Plan Overview

A Data Management Plan created using DMP Tool

DMP ID: https://doi.org/10.48321/D1A2A9CFDF

Title: Wheels on the Bus – Real-Time Data (RTD)

Creator: Ian Lin - **ORCID:** 0009-0000-7355-8097

Affiliation: United States Department of Transportation (DOT) (transportation.gov)

Principal Investigator: Lauren Ledbetter, Adam Burger, Ian Lin

Data Manager: Lauren Ledbetter, Adam Burger, Ian Lin

Project Administrator: Ian Lin

Funder: United States Department of Transportation (DOT) (transportation.gov)

Funding opportunity number: NOFO #20.941

Grant: https://www.transportation.gov/grants/smart/2022

Template: SMART Grants Stage 1 Data Management Plan (DMP)

Project abstract:

The Wheels on the Bus – Real Time Data Project explores whether occupancy information for front-mounted bicycle racks on buses and occupancy information for internal priority seating areas can be provided in real-time to riders. If that information can be provided, transit riders who travel with bicycles or by use of wheelchairs or mobility devices may be able to make informed decisions about whether or when to use transit and avoid being denied boarding because there is no space for their wheels on the bus. A secondary goal is to use time and location-stamped occupancy data to assist VTA in reducing denied boardings and responding to customer concerns in real-time.

The project is a planning/prototyping effort and would not implement any systemwide solutions. Rather, it would explore the technological, operational, logistical, policy and cost challenges of implementing a systemwide solution.

Start date: 09-15-2023

End date: 07-15-2025

Last modified: 09-10-2025

Copyright information:

Wheels on the Bus – Real-Time Data (RTD)

Project Contact Information

Please provide as much of the the following information as possible:

1. **Name of the project;**
2. **Grant number;**
3. **Name of the person submitting this DMP;**
4. **ORCiD of the person submitting this DMP (need an ORCID? Register here: https://orcid.org/);**
5. **Email and phone number of the person submitting this DMP;**
6. **Name of the organization for which the person submitting this DMP is working;**
7. **Email and phone number for the organization;**
8. **Link to organization or project website, if applicable; and,**
9. **Date the DMP was written.**

1. Wheels on the Bus - Real-Time Data (RTD)
2. SMARTFY22N1P1G12
3. Ian Lin
4. 0009-0000-7355-8097
5. ian.lin@vta.org, 408-321-5809
6. Santa Clara Valley Transportation Authority
7. customer.service@vta.org, (408) 321-2300
8. https://www.vta.org/projects/wheels-bus-real-time-data
9. 8/20/2025

Data Description

Please provide as much information as possible:

1. **Provide a description of the data that you will be gathering in the course of your project or data from a third party that you will re-use, if any;**
    1. **If there will be no data collected or re-used from another source, state that this is case;**
        1. **If you answered "No data" above, then you are finished and may skip the most of the steps and sections below, after you:**
            1. **Save your DMP as it exists;**
            2. **Submit it to your Grant Manager or the NTL staff for review.**
2. **Address the expected nature, scope, and scale of the data that will be collected, as best as you can at this stage;**
3. **As best as you can, describe the characteristics of the data, their relationship to other data, and provide sufficient detail so that reviewers will understand any disclosure risks that may apply;**
    1. **If data might be sensitive, please describe how you will protect privacy and security, if you know that now;**
    2. **You may need to update your DMP later to add more detail;**
4. **Discuss the expected value of the data over the long-term.**

The following data sets are available through the Wheels on the Bus Project.

1 Occupancy Status - Sportworks

VTA retrieves data about when and where bicycles are added and removed from the bike rack. The bike rack vendor, Sportworks, provides this occupancy data via APIs available through its Velolink product. The date, time, location, bus route, direction, stop, and occupancy status is available for each use of the bicycle rack.

Sportworks post-processes the raw location data for each bus against the GTFS-RT feed to identify the scheduled block the bus is currently running. Using the RT feed, Sportworks provides the route and bus stop at which the bicycle was added or removed.

As a result of this project, Sportworks provides occupancy data on ADA priority seats as well, indicating wheelchair usage (where type = priority). The data structure is identify for priority seats and bicycle rack occupancy. See the implementation report for more technical detail on how this data is made available.

Sportworks has provided two APIs relevant for this project - Historical and Journeys. Each record in the Historical API provides the state of the racks at a given time stamp. Meanwhile, the 'journeys' API is oriented from the passenger's lens; each record here reflects a passenger journey, showing origin and destination. The identifier for the journey is an anonymized journeyID, obscuring any PII about the passenger, aside from their origin and destination stop.

At one instance, for example, Historical data provided from Sportworks could indicate that at 7am, two bicycle occupied the racks at 2nd & Santa Clara on the 23 toward De Anza College. Suppose the next record is ten stops later, at 7:30am, when there may just be one bicycle occupying the rack. Journeys data would show one record with these two start and end stops for that given journeyID.

2 Vehicle State File - Clever Devices (Priority Seating Area)

One priority seat occupancy technology piloted in the project was through Clever Devices, transmitting data through the vehicle's State File. This data records the date, time, location, bus route, direction, and occupancy status, along with many other parameters about the vehicle state (e.g. odometer, speed, audio files, etc.).

Vehicle state file data showing actual occupancy is only available from the bus on the box. Clever Devices indicated that the state file from the bus running revenue service was not configured to provide occupancy data by the end of the project.

Dataset included records in August 2024.

3 CCTV-Based Priority Seat Occupancy Footage - alwaysAI

As part of a proof-of-concept project supported by grant funding, alwaysAI collaborated with the Santa Clara Valley Transportation Authority (VTA) to develop and deploy a computer vision solution leveraging onboard CCTV footage. The objective was to enhance visibility into onboard bus occupancy trends, specifically focusing on priority seat availability and the types of “wheels” occupying those seats (e.g., wheelchairs, strollers, carts, or bicycles).

The dataset consists of CCTV video captured from VTA’s existing onboard camera infrastructure. These videos were processed using alwaysAI’s proprietary edgeIQ platform, which allows for real-time, edge-based computer vision applications to be deployed and managed directly on the vehicles,

without the need for extensive new hardware. The CCTV-based occupancy system can detect multiple types of wheels on board and recognizing each passenger's journey.

The footage in Stage 1 includes annotation boxes around the priority seat areas, identifying whether there is an "object" currently occupying this space. Stage 2 aims to refine this algorithm, training a model to differentiate between types of wheels.

The videos include recording from November 2024 to June 2025.

4 Wheelchair Deployment Dataset - Clever Devices

This project piloted occupancy technology on just one bus. Because trends on occupancy rely on larger sample size, the Project Team used the thousands of wheelchair-ramp deployments already logged by the CAD/AVL system as a proxy dataset. The ramp is deployed for a variety of reasons (e.g. testing or for any passenger needing better access to the curb). However, it can be used as an approximation of wheelchair usage in the absence of systemwide data. This data was used as a framework and proof of concept; to demonstrate the types of KPIs VTA will be able to track once sensor data is available.

The dataset includes raw ramp-deployment counts, route, block, time of day and total dwelling duration. The only sensitive information this dataset included was operators' names but these were removed.

Dataset included records from years 2023 and 2024.

Data Format and Metadata Standards Employed

Please provide as much information as you can:

1. **Describe the anticipated file formats of your data and related files;**
2. **To the maximum extent practicable, your DMP should address how you will use platform-independent and non-proprietary formats to ensure maximum utility of the data in the future;**
    1. **If you are unable to use platform-independent and non-proprietary formats, you should specify the standards and formats that will be used and the rationale for using those standards and formats.**
3. **Identify the metadata standards you will use to describe the data.**
    1. **At least one metadata file should be a DCAT-US v1.1 (https://resources.data.gov/resources/dcat-us/) .JSON file, the federal standard for data search and discovery.**

Data is available in platform-independent file types such as .csv, .mp4, .mov, and .word.

VTA's Open Data Policy (310.002, dated 3/21/18) sets Open Data as the default at VTA, meaning that VTA practices a phased and incremental approach to proactively share VTA Data publicly on the web in a machine-readable, accessible, and easy-to-use format. The policy establishes an Open Data Committee, which reviews and approves Open Data releases, including ensuring metadata is included. It establishes that VTA General Counsel's Office will review proposed Open Data releases to ensure consistency with applicable laws and to protect against disclosure of data that is confidential, privileged, or otherwise exempt from or inappropriate for disclosure.

The CCTV video data is available in standard, platform-independent formats such as MP4 or AVI, which are widely supported and non-proprietary to ensure future accessibility and usability.

The metadata is a .JSON file in consistent with DCAT-US v1.1.

Access Policies

In general, data from DOT-funded projects must be made publicly accessible. Exceptions to this policy are: data that contain personally identifiable information (PII) that cannot be anonymized; confidential business information; or classified information. Protecting research participants and guarding against the disclosure of identities and/or confidential business information is an essential norm in scientific research. Your DMP should address these issues and outline the efforts you will take to provide informed consent statements to participants, the steps you will take the protect privacy and confidentiality prior to archiving your data, and any additional concerns. In general, in matters of human subject research, your DMP should describe how your 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 privacy and confidentiality. Additionally, when working with, or conducting research that includes Indigenous populations or Tribal communities, researcher will adhere to the CARE Principles for Indigenous Data Governance https://www.gida-global.org/care and make an explicit statement to that effect in this portion of the DMP.

Please provide as much information as possible:

1. **Describe any sensitive data that may be collected or used;**
2. **Describe how you will protect PII or other sensitive data, including IRB review, application of CARE Principles guidelines, or other ethical norms and practices;**
    1. **If you will not be able to deidentify the data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset, you should describe the necessary restrictions on access and use;**
3. **Describe any access restrictions that may apply to your data;**
4. **If necessary, describe any division of responsibilities for stewarding and protecting the data among Principal Investigators or other project staff.**

This project's data collection methodology is in line with VTA's current policies of collecting ridership data by type of transit rider. Cameras use aligns with VTA's current use of internal cameras inside buses for security reasons. Faces are blurred in the footage analyzed through this pilot. As such, no ethical issues are anticipated. Focus groups and surveys helped the Project Team understand end user needs and evaluate how well the pilot meets these needs. Based on the National Institute of Health's online Human Subjects decision tool, we believe these focus groups and surveys are exempt from human subject's regulations (category 2, Exemption 2). The implementation report describes this stakeholder engagement approach.

Sensitive Security Information is information that, if publicly released, would be detrimental to transportation security, and is defined and regulated by 49 Code of Federal Regulations (CFR) parts 15 and 1520 et seq. We do not believe that the data collected for this pilot will meet that definition.

The Project Manager, Adam Burger, is responsible for implementing the Data Management Plan with technical assistance from VTA's Information Technology Department and the consultant team from

WSP. Data collection was a team effort with VTA assuming responsibility of data ownership, metadata production, data quality, storage and backup, data archiving and data sharing.

Re-use, Redistribution, and Derivatives Products Policies

Recipients are reminded:

1. **Data, as a collection of facts, cannot be copyrighted under US copyright law;**
2. **Projects carried out under a US DOT SMART Grants is federally funded; therefore, as stated in grant language:**
    1. **Recipients must comply with the US DOT Public Access Plan, meaning, among other requirements, project data must be shared with the public, either by the researchers or by US DOT;**
    2. **That by accepting US DOT funding through this grant, recipients have granted to US DOT a comprehensive non-exclusive, paid-up, royalty-free copyright license for all project outputs (publications, datasets, software, code, etc.). This includes all rights under copyright, including, but not limited to the rights to copy, distribute, prepare derivative works, and the right to display and/or perform a work in public; and,**
    3. **In accordance with Chapter 18 of Title 35 of the United States Code, also known as the Bayh-Dole Act, where grant recipients elect to retain title to any invention developed under this grant, US DOT retains a statutory nonexclusive, nontransferrable, irrevocable, paid-up license to practice or have practiced for or on behalf of the United States any such invention throughout the world.**

Please provide as much information as possible:

1. **Describe who will hold the intellectual property rights for the data created or used during the project;**
2. **Describe whether you will transfer those rights to a data archive, if appropriate;**
3. **Identify whether any licenses apply to the data;**
    1. **If you will be enforcing terms of use or a requirement for data citation through a license, indicate as much in your DMP;**
4. **Describe any other legal requirements that might need to be addressed.**

All data collected during the SMART Grant project is owned or co-owned by VTA, consistent with typical arrangements, unless explicitly restricted by vendor agreements. Proprietary software or algorithms used to generate the data remain the IP of the vendors.

For large-scale implementation in future phases, real-time data would be made available to the public through VTA's online data portal, which currently hosts real-time arrival data and real-time passenger volume data.

Details on the data described below are available in the data description and dataset README.

1. Intellectual Property Rights

All vendors retain their software IP, while datasets have no such restrictions. VTA has provided raw camera footage, which alwaysAI annotated through their proprietary computer vision algorithms. The

output .mov footage is not IP-restricted.

2. Transfer of Rights to Archive

All datasets (vehicle state file, ramp deployments, CCTV-derived occupancy, bike rack occupancy) will be deposited in a public repository as required by DOT Public Access Plan. Only derived data is shared; the IP for vendor's proprietary software is not transferred.

3. Licensing

Consistent with DOT requirements, all shared datasets will be released under the non-restrictive, royalty-free Creative Commons Attribution 4.0 License.

4. Other Legal Requirements

All datasets comply with federal rules prohibiting copyright on factual data. Any references to software or proprietary tools have been clearly cited but not shared or redistributed. Additionally, no personally identifiable information (PII) is present; footage is entirely anonymized.

Archiving and Preservation Plan

Please provide as much information as possible:

1. **State where you intend to archive your data and why you have chosen that particular option;**
2. **Provide a link to the repository;**
3. **You must describe the dataset that is being archived with a minimum amount of metadata that ensures its discoverability;**
    1. **Whatever archive option you choose, that archive should support the capture and provision of the US Federal Government DCAT-US Metadata Schema https://resources.data.gov/resources/dcat-us/**
4. **In addition, the archive you choose should support the creation and maintenance of persistent identifiers (e.g., DOIs, handles, etc.) and must provide for maintenance of those identifiers throughout the preservation lifecycle of the data;**
5. **Your plan should address how your archiving and preservation choices meet these requirements.**

Data collected through the project was available through .csv exports. Sportworks is still developing its API suite to enable agencies to regularly extract and load into their services (cloud or on-prem). Data from Clever Devices continues to reside in their source systems, independent of this project. alwaysAI annotated footage for a sample bus, which is available through this project repository.

VTAs servers are backed up regularly to protect against data loss. VTA's Information Technology Department will assist with data recovery, if necessary.

Per US DOT's public data access plan, VTA has provided a DCAT-US compliant metadata file for the repository along with persistent identifiers (reference VTA's open data policies, Section 4.2).

Data collected during the prototyping and testing phase will be stored a minimum of 7 years, per VTA's document retention policies. Data collected during a wider-implementation would be stored in perpetuity for historical research purposes.

These datasets have the potential to inform other analyses related to priority seating area policy and planning transit service with greater consideration of combination bicycle/bus trips. The former features a complication in that the flexible use of space allows non-disabled riders to use the space, including those traveling with carts or strollers who cannot move to another part of the bus when the space is needed by wheelchair/mobility device users. The latter is informed only by anecdotal accounts of bicycle usage on transit. Solutions like increasing the number of priority seating areas or flexible space areas on the bus as well as increasing transit service levels where bicycle use is frequent may be solutions VTA could explore with this data.
