Appendix B

IMI Data Management Plan

TriMet IMI Demonstration Program Project

STEPS to MOD & MPI: Opening the door for Safe, Total, Equitable, Personalized, and Seamless payment and complete trips for all.

Preliminary Data Management Plan

1. Project Overview

- Project Title: STEPS to Mobility on Demand & Mobility Payment Integration Demonstration Project
- Proposed IMI Area(s) of Inquiry:
    - Mobility on Demand (MOD) Sandbox Demonstration
    - Mobility Payment Integration
- Project Goals and Objectives:
    - Exploring new business approaches and emerging technology solutions that support transformational mobility services;
    - Enabling communities to adopt innovative mobility solutions that enhance transportation efficiency and effectiveness; and
    - Facilitating the widespread deployment of proven mobility solutions that foster expanded personal mobility.
    - Advance the vision of a Complete Trip for All, which reflects understanding that a person's travel comprises a chain that has consequences if broken. Steps begin with:
        - an often-spontaneous decision to make a trip through to planning an itinerary,
        - traversing the built environment and its transportation networks (with or without a vehicle)

- navigating streets,intersections, facilities, stations, and stops to their destination
- safely, efficiently, and carefree.

2. Data Overview

Please provide a description of the nature, scope, and scale of the data that will be collected and/or produced. To the extent possible, please describe the datasets that are anticipated to be collected throughout the research project and any details for these datasets.

Suggested elements for this section include, but are not limited to:

- Data Title: Provide the anticipated data and/or dataset title(s) for the research project.
- Description: Briefly describe the purpose of the data and/or datasets that are anticipated to be collected during the research project.
- Type/Scale: Specify the type and scale of the data that will be generated (e.g., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.) for the research project's data and/or datasets.
- Collection Method: Describe the methods for collecting and/or acquiring the data (e.g., simulated; observed; experimental; software; physical collections; sensors; satellite; enforcement activities; researcher-generated databases, tables, and/or spreadsheets; instrument generated digital data output such as images and video; etc.), and specify by dataset if appropriate.
- Data File Format(s): Provide the data file format(s) the data and/or datasets are anticipated to be in when they are made accessible to the U.S. DOT (e.g. .csv, .txt, json, newline json, etc.). If not yet known, explain to the best of your ability.
- Metadata: Indicate what metadata schema you will use to describe the data and/or dataset in various contexts. If not yet known, explain to the best of your ability. U.S. DOT expects metadata in several contexts, including:
    - At the field level, describing the elements in the data and/or dataset
    - At the asset level, describing the contents of the data or data set generally, in accordance with the Project Open Data Metadata Schema or another appropriate standard

- At any other level that may be appropriate for the data and/or dataset

Data Title	Description	Type/Scale	Collection Method	Data File Format(s)	Metadata	Associated Task(s) Ex: 1.2
Existing Hop Fastpass System Data	Data created and/or logged by the Hop Fastpass fare collection system	Includes but not limited to user account data, transaction records, and other records of customer interaction with the Hop Fastpass System.	See Hop Fastpass System Documentation	Various	See Hop Fastpass System Documentation	1, 3
CAD/AVL Data	Input vehicle location, on-time performance, etc data from TriMet's CAD/AVL system	Transit vehicle location data feed	INIT CAD/AVL system	SIRI data feed	See existing CAD/AVL system documentation	2, 3

Traffic Data	Data about traffic congestion that impacts TriMet bus service reliability	Realtime traffic information via API	Third party traffic data vendor TBD	API	Will be determined in coordination with selected traffic data provider	2, 3
Arrival/Departure Predictions	The time when a TriMet transit vehicle will arrive/depart from a stop	Real-time arrival/departure prediction datafeed	Existing TransitTracker prediction engine -TriMet-produced machine learning prediction engine	GTFS realtime feeds -APIs	See existing documentation for TriMet's real-time passenger information	2,3
Vehicle Locations	Real-time locations of TriMet transit vehicles	Real time vehicle location data feed	Existing TriMet traveler information system	GTFS-realtime feed -APIs	See existing documentation for TriMet's real-time passenger information	2, 3
Validated Incentive Program User Data	Data on users and usage of TriMet's pilot incentive program for transit	TBD	TBD	TBD	Will be developed in coordination with Moovel (the provider of the	2, 3

	fares with Validated				Validated platform)	
Transportation Incentive Program User Account Data	Details on user accounts for participants in the City of Portland Transportation Incentive Program	TBD	TBD	TBD	TBD	2, 3
Transportation Incentive Program Evaluation Data	Data on the performance of the City of Portland Transportation Incentive Program	TBD	TBD	TBD	TBD	2, 3
Uber Origin/Destination Dataset	Aggregated origin/destination	Datatype TBD, Scale of TriMet service area, aggregation level TBD	API	TBD	Available via API	3
Evaluation Surveys	Customer surveys for program evaluation	TriMet and PSU will conduct surveys in	In-field	CSV	Predefined	3

		TriMet service area				

3. Data Stewardship

Please provide details around data stewardship. Data stewardship involves proper management of data throughout the data lifecycle. This includes, but is not limited to, maintaining data quality and safeguarding the data.

3.1 Data Owner and Steward

Please define the data owner and data steward for the anticipated data and/or datasets that will be produced as part of the research project. It is acceptable for multiple datasets to have the same data owner and/or data steward.

Suggested elements for this section include, but are not limited to:

- Dataset Title: Please provide the data and/or datasets that will be assigned the designated data owner and/or data steward. You may break out data and/or datasets that have separate data owners and/or data stewards.
- Data Owner: List who the data owner of the data or dataset is. The data owner is the person or organization that has 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. Please note the data owner may be the FTA for FTA-funded research projects.
- Data Steward: List who the data steward for the data or dataset is. The data steward is, at the direction of the data owner, 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.

Dataset Title	Data Owner	Data Steward
Existing Hop Fastpass System Data	TriMet	TriMet
CAD/AVL Data	TriMet	TriMet
Traffic Data	Provider to be determined	Dependent on agreement
Arrival/Departure Predictions	TriMet	TriMet
Vehicle Locations	TriMet	TriMet
Validated Incentive Program User Data	Moovel	Moovel
Transportation Incentive Program User Account Data	PBOT	PBOT
Transportation Incentive Program Evaluation Data	PBOT	PBOT

Aggregated Uber Origin / Destination Data	TriMet, Uber, 3rd party	Dependent on agreement
Evaluation Surveys	TriMet	TriMet

3.2 Access Level

Providing appropriate data access to safeguard data is a key aspect of data stewardship. In accordance with the OPEN Government Data Act, datasets should be made publicly accessible unless specific concerns require the data to have controlled-access. Please detail which data may require controlled-access, reasons for controlling access to data, and plans for safeguarding these data.

Controlled-access is defined as restricting access to certain groups of persons due to data containing personally identifiable information (PII), information that threatens the privacy of an individual or group, information that threatens confidentiality of a person or group, and/or contains confidential business information (CBI).1

Suggested elements for this section include, but are not limited to:

- Access Level: Can all of the data from this project be shared with the public, or is controlled-access required for at least some of the data? Please detail whether the data are:
    - All Public Access: There is no anticipated data requiring controlled-access
    - Some/All Controlled-Access: There is anticipated data in one or more datasets requiring controlled-access (e.g. Personally Identifiable Information [PII] or Confidential Business Information [CBI])
- Dataset Title: Provide the anticipated data/dataset(s) that will be at the specified access level.
- Reasons for Controlled Access: For any data requiring controlled-access, please detail what in the data is anticipated to require controlled-access and why controlled access may be required for those elements. Please provide details on all elements of the data and/or datasets that are anticipated to require controlled- access.

- Safeguarding Methods and Processes: describe the planned methods and processes that will be taken to safeguard the data requiring controlled-access, such as redacting data before providing public access to the dataset. Please elaborate on the specific methods, for example blurring license plates in videos, geofencing sensor data, removing persistent IDs, etc. Any privacy, ethical or confidentiality concerns raised due to data sharing should be described.
- Informed Consent: If your project contains human subject research, please describe how informed consent forms will permit sharing with the research community and whether additional steps, such as an Institutional Review Board (IRB) may be used to protect the privacy and confidentiality.

Dataset Title	**Access Level**	**Reasons for Controlled Access**	**Safeguarding Methods and Processes**	**Informed Consent**
Existing Hop Fastpass System Data	Some Controlled-Access	PII, CBI	Hashed identifiers and/or anonymized/aggregated records	N/A
CAD/AVL Data	Some Controlled-Access	CBI	Aggregation	N/A
Traffic Data	Some Controlled-Access	CBI	TBD, depending on data provider selected	N/A

Arrival/Departure Predictions	All Public Access	N/A	N/A	N/A
Vehicle Locations	All Public Access	N/A	N/A	N/A
Validated Incentive Program User Data	Some Controlled-Access	PII,CBD	TBD, will be determined during project design phase.	N/A
Transportation Incentive Program User Account Data	Some Controlled-Access	PII,CBD	TBD, will be determined during project design phase.	N/A
Transportation Incentive Program Evaluation Data	Some Controlled-Access	PII,CBD	TBD, will be determined during project design phase.	N/A
Uber Origin / Destination Data	Some Controlled-Access	PII, CBI	Hashed identifiers and/or anonymized/aggregated records	N/A
Micro-mobility data	Some Controlled-Access	PII, CBI	TBD, will be determined during project design phase.	N/A

Transit Stop Inventory	All Public Access	N/A	N/A	N/A
Evaluation Surveys	All Public Access	N/A	N/A	N/A

3.3 Re-Use, Redistribution, and Derivative Products Policies

The U.S. DOT is interested in who holds the intellectual property right(s) for the research project data, whether these rights will transfer to a data archive, whether any copyrights apply to the data, and other aspects that apply to the rights to use the data. Please provide details on these elements in this section.

In accordance with the OPEN Government Data Act, to the extent practicable data must be made available through an open license that is available at no cost to the public and with no restrictions on copying, publishing, distributing, transmitting, citing or adapting such asset. In this section, please provide the applicable licenses for each dataset and specify if an open license will be used. If the project is U.S. DOT-funded, an open license should be used for the dataset. If a non-open license is anticipated to be used for this data, then include a justification for why a non-open license will be used.

Suggested elements for this section include, but are not limited to:

- License Used: List the license that is anticipated to be used for this dataset (e.g. Creative Commons – BY version X.0) and provide a URL (or several) with more information on the license. Specify whether the license is an open license or a non-open license. An "'open license' means a legal guarantee that a data asset is made available (A) at no cost to the public and (B) with no restrictions on copying publishing, distributing, transmitting, citing or adapting such asset.3" A non-open license does not adhere to all of the open license criteria.
- Reasons for a Non-Open License: If a non-open license is anticipated to be used for this dataset, provide the reason for using a non-open license.

Dataset	License Used	Reason(s) for Non-Open License
Existing Hop Fastpass System Data	Contract with fare collection system vendor (INIT)	PII, CBI
CAD/AVL Data	Contract with CAD/AVL vendor	CBI
Traffic Data	Dependent on provider	CBI
Arrival/Departure Predictions	Agency terms of use (GTFS-RT)	N/A
Vehicle Locations	Agency terms of use (GTFS-RT)	N/A
Validated Incentive Program User Data	Contract with Moovel	PII, CBI
Transportation Incentive Program User Account Data		PII
Transportation Incentive Program Evaluation Data		PII

Uber Origin / Destination Data	Agreement between TriMet and Uber	PII, CBI
Micro-mobility data	Dependent on provider	PII
Transit Stop Inventory	TriMet Open Data	N/A
Evaluation Surveys	TriMet Open Data (Anonymized Results Only)	PII

1 https://www.its.dot.gov/data/public-access/

2 https://ntl.bts.gov/public-access/creating-data-management-plans-extramural-research

3.4 Data Storage, Archiving and Preservation

Storing and retaining the data is a key part of the data steward's responsibilities to manage, control, and maintain the quality of a data asset throughout the data lifecycle.

In this section, list all of the data storage systems that will be used to store the research project's data, provide details on those data systems, and specify how long the data will be stored in each system. Please note that IMI NOFO recipients "must ensure the appropriate data are accessible to the FTA and/or the public for a minimum of five years after the award period of performance expires."4 The U.S. DOT understands that dates provided in this section are anticipated dates and will be updated after award.

A public access system provides full data access to the public. A controlled-access system restricts access to certain groups of persons due to data containing personally identifiable information (PII), information that threatens privacy of an individual or group, information that threatens confidentiality of a person or group, and/or contains confidential business information (CBI).5 An applicant system is managed by the applicant, a U.S. DOT system is managed by the U.S. DOT, and a third-party system is a system that is managed by a person, group or organization other than the U.S. DOT or the applicant.

Suggested elements for this section include, but are not limited to:

- Data Storage System Name: Provide the name of the anticipated data storage system(s) the dataset(s) will be stored in. Provide a URL/URLs or other means of accessing the data storage system.
- Data Storage System(s) Type: For each data storage system, state whether the data storage system is:
    - "Applicant – Public System"
    - "Applicant – Controlled System"
        - If you are using an applicant-managed system, please state whether the system is currently operational, under development, or will be developed as part of this project.
    - "U.S. DOT-managed – Public System"

    - "U.S. DOT-managed – Controlled-Access System"
    - "Third Party – Public System"
    - "Third Party – Controlled-Access System"
- Data Storage System Description: Provide a description of the data storage system, including any URLs or other means that provide additional details and documentation of the system; details on data backup of the system; details on access and controlled access, if applicable. If using an applicant or third-party data storage system, please provide additional details that demonstrate the system adheres to the U.S. DOT's policies for data storage systems, including:
    - Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan
    - For Controlled-Access Systems, describe how the system will enable privacy protection, controlled access and collaboration at least equal to U.S. DOT's controlled-access system, the U.S. DOT Secure Data Commons
- Cybersecurity Policies: State any cybersecurity policies that apply to the data storage system or the entirety of this DMP.
- Dataset Title(s): Provide all data and/or datasets that are anticipated to be stored in each data storage system.
- Initial Storage Date: Provide the anticipated initial date that data will be available in each data storage system. If you are live-streaming data, this may be the initial ingest date. You can write this date as a period after award (e.g. three months after award, six months after award, etc.)
- Frequency of Update: Please provide how frequently the data will be updated in the data storage system once ingestion begins. Please note the U.S. DOT prefers obtaining the data as quickly as possible after collection, preferably in near real-time and/or through automated uploads. This is in accordance with the OPEN Government Data Act's requirement to maximize accessibility to federally-funded datasets.
- Archiving and Preservation Period: Provide the duration for which the data and/or dataset will be maintained in each data storage system.

Data Storage System Name	Data Storage System(s) Type	Data Storage System Description	Cybersecurity Policies	Dataset Title(s)	Initial Storage Date	Frequency of Update	Archiving and Preservation Period
Portland Urban Data Lake (PUDL)	Applicant - Controlled System	NoSQL data lake for transportation and other municipal data in the Portland region. Open data and controlled access data depending on the specific dataset	TBD as part of design phase of the project	Arrival/ Departure Predictions	TBD as part of design phase of the project	TBD as part of design phase of the project	TBD as part of design phase of the project

3 https://www.congress.gov/bill/115th-congress/house-bill/1770/text

4 https://www.grants.gov/web/grants/view-opportunity.html?oppId=315743

[5] https://www.its.dot.gov/data/public-access/