COMPLETE TRIP

ITS4US

Task 3 Training: Data Management Plan (DMP)
Kate Hartman
Chief – Research, Evaluation, and Program Management
ITS JPO
Agenda

- Briefing Purpose and Outcomes
- Brief Program Overview
- Data Management Plan (DMP) Template Walkthrough
  - DMP Overview
  - DMP Sections
- Resources
  - Useful References
  - Stay Connected
Purpose and Outcomes

- **Purpose:**
  - The aim of this presentation is to review the DMP template provided by USDOT to ensure that project teams can properly address all requirements and instructions in drafting their DMPs.

- **Outcomes:**
  - Understanding of how DMP fits in to the broader set of project deliverables.
  - Clear understanding of the sections of the DMP Template.
Brief Program Overview
Complete Trip - ITS4US Deployment Program

- A USDOT Multimodal Deployment effort, led by ITSJPO and supported by OST, FHWA and FTA
- Supports multiple large-scale replicable deployments to address the challenges of planning and executing all segments of a complete trip

**Vision**

Innovative and integrated complete trip deployments to support seamless travel for all users across all modes, regardless of location, income, or disability.
Program Goals

- Spur high-impact integrated Complete Trip deployments nationwide
- Identify needs and challenges by populations
- Develop and deploy mobility solutions that meet user needs
- Measure impact of integrated deployments
- Identify replicable solutions and disseminate lessons learned
Complete Trip Phase 1 Awardees

- University of Washington
  OR, WA, MD

- California Association of Coordinated Transportation
  CA, OR, and WA

- Heart of Iowa Regional Transit Agency
  Dallas County, IA

- ICF
  Buffalo, NY

- Atlanta Regional Commission
  Gwinnett County, GA
Deployment Phases

PHASE 1: Concept Development
- Concept Development for Complete Trip Deployment
- Establish Cohort Roundtables

PHASE 2: Design & Test
- Design, Test and Deploy Complete Trip Solutions
- Evaluation Framework and Planning

PHASE 3: Operate & Evaluate
- Demonstrate Multiple Large-Scale Deployments
- Evaluate Deployments
- Share Data & Lessons Learned

Operations Maintenance
- Sustain operations for a minimum period of five years after the program is completed with no supplementary federal funds

Deployment
- Up to 12 months
- Up to 24 months
- Minimum of 18 months

Post-Deployment
- 5 years
Data Management Plan (DMP) Overview
A Data Management Plan (DMP) is a document that describes the data you expect to acquire or generate during the course of the ITS4US project, how you will manage, describe, analyze, protect, and store those data, and what mechanisms you will use during your project to share and preserve your data.

**Deliverables**

1. Draft Data Management Plan – Kick-Off + 22 weeks (July 26)
2. Final Data Management Plan* - Kick-Off + 26 weeks (Aug. 23)

*508 Compliant Deliverable
Task 3 Overview

- The **Data Management Plan (DMP):**
  - Describes needs related to protecting the privacy of users
  - Assist future researchers and deployers with understanding and using the data
  - Aimed at people with some technical background in data collection and analysis
  - Living document that will be updated several times during the lifecycle of the project

- The Data Management Plan does **not:**
  - Supersede other project documentation: PMP, ConOps, HUA, & SEMP
  - Need to include all of the details that will not be known until Phase 2 or 3
# DMP Schedule

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>User Needs</td>
</tr>
<tr>
<td>Task 2</td>
<td>Concept of Operations</td>
</tr>
<tr>
<td>Task 3</td>
<td>Data Management Plan</td>
</tr>
<tr>
<td>Task 4</td>
<td>Safety Plan</td>
</tr>
<tr>
<td>Task 5</td>
<td>Performance Measurement</td>
</tr>
<tr>
<td>Task 6</td>
<td>System Requirements</td>
</tr>
<tr>
<td>Task 7</td>
<td>Tech Readiness</td>
</tr>
<tr>
<td>Task 8</td>
<td>Human Use Approval</td>
</tr>
<tr>
<td>Task 9</td>
<td>Training Plan</td>
</tr>
<tr>
<td>Task 10</td>
<td>Institutional, Partnership, and Financial Plan</td>
</tr>
<tr>
<td>Task 11</td>
<td>Outreach Plan</td>
</tr>
<tr>
<td>Task 12</td>
<td>SEMP</td>
</tr>
<tr>
<td>Task 13</td>
<td>Deployment Plan</td>
</tr>
<tr>
<td>Task 14</td>
<td>Deployment Readiness Summary</td>
</tr>
</tbody>
</table>
# DMP Major Components

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Summary</td>
<td>Summary of the types and nature, scope and scale of data</td>
</tr>
<tr>
<td>PII Information</td>
<td>Document all PII data elements and how they will be handled during the task</td>
</tr>
<tr>
<td>System(s)</td>
<td>Document the system or systems that will be used for collecting, monitoring and storing data</td>
</tr>
<tr>
<td>Security</td>
<td>Document how the system will provide Security and Privacy controls</td>
</tr>
<tr>
<td>Context Diagram</td>
<td>Add Data Flows to the Context Diagram from the ConOps</td>
</tr>
<tr>
<td>Standards</td>
<td>Document any standards used for collection, storage or transport of data</td>
</tr>
<tr>
<td>Metadata</td>
<td>Provide Metadata to address USDOT needs</td>
</tr>
<tr>
<td>Data License</td>
<td>The data created is covered under a documented license</td>
</tr>
</tbody>
</table>
Data Management Plan Interdependencies

**Inputs**
- Data Needs
  - Task 2: ConOps
- Context Diagram
  - Task 2: ConOps

**Data Management Plan**
- Data Summary
- PII
- Systems
- Security
- Context Diagram
- Standards
- Metadata
- Data License

**Outputs**
**Data Definitions**
- Task 4: Safety Mgmt.
- Task 6: SyRS
- Task 9: Training Plan
- Task 13: ICTD Plan
- Task 14: Dep. Briefing
Deployment Phases and DMP Development

PHASE 1: Concept Development
- Initial assessment of internal and external data format and sources
- Potential PII is determined
- Data management process should be clear
- Data agreements may not be confirmed
- IRB requirements may not be known

PHASE 2: Design & Test
- Sample Data is collected and provided to USDOT
- Data schema and Metadata are defined
- Data Agreements are confirmed
- IRB requirements included
- Systems are fully defined
- Baseline data may be collected

PHASE 3: Operate & Evaluate
- Live Data is collected
- Public Data and Metadata provided to USDOT
- DMP is finalized

Data Management Details

Rough
- Defined
USDOT & Open Data
What is Open Data?

- **Open Data** is data that is freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control.
  - **Technically Open** - Available in a machine-readable non-proprietary standard format
  - **Legally Open** - Explicitly licensed in a way that permits commercial and non-commercial use and re-use without restrictions.
USDOT’s Interest in Open Data

- Allows others to build upon USDOT funded development work
- Provides transparency into development of resources to support applications/software
- Promotes collaboration on development activities
- Facilitates sharing of common code across projects/deployments
- It’s the Law:
  - Foundations for Evidence-Based Policymaking Act of 2018: Title 2 (OPEN Government Data Act)
USDOT Data Sharing

- Data including data provided by partners from the project shall be provided for public access to the data collected by default, unless specific privacy, confidentiality, security, or other valid restrictions are identified and documented to the USDOT.

- Some of the data must be made available to the public at least at an aggregate level or anonymized format.

- Data rights for data generated/created/captured by project partners should be determined and documented early in the process.

- Data must also include proper documentation and metadata.
DMP Sections
Template Sections

1. Introduction
2. Project Overview
3. Data Overview
4. Data Stewardship
5. Data Standards
6. Glossary of Terms
Section 2: Project Overview

- Provides summary information about the research project and its goals, as well as how the data helps achieve USDOT’s research goals.

- Subsections:
  - **2.1: Change control** - Describes plans for modifications and updates to the DMP and Include plans for how changes in any of the data will be logged.
  - **2.2: Relevant sources** - Lists any reference documents or sources with information relevant to data management.
  - **2.3: Data Schedule** - Provides schedule documenting key milestones pertaining to data.
Section 3: Data Overview

- Provides a summary of the data flows at a high-level and documents all the different datasets planned for the system

- Subsections:
  - **3.1: Data Needs Summary** - High-level extension of the ConOps Context Diagram showing data flows
  - **3.2: Data Overview** - Provides a description of the nature, scope, and scale of the data that is collected and/or produced.
Initial Context Diagram [Example]

Original Context Diagram Taken from your ConOps
Data Needs Summary [Example]

- Extends the context diagram at a high-level a providing a summary of the types, nature, scope, and scale of the data expected to flow among the System entities
- Provides a single location for a high-level view for data flows

Examples of Data Flows:

1. Mobility data (travel times, service vehicle locations, mapping data) - PII
2. Transit data (bus/rail stops, next bus/rail timing, fare information)
3. Internal web update information in proprietary format only used by system
3.2: Data Overview

- Provides a description of the nature, scope, and scale of the data that is collected and/or produced.
- Each *unique* dataset should be included in this section.
- Recommended elements for this table:
  - **Dataset Title**
  - **Description** (including purpose, externality, value, and relevance to performance measures)
  - **Type of Data**
  - **Collection Method**
  - **Data File Format(s)**
### 3.2 Data Overview [Example]

<table>
<thead>
<tr>
<th>Dataset Title</th>
<th>Description</th>
<th>Type / Scale</th>
<th>Collection Method</th>
<th>Data File Format(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connected Vehicle (CV) Pilot Basic Safety Message (BSM)</td>
<td>This data consists of Basic Safety Messages (BSMs) generated by participant and public transportation vehicles onboard units (OBU) and transmitted to road-side units (RSU) located throughout the Project Study area. This data will be used for PM to track vehicle to determine difference in travel times and other variables.</td>
<td>Numerical data, text sequences, positional data (e.g., latitude and longitude)</td>
<td>Experimental with sensors placed throughout the test area and on the car collecting daily information.</td>
<td>.csv files</td>
</tr>
</tbody>
</table>
## Changes that Create Unique Datasets

<table>
<thead>
<tr>
<th>Different Data Types but Same Data</th>
<th>Different Aggregation Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User Profile</strong></td>
<td><strong>5-minute Weather</strong></td>
</tr>
<tr>
<td>Account information</td>
<td>Precipitation data;</td>
</tr>
<tr>
<td></td>
<td>Recorded every 5 minutes</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>User Profile</strong></td>
<td><strong>Daily Weather</strong></td>
</tr>
<tr>
<td>Account information; Time</td>
<td>Precipitation data;</td>
</tr>
<tr>
<td>information coded as strings and</td>
<td>Averaged by hour</td>
</tr>
<tr>
<td>not date formats</td>
<td></td>
</tr>
</tbody>
</table>

### Update In Format

<table>
<thead>
<tr>
<th>Warning Log</th>
<th>Updated Warning Log</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warning logs presented to user</td>
<td>Warning logs presented to</td>
</tr>
<tr>
<td></td>
<td>user; Includes new fields</td>
</tr>
</tbody>
</table>

**Notes:**
- Different Data Types but Same Data: User Profile features account information, with time information coded as strings and not date formats.
- Different Aggregation Levels: 5-minute Weather provides precipitation data recorded every 5 minutes, while Daily Weather offers averaged precipitation data by hour.
- Update In Format: Warning logs are presented to the user, with the updated version including new fields.
Section 4: Data Stewardship

- Provides details concerning data stewardship, e.g., maintaining data quality and safeguarding data.

- Subsections:
  - 4.1: Data Owner and Stewardship
  - 4.2: Access Level
  - 4.3: Re-Use, Redistribution, and Derivative Products Policies
  - 4.4: Data Storage and Retention
4.2: Access Level

- Brief summary of different access levels for each of the different datasets, relating back to context diagram where possible.

Subcategories:

- 4.2.1: Private Datasets
- 4.2.2: Access Request
- 4.2.3: Related Tools, Software, and/or Code
- 4.2.4: Relevant Privacy and/or Security Agreements
4.2.1 Private Datasets

Reasons for restricting Data Access:

- **Data contains PII**
  - SSN, Personal Location, Etc..

- **Data contains Confidential Business Information (CBI).**
  - Delivery location for Business
  - 3\textsuperscript{rd} Party Data with licensing that cannot be shared outside of the project

- **Data contains any information that may threaten privacy or security of any individual or group**
  - Location of explosive Material
  - Location of Private religious centers
## Personally Identifiable Information (PII)

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-PII</strong></td>
<td>Traffic count information, general trends on network conditions, date, time, and weather</td>
</tr>
<tr>
<td><strong>Potential PII</strong></td>
<td>Internet cookies, IP addresses, and vehicle characteristics (size, color, and make/model)</td>
</tr>
<tr>
<td><strong>Actual PII</strong></td>
<td>Names, addresses, telephone numbers, and vehicle identification numbers (VIN)</td>
</tr>
<tr>
<td><strong>Locational PII</strong></td>
<td>GPS tracking information (Lat./Long.), roadway video data, video of faces, and in-vehicle video</td>
</tr>
<tr>
<td><strong>Sensitive PII</strong></td>
<td>Medical records/information; Social Security, bank account, and passport numbers</td>
</tr>
</tbody>
</table>
PII Challenges

- **Survey Data**
  - Issue: Can include PII data such as name, home address, etc.
  - Possible Strategy: Get IRB approval and keep data separate from research data.

- **GPS Trajectories**
  - Issue: Trajectories can identify an individual and where they live/work.
  - Possible Strategy: De-identify sensitive locations.

- **Personally Identifiable Information**
  - Issue: Tracking an individual or stealing their identity can be accomplished through stolen PII.
  - Possible Strategy: All data collection needs to be justified and protected.

- **Agreements Covering 3rd Parties’ Data**
  - Issue: It is often unclear how much or at what level a 3rd party's data will be shared for a project.
  - Possible Strategy: Discuss data sharing up front and make sure to have a written agreement with the 3rd party early in the project.
4.3: Re-Use, Redistribution, and Derivative Products Policies

- Must assign open licenses to federally-funded data and custom-developed source code.
  - USDOT recommends Creative Commons Attribution 4.0 International (CC BY 4.0)
    - [https://creativecommons.org/licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)
- Suggested elements include:
  - Dataset title
  - License(s) Used
  - Reasons for Non-Open License, if applicable
4.4: Data Storage and Retention

- List of all data storage systems that will be used to store the project’s data, with details of those systems, and specifying how long the data will be stored in each system.

- Where possible, reference the Data Needs Summary Diagram to provide additional context.

- Subsections:
  - 4.4.1: Storage Systems
  - 4.4.2: Data Storage System Description
  - 4.4.3: Cybersecurity Policies
  - 4.4.4: Data Security Policies and Procedures
  - 4.4.5: Back-up and Recovery Policies and Procedures
# 4.4.1 Storage Systems

Each unique dataset could be stored in different systems, and/or location, and be updated at varying frequencies.

<table>
<thead>
<tr>
<th>Data Storage System Type</th>
<th>Dataset Title(s)</th>
<th>Initial Storage Date</th>
<th>Frequency of Update</th>
<th>Archiving and Preservation Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contractor FedRamp AWS DB System</td>
<td>CV BSM</td>
<td>One month after sample data provided in Phase 2</td>
<td>Continuous during testing</td>
<td>Through POP</td>
</tr>
<tr>
<td>Contractor FedRamp AWS DB System</td>
<td>CV SPaT</td>
<td>Two months after sample data provided in Phase 2</td>
<td>Continuous during testing</td>
<td>Through POP</td>
</tr>
<tr>
<td>U.S. DOT-managed – Public System</td>
<td>CV BSM</td>
<td>Six months after data collection starts</td>
<td>Daily</td>
<td>Five years</td>
</tr>
<tr>
<td>U.S. DOT-managed – Public System</td>
<td>CV SPaT</td>
<td>Three months after data collection starts</td>
<td>Daily</td>
<td>Five years</td>
</tr>
</tbody>
</table>
Security Needs

Confidentiality: Data is not disclosed to unauthorized users or systems

Availability: Data is available, functioning at a required time

Integrity: Data is accurate and consistent to meet the system needs

Authenticity: Data source can be confirmed, and log what has been sent and received
Section 5: Data Standards

- Discusses the standards that will be used for the data, as well as detailing the support documents related to data analysis.

- Subsections:
  - 5.1: Data Standards
  - 5.2: Versioning
  - 5.3: Metadata and Data Dictionary
5.1: Data Standards Introduction

- Provides details on the data standard(s) used for each dataset

- Suggested elements:
  - Dataset Title
  - Data Standard(s)
  - Open or Proprietary
  - Data Standard Rationale
5.2: Versioning

- **Versioning**
  - Outline procedures for version control
  - Document how older data will be updated if required
  - Document how data changes will be recorded for scheduled and unscheduled events
5.3: Metadata Types

- **Business Metadata**
- **Discovery**
- **Licensing**

- **Technical Metadata**
- **Schema**
- **Processing**
- **Impact Log**
- **Static**
Data Management Plan Challenges

- **Ensuring Proper Amount of Data is Collected**
  - **Issue:** Data collection can be disrupted by various items reducing the amount of data collected.
  - **Possible Strategy:** Provide data buffering for both the before and after case data to ensure adequate data is collected. Monitor data processes for changes or disruptions.

- **Ensure Current Data Information is Shared**
  - **Issue:** Sometimes data documentation lags behind collection which can cause issues with analysis and research on the data collected by the project.
  - **Possible Strategy:** Have a set plan for updating the DMP and other data related documentations which includes notification to users working with the data.
Useful References


Note: FIPS PUBS and NIST Special Publications provide invaluable guides for use by state and local governments as well as the private sector, but their use is not mandatory for non-Federal systems.
Stay Connected

For more information please contact:

Elina Zlotchenko Program Manager, ITS JPO
Elina.Zlotchenko@dot.gov

Kate Hartman, ITS JPO
Chief – Research, Evaluation, and Program Management
kate.hartman@dot.gov

Visit the Complete Trip - ITS4US Deployment Program Website and FAQs:
https://its.dot.gov/its4us/
https://www.its.dot.gov/its4us/its4us_faq.htm
Any questions?