How to Share Publications and Datasets under the USDOT Public Data Access Plan

Leighton L Christiansen 0000-0002-0534-4268
Data Curator, National Transportation Library
ntldatcurator@dot.gov

Laura Farley 0000-0002-9980-6850
Fellow, National Transportation Library
Laura.Farley.ctr@dot.gov

Jacky Hart 0000-0003-4333-7511
Fellow, National Transportation Library
Jacky.Hart.ctr@dot.gov

Presentation to:
FTA
2018-02-07
Agenda

• Data Management and the USDOT Public Data Access Plan
• Writing Data Management Plans
• Monitoring and Updating Data Management Plans
• Questions
Data Management and the USDOT Public Data Access Plan
What is Data Management?

• Data Management\(^1\)
  – deliberate planning, creation, storage, access and preservation of data produced from a given investigation\(^1\)
Managing Data Protects it from Loss

Data Availability Declines at 17% per year

Why manage data?

• Preserve data from loss
• Increase research efficiency
• Improve research integrity
• Increase visibility of your research and profile
• Enable and attract collaboration
• Promote new discoveries
• Meet grant requirements
My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government.

― PRESIDENT OBAMA, 05/21/09

**H.R. 4174**

115th CONGRESS 1st Session

To expand the Government’s use and administration of data to facilitate transparency, effective governance, and innovation, and for other purposes.

**S. 760**

115th CONGRESS 1st Session

[Report No. 115–134]

This Act may be cited as the “Open, Public, Electronic, and Necessary Government Data Act” or the “OPEN Government Data Act”.

[Image of a webpage titled Data.gov]
Increasing Access to the Results of Federally Funded Scientific Research

February 22, 2013

- Departments with > $100 million in yearly R&D must improve public access to federally funded research results
- Applies to researchers both within the federal government and who receive federal funding

http://obamawhitehouse.archives.gov/blog/2016/02/22/increasing-access-results-federally-funded-science
Requirements for Agencies

• Each sponsoring agency’s plan must contain:
  – a strategy for leveraging existing archives, where appropriate;
  – a strategy for improving the public’s ability to locate and access digital data;
  – an approach for optimizing search, archival, and dissemination features, while ensuring long-term stewardship of the results;
  – a plan for notifying awardees and other federally funded scientific researchers of their obligations (e.g., through guidance, conditions of awards, and/or regulatory changes); and
  – an agency strategy for measuring and enforcing compliance with its plan.
DOT Response

- Information drives advances in safety, state of good repair, livable communities, economic competitiveness, and environmental sustainability
- Enhance tracking of the complete research lifecycle at project level
- Ensure public access to Publications and Data Sets from DOT-managed programs
- Scale and institutionalize data sharing practices across DOT

What is Public Access?

- The public is **aware** of the data generated through federally funded scientific research
- The public is able to **download** and **analyze** unclassified publications and data sets
  - Unless precluded by privacy, confidentiality, or security concerns
  - If the information is sensitive, public access may be restricted to subsets of the public
DOT Public Access Plan: Objectives

• Affirm DOT’s commitment to public access and the reproducibility of research
• Ensure continuous access to and reliably preserve DOT-funded research (publications, data, and projects)
• Enhance scientific discovery and deployment of results
• Promote innovation and economic competitiveness
Requirements

• Project records
  – Research Project Record process will be used to connect publications and data sets and as a compliance and reporting mechanism

• Data
  – Submit a Data Management Plan for approval, to include metadata and terms of access and use
  – Deposit final data sets in a repository for public access

• Publications
  – Submit final manuscripts accepted for publication to NTL, the central repository for USDOT research
Exceptions

• Public access to publications and data must honor and protect:
  – confidentiality and personal privacy;
  – proprietary interests and business confidential information;
  – intellectual property rights;
  – national and homeland security; and
  – other exemptions and protections provided by the Freedom of Information Act (FOIA).
DOT Public Access Guidance Pages

http://ntl.bts.gov/public-access

DOT Public Access

The USDOT Public Access Plan is issued in response to the 2013 Office of Science and Technology Policy (OSTP) memorandum, which directed all Executive Departments and Agencies to develop plans for improving the public access to the results of federally funded scientific research. The DOT Public Access Plan is designed to ensure that DOT-generated data and digital data sets arising from DOT-managed research and development (R&D) programs are made available in a timely manner.

This plan establishes objectives to ensure public access to data and digital data sets arising from DOT-managed research and development (R&D) programs. DOT already provides intramural and extramural research in progress and completed research.
What You Need to Know

• Applicants for USDOT research funds must submit:
  – 2-3 page Data Management Plan (with proposal);
  – ORCiDs for each author;
  – Final peer-reviewed manuscript;
  – Any other written outputs (metadata, data documentation, final reports, technical reports, tech summaries, etc. – the Data Package); and
  – Final Digital Datasets or a repository link
## Public Access and Research

### Submit Research Proposal
- Obtain ORCID for each researcher
- Include and obtain approval of DMP

### Perform Research
- Follow data management practices in approved DMP
- Submit research description and updates to TRB’s Research in Progress (RiP) database
- Project records automatically appear in Research Hub (RH)

### Package Research Results
- Archive final dataset(s)
- Include ORCID and funding agreement number with data and on tech report documentation page

### Report Research Results to DOT
- Submit research results (files or URLs) with ORCID, RH Display ID, and funding agreement number
Writing Data Management Plans
DOT DMP: Overview

Data management plan (DMP): Written plan of how digital data will be handled during and after a project.

2-3 page narrative description in 5 sections:
1. Data Description
2. Standards Used
3. Access Policy
4. Re-Use, Redistribution, and Derivative Products Policies
5. Archiving and Preservation Plans
DOT DMP: Start Here

Go To:

http://ntl.bts.gov/publicaccess

- Info. For Researchers
- Creating Data Management Plan
DOT DMP: Extramural vs. Intramural

Extramural Research
• Grant funded
• Contract and/or cooperative agreement

Intramural Research
• Federal employee led
• Funding from DOT budget
  • Includes salaries, laboratories, technical research centers, and others

To find the DMP for your situation, go to: https://ntl.bts.gov/public-access/creating-data-management-plans
DOT DMP Suggestion: 5 Sections

1. Data Description
2. Standards Used
3. Access Policies
4. Re-Use, Redistribution, and Derivative Products Policies
5. Archiving and Preservation Plans

Data creation

Data’s life beyond the original project
DMP Section 1: Data Description

• Nature, scope, and scale of the data gathered in the course of the project
• Characteristics and relationship of the data to other data
• Sufficient detail for reviewers about any disclosure risks
• Long-term value of the data
DMP Section 2: Standards Used

• Anticipated formats of data and related files
• Use of **platform-independent** and **non-proprietary formats**
  - What’s practicable and in accordance with practices in your field?
  - Why platform-independent and non-proprietary formats? Ensures maximum utility of the data in the future
• For any **non-platform-independent** and **proprietary** formats specify the standards, formats, and rationale for use
• Metadata standards used for data
DMP Section 3: Access Policies

• Data from DOT funded research projects must be made publicly accessible
  • Exceptions: personally identifiable information (PII), confidential business information, or classified information
• Restrictions for access and use are acceptable if deidentifying the data cannot be done while maintaining the utility of the dataset
DMP Section 3: Access Policies Continued

• Protection of research participants’ identities and/or confidential business information
  • How informed consent statements will be provided to participants, the steps taken to protect privacy and confidentiality prior to archiving data, and additional concerns like embargo periods for data
  • When applicable describe division of responsibilities for stewarding and protecting the data among PIs or other project staff
• For human subject research describe how the informed consent forms will permit sharing
  • Are additional steps, such as an Institutional Review Board (IRB), required to protect privacy and confidentiality
DMP Section 4: Re-Use, Redistribution...

- Who will hold the intellectual property rights for the data created by the project
- Will those rights be transferred to a data archive (if applicable)
- Does copyright apply to the data,
  - May be the case when using copyrighted instruments
- Indicate any enforcement of terms of use or a requirement for data citation through a license
- Any legal requirements
DMP Section 5: Archiving and Preservation

• How data will be archived and why was that archive was chosen?
  • Archive options include but are not limited to:
    • Use of an institutional repository
    • Use of an archive or other community-accepted data storage facility
    • Self-dissemination
• Datasets must be described with essential metadata to ensure discoverability
• The chosen archive must support:
  • The capture and provision of the US Federal Government Project Open Data Metadata Schema
  • The creation and maintenance of persistent identifiers (e.g., DOIs, handles, etc.) and maintenance of those identifiers throughout the preservation lifecycle
DMP Required Description: Prompts

Need more guidance on what to write?
We’ve got you covered.

Prompts for all 5 DMP sections:
Example DMP

- 2016 National Census of Ferry Operators

- About this Dataset

### Common Core

<table>
<thead>
<tr>
<th>Common Core</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact Name</td>
<td>Janine McFadden</td>
</tr>
<tr>
<td>Temporal Applicability</td>
<td>2015-01-01/2015-12-31</td>
</tr>
<tr>
<td>License</td>
<td>Other</td>
</tr>
<tr>
<td>Program Code</td>
<td>021:053</td>
</tr>
<tr>
<td>Publisher</td>
<td>Bureau of Transportation Statistics</td>
</tr>
<tr>
<td>Is Quality Data</td>
<td>True</td>
</tr>
<tr>
<td>Bureau Code</td>
<td>021:04</td>
</tr>
<tr>
<td>Contact Email</td>
<td><a href="mailto:janine.mcfadden@dot.gov">janine.mcfadden@dot.gov</a></td>
</tr>
<tr>
<td>Last Update</td>
<td>9/29/2017</td>
</tr>
<tr>
<td>Theme</td>
<td>Ferry Transit</td>
</tr>
<tr>
<td>Geographic Coverage</td>
<td>U.S., U.S. Territories, and applicable Canadian providers.</td>
</tr>
<tr>
<td>Access Level Comment</td>
<td>Business-sensitive data has been removed from this dataset.</td>
</tr>
<tr>
<td>Public Access Level</td>
<td>public</td>
</tr>
</tbody>
</table>

### Attachments

- bts_osp_national_census_ferry_operators_2016_DMP_2017_10_26.docx
- bts_osp_national_census_ferry_operators_2016_DMP_2017_10_26.txt
- bts_osp_national_census_ferry_operators_2016_Metadata_2017_10_26.json
Monitoring and Updating Data Management Plans
DMPs are Guides to Action

- Researcher commitment to making data accessible
- Assist and improve data collection and preservation
- Give research team guidance from day one
DMPs are Living Documents

Review DMPs frequently:

• Quarterly report to FTA
• Project Change
• Data Environment Change
• Unexpected Sensitive Data
DMP Version Control

Version Control in Document:
Data Management Plan (DMP) for “National Census of Ferry Operators (NCFO)” dataset.
Office of Survey Programs (OSP),
2017-10-26

Change log:
2017-10-26: Initial DMP written

Version Control in file name:
bts_osp_national_census_ferry_operators_2016_DMP_2017_10_26.txt
The Future of DMPs

• Currently DMPs are document-based
• Move towards web-interface
• Move towards “machine-actionable” DMPs
  – Easier updates
  – Automatic information collection
Review

• Data Management and the USDOT Public Data Access Plan
• Writing Data Management Plans
• Monitoring and Updating Data Management Plans
Questions?

ntldatcurator@dot.gov
Resources


Frequently Asked Questions:

What is DOT's Public Access Plan?

A. The White House Office of Science & Technology Policy's (OSTP's) Feb. 22, 2013 memorandum entitled Increasing Access to the Results of Federally Funded Scientific Research describes new requirements for providing the public access to the publications and digital data sets resulting from federally funded scientific research. USDOT's Public Access Plan is a framework to ensure the Department achieves the memorandum requirements. It covers any written deliverable funded, fully or partially, through a DOT-managed contract, grant, or other funding agreement, including all final and technical reports, and all final peer reviewed manuscripts accepted for publication. It also provides a strategy for improving the public's ability to locate and access the digital data supporting the research, where available.
Frequently Asked Questions:

What is the difference between Public Access and Open Access?

A. DOT's Public Access Plan ensures that the public has access to publications and digital data sets arising from DOT-funded scientific research programs through the National Transportation Library's digital repository. The public is able to freely search, download, and analyze unclassified publications and digital data sets unless specifically precluded by privacy, confidentiality or other security concerns. Open Access is unrestricted access and unrestricted reuse of documents copyrighted under a Creative Commons or similar license-type agreement. DOT's Public Access Plan covers final peer reviewed manuscripts accepted for publication, but not published articles. Because most publishers own the rights to the published articles in their journals, users are required to pay for access and request permission to reuse.
What is scientific research as defined in the DOT Public Access Plan?

- Scientific Research in the plan is activities comprising creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society. Research includes:
  - **Basic research** (without specific application),
  - **Applied research** (for a specific need), and
  - **Developmental research** (design, development, and improvements of prototypes and processes, including demonstration projects and other related activities associated with research and development activities).

- The following activities are NOT Scientific Research under the plan:
  - R&D Facilities (funding for the construction and rehabilitation of research and development facilities. Includes the acquisition, design, and construction of, or major repairs or alterations to, all physical facilities for use in R&D activities); IT Support and funding for IT equipment (hardware/software); Grant support / contract support / program support activities; Administrative activities; Conference / workshop funding; Funding for training / capacity building; Outreach activities; and Travel.
Frequently Asked Questions:
How does the National Transportation Library enable public access to my research results?

A. The National Transportation Library has a MAP-21 (49 USC 6304) mandate to be the central repository for USDOT research and technical reports and a clearinghouse for Government transportation data. NTL maintains an OAI-PMH compliant repository that serves as its central clearinghouse and archive for public access to transportation information. In many cases, publications arising from DOT-funded scientific research are already made freely available to the public through the NTL repository. Under DOT’s Public Access Plan, all final peer reviewed manuscripts accepted for publication will now be archived in the NTL repository. Additionally, when you apply for research funds from the USDOT under a contract, grant, cooperative agreement, or other funding agreement, you will be required to submit a data management plan for approval. Your plan must identify a repository for the preservation of your data that is accessible by NTL. Your dataset's metadata will be included in DOT Enterprise Data Inventory. Through these mechanisms, datasets will be discoverable through data.gov, NTL, internet search engines and other tools leveraging open formats and standards.
Frequently Asked Questions:
What is the USDOT Research Hub?

A. The USDOT Research Hub is a searchable database of USDOT-sponsored research, development, and technology project records. The database acts as a central repository for information on active and recently completed projects from nine USDOT Operating Administrations and the Office of the Secretary, providing a comprehensive account of the Department's research portfolio at the project level.
Frequently Asked Questions:
What is an ORCID and how do I create one?

A. ORCID stands for Open Researcher and Contributor ID. ORCID.org provides a registry of persistent unique identifiers for researchers and scholars and automating linkages to research objects such as publications, grants, and patents. Registration is free and takes about 5 minutes. A quick overview is provided on the by NTL on creating an ORCID. You can find more information about ORCID in the User section at http://support.orcid.org/knowledgebase
Frequently Asked Questions:

What, if any, funding programs are exempt from DOT's Public Access Plan?

A. Federal Aid programs flowing funding to states, such as State Planning & Research (SP&R) and National Cooperative Highway Research Program (NCHRP), as well as Small Business Innovation Research (SBIR) programs are exempt from the requirements of the plan.
Frequently Asked Questions:

SP&R funds are considered “federal funds” for purposes such as the required 80/20 match, auditing, and other federal standards, but not for the Public Access plan. How can they be considered federal funds for some purposes and considered state funds for other purposes?

A. The requirements of the DOT Public Access plan apply to recipients of funds obtained directly from USDOT through grants, contracts, cooperative agreements, or other funding agreements. SP&R funds are a set-aside of Federal-aid funds apportioned to the States. They are not provided to the States through any of the funding mechanisms specified in the DOT Public Access plan. Therefore, for purposes of the plan, they are considered state funds.
Frequently Asked Questions:

When does the "clock start ticking" on the embargo period?

A. Often there is a significant time lag between the final manuscript being provided to the publisher and it actually being published. The "clock starts ticking" when the article is published. The official date of publication is determined by the publisher.
Frequently Asked Questions:

What are the criteria for determining if a journal article was federally funded?

A. If a publication is fully or partially federally funded and is a deliverable under the funding agreement managed by USDOT, then it falls under the purview of the plan.
Frequently Asked Questions:

Are theses and dissertations included in DOT's Public Access Plan?

A. Theses and dissertations would only be included if they met the criteria for journal articles: The work is a specified deliverable of a fully or partially federally funded project.
Frequently Asked Questions:  
What about copyright?

A. As a new term and condition of all DOT funding agreements, the DOT Public Access Plan requires that a license to "all rights under copyright" is granted to DOT for any written deliverables, including the supporting data. The copyright license will be non-exclusive, non-transferrable, and royalty free. Contractor and grantee's rights to other forms of intellectual property will continue to be governed by statute and regulation. The copyright license will provide DOT the same rights as it would otherwise have under its rights in data provisions included in the Federal Acquisition Regulations: the rights to copy, distribute, publicly display and perform, and create derivative works, or to have others do so on behalf of DOT.
What is a data management plan?

A. A data management plan (DMP) is a document created and submitted for approval as part of the research proposal. It describes your proposed plan for how to handle the final dataset generated during your research. DMPs will describe how the research proposal conforms to DOT policy on the dissemination and sharing of research results and will include:

- The final research data to be produced in the course of the project;
- The standards to be used for data and metadata format and content;
- Policies for access and sharing the final research data, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, and other rights or requirements;
- Policies and provisions for re-use, re-distribution, and the production of derivatives; and,
- Plans for archiving final research data and other research products, and for preservation of access to them.
Frequently Asked Questions:

Could we consider including an embargo period for the data as well as for the publications? This would give the researchers that made the effort to collect the data the time to correct any errors, leverage its value, and publish further research before it was made publicly available.

A. No. DOT expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. This time point will be influenced by the nature of the data collected. Data from small studies can be analyzed and submitted for publication relatively quickly. If data from large or longitudinal studies are collected over several discrete time periods or waves, data should be released in waves as data become available or main findings from waves of the data are published. DOT recognizes that the investigators who collected the data have a legitimate interest in benefiting from their investment of time and effort. DOT expects that the initial investigators may benefit from the first and continuing use, but not from prolonged exclusive use. While DOT also understands that an institution's desire to exercise its intellectual property rights may justify a need to delay disclosure of research findings, a delay of 90 to 120 days is generally viewed as a reasonable period for such activity.
Frequently Asked Questions:
What constitutes "data" covered by a DMP?

A. 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. As part of your research, you may also generate unique data, which are data that cannot be readily replicated. Examples of studies producing unique data include: large surveys that are too expensive to replicate; studies of unique populations, such as centenarians; studies conducted at unique times, such as a natural disaster; studies of rare phenomena, such as rare metabolic diseases. Your DMP should also address unique data that may arise from your research.
Frequently Asked Questions:

What kinds of data are candidates for sharing?

A. Potentially all kinds of data are candidates for sharing, but unique data are especially important. Some sciences already have data-sharing plans in place, such as genetic mapping. But other basic science data are also amenable to sharing. Data from human subjects (e.g., surveys, clinical studies) also can be shared if the identity and privacy of research participants can be protected.
Frequently Asked Questions:

Am I required to deposit my data in a public database?

A. Researchers are encouraged to use publicly accessible repositories for the deposit of their data, where appropriate and available. This may include university institutional repositories, or other “free standing” repositories.

For guidance on how to locate a repository please reference our “Data Repositories Conformant with the DOT Public Access Plan” web page at:

Frequently Asked Questions:

Should the budget and its justification specifically address the costs of implementing the DMP?

A. Yes. As long as the costs are allowable in accordance with the applicable cost principles, and necessary to implement the DMP, such costs may be included of the proposal budget, and justified in the budget justification.
Frequently Asked Questions:

My institution's policy is that the data and all supporting materials from all research are owned and must remain with the institution if I leave. How does this policy affect what I can say about data management?

A. Data management by an institution is one avenue by which data preservation and access can be achieved. However, the DMP plan must address the institutional strategy for providing access to relevant data and supporting materials.
Frequently Asked Questions:

Should I consider contributing my research data to a data archive?

A. Maybe. Archives are organizations that collect and distribute data. They understand what is needed to prepare data for wider distribution and documentation for users. They provide stable, reliable, and cost-effective means for distributing data. They also provide protections for the dataset and technical assistance for requestors.
Frequently Asked Questions:

How soon after data collection am I obliged to share the final data?

A. Recognizing that the value of data often depends on their timeliness, data sharing should occur in a timely fashion. DOT expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. This time point will be influenced by the nature of the data collected. Data from small studies can be analyzed and submitted for publication relatively quickly. If data from large or longitudinal studies are collected over several discrete time periods or waves, data should be released in waves as data become available or main findings from waves of the data are published. DOT recognizes that the investigators who collected the data have a legitimate interest in benefiting from their investment of time and effort. DOT expects that the initial investigators may benefit from the first and continuing use, but not from prolonged exclusive use. While DOT also understands that an institution's desire to exercise its intellectual property rights may justify a need to delay disclosure of research findings, a delay of 90 to 120 days is generally viewed as a reasonable period for such activity.
Frequently Asked Questions:

What are DOT's expectations regarding the release of data that include sensitive information (e.g., information about individuals or locations of endangered species)?

A. Such data must be maintained and released in accordance with appropriate standards for protecting privacy rights and maintaining the confidentiality of respondents. Within legal constraints, what constitutes reasonable data access will be determined by the community of interest through the process of peer review and DOT program management.
Frequently Asked Questions:

My research, which seeks support from both the public and private sectors, will involve proprietary data. How do I deal with the data-sharing issue in my application?

A. DOT recognizes that there may be circumstances where a co-funder has requested restrictions on data sharing as a condition of funding. These restrictions should be identified in the DMP and a proposal made about how data from the co-funded project will be shared. Should you believe that you are unable to share any of the data, your justification will be considered by DOT program staff.
Frequently Asked Questions:

If I participate in a collaborative international research project, do I need to be concerned with data management policies established by institutions outside the United States?

A. Yes. There may be cases where data management plans are affected by formal data protocols established by large international research consortia or set forth in formal science and technology agreements signed by the United States Government and foreign counterparts. You should address these issues in your DMP. Be sure to discuss this issue with your sponsored projects office (or equivalent) and your international research partner when first planning your collaboration.
Frequently Asked Questions:

Is Project Open Data Metadata Schema v1.1 the only metadata schema that can be used in order to be compliant with the plan?

A. The metadata requirement in the plan is for study-level data. DOT recognizes that there are sector-specific metadata standards for differing fields of research or data. Researchers may use standards other than the Project Open Data Metadata Schema v1.1 provided that the chosen metadata standard can supply and be mapped to the data elements required by Project Open Data Metadata Schema v1.1. Project Open Data provides resources to facilitate such crosswalks (see: https://project-open-data.cio.gov/v1.1/metadata-resources/#field-mappings) and encourages the development of additional crosswalks as needed by offering the opportunity to contribute to the body of knowledge (see: https://github.com/project-open-data/project-open-data.github.io/blob/master/CONTRIBUTING.md).

Additional guidance on DMPs is located on the "Create a Data Management Plan" page at: https://ntl.bts.gov/public-access/creating-data-management-plans.
Frequently Asked Questions:
What do I have to do to comply with DOT's Public Access Plan?

A. See the "How to Comply" page at: https://ntl.bts.gov/public-access/how-comply
Frequently Asked Questions:

What could happen if investigators and institutions do not comply with DOT's Public Access Plan?

A. DOT will contact institutions to request outstanding requirements be fulfilled or an explanation of why an institution feels that the public access policy does not apply. Non-compliance will be taken into consideration in evaluation of future grants and awards. Non-compliance alone is not a reason to be denied a future funding agreement, for example a grant, contract or cooperative agreement.