This questionnaire should not be filled out until the Data Access Center of Excellence (CoE) team asks for it closer to the date of data ingestion.

Background

The following questionnaire supports the Intelligent Transportation Systems (ITS) Joint Program Office's (JPO's) effort in ensuring all associated datasets are catalogued and discoverable through the ITS DataHub. The information provided in this questionnaire will help ensure information is properly categorized, tagged, and reported within ITS DataHub.

This questionnaire was created in adherence with the 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 dataset they produce. If a data provider's project produces several separate datasets, a metadata questionnaire must be filled out for each dataset.

All fields in this questionnaire are required unless otherwise noted. Some fields may have prepopulated fields. Please change these field values if they are not the best values for your dataset. 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 Guidelines. These exemptions must be approved by the data provider's U.S. DOT Project Manager. ITS JPO has the right to deem these exemptions inappropriate and require the data provider to complete all required fields.

Glossary

Definition of key terms mentioned in this metadata questionnaire.

Access Level: the degree to which an asset can be made publicly-available, regardless of whether is is already available. There are three types of access levels:

- **Public**: Asset is or could be made publicly available to all without restrictions.
- **Restricted Public**: Asset is available under certain use restrictions.
- **Non-Public:** Asset is not available to members of the public.

Accessible: data is made available in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched. Formats should be machine-readable (i.e., data are reasonably structured to allow automated processing).

Confidential business information (CBI): trade secrets and commercial or financial information obtained from a person and privileged or confidential.

Controlled Access: restricting 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).

Data Access Center of Excellence (CoE): the Data Access CoE's mission is to promote data access and use across the ITS research portfolio, operationalize existing data access policies and systems, and promote best practices, standards, templates, and data storage and discovery capabilities.

Data Dictionary: a machine readable file that contains the most up to date version of metadata which describes and defines a dataset's elements, fields, or column headings. This metadata describes the meaning and purposes of each data element, provides a list of valid values for each variable, and helps one maintain consistency in how the data is described and categorized.

Data Owner: the person or organization that has the authority, ability, and responsibility to access, create, modify, store, use, share, and protect data. Data owners have the right to delegate these privileges and responsibilities to other parties.

(Glossary continued)

Data Steward: at the direction of the data owner, the person or organization that has delegated the privileges and responsibilities to manage, control, and maintain the quality of a dataset throughout the data lifecycle. The data steward may also apply appropriate protections, restrictions, and other safeguards depending on the nature of the data, subject ot the direction of the data owner.

Derived Research Data: data derived through analysis of other data and shared as part of a final report to communicate the methodology and findings of a research project.

Extramural: research activities undertaken as the result of an award of a grant, contract or cooperative agreement to an outside institution or individual, partially or fully funded by the U.S. DOT.

Intramural: research activities led by federal government employees and funded directly as a part of the U.S. DOT's budget, including salaries, laboratories, technical research centers, and other resources.

ITS DataHub: website maintained by the Data Access CoE that provides a single point of entry to discover available U.S. DOT ITS research data. Currently accessible at https://its.dot.gov/data/.

Metadata: information that describes, explains, locates, or otherwise make it easier to retrieve, use, or manage an information resource.

Personally Identifiable Information (PII): information that alone or in conjunction with additional information can threaten the privacy or identify of an individual.

Primary Research Data: original and/or raw data collected that can be used for further analysis.

Spatial: relating to a certain geographic area.

Temporal: relating to a certain time period.

Third-Party: a person, group, or organization other than the U.S. DOT or the data provider.

Section A. Access Level & General Information

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

Title

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

For example:

- *Next Generation Simulation (NGSIM) Vehicle Trajectories and Supporting Data*
- *Safety Pilot Model Deployment Data*
- *Tampa CV Pilot Basic Safety Message (BSM) Sample*
- *Active Transportation Demand Management (ATDM) Trajectory Level Validation*

SampleTitle

Access Level

The degree to which this dataset can be made publicly-available, regardless of whether it is already available.

Please indicate the access level of the dataset:

Public

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

For an access level other than “Public”, sufficient justification and U.S. DOT approval must be provided in this document. ITS JPO has the right to rescind "Restricted Public" and "Non-Public" access levels and make a dataset's access level "Public" if it considers a justification for restricting access to be insufficient or if U.S. DOT approval is not provided.

Controlled Access Details

Please provide details explaining why an dataset is not "Public," including restrictions based on privacy, security, or other policies. Additionally, please provide instructions for how to access a restricted file. Text limit is 255 characters.

Data Dictionary Access Level

The degree to which the data dictionary of the dataset can be made publicly-available, regardless of the dataset's overall "Access Level."

Please indicate the access level of the dataset's data dictionary:

Public

Please provide details explaining why a dataset's data dictionary is not "Public," including restrictions based on privacy, security, or other policies, and providing proof of U.S. DOT Project Manager's approval in the attachment. Additionally, please provide instructions for how to access a restricted data dictionary file. Text limit is 255 characters. *Required if data dictionary's access level is not "Public."*

Will the dataset be hosted in system managed by U.S. DOT, data provider, or third-party?

- **U.S. DOT-managed:** This currently includes Secure Data Commons (SDC), National Transportation Library (NTL), data.transportation.gov (DTG), datahub.transportation.gov (DataHub), and ITS Sandbox. Where the dataset resides will depend on the dataset's type, access level, size, maturity, and target use case. If the dataset will be hosted in a U.S. DOT-managed system, the Data Access CoE team will work with the data provider to ingest their dataset into the appropriate data storage system(s).

- **Data provider or third-party-managed:** If the dataset will be hosted in a data storage system managed by the data provider or by a third-party, the data provider will be responsible for hosting both the dataset and the data dictionary, but the Data Access CoE team will work with the data provider to make sure their dataset is discoverable through ITS DataHub.

If the dataset will not be hosted in "U.S. DOT-managed systems," please complete this shaded section.

Please provide one or more publicly accessible link(s) to the dataset in the third-party system. *Required if the dataset's "Access Level" is "Public" or "Restricted Public."*

- **Download URL:** URL providing direct access to a downloadable file of a dataset.

- **Access URL:** URL providing indirect access to a dataset, for example, via API or a graphical interface.

It is required that the data provider also hosts the data dictionary when the dataset will be hosted in applicant or third-party managed 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(s) below in sync with the dataset that can be accessed through the link(s) above.

Please provide link(s) to the data dictionary.

Please provide the machine-readable file format (IANA Media Type, also known as MIME Type) of the dataset's data dictionary. *Required if this dataset's data dictionary is not an HTML web page.*

Contacts

Contacts for the dataset, including the contact's full name and email. At least one contact must be listed for each of the roles unless specified otherwise. Organizational points of contacts and shared email boxes are acceptable when applicable. Please separate information for each contact with a comma. For each contact, list the full name followed by the contact email, separated by a colon.

For example: *Jane Doe: jane@dot.gov, John Smith: jane@dot.gov*

- **Data Owner:** The person or organization that has the authority, ability, and responsibility to access, create, modify, store, use, share, and protect data. Data owners have the right to delegate these privileges and responsibilities to other parties. Please note, for U.S. DOT-funded projects this may be "U.S. DOT."

  JaneDoe:jane@dot.gov

- **Data Steward:** 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 dataset throughout the data life cycle. The data steward may also apply appropriate protections, restrictions, and other safeguards depending on the nature of the dataset, subject to the direction of the data owner. The data steward is responsible for answering technical questions about the dataset.

  ITSJPO:itsjpo@dot.gov,JoanSmith:joan@dot.gov

- **Federal Sponsor:** The U.S. DOT Project Manager who is assigned to oversee the research project. The federal sponsor's main responsibilities include: coordinating execution of the project with the U.S. DOT-funded project team; receiving, reviewing, and approving the project's Data Management Plan; and coordinating with the U.S. DOT on research project execution and data submission. If project is not federally sponsored research, put "N/A."

  JohnSmith:john@dot.gov

Creator

The person or organization primarily responsible for creating the dataset. In certain cases, this may be the same as the "Contact." For U.S. DOT intramural datasets, please provide the primary modal agency responsible for creating the dataset, not the name of a specific person.

Publisher

The entity responsible for making the dataset accessible, or available in convenient, modifiable, and open formats that can be be retrieved, downloaded, indexed, and searched by the public or designated users.

USDOT

Bureau Code

List the combined agency and bureau code from OMB Circular A-11, Appendix C (PDF, CSV) of the federal bureau that is funding the research project that produced this dataset. Use the format of 015:11. Comma delimited list if more than one code.

021:00

Program Code

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

021:013

License

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

https://creativecommons.org/licenses/by-sa/4.0/

Description

Human-readable description of the dataset (e.g. an abstract) with sufficient detail to enable a user to quickly understand whether the dataset is of interest. Text limit is 4000 characters.

Spatial

The range of spatial applicability of a dataset. Could include a spatial region like a bounding box or a named place. *Required if there is a spatial component to the dataset.*

Temporal

The start and end date of data collection. Formatted as pairs of startDate/endDate, with each date in the format of YYYY-MM-DD. For example: 2018-01-01/2018-12-31. *Required If there is a temporal component to the dataset.*

Frequency

The frequency with which the dataset is published. Must be an ISO 8601 repeating duration unless this is not possible because the frequency is completely irregular. The value should not include a start or end date but rather simply express the duration of time between data publishing. *Required if the dataset is expected to be updated over the life of the project (e.g. published more than once).*

Some common examples are listed below. Further examples and documentation can be found on the ISO 8601 Guidance page.

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

Category

State what applicable categories apply (Railroads, Roadways & Bridges, Pipelines & HAZMAT, Trucking & Motorcoaches, Aviation, Public Transit, Automobiles, Maritime & Waterways, Research & Statistics, Bicycles & Pedestrians), separated by commas.

Automobiles

Tags

Tags (or keywords) help users discover datasets. Please include terms that would be used by technical and non-technical users and keep this at a maximum of 8-12 keywords per dataset, separated by commas.

Tags should follow the format below and be **lower case** with **acronyms spelled out**. If dataset is derived data that will not be used in further analysis, please select only terms from the Transportation Research Thesaurus (TRT).

- 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 categories it fits: 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: arterial, freeway, transit (bus), freight
- 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))

Version

Edition or version number of the dataset. *Required if the dataset is derived research data, but recommended for all datasets.*

Release Date

The date that the dataset is made available through its designated U.S. DOT, data provider, or third-party-managed "Data Storage System" in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched, in YYYY-MM-DD format.

Homepage URL

This field is not intended for an agency's homepage (e.g. www.agency.gov), but rather if a dataset has a human-friendly hub or landing page that users can be directed to for all resources tied to a dataset. *Required if the dataset is derived research data, but recommended for all datasets.*

Identifiers

List any kind of identifiers for the dataset, such as ISBNs, DOIs, ORCIDs, etc. Please list the applicable identifiers in a comma delimited list below. For each identifier, start with the acronym of the type of identifier, followed by colon and the textual string or URL (URI) link of the identifier. For example: ISBN: xxx, DOI: xxxx.

A sample list of identifier types are listed below, with their acronyms in parenthesis.

- Open Researcher and Contributor ID numbers (ORCID)
- Digital Object Identifiers (DOI)
- International Standard Book Number (ISBN)

Award Identifier

Any and all contract, grant, or other fund identifiers associated with the dataset, separated by commas. *Required if dataset is funded by a U.S. DOT contract, grant, or other funding agreement.*

Section B. Data Ingestion

Section B focuses on information that is needed to begin ingestion of the dataset.

Section B can be skipped if both conditions are met:

- **If the dataset will be hosted in third-party or data provider-managed systems and not in a U.S. DOT-managed system AND**
- **If there is no interest in getting a sample of the data into data.transportation.gov (DTG). If integrated with DTG, community members will be able to create visualizations and filters using the dataset.**

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

columnar and flat

Will the dataset be populated by a one-time data upload, continuous data stream, or scheduled batch uploads?

continuous data stream

Data Size Estimates

Please provide estimates on the size of the dataset: an **overall estimate** if populated by a one-time data upload, a **per day estimate** if populated by a continuous data stream, or a **per batch estimate** if populated by scheduled batch uploads.

Data Volume (number of records):

Data Size:

One-Time Data Upload Ingestion

Required if "one-time data upload" was selected.

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

- Load a join of tables X and Y into the interactive data.transportation.gov (DTG) table, and add table Z to the DTG dataset as attachment.
- Load table X into the interactive DTG table, and add tables Y and Z to the DTG dataset as attachments.
- Add tables X, Y, and Z to the DTG dataset as attachments.
- If the tables should be put into separate DTG datasets, please fill out a separate Metadata Questionnaire for each of the tables (e.g. Wyoming Department of Transportation Connected Vehicle Pilot has separate datasets for Basic Safety Messages (BSM) data and Traveler Information Messages (TIM) data).

Batched/Streaming Data Ingestion

Required if "continuous data stream" or "scheduled batch uploads" was selected.

When is data submission expected to start? This is the date by which the data provider should be ready to deposit data into a U.S. DOT-managed system. Use format YYYY-MM-DD.

Section C. File Attachments

Naming of File Attachments

All project reports, datasets, metadata files, zip files, etc. must be named descriptively and consistently in order to 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 and acronym (whether report, data, data management plan (dmp), 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

Files to Include

- **Data Dictionary** for each of the data files
  - Provide the data dictionary in Data Access CoE's standardized format, following the guidance at Draft Data Dictionary Template.
  - Please make sure that the all possible fields that will eventually be sent are included in the data dictionary.
- **Data Files**
  - If the dataset will be populated by a one-time data dump, please attach all data files.
  - If the dataset will be populated by a continuous data stream or scheduled batch uploads, please attach representative samples of each data type that will be ingested for this dataset.
- **Related documents**
  - **Additional technical documentations:** Please include related documents such as technical information about a dataset (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 dataset.
  - **Treatment of PII:** If there is personally identifiable information (PII) in the dataset, please provide documentation on methods that will be used to provide a redacted copy to the U.S. DOT. If the plan is to provide a non-redacted version of the dataset to the SDC, ITS JPO's controlled access secure data storage environment, and a redacted version to a non-controlled access environment, such as the ITS DataHub Sandbox, please also describe that in the document.
  - **Distinction of baseline data:** If the dataset 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 dataset. 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 dataset.

Please list all file attachments in the textbox below and provide URLs, links, attachments, or other means of accessing these files, 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_cvpilot_bsm_2019_dmp_20191026.pdf: Data Management Plan