

Data Management Plan (DMP) for “Commodity Flow Survey (CFS)” dataset.

Office of Survey Programs (OSP),
Bureau of Transportation Statistics (BTS),
U.S. Department of Transportation (USDOT)
October 17, 2018

CONTENTS

- 0. Basic 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. Basic Information

- 0.01 Lead researcher, or lead staff name: Ryan Grube
- 0.02 Lead researcher, or lead staff ORCID or other identifier: 0000-0002-9074-0509
- 0.03 Lead researcher contact information: 1200 New Jersey Avenue, SE Washington, DC 20590, E32-317, ryan.grube@dot.gov, 202-366-0634
- 0.04 Organization: U.S. Department of Transportation, Bureau of Transportation Statistics, Office of Survey Programs
- 0.05 Other researchers: CFS Team Members (Joy Sharp, Julie Parker)
- 0.06 Title of Research Proposal/Project: Commodity Flow Survey (CFS)
- 0.07 URL: <https://www.bts.gov/surveys/commodity-flow-survey/commodity-flow-survey-overview>
- 0.08 This is an ☐ initial DMP or a ☒ revised DMP.

- 1. **Data Description:** The Commodity Flow Survey (CFS), a component of the Economic Census, is conducted every five years by the U.S. Census Bureau in partnership with the U.S. Department of Transportation’s Bureau of Transportation Statistics. The 2012 CFS is the fifth survey in the program that began in 1993. The CFS is a shipper survey of approximately 100,000 establishments from the industries of mining, manufacturing, wholesale trade, auxiliaries (i.e. warehouses and distribution centers), and select retail and service trade industries that ship commodities. Data requested by the CFS includes the type of commodities shipped, their origin and destination, their value and weight, and mode(s) of transport. The CFS provides a comprehensive multimodal picture of national freight flows and represents the only publicly available source of data for the highway mode. Results from the CFS are used to analyze trends in the movement of goods, mapping spatial patterns of commodity and vehicle flows, forecasting demands for the movement of goods, and for guiding management and investment decisions on transportation infrastructure. Researchers, transportation planners, transportation advocacy and non-profit groups, policymakers, state DOT’s, etc. represent some of the data users who benefit from the CFS. Additionally, roughly 70% of the FHWA Freight Analysis Framework (FAF) is derived from CFS data.

The survey year represents the year of data collection. Data is collected from shippers during the survey year on a quarterly basis using a questionnaire. The questionnaire asks for a sample of their

shipments for a specified reporting week within that quarter. The respondents are given the option to report either electronically or through paper form. In past cycles. The data can be accessed electronically through the Census Bureau's data platform (currently American FactFinder but is transitioning to a new platform), through the BTS CFS webpage, and through print publications. The printed reports can be accessed from the CFS webpage via PDF download and the tables in the report can be accessed in Excel, CSV, or HTML format. CFS data, from all cycles, can be accessed under the Data and Reports_<<https://www.bts.gov/topics/commodity-flow-survey-reports>> tab of the CFS webpage.

For the 2012 CFS, a public-use microdata file (PUM) was released for the first time in the history of the survey. This file, accessible in SAS and CSV format, contains approximately 4.5 million shipments obtained from businesses included in the 2012 CFS. For the PUM file, additional disclosure avoidance techniques were applied to protect the confidentiality of the shippers. The file can be accessed from the CFS webpage and is featured through four Tableau data visualizations on the CFS landing page <<https://www.bts.gov/surveys/commodity-flow-survey/commodity-flow-survey-overview>>.

The CFS team in the Office of Survey Programs (OSP) manages the data on the webpage. Ryan Grube is the current team lead for data management of the CFS webpage. The CFS data management plan will be reviewed regularly and updated as needed.

2. **Standards Employed:** The CFS is administered as a mail-out, mail-back survey. The Census Bureau is responsible for the mailing of survey forms (either electronically or via paper mail) and the data collection. The questionnaire is mailed on a quarterly basis during the survey year. Returned questionnaires are received, edited, and processed by the Census Bureau. Respondents have the option to report electronically or by paper. Collected data, the raw data, is stored in a secure environment at the Census Bureau to protect the confidentiality of the data. Only the aggregated data, absent of any establishment identifiers, are released outside the Census Bureau.

For BTS, the final estimates (published, aggregated data) are made available on the CFS webpage as Microsoft Excel spreadsheet and HTML tables. The Methodology <https://www.bts.gov/archive/publications/commodity_flow_survey/methodology_2012> tab on the CFS webpage provides a thorough overview of the survey and the methodology used. The methodology is updated as needed with each new survey while older versions can still be referenced. Additionally, the CFS webpage provides an FAQ page <<https://www.bts.gov/surveys/commodity-flow-survey/commodity-flow-survey-cfs-help-data>> and a Key Terms page <https://www.bts.gov/archive/publications/commodity_flow_survey/def_terms>. Both pages are updated as needed. To view CFS data made available online, users would need an internet connection. If that's not available, CFS print reports can be requested.

Regarding quality control, before each data release the data is closely analyzed and measured against outside data sources for comparability purposes. The webpage is monitored on a regular basis to ensure access to the data, metadata, and supporting materials is not interrupted.

3. **Access Policies:** Access to the raw CFS data (shipment records that identify the establishment) is only permitted by individuals who have Census Bureau Special Sworn Status (SSS). To receive and maintain this status, annual Title 13 and Title 26 training is required. After SSS is received, access to

raw data can only be accessed within a secure environment at the Census Bureau. The aggregated data, the final estimates, can be shared with the public. CFS data can be accessed through the Census Bureau's data platform, through the BTS CFS webpage, and through print publications. The Census Bureau is ultimately responsible for protecting the confidential data that only they receive. Sharing of raw data only occurs at the Census Bureau. Sharing of the confidentiality-protected aggregated data typically occurs through email or a secure FTP channel between Census and BTS.

4. **Re-Use, Redistribution, and Derivative Products Policies:** CFS data, metadata, and supporting materials accessible on the CFS webpage will be managed by the CFS team members (Ryan Grube, Julie Parker, and Joy Sharp). Ryan Grube will be the lead and contact person for this effort. The final dataset is in the public domain and has no copyright restrictions. The dataset will also be preserved in the NTL repository, where it will remain under public domain. The data are under public domain and free for reuse, redistribution, and derivative products.
5. **Archiving and Preservation Plans:** The CFS dataset will be archived in the National Transportation Library Repository and Open Science Access Portal (ROSA P). The dataset will be updated in ROSA P within 30 days of each update. 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 aggregated and anonymized CFS 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. 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>>.
6. **Policies Affecting this Data Management Plan:** This data management plan was created to meet the requirements enumerated in the U.S. Department of Transportation's "Plan to Increase Public Access to the Results of Federally-Funded Scientific Research" Version 1.1 <<https://www.transportation.gov/sites/dot.gov/files/docs/Official%20DOT%20Public%20Access%20Plan%20ver%201.1.pdf> > and guidelines suggested by the DOT Public Access website <<https://ntl.bts.gov/publicaccess/>>, in effect and current as of May 2018.

Change log:

2018-07-25: Initial DMP written

2018-08-07: Revisions sent to Ryan Grube by Laura Farley

2018-08-15: Draft returned from Ryan Grube to Laura Farley