices

The CALACT ITS4US projects intends to make public and reusable as much project data as possible. Table 5 specifies the license types for each dataset:

Table 6. Re-Use, Redistribution, and Derivative Products Licensing for CALACT ITS4US Data

Dataset Title	License Used	Reason(s) for Non-Open License
State of California Transit Services Directory	Creative Commons (CC) BY 4.0	--

Dataset Title	License Used	Reason(s) for Non-Open License
State of Oregon Transit Services Directory	Creative Commons (CC) BY 4.0	--
State of Washington Transit Services Directory	Creative Commons (CC) BY 4.0	--
State of California Eligibility Verification Projects List	Creative Commons (CC) BY 4.0	--
State of Oregon Eligibility Verification Projects List	Creative Commons (CC) BY 4.0	--
State of Washington Eligibility Verification Projects List	Creative Commons (CC) BY 4.0	--
State of California Electronic Payment Projects List	Creative Commons (CC) BY 4.0	--
State of Oregon Electronic Payment Projects List	Creative Commons (CC) BY 4.0	--
State of Washington Electronic Payment Projects List	Creative Commons (CC) BY 4.0	--
State of California Wayfinding and Accessibility Projects List	Creative Commons (CC) BY 4.0	--
State of Oregon Wayfinding and Accessibility Projects List	Creative Commons (CC) BY 4.0	--
State of Washington Wayfinding and	Creative Commons (CC) BY 4.0	--

Dataset Title	License Used	Reason(s) for Non-Open License
Accessibility Projects List		
Website Directory of Transit Services	Creative Commons (CC) BY-SA 4.0	--
Transit Directory Website Analytics	Open licensing may have to be negotiated with the analytics provider; the CALACT ITS4US project is committed to allowing for unrestricted reuse to the extent possible. This will be updated when the project has conducted more planning and has more information on its implementation partners.	
Website Help Desk Exchange Records	Open licensing may have to be negotiated with the help desk provider; the CALACT ITS4US project is committed to allowing for unrestricted reuse to the extent possible. This will be updated when the project has conducted more planning and has more information on implementation and its implementation partners.	
Rider Survey(s)	Open licensing may have to be negotiated with the data owner; the CALACT ITS4US project is committed to allowing for unrestricted reuse to the extent possible. This will be updated when the project has conducted more planning and has more information on implementation and its implementation partners.	

Dataset Title	License Used	Reason(s) for Non-Open License
Scheduling Application Booking Statistics	Open licensing may have to be negotiated with the data owner; the CALACT ITS4US project is committed to allowing for unrestricted reuse to the extent possible. This will be updated when the project has conducted more planning and has more information on implementation and its implementation partners.	
Rider Application User Statistics	Open licensing may have to be negotiated with the data owner; the CALACT ITS4US project is committed to allowing for unrestricted reuse to the extent possible. This will be updated when the project has conducted more planning and has more information on implementation and its implementation partners.	

4.4. Data Storage and Retention

There are two key data owners that have responsibility for storing and retaining raw, unredacted, and unmanipulated/untransformed data generated or used by the CALACT ITS4US project: State DOTs (e.g., to create lists of providers/services and projects) and software vendors the project contracts with to support the deployment of system components (e.g., help desk provider). While there are still details to be determined about exact storage systems, Table 6 provides what is expected given information that is known at this time in the project.

All data provided to the USDOT for public use will be discoverable through a USDOT-managed public system and be made available through the performance period of the project or through the ongoing operations period as appropriate. During Phase 2 and 3 additional information will be provided on the specific USDOT-managed public system to be used.

Table 7. CALACT ITS4US Storage Systems

Data Storage System Type	Dataset Title(s)	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
State of California – Public System	Transit Services List Eligibility Verification Projects List Electronic Payment Projects List Wayfinding and Accessibility Projects List	Phase 2	As needed	Through February 2031
State of Oregon – Public System	Transit Services List Eligibility Verification Projects List Electronic Payment Projects List Wayfinding and Accessibility Projects List	Phase 2	As needed	Through February 2031
State of Washington – Public System	Transit Services List Eligibility Verification Projects List Electronic Payment Projects List Wayfinding and Accessibility Projects List	Phase 2	As needed	Through February 2031

Data Storage System Type	Dataset Title(s)	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
Third-party – Restricted System (vendor to be determined)	Directory Analysis Frontend Official List of Transportation Services Data APIs	Phase 3	As needed	Through the performance period
Third-party – Restricted System (vendor to be determined)	Transit Directory Website Analytics	Phase 3	Daily	Through the performance period
Third-party – Restricted System (vendor to be determined)	Website Help Desk Exchange Records	Phase 3	Daily	Through the performance period
Third-party – Restricted System (vendor to be determined)	Rider Survey(s)	Phase 3	Continuous during testing	Through the performance period
Third-party – Restricted System (vendor to be determined)	Scheduling Application Booking Statistics	Phase 2/3	Monthly	Through the performance period
Third-party – Restricted System (vendor to be determined)	Rider Application User Statistics	Phase 2/3	Monthly	Through the performance period

Data Storage System Type	Dataset Title(s)	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
USDOT-managed – Public System	Transit Services Directories Eligibility Verification Projects Lists Electronic Payment Projects Lists Wayfinding and Accessibility Projects Lists Transit Directory Website Analytics Website Help Desk Exchange Records Rider Survey(s) Rider Application User Statistics Scheduling Application Booking Statistics	Phase 3	Monthly	5 years after the performance period

4.4.1. Data Storage System Description

Each state DOT follows information management policies and procedures set by the authority for information technology within their state's government, including for determining the appropriate data system for any given category of data asset, with oversight and compliance with those policies and procedures also provided by the appropriate personnel and processes determined by the state government (e.g., designees within each DOT such as a Chief Information Officer). As each state is considered the owner of its respective project-related datasets (see Section 4.1), each state will determine the appropriate data system, including those appropriate for data in the public domain or open data, and subsequent details follow from those determinations.

The specific data storage systems that will be used by vendors for project-related data that are not state-owned are unknown at this point in the project's development. The project, however, does expect that vendors will use data systems that meet minimum security, back-up, access, and related standards for work with government partners.

4.4.2. Cybersecurity Polices

At this point, the project does not intend to mandate any project partner to apply one specific cybersecurity policy or framework to the data it will manage. Each state DOT that is a partner on the CALACT ITS4US project will follow information management and cybersecurity policies and procedures set by the authority for its jurisdiction. Some variation across state DOTs in their policies is expected.

The project intends to execute agreements with vendor partners that base terms and conditions on current and emerging standards for privacy and security, such as NIST 800-53, ISO 27001, SAML, and other well-governed, vetted guidelines, and in the spirit of the California Consumer Privacy Act and similar privacy protection laws. To the extent necessary for government partners involved in the project, the project may also ensure software vendors have recently completed a standards-based security audit, such as SOC 2 Type 1 or 2 or can document compliance with industry-based cybersecurity standards.

4.4.3. Data Security Policies and Procedures

Each state DOT follows information management policies and procedures set by the authority for information technology within their state's government, including for the setting of data policies and procedures, with oversight and compliance with those policies and procedures also provided by the appropriate personnel and processes determined by the state's information technology authority (e.g., designees within each DOT such as a Chief Information Officer). As each state is considered the owner of its respective project-related datasets (see Section 4.1), each state will determine the appropriate data security policies and procedures that are to be applied to the relevant data. Each state's data security policies and procedures are generally designed to ensure confidentiality, availability, integrity, and authenticity.

The specific data storage systems that will be used by vendors for project-related data that are not state-owned are unknown at this point in the project's development. The project, however, does expect that vendors will use data systems that meet minimum standards to ensure confidentiality, availability, integrity, and authenticity, and those expectations will be formally conveyed in the appropriate agreements.

4.4.4. Back-up and Recovery Policies and Procedures

Back-up and recovery policies and procedures for state DOT-owned datasets will be set at the direction of each state DOT, following the rules and processes determined by the role, function, or agency responsible for those policies and procedures within the state, such as the department of technology or the Office of the Chief Information Officer.

Additional information on the typical back-up and recovery policies and procedures of software vendors for their relevant datasets will be provided in the DMP as the project progresses in its development through Phases 2 and 3. Back-up and disaster recovery of data are expected and will be stipulated to in the appropriate agreements with vendors.

5. Data Standards

5.1. Data Standards

Table 8 provides details on the data standards used for each dataset named in Table 3. The CALACT ITS4US project shares the perspective provided under the banner of the federal government's open data efforts (resources.data.gov) and uses a broad, multifaceted definition of data standards, including both the concept of a formal, governed data specification that defines one and more data elements and how they can be represented as well as the technical formats for structuring data so that they can be read and interpreted by a computer/machine. While so much of the CALACT ITS4US project is rooted in the extension and large-scale adoptions of formally governed data standards for transit services (e.g., GTFS; GOFS), much less specification or guidance exists for implementing a data vocabulary. Therefore, the project uses technical data standards for structuring project datasets.

Table 8. Data Standards for Datasets

Dataset Title	Data Standard(s)	Data Standard(s) Digital Object Identifier(s) (DOI[s])	Open or Proprietary	Data Standard(s) Rationale
State of California Transit Services Directory	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows. For example, California currently uses the .csv format for its statewide transit providers dataset.

Dataset Title	Data Standard(s)	Data Standard(s) Digital Object Identifier(s) (DOI[s])	Open or Proprietary	Data Standard(s) Rationale
State of Oregon Transit Services Directory	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows.
State of Washington Transit Services Directory	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows.
State of California Eligibility Verification Projects List	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows.

Dataset Title	Data Standard(s)	Data Standard(s) Digital Object Identifier(s) (DOI[s])	Open or Proprietary	Data Standard(s) Rationale
State of Oregon Eligibility Verification Projects List	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows.
State of Washington Eligibility Verification Projects List	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows.
State of California Electronic Payment Projects List	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows.

Dataset Title	Data Standard(s)	Data Standard(s) Digital Object Identifier(s) (DOI[s])	Open or Proprietary	Data Standard(s) Rationale
State of Oregon Electronic Payment Projects List	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows.
State of Washington Electronic Payment Projects List	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows.
State of California Wayfinding and Accessibility Projects List	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows.

Dataset Title	Data Standard(s)	Data Standard(s) Digital Object Identifier(s) (DOI[s])	Open or Proprietary	Data Standard(s) Rationale
State of Oregon Wayfinding and Accessibility Projects List	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows.
State of Washington Wayfinding and Accessibility Projects List	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows.
Website Directory of Transit Services	RFC 4180 (.CSV) or similar	https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml	Open	RFC 4180 (.CSV) is platform, tool, and software independent, allowing data to be read and ingested across many contexts. This standard provides the optimal structure for organizing data expected to take the form of columns and rows. This dataset is a combined subset of the CA, OR, and WA Transit Service Directories.

Dataset Title	Data Standard(s)	Data Standard(s) Digital Object Identifier(s) (DOI[s])	Open or Proprietary	Data Standard(s) Rationale
Transit Directory Website Analytics	Unknown at this time; analytics provider will not be selected until later in Phase 1.			Depends on analytics provider, which has not been selected at this point in Phase I.
Website Help Desk Exchange Records	Unknown at this time; help desk software provider will not be selected until later in Phase I.			Depends on help desk provider, which has not been selected at this point in Phase I.
Rider Survey(s)	Unknown at this time; collection method will not be determined until later in Phase I and that will set in motion more information on what to expect structure and format wise.			Generally expected to conform to a .CSV or similar standard
Scheduling Application Booking Statistics	Unknown at this time; collection method will not be determined until Phase 2 and that will set in motion more information on what to expect structure and format wise.			
Rider Application User Statistics	Unknown at this time; collection method will not be determined until Phase 2 and that will set in motion more information on what to expect structure and format wise.			

5.2. Versioning

Versioning of state-specific datasets (i.e., transit service directories and project lists) will be handled through file versioning, either manually through file name changes (e.g., appending version number to the end of a file name) and revision/change history notes, or a software application that satisfies the same intent will allow the project to track what in the dataset has changed and when the change was made. The same approach will be taken with versioning of the website directory of transit services.

Additional information on the approach to versioning recommended or taken by software will be provided in an update of the DMP.

5.3. Metadata

5.3.1. Metadata Types

The CALACT ITS4US project anticipates that the primary metadata that will be included with project datasets will support discovery, providing descriptive and reference information to potential data users. The metadata in this regard is “business metadata” that is intended to provide information on the content of the data to support potential users in understanding what is included in the datasets and governance of the datasets and access rules. These metadata will describe the datasets in a way that is aligned with existing and popular vocabularies and schema, such as the Data Catalog Vocabulary (DCAT) that is used by federal government agencies. The metadata will include data dictionaries and/or readme files that provide instruction on elements within each dataset to support potential users’ work with the data.

5.3.2. Metadata Structure

Table 9 provides the anticipated structure of the metadata the CALACT ITS4US project will use when it is provided to USDOT.

Table 9. CALACT ITS4US Metadata Structure

Folder	Files	Notes
CALACT ITS4US Project	readme.txt	Readme file includes descriptive information for all project datasets and could include links to external resources, such as Github, as appropriate.
State of California Transit Services Directory	readme.txt datadictionary.csv catransitdirectory.csv	Readme file will include, among other details, license information for the dataset.
State of Oregon Transit Services Directory	readme.txt datadictionary.csv ortransitdirectory.csv	Readme file will include, among other details, license information for the dataset.
State of Washington Transit Services Directory	readme.txt datadictionary.csv watransitdirectory.csv	Readme file will include, among other details, license information for the dataset.

Folder	Files	Notes
State of California Eligibility Verification Projects List	readme.txt datadictionary.csv caeligibilityprojects.csv	Readme file will include, among other details, license information for the dataset.
State of Oregon Eligibility Verification Projects List	readme.txt datadictionary.csv oreligibilityprojects.csv	Readme file will include, among other details, license information for the dataset.
State of Washington Eligibility Verification Projects List	readme.txt datadictionary.csv waeligibilityprojects.csv	Readme file will include, among other details, license information for the dataset.
State of California Electronic Payment Projects List	readme.txt datadictionary.csv capaymentsprojects.csv	Readme file will include, among other details, license information for the dataset.
State of Oregon Electronic Payment Projects List	readme.txt datadictionary.csv orpaymentsprojects.csv	Readme file will include, among other details, license information for the dataset.
State of Washington Electronic Payment Projects List	readme.txt datadictionary.csv wapaymentsprojects.csv	Readme file will include, among other details, license information for the dataset.
State of California Wayfinding and Accessibility Projects List	readme.txt datadictionary.csv caaccessibilityprojects.csv	Readme file will include, among other details, license information for the dataset.
State of Oregon Wayfinding and Accessibility Projects List	readme.txt datadictionary.csv oraccessibilityprojects.csv	Readme file will include, among other details, license information for the dataset.
State of Washington Wayfinding and Accessibility Projects List	readme.txt datadictionary.csv waaccessibilityprojects.csv	Readme file will include, among other details, license information for the dataset.
Website Directory of Transit Services	readme.txt datadictionary.csv threestatetransitdirectory.csv	Readme file will include, among other details such as how the state files have been combined (see Table 7), license information for the dataset.
Transit Directory Website Analytics	To be determined	Given the unknowns at this point around this dataset, the CALACT ITS4US project will wait until it has more information before providing information on metadata content and structure.
Website Help Desk Exchange Records	To be determined	Given the unknowns at this point around this dataset, the CALACT ITS4US project will wait until it has more information before providing information on metadata content and structure.
Rider Survey(s)	To be determined	Given the unknowns at this point around this dataset, the CALACT ITS4US project will wait until it has more information before providing

Folder	Files	Notes
		information on metadata content and structure.
Scheduling Application Booking Statistics	To be determined	Given the unknowns at this point around this dataset, the CALACT ITS4US project will wait until it has more information before providing information on metadata content and structure.
Rider Application User Statistics	To be determined	Given the unknowns at this point around this dataset, the CALACT ITS4US project will wait until it has more information before providing information on metadata content and structure.

5.3.3. Metadata Update Process

The CALACT ITS4US project anticipates that it will transfer datasets and corresponding metadata to the USDOT at the end of the project on a one-time basis. Therefore, the project does not anticipate a process for updating metadata provided to the USDOT.

Appendix A. Acronyms and Glossary

API – Application Programming Interface

CALACT – California Association for Coordinated Transportation

Caltrans – California Department of Transportation

ConOps – Concept of Operations

CSV – Comma-separated values

DCAT – Data Catalogue Vocabulary

Demand-responsive transit – Transit services which provide trips at a location and/or time that is requested by a rider. Generally, any transit service that is not Fixed-route is considered a type of Demand-responsive transit for the purposes of this document, including general public DAR, ADA paratransit, and other transit models.

DMP – Data Management Plan

DOT – Department of Transportation

DSA – Data sharing agreement

GOFS – General On-Demand Feed Specification

GTFS – General Transit Feed Specification

IE – Independent Evaluator

IRB – Institutional Review Board

ISO – International Organization for Standardization

ITS – Intelligence Transportation Systems

JPO – Joint Program Office

LEP – Limited English Proficiency

MOU – Memorandum of Understanding

NIST – National Institute of Standards and Technology

NTD – National Transit Directory

ODOT – Oregon Department of Transportation

PII – Personally Identifiable Information

PMESP – Performance Measurement and Evaluation Strategy Plan

PMO – Project Management Organization

PMT – Project Management Team

SCC – System Coordination Committee

SOC - System and Organization Controls

TNC – Transportation Network Company

USDOT – U.S. Department of Transportation

WSDOT – Washington Department of Transportation

WSTA – Washington State Transit Association

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- JPO-21-866

U.S. Department of Transportation