



# Evaluating Data Management Plans (DMPs)

Leighton L Christiansen 0000-0002-0534-4268  
Data Curator,  
National Transportation Library,  
Bureau of Transportation Statistics,  
OST-R, US Department of Transportation  
leighton.christiansen@dot.gov  
ntldatacurator@dot.gov

Jesse Long 0000-0002-4962-1380  
Data Curation & Data Management Fellow,  
National Transportation Library,  
Bureau of Transportation Statistics,  
OST-R, US Department of Transportation  
jesse.long.ctr@dot.gov

Zoe Mann 0000-0002-0198-0915  
STIPDG Intern,  
National Transportation Library,  
Bureau of Transportation Statistics,  
OST-R, US Department of Transportation  
zoe.mann.ctr@dot.gov

Researchers' Development of a Data Management Plan: Session 2  
For Federal Aviation Administration  
2021-08-04



# Workshop Schedule

---

**Wednesday, July 28:**

**Workshop 1:**

Data Management Plans (DMPs) for Research Proposals

**Wednesday, July 28 to Monday August 2:**

**Homework:**

Writing Your DMP

**Wednesday, August 4:**

***Workshop 2:***

*Evaluating Data Management Plans (DMPs)*



# Workshop 2 Overview

---

- **Discussion: Writing a DMP was harder than I thought (20 min)**
- **How to Evaluate a DMP (10 min)**
- **Evaluating CFS 2017 DMP (20 min)**
  - **Each Breakout Room, working as a group on assigned section of DMP**
- **Break (10 minutes) (0.50 to 1.00)**
- **Report back (25 min)**
  - **Each Breakout Room (5 minutes each)**
- **DMP Self-Evaluation (25 min)**
- **Break (10 minutes) (1.50 to 2.00)**
- **Volunteers Sharing Results (25 min) (2.25)**
- **Questions & Discussion from Session 1 & 2**



# Discussion: Writing a DMP was harder than I thought

## Participant feedback:

- There were many prompts in the DMP I could not answer. What do I do?
  - That is expected. DMP writing may be a team sport.
  - You may not have all of the answers at proposal writing time, but in order to be a good data manager, you should think through the questions at proposal time.
  - Persistent links may not exist until near end of project, once archive record is created.
- DMP Guidance web pages useful. More examples would be nice.
  - Noted. Site redesign starting Fall 2021
- DMPTool has many templates, but none for my discipline.
  - DMPTool is NOT a DOT-managed system. Has been used for years by other disciplines.
  - But we can craft and upload templates that make sense for DOT research units.
- In DMPTool, how do I...?
  - We will add DMPTool training and tips to Guidance website.
- Others from the live audience?



# How to Evaluate a DMP

- Read over CFS 2017 DMP
- Use supplied checklist to check each section
- Tally scores
- Communicate results

## 1. Data Description:

The **Data Description** section should:

- Include a description of the data that will be gathered during the project;
- Address the nature, scope, and scale of the data to be collected;
- 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; and,
- Discuss value of the data over the long-term.

Data Description Narrative Evaluation Prompts		Explained Fully	Partially Explained	No information	Not applicable
1.01	The DMP names the data, data collection project, or program.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.02	The DMP describes the purpose of the research or data collection.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.03	The DMP describes the data generated in terms of nature and scale (e.g., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.).	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.04	The DMP describe methods for creating the data (e.g., simulated; observed; experimental; software; physical collections; sensors; satellite; enforcement activities; researcher-generated databases, tables, and/or spreadsheets; digital imagery).	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.05	The DMP discusses the period of time for which data will be collected and frequency of update.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
1.06	The DMP describes the relationship between the new data collected for this effort and any existing data also used.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
1.07	The DMP lists potential users of the data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.08	The DMP discusses the potential value the data have over the long-term, not only for USDOT, but also for the public.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
1.09	If the DMP contains a request to not make the data publicly accessible, it explains the rationale for lack of public access.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
1.10	The DMP indicates the party responsible for managing the data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.11	The DMP describes how project leads will check for adherence to this data management plan.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Total of checked boxes for each column, out of 11:		6	1	2	2

### Evaluation questions:

- Did a majority of the prompts rate "Explained Fully"?
- Do you have a complete picture of: What data will be gathered; How much data to expect; Who is responsible for managing data; and, How the data will be managed?

If not, this section may not be sufficiently detailed.

Please note any outstanding questions you have about how the data is described, in Section 6. **Notes and Questions**



# The Art of Evaluating a DMP

Standards Employed Narrative Evaluation Prompts		Explained Fully	Partially Explained	No information	Not applicable
2.01	The DMP lists the format(s) in which the data will be collected, and indicates if they are open or proprietary.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.02	The DMP includes a rationale for using proprietary data formats, if applicable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2.03	The DMP describes how versions of data will be controlled, including version file naming conventions.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.04	The DMP describes how the researchers will document non-standard file formats, if applicable.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

- **Some science; some art**
- **Not all apply**
- **Quantitative tool for qualitative decision**



# Evaluation Exercise

---

- **Evaluate CFS 2017 DMP (20 min)**
  - **Use Sufficiency Eval form (1 note taker per breakout room)**
  - **Each breakout room, working as a group**
    - **Room 1: Section 1**
    - **Room 2: Section 2**
    - **Room 3: Section 3**
    - **Room 4: Section 4**
    - **Room 5: Section 5**
  - **Record consensus answer; note debate**
  - **Some section are longer; you may not finish**



# Commodity Flow Survey 2017 [supporting dataset]

The data from the 2017 CFS for hazardous material shipments are aggregated to these nine classes, as well as their subcategories known as divisions. Data are also shown for selected UN/NA codes.

For the 2017 CFS, 26 Standard Classification of Transported Goods (SCTG) codes were identified as always being hazardous materials. Even if the respondent left the UN/NA code blank, we assigned the shipment to the appropriate UN/NA code. For example, every shipment of gasoline (SCTG 17110) was assigned a UN/NA code of 1203 either by the respondent or during our tabulation process. When an SCTG could have translated to more than one UN/NA code, the shipment was reviewed and the appropriate UN/NA code was chosen.

Please note that because of the industry coverage and shipment definitions of the CFS, certain hazardous materials such as infectious substances or radioactive wastes were not well represented in the CFS data.

The UN classification system has been adopted for worldwide use by the United Nations Committee of Experts on the Transport of Dangerous Goods. The UN system was incorporated into the Federal Code of Regulations.

## Data Management Plan (DMP) for Commodity Flow Survey (CFS) 2017 Dataset

Bureau of Transportation Statistics (BTS)  
U.S. Department of Transportation (USDOT)  
2021-06-01

Persistent link: <https://doi.org/10.21949/1522565>

### 2. Standards Employed:

The data files collected here are saved in the ubiquitous Documentation will include this data management plan. Documentation will also include the variable definitions created alongside the data from 2017.

A Project Open Data Version 1.1 .json metadata file will and that .json file will be uploaded to data.gov and trans Necessary software tools: The file formats found in the opened using any text editor; .xls files, which can be opened software, such as OpenRefine; .json files, which can be which can be opened with PDF readers.

### 3. Access Policies:

These data files are in the public domain, and can be shared information.

### 4. Re-Use, Redistribution, and Derivative Products P

These data are managed by the Bureau of Transportation be re-use without restriction.

Citation of the data is appreciated. Please use the following: U.S. Department of Transportation, Bureau of Transportation Survey (CFS) 2017 [datasets]. <https://doi.org/10.21949/1522565>

### 5. Archiving and Preservation Plans:

The dataset will be archived in the National Transportation (ROSA P). Prior to archiving, the data are stored on the nightly. The US DOT systems are secured from outside

Files in ROSA P are backed up in NTL drives at US DOT

### Recommended Citation:

U.S. Department of Transportation, Bureau of Transportation Statistics. (2020). Commodity Flow Survey (CFS) 2017 [datasets]. <https://doi.org/10.21949/1522565>

### Change log:

2021-06-01: Initial DMP written

### CONTENTS

0. Dataset and Contact Information
1. Data Description
2. Standards Employed
3. Access Policies
4. Re-Use, Redistribution, and Derivative Products Policies
5. Archiving and Preservation Plans
6. Policies Affecting this Data Management Plan

### 0. Dataset and Contact Information

Staff lead: Jesse A. Long  
Staff lead ORCID: <https://orcid.org/0000-0002-4962-1380>  
Contact information: 1200 New Jersey Avenue, SE Washington, DC 20590, E34-471, [jesse.long.ctr@dot.gov](mailto:jesse.long.ctr@dot.gov) or [infodatacurator@dot.gov](mailto:infodatacurator@dot.gov), 202-366-8951  
U.S. Department of Transportation, Bureau of Transportation Statistics, National Transportation Library  
Title of Dataset: Commodity Flow Survey (CFS) 2017 Dataset  
URL: <https://doi.org/10.21949/1522565>  
This is an  initial DMP or a  revised DMP.

### Organizational Contact Information

Name: Commodity Flow Survey  
Institution: Office of Data Development and Standards, Bureau of Transportation Statistics, U.S. Department of Transportation  
Address: 1200 New Jersey Ave SE, Washington D.C. 20590  
Email: [cfs@dot.gov](mailto:cfs@dot.gov)

### 1. Data Description:

#### General:

The Commodity Flow Survey (CFS) is a joint effort by the Bureau of Transportation Statistics (BTS) and the U.S. Census Bureau, U.S. Department of Commerce. The survey is the primary source of national and state-level data on domestic freight shipments by establishments in mining, manufacturing, wholesale, auxiliaries,

### 1. Data Description:

The Data Description section should:

- Include a description of the data that will be gathered during the project;
- Address the nature, scope, and scale of the data to be collected;
- 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; and,
- Discuss value of the data over the long-term.

### Data Description Narrative Evaluation Prompts

		Explained Fully	Partially Explained	No Information	Not Applicable
1.01	The DMP names the data, data collection project, or program.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.02	The DMP describes the purpose of the research or data collection.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.03	The DMP describes the data generated in terms of nature and scale (e.g., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.).	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.04	The DMP describe methods for creating the data (e.g., simulated; observed; experimental; software; physical collections; sensors; satellite; enforcement activities; researcher-generated databases, tables, and/or spreadsheets; digital imagery).	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.05	The DMP discusses the period of time for which data will be collected and frequency of update.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.06	The DMP describes the relationship between the new data collected for this effort and any existing data also used.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
1.07	The DMP lists potential users of the data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.08	The DMP discusses the potential value the data have over the long-term, not only for USDOT, but also for the public.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.09	If the DMP contains a request to not make the data publicly accessible, it explains the rationale for lack of public access.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
1.10	The DMP indicates the party responsible for managing the data.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.11	The DMP describes how project leads will check for adherence to this data management plan.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Total of checked boxes for each column, out of 11:		7	1	2	1

### Evaluation questions:

- Did a majority of the prompts rate "Explained Fully"?
- Do you have a complete picture of: What data will be gathered; How much data to expect; Who is responsible for managing data; and, How the data will be managed?

If not, this section may not be sufficiently detailed.

Please note any outstanding questions you have about how the data is described, in Section 6. Notes and Questions.





---

**Break: 10 Min**



# Report Back

---

## CFS 2017

- **Room 1: Section 1**
- **Room 2: Section 2**
- **Room 3: Section 3**
- **Room 4: Section 4**
- **Room 5: Section 5**



# CFS 2017: 1) Data Description

The Commodity Flow Survey (CFS) is a joint effort by the Bureau of Transportation Statistics (BTS) and the U.S. Census Bureau, U.S. Department of Commerce. The survey is the primary source of national and state-level data on domestic freight shipments by establishments in mining, manufacturing, wholesale, auxiliaries, and selected retail and services trade industries located in the 50 states and the District of Columbia. Data are provided on the type, origin and destination, value, weight, modes of transportation, distance shipped, and ton-miles of commodities shipped. The CFS is conducted every 5 years as part of the economic census. It provides a modal picture of national freight flows and represents the only publicly available source of commodity flow data for the highway mode. The CFS was conducted in 1993, 1997, 2002, 2007, 2012, and most recently in 2017.

The CFS assesses the demand for transportation facilities and services, energy use, and safety risk and environmental concerns. CFS data are used by policy makers and transportation planners in various federal, state, and local agencies. Additionally, business owners, private researchers, and analysts use the CFS data for analyzing trends in the movement of goods, mapping spatial patterns of commodity and vehicle flows, forecasting demands for the movement of goods, and determining needs for associated infrastructure and equipment.

The CFS publication provides data by shipment characteristics of commodities transported in the U.S. The publication data series include the geographic area series, temperature control series, exports series, and hazardous materials series.

Data Description Narrative Evaluation Prompts		Explained Fully	Partially Explained	No Information	Not Applicable
1.01	The DMP names the data, data collection project, or program.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.02	The DMP describes the purpose of the research or data collection.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.03	The DMP describes the data generated in terms of nature and scale (e.g., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.).	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.04	The DMP describe methods for creating the data (e.g., simulated; observed; experimental; software; physical collections; sensors; satellite; enforcement activities; researcher-generated databases, tables, and/or spreadsheets; digital imagery).	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.05	The DMP discusses the <u>period of time</u> for which data will be collected and frequency of update.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.06	The DMP describes the relationship between the new data collected for this effort and any existing data also used.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
1.07	The DMP lists potential users of the data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.08	The DMP discusses the potential value the data have over the long-term, not only for USDOT, but also for the public.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.09	If the DMP contains a request to not make the data publicly accessible, it explains the rationale for lack of public access.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
1.10	The DMP indicates the party responsible for managing the data.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.11	The DMP describes how project leads will check for adherence to this data management plan.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Total of checked boxes for each column, out of 11:		7	1	2	1

# CFS 2017: 2) Standards Employed

The data files collected here are saved in the ubiquitous and common .csv file format. Documentation will include this data management plan, and the metadata and readme files created in 2021. Documentation will also include the variable definitions, tables, data dictionary, and relevant supporting files created alongside the data from 2017. A Project Open Data Version 1.1 .json metadata file will be created to describe the archival location of this data, and that .json file will be uploaded to data.gov and transportation.data.gov

Necessary software tools: The file formats found in the zip files include: .txt files and .csv files, which can be opened using any text editor; .xls files, which can be opened with Microsoft Excel, and other free available software, such as OpenRefine; .json files, which can be opened in text editors or xml editors; and, .pdf files which can be opened with PDF readers.

Standards Employed Narrative Evaluation Prompts		Explained Fully	Partially Explained	No Information	Not Applicable
2.01	The DMP lists the format(s) in which the data will be <u>collected</u> , and indicates if they are open or proprietary.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.02	The DMP includes a rationale for using proprietary data formats, if applicable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2.03	The DMP describes how versions of data will be controlled, including version file naming conventions.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2.04	The DMP describes how the researchers will document non-standard file formats, if applicable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2.05	The DMP lists the documentation (or metadata) that will be created <u>in order to</u> make the data understandable by other researchers.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.06	The DMP indicates which metadata schema will be used to describe the data. If the metadata schema is not one standard for this field, the DMP discusses the rationale for using that schema.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.07	The DMP describes how will the metadata be managed and stored.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.08	The DMP indicate what tools or software is required to read or view the data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.09	The DMP describes quality control measures employed.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Total of checked boxes for each column, out of 9:		5	0	2	2

# CFS 2017: 3) Access Policies

These data files are in the public domain, and can be shared without restriction. The data files contain no sensitive information.

Access Policies Narrative Evaluation Prompts		Explained Fully	Partially Explained	No Information	Not Applicable
3.01	The DMP lists the roles data-creation team members have in data management, including any limitations on team member access due to the presence of personal or confidential information.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3.02	The DMP states whether the data can be shared with the public.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.03	The DMP describes what data will be shared, how data files will be shared, and how others will access them.	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3.04	The DMP indicate whether the data contain private or confidential information.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.05	The DMP discusses how researchers will guard against disclosure of identities and/or sensitive information, if applicable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3.06	The DMP list what processes the researchers will follow to provide informed consent to participants, if applicable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3.07	The DMP state the party(ies) responsible for protecting the data.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3.08	The DMP describes what, if any, privacy, ethical, or confidentiality concerns are raised due to data sharing.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3.09	If applicable, the DMP describes how data will be deidentified before sharing.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3.10	The DMP identifies what restrictions on access and use you will place on the data, if applicable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3.11	If applicable, the DMP discusses additional steps, if any, needed to protect privacy and confidentiality.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Total of checked boxes for each column, out of 11:		2	0	2	7

# CFS 2017: 4) Reuse Policies

Re-Use, Redistribution, and Derivative Products Policies Narrative Evaluation Prompts		Explained Fully	Partially Explained	No Information	Not Applicable
4.01	The DMP names the party who has the right to manage the data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.02	The DMP indicates who holds the intellectual property rights to the data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.03	The DMP lists any copyrights to the data, and indicates who owns them, if applicable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
4.04	The DMP discusses any rights be transferred to a data archive.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
4.05	The DMP describes how the data will be licensed for reuse	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Total of checked boxes for each column, out of 5:		2	0	0	3

**These data are managed by the Bureau of Transportation Statistics. The data are in the public domain and may be re-used without restriction.**

Citation of the data is appreciated. Please use the following recommended citation:  
U.S. Department of Transportation, Bureau of Transportation Statistics. (2020).  
Commodity Flow Survey (CFS) 2017 [datasets]. <https://doi.org/10.21949/1522565>

# CFS 2017: 5) Archiving

The dataset will be archived in the National Transportation Library Repository and Open Science Access Portal (ROSA P). ~~Prior to archiving, the data are stored on the secured BTS networks and drives, which are backed up nightly. The US DOT systems are secured from outside users and backed up daily.~~

Files in ROSA P are backed up in NTL drives at US DOT, daily; at the Centers for Disease Control, the repository managing facility, daily; and in Amazon Web Service Cloud servers in Virginia and Oregon daily.

The dataset will be retained in perpetuity.

NTL staff will mint persistent Digital Object Identifiers (DOIs) for each dataset stored in ROSA P. These DOIs will be associated with dataset documentation as soon as they become available for use.

The DOIs associated with this dataset include:  
<https://doi.org/10.21949/1522565>

The assigned DOI resolves to the repository landing page for the “Commodity Flow Survey (CFS) 2017” dataset, so that users may locate associated metadata and supporting files.

ROSA P meets all the criteria outlined on the “Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan” page:  
<https://ntl.bts.gov/publicaccess/evaluatingrepositories.html>

Archiving and Preservation Plans Narrative Evaluation Prompts		Explained Fully	Partially Explained	No Information	Not Applicable
5.01	The DMP discusses how& where the data will be archived.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.02	The DMP indicates the approximate <u>time period</u> between data collection and submission to the archive.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
5.03	The DMP identifies where data will be stored prior to archiving.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
5.04	The DMP describes how back up, disaster recovery, off-site data storage, and other redundant storage strategies will be used to ensure the data's security and integrity.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.05	The DMP describe how data will be protected from accidental or malicious modification or deletion prior to receipt by the archive.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
5.06	The DMP discusses the chosen data archive's policies and practices for back up, disaster recovery, off-site data storage, and other redundant storage strategies to ensure the data's security and integrity for the long-term.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.07	The DMP indicates how long the chosen archive will retain data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.08	The DMP indicates if the chosen archive employs, or allows for the recording of, persistent identifiers linked to the data.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.09	The DMP discusses how the chosen data repository meets the criteria outlined on the <a href="#">Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan</a> page.	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Total of checked boxes for each column, out of 9:		6	0	0	3

# **DMP Self-Evaluation (25 min)**

---

- **Each Attendee, working alone or with co-authors**
  - **No DMP? Use one from**  
[https://rosap.ntl.bts.gov/collection\\_pa\\_dmp](https://rosap.ntl.bts.gov/collection_pa_dmp)
- **Use Sufficiency Eval form**
- **Be objective, but gentle**
- **Get as far as you can**
- **Feedback to NTL on checklist is welcome**



---

**Break: 10 Min**



# **Volunteer Sharing Results (25 min)**

## **Volunteers for:**

- **Section 1**
- **Section 2**
- **Section 3**
- **Section 4**
- **Section 5**



# **Public Access Implementation Working Group (PAIWG)**

---

- **PAIWG is an OST-R RD&T Planning Team Topical Research Working Group (TRWG).**
- **PAIWG special role: helping modes implement Public Access Plan.**
  - **Task forces:**
    - **Public Access**
    - **Publications**
    - **Data Access**
- **PAIWG has a number of FAA participants.**
- **We would like to have you, too.**

# NTL To Do List

---

1. **Archive workshop materials in ROSA P**
2. **Email completed DMP sufficiency evals to FAA attendees**
3. **Gather questions and suggestions**
4. **Send answer through Mike and Anthony**
5. **Update and add Public Access Guidance pages**
6. **Modify DOT DMP Template to match Public Access Plan update and federal shared sections**
7. **Match Template and eval tool**
8. **Plan for next workshop (Data dictionary and README best practices?)**



# Questions & Discussion from Sessions 1 & 2

---

**public.access@dot.gov**



# Questions From Session 1: What is Data?

---

**Question:** Who or what documents determine what qualifies as "Data" explicitly? I see datasets, reports, publications, etc. being listed, but who or what ultimately determines if it's "data"?

**Response:** As defined by the US DOT Public Access Plan <https://doi.org/10.21949/1520559>, "Digital Data Sets," for the purpose of this plan, will be defined as all scientific data collected through research projects funded, either fully or partially, by federal funds awarded through a DOT contract, grant or other agreement or collected by DOT employees. Such scientific data are the digitally recorded factual materials resulting from research that is necessary to validate research findings." This data can be recorded in any number of formats.

As our FAQ <https://doi.org/10.21949/1520567> adds: What constitutes such data will be determined by the community of interest through the process of peer review and program management. This may include, but is not limited to: *data, samples, physical collections, software and models*. In general, your plan should address final research data. This includes recorded factual material commonly accepted in the scientific community as necessary to validate research findings. Final research data do not include laboratory notebooks, partial datasets, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as gels or laboratory specimens.

For more Questions and Answers, see the Public Access Plan FAQ page at <https://doi.org/10.21949/1520567>



# Questions From Session 1: Final Dataset?

---

**Question:** How do we define "final dataset"? Would this be the post-wrangled data with duplicates/invalid data removed? What about outliers? Where in the "data cleaning" process should we be considering the data as required for submission?

**Response:** The final dataset is that dataset that used to support the conclusions and analysis of research project and any research reports, journal articles, etc. This can be the cleaned dataset.

For the purposes of Public Access and the greatest sharing, this data should not contain any sensitive information, such as personally identifiable information, business intellectual property, or data that might compromise national or homeland security.

If the data must contain some sensitive information, then access should be limited. However, federal law requires the greatest possible sharing of non-sensitive data. This might mean the creation of a "Public Use File" to share, while a "limited access version" is kept secure at DOT for more robust research and analysis.

You may also find it useful to keep the raw data. If so, that raw data may need its own data management planning, as it may not be shared in a public archive. It will likely have different preservation needs.

For more Questions and Answers, see the Public Access Plan FAQ page at <https://doi.org/10.21949/1520567>



# Questions From Session 1: Licensing and Rights?

---

**Question Part A:** Do we have a standard data sharing license for our data?

**Response Part A:** By law, the research outputs of U.S. Government employees is in the Public Domain.

If the research is funded by US DOT and carried out by a third party, the DOT retains non-exclusive (joint) copyright of the research outputs.

In order to promote research sharing, we encourage funded researchers to use open licenses, such as the Creative Commons CC-BY Attribution license.

Please see our “Managing Rights” page at <https://doi.org/10.21949/1520564> for more information.

**Question Part B:** Is there legal assistance available for completing the DMP section on Re-use, Redistribution, and Derivative Product Policies?

**Response Part B:** Yes. The DOT Office of General Counsel can help.

We also have training materials such as:

- “Managing Rights” <https://doi.org/10.21949/1520564>
- TRB Webinar: USDOT Public Access Plan and Data Management Primer <https://doi.org/10.21949/1520568>
- U.S. DOT Public Access and Data Management Review <https://doi.org/10.21949/1503909>

For more Questions and Answers, see the Public Access Plan FAQ page at <https://doi.org/10.21949/1520567>





# Questions From Session 1: Training for CORs?

---

**Question:** Is there any similar material or training for Legal or Acquisitions (CORs, COs)?

**Response:** Yes. NTL staff gave a training entitled “How Acquisition Professionals Fit into the USDOT Public Access Plan” at the 7th Annual DOT Acquisition & Financial Assistance Conference in 2018.

This training can be updated and presented as needed.

Please contact [public.access@dot.gov](mailto:public.access@dot.gov) to schedule a training.

For more Questions and Answers, see the Public Access Plan FAQ page at <https://doi.org/10.21949/1520567>



# Questions From Session 1: Lit Reviews?

---

**Question Part A:** Does a narrative review of the scientific literature require a DMP?

**Response:** Probably not, if it is only a narrative. However, you are performing text mining, natural language processing, or other computational actions, which might lead to a machine-readable corpus or database, you may want to consider a DMP.

A DMP for that kind of data can help protect against loss.

Further, sharing a large textual corpus may have interest to other transportation researchers.

**Question Part B:** To extend the question above, if there is a more structured output from a literature review (think a spreadsheet containing citations, keywords, and a 1-line summary) would this be something we should submit as "data" for the sake of DMP/Public Access?

**Response Part B:** Again, it depends on use. Just a spreadsheet bibliography, probably not. However, as an analyzed or analyzable corpus for research trend analysis, then yes.

For more Questions and Answers, see the Public Access Plan FAQ page at <https://doi.org/10.21949/1520567>



# Questions From Session 1: Continuous Data?

---

**Question:** We collect a ton of data monitoring displacement gages over months of testing. If we publish a plot generated via that data, is it expected that all raw data is stored in a public location?

**Response:** Is this research data? The Public Access Plan would only require the sharing of the data used to generate the specific plot as part of a research project or as a research output.

Non-research data should follow the DOT Data Release Policy DOT 1351.34 or any orders or policy that might supersede it.

Your office or program may consider that public access to the continuous data would be in the interest of the DOT and the public. You should contact the DOT Chief Data Officer for that discussion, as it is outside the scope of the Public Access Plan.

For more Questions and Answers, see the Public Access Plan FAQ page at <https://doi.org/10.21949/1520567>

# Questions From Session 1: How soon does data need to be released?

---

**Question:** If you plan to do a follow-up analysis on the data, to what extent is it permissible to delay data release, to avoid having duplication of efforts by others on work already being conducted in the DOT?

**Response:** The DOT Public Access Plan <https://doi.org/10.21949/1520559> Section 4.2 specifically directs DOT employees to the Departmental Data Release Policy: <https://www.transportation.gov/digitalstrategy/policyarchive/Departmental-Data-Release-Policy> Basically, non-classified data should be released as quickly as possible.

For more Questions and Answers, see the Public Access Plan FAQ page at <https://doi.org/10.21949/1520567>



# Questions From Session 1: Limiting Data Access?

---

**Question:** Regarding informed consent - is the “Access Policies” section where we could describe that human subjects data would be shared only with researchers performing research aligning with the original purpose to which consent was given?

**Response:** Yes. As the “Access Policies” guidance section of the **Creating Data Management Plans for Intramural Research** page <https://doi.org/10.21949/1520572> notes, this is the part of a Data Management Plan where any restrictions on data access should be explained. Therefore, you can be in compliance with the DOT Public Access Plan while limiting access to research data.

However, U.S. government policy is to share as much data as possible with the public. Therefore, you may consider creating a “public use data” file, that has all sensitive data anonymized, while still limiting access to sensitive data.

U.S. government policy can be described as making data as open as possible, while still protecting personal, business, and national/homeland security information and data.

For more Questions and Answers, see the Public Access Plan FAQ page at <https://doi.org/10.21949/1520567>

