

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

- [Background](#)
- [Instructions](#)
- [Section A. Access Level & General Information](#)
- [Section B. Primary vs Derived Research Data](#)
- [Section C. Data Ingestion](#)
  - [One-Time Data Ingestion](#)
  - [Batched/Streaming Data Ingestion](#)
- [Section D. File Attachments](#)
  - [Naming of File Attachments](#)
  - [Files to Include](#)

## Background

The following questionnaire supports the Intelligent Transportation Systems (ITS) Joint Program Office's (JPO's) mission of ensuring all metadata and associated data sets are discoverable through the [DataHub microsite](#) and cataloged within the United States Department of Transportation's (USDOT's) Data.json file, regardless of level of controlled access. The information provided will help ensure information is properly categorized, tagged, and reported within DataHub.

This questionnaire was created in adherence with [Open, Public, Electronic, and Necessary \(OPEN\) Government Data Act](#), Title II of Foundations for Evidence-Based Policymaking Act (H.R. 4174) and the [Project Open Data Metadata Schema](#).

## Instructions

The data provider must fill out one metadata questionnaire for each data set they produce. If a data provider's project produces several separate data sets, a metadata questionnaire must be filled out for each data set.

All fields in this questionnaire are required unless otherwise noted. If the data provider believes they are exempt from completing a required field, the data provider must provide a detailed description of their believed exemption in accordance with the requirements of the ITS JPO Data Access Policy. These exemptions must be approved by the data provider's USDOT Project Manager. ITS JPO has the right to deem these exemptions inappropriate and require the data provider to complete all required fields.

# Section A. Access Level & General Information

Section A focuses on questions relating to the access level of the data set and general information about the data set that is required by data.gov and/or ITS JPO.

## Title

Human-readable name of the data set. Must be in plain English and include sufficient detail to facilitate search and discovery.

## Public Access Level

There are three types of public access level:

- **Public:** data set can be made publicly available to the general public without restrictions.
- **Restricted Public:** data set is only available under certain conditions or to certain audiences.
- **Non-Public:** data set can never be made available to members of the public for privacy, security or other reasons as determined by the USDOT.

Please describe the public access level of the data set.

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

**If "Public Access Level" is not "Public", please complete this shaded section.**

For public access level other than "Public", sufficient justification and USDOT approval must be provided in this document. ITS JPO has the right to rescind "Restricted Public" and "Non-Public" access levels and make a data set's access level "Public" if it considers a justification for restricting access to be insufficient or if USDOT approval is not provided.

**Rights / Access Level Justification**

Please provide explanations for the selected "Public Access Level," as well as instructions for how to access the restricted data set and descriptions of who may access the restricted data set. Text, max 255 characters.

**Discoverability**

A data set with restricted access will still be discoverable unless specific justification and USDOT approval is given to make this data set non-discoverable. Please note that discoverability does not provide access to the data, but allows public users to see the data set's title, metadata, and data dictionary. If there is a specific reason for not allowing discoverability for this data set's metadata, please provide a justification here along with the FOIA code that corresponds to this restriction and proof of USDOT Project Manager's approval.

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

Is the data set's metadata discoverable?

If no, please provide justification and proof of Project Manager's approval below:

Is the data set's data dictionary discoverable?

If no, please provide justification and proof of Project Manager's approval below:

### Will the data be hosted in ITS DataHub or third-party systems?

- **ITS DataHub:** DataHub currently connects to several data storage systems, including Secure Data Commons (SDC), National Transportation Library (NTL), data.transportation.gov (DTG), and S3 sandbox bucket. Where the data set resides within DataHub will depend on the data set's type, access level, size, and maturity. If this option is chosen, DataHub will work with the data provider to ingest their data into the appropriate data storage system(s).
- **Third-party systems:** If this option is chosen, the data provider will be responsible for hosting both the data and the data dictionary, but DataHub will work with the data provider to make sure their data set is discoverable through DataHub.

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

**If the data will be hosted in “Third-party systems”, please complete this shaded section.**

We ask that the data provider also hosts the data dictionary when the data will be stored on third-party systems. The hosted data dictionary should include elements that are listed in the [Draft Data Dictionary Template](#). The data provider will be responsible for keeping the data dictionary that is hosted at the link below in sync with the data that can be accessed through the link below.

Please provide a publicly accessible link to the data in the third-party system below:

Please provide a publicly accessible link(s) to the data dictionary below:

Please indicate if there is interest in having a sample of the data integrated into a Socrata data set on [data.transportation.gov](http://data.transportation.gov). If integrated, community members will be able to create Socrata visualizations and filters using the data.

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### Data Provider

The group that provides the data set. For example, Wyoming Department of Transportation (WYDOT).

### Technical Point of Contact

Name and email of the person who can answer technical questions about the data set.

Name:

Email:

### License (If Applicable)

The license or non-license (i.e. Public Domain) status with which the data set or API has been published. See [Open Licenses](#) for more information.

### Copyright Holder

Organization entity with legal rights to release this data set under the above license. USDOT is currently working to finalize an official definition and guideline for this field.

### USDOT Point of Contact

Name and email of the USDOT Project Manager for this data set.

Name:

Email:

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### **Publisher**

The publishing entity that is responsible for making the data set available.

### **Publisher's parent organization (Optional)**

A parent organizational entity described using the same publisher object fields.

### **Bureau Code**

Federal agencies; combined agency and bureau code from OMB Circular A-11, Appendix C (PDF, CSV) in the format of 015:11. Comma delimited list if more than one code.

### **Program Code**

Federal agencies; list the primary program related to this data set, from the Federal Program Inventory. Use the format of 015:001. Comma delimited list if more than one code.

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### Description

Brief, human-readable description (e.g. an abstract) with sufficient detail to enable a user to quickly understand whether the data set is of interest.

**Spatial** (Required If data set was collected in a specific geographical area)

The range of spatial applicability of a data set. Could include a spatial region like a bounding box or a named place.

**Temporal** (Required If data set was collected within a specified time frame)

The range of temporal applicability of a data set (i.e., a start and end date of applicability for the data set).



# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### Frequency

The frequency with which data set is published. Must be an ISO 8601 repeating duration unless this is not possible because the accrual periodicity is completely irregular. The value should not include a start or end date but rather simply express the duration of time between data publishing. Some common examples are listed below. Further examples and documentation [can be found here](#).

- Irregular frequency: “irregular”
- Continuously or streaming: “R/PT1S”
- Daily: “R/P1D”
- Weekly: “R/P1W”
- Every three months: “R/P3M”
- Annually: “R/P1Y”

--

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### Tags

Tags or keywords help users discover data sets. Please include terms that would be used by technical and non-technical users using a search engine to look for this data set and provide the tags in a comma delimited list.

Tags should follow this format and be **lower case** with **acronyms spelled out**:

- **Required for all data sets:** intelligent transportation systems (its)
- **Required for all data sets:** its joint program office (jpo)
- The project name written out and with acronym (e.g. multi modal intelligent transportation signal system (mmitss))
- The type of the data written out and with acronym (e.g. basic safety message (bsm))
- The microsite category name it fits (e.g. connected vehicle message, application message, trajectories, field test, sensor data, research results, connected equipment, weather)
- The location where the data was collected (e.g. seattle, washington)
- The type of facility (e.g. arterial, freeway, transit (bus), freight)
- Any DOT data program/data management area it fits (see [DOT data program/data management area inventory list](#))
- The author(s) of the data written out and with acronym (e.g. wyoming department of transportation (wydot))
- If application related data, the name of the application and acronym
- If the data was collected or processed by connected equipment, the name of equipment and acronym (ex. roadside equipment (rse))

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### TRT Terms

Terms from [Transportation Research Thesaurus](#) (TRT) that describe the data set. Comma delimited list if more than one term.

### Category (Optional)

Categories that describe the data set, including but not limited to International Organization for Standards (ISO) Topic Categories. Comma delimited list if more than one category.

### Language (Optional)

The language of the data set. Comma-delimited list if more than one language.

### Homepage URL (Optional)

This field is not intended for an agency's homepage (e.g. [www.agency.gov](#)), but rather if a data set has a human-friendly hub or landing page that users can be directed to for all resources tied to the data set.

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### Reference Links

Any reports or external resources that the data set should link to. Please list one resource per line. For each resource, start with the name of the resource, followed by colon and a link to the resource.

For example:

*J2945 Standard: [https://www.sae.org/standards/content/j2945/1\\_201603/](https://www.sae.org/standards/content/j2945/1_201603/)*

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### Unique Identifiers

All unique identifiers for this data set that identify it in a discoverable research database or other web location. Identifiers that must be listed include, but are not limited to:

- Primary IT Investment Unique Investment Identifier (UII)
- Open Researcher and Contributor ID (ORCID) numbers
- The creation and maintenance of a research project record in the Transportation Research Board's (TRB) Research in Progress (RiP) database
- Digital Object Identifiers (DOI)

Please list the applicable unique identifiers in a comma delimited list below. For each identifier, start with the acronym of the type of identifier, followed by colon and the actual unique identifier.

For example:

*ISBN: xxxxx, DOI: xxxxx*

### Data Quality (Optional)

Does the data set meet the agency's Information Quality Guidelines?

## Section B. Primary vs Derived Research Data

Section B focuses on whether the data set is primary or derived research data, and collects additional metadata information if the data set is derived research data.

**Primary Research Data:** data that is collected during pilot deployment, research, or testbed, data generated from simulations, or larger data sets. (e.g. SHRP 2, ICM, CV Pilots).

**Derived Research Data:** data that is final research results as a result of thorough analysis, such as evaluation data. Useful in context with a final research report or analysis documenting methodology. This data tends to be from smaller data sets. (e.g. SHRP2 Subprojects, AMS)

Is the data primary research data or derived research data?

**If the data is “Derived Research Data”, please complete this shaded section.**

**Contract Number** (Required if applicable)

Contract, grant, or other agreement number associated with the data set. This field is required if the data is from extramural research, meaning research undertaken as the result of a grant, contract, or cooperative agreement to an outside institution or individual, partially or fully funded by the USDOT. Comma delimited list if more than one number.

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### ISBN (Optional)

The International Standard Book Number for associated documentation.

### Agency/Contractor Name(s)

Federal agency name(s) or contractor name(s) (e.g. company or research institution) responsible for collecting and/or publishing the data set. Comma delimited list if more than one name.

### Version

Edition or version number of the data for static data updates.

### Personal Contributor

Names of researchers from the project who contributed to data generation, documentation, and sharing. Comma delimited list if more than one name.

### Corporate Contributor (Optional)

Company name from the project who contributed to data generation, documentation, and sharing (if not done by individual researchers). Comma delimited list if more than one name.

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### Publication Date

Date of formal publication in YYYY-MM-DD format.

### Copyright Date (Optional)

Copyright date of the resource in YYYY-MM-DD format.



### Section C. Data Ingestion

Section C focuses on information that is needed to begin ingestion of the data.

**Section C can be skipped if both conditions are met:**

- If the data will be hosted in third-party systems and not in ITS DataHub and/or Secure Data Commons (SDC) **AND**
- If there is no interest in getting a sample of the data into Socrata

Is the data a one-time data upload, continuous data stream, or scheduled batch uploads?

#### Data Size Estimate

Please provide an estimate on the size of the data. Please provide an **overall estimate** if it is a one-time data upload, a **per day estimate** if it is a continuous data stream, or a **per batch estimate** if it is a scheduled batch upload.

Data Volume (number of records):

Data Size:

#### Additional Technical Contact

Technical contact we can work with to set up the ingestion pipeline, if different than the Technical Point of Contact provided in Section A.

Name:

Email:

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

Is the data columnar and flat (e.g. CSV) or does the data contain nested complex objects (e.g. nested JSON objects)?

### One-Time Data Ingestion

If there are multiple tables in this data set, please describe how the data set should be uploaded. If there is only one table, write "Not Applicable". Some examples of how DataHub had handled multiple tables are below:

- Load a join of tables X and Y into the interactive Socrata table, and add table Z to the Socrata data set as attachment.
- Load table X into the interactive Socrata table, and add tables Y and Z to the Socrata data set as attachments.
- Add tables X, Y, and Z to the Socrata data set as attachments.
- If the tables should be put into separate Socrata data sets, please fill out a separate Metadata Questionnaire for each of the table (e.g. WYDOT CV Pilot has separate data sets for BSM and TIM data).

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### Batched/Streaming Data Ingestion

If the data will be updated periodically through a continuous data stream or scheduled batch uploads, please fill out the following:

When is data submission expected to start? This is the date by which the data pipeline should be ready to deposit data into DataHub's ingestion S3 bucket. Use format YYYY-MM-DD.

## Section D. File Attachments

### Naming of File Attachments

All project reports, data sets, metadata files, zip files, etc., must be named descriptively and consistently, in order to help keep data and publications linked for discovery. File names must **use only lower-case letters**, and **the only punctuations allowed are: underscores between each word or series of numbers, and one period before the file extension**. Files belonging to the same project must be named nearly identically, utilizing the following elements:

- Agency or Organization name or acronym (e.g. its\_jpo)
- Project name
- Year of publication (if final report) or data collection (if data collection is ongoing, leave this out)
- Type of file (whether report, data, data management plan (dmp), metadata file, data dictionary or readme file, etc.)
- Date stamp of the current version, entered as YYYYMMDD (Year Month Day), using only numerals
- File extension (e.g. .pdf, .doc, etc.)

For example:

*its\_jpo\_tampa\_connected\_vehicle\_pilot\_bsm\_2019\_dmp\_20191026.pdf*

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

### Files to Include

- **Data Dictionary** for each of the data files
  - Provide the data dictionary in DataHub's standardized format, following the guidance at [Draft Data Dictionary Template](#).
  - Please make sure that the all possible fields are included in the data dictionary, even if the data pipeline does not currently send it.
- **Data Files**
  - If the data is a one-time data dump, please attach all data files.
  - If the data is a continuous data stream or scheduled batch uploads, please attach representative samples of each data type that will be ingested.
- **Related documents**
  - **Additional technical documentations:** Please include related documents such as technical information about a data set (e.g. description of the sensor types used to collect the data, vehicles involved, weather stations, etc.), developer documentation, documentation on any quality assurance or control mechanisms in place to ensure correctly formatted data, and documentation providing additional context of the data.
  - **Treatment of PII:** If there is personally identifiable information (PII) in the data set, please provide document on methods that will be used to provide a redacted copy to the USDOT. If the plan is to provide a non-redacted version of the data to the Secure Data Commons (SDC), ITS JPO's controlled access secure data storage environment, and a redacted version to a non-controlled access environment, such as the ITS JPO Research Data Sandbox, please also describe that in the document. PII is defined as information that alone or in conjunction with additional information can threaten the privacy or identity of an individual.
  - **Distinction of baseline data:** If data set includes baseline data and experimental data, please provide document that speaks to what indicator might be used to distinguish baseline data from experimental or test data within the data set. If there is no distinction, please describe why in the document.
  - Any other information that may be important for users to know in order to make full use of this data set.

# ITS DataHub – Data Management Plan

## Appendix A – Metadata Questionnaire

Please list all file attachments in the textbox below, with one file per line. For each file, start with the file name, followed by colon and a brief description of the file.

For example:

*its\_jpo\_tampa\_cv\_pilot\_bsm\_2019\_dmp\_20191026.pdf: Data Management Plan*