


Slide 1



[Announcer] Thank you for joining us for this direct-to-recording Special Transportation Librarians Roundtable: U.S. DOT Public Access and Data Management Review. Our speakers are Leighton Christiansen and Charles Ducker from the United States Department of Transportation.

[NEXT SLIDE]



U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology

U.S. DOT Public Access and Data Management Review

Leighton L. Christiansen <https://orcid.org/0000-0002-0543-4268>
Data Curator, National Transportation Library
leighton.christiansen@dot.gov

Charles Ducker, Jr. <https://orcid.org/0000-0002-7535-5258>
Sr. Intellectual Property Counsel, Office of General Counsel
charles.ducker@dot.gov

Persistent link to this webinar:
<https://doi.org/10.21949/1503909>

[Leighton] Welcome to the “U.S. DOT Public Access and Data Management Review” webinar.

I am Leighton Christiansen, Data Curator at the National Transportation Library.

[Charles] And I am Charles Ducker, Senior Intellectual Property Counsel in the Office of General Counsel.

[Leighton] We hope you find this webinar useful. You may return to the webinar, and share it with others, by using the persistent link to this webinar:
<https://doi.org/10.21949/1503909>

[NEXT SLIDE]

Topics to Cover

- Opening U.S. Government-Funded Research Data
- U.S. DOT Public Access Review
- Submitting Final Reports and Final Datasets
- Benefits of Data Management
- Writing Data Management Plans
- Implementing Data Management Plans

3

[Charles] Since the U.S. DOT Public Access Plan went into effect on December 15, 2015, we have been working closely with the transportation research community to socialize the plan and ease implementation. This has included hosting large workshops, webinar and conference presentations, as well as individual emails and phone conversations. In this webinar we will answer some of the most frequently asked questions relating to the Public Access Plan and data management.

[Leighton] In this webinar, we will cover the following topics:

- Opening U.S. Government-Funded Research Data
- U.S. DOT Public Access Review
- Submitting Final Reports and Final Datasets
- Benefits of Data Management
- Writing Data Management Plans
- Implementing Data Management Plans

Slide 4

Opening U.S. Government-Funded Research Data

4

Transition slide Title: Opening U.S. Government-Funded Research Data

[NEXT SLIDE]

Opening U.S. Government-Funded Research Data

DATA.GOV DATA TOPICS IMPACT APPLICATIONS DEVELOPERS CONTACT

The home of the U.S. Government's open data
Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more.

115TH CONGRESS 1ST SESSION **H. R. 4174**

IN THE SENATE

115TH CONGRESS 1ST SESSION **S. 760**
[Report No. 115-134]

To expand the Government's use and administration of data to facilitate research.

SECTION 1. SHORT TITLE.

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

ONE HUNDRED FIFTIETH CONGRESS OF THE UNITED STATES OF AMERICA

AT THE SECOND SESSION

Approved on the 15th day of September, 2017.
An Act of Congress, for the purpose and purposes hereinafter expressed.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

SEC. 1001. SHORT TITLE; TABLE OF CONTENTS.

5

[Charles] Leighton, do you think you can provide our viewers with a brief history of the move to Open U.S. Government-Funded Research Data?

[Leighton] Yes Charles, I would be happy to. Over the past 10 years we have seen a number of White House executive orders, policies, and laws, that seek to make the operations of the government more transparent. This included executive orders calling for increased public access to federally funded publications, research, and data, so that citizens have as much access as possible to the products they fund through taxes. It is hoped that opening data to broader public use may also have social, economic, and research benefits, especially as data is re-used in novel ways, perhaps not considered by the original data creator.

Since 2009, U.S. agencies, following leads given by the White House, have been moving forward towards openness, and we have seen data portals such as DATA.gov go live. Data.gov is intended to increase public access to data across all federal agencies. Data.gov harvests its information from all governmental agency data catalogs, giving the public a "one-stop shop" for government data. As of June 2019, Data.gov lists nearly 250,000 datasets. The U.S. DOT data catalog from which data.gov pulls is called DATA.TRANSPORTATION.gov, a data catalog, data warehouse, and data visualization suite. As of June 2019,

Data.transportation.gov contains records of more than 4000 datasets. Most of these are available to the public for download.

Over the last decade, the U.S. Congress has put forward open data bills in the House and the Senate. In early 2019, the Congress passed, and the President signed, the **Foundations for Evidence-Based Policymaking Act of 2018 (HR 4174)**. Title II of the Act, includes the Open, Public, Electronic, and Necessary (OPEN) Government Data Act, which has been moving through Congress since 2017. The OPEN Government Data Act requires that non-restricted U.S. government data be available in machine-readable formats. This act is consistent with the spirit of the U.S. DOT Public Access Plan, which we will review next.

[Next slide]

Slide 6

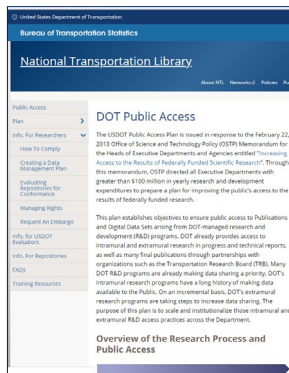
U.S. DOT Public Access Review

6

Transition slide Title: U.S. DOT Public Access Review

[NEXT SLIDE]

U.S. DOT Public Access Plan Guidance Website



<http://ntl.bts.gov/publicaccess/>

[Leighton] Before we review the U.S. DOT Public Access Plan, I want to take a moment to remind folks that the Plan, descriptions of data management plans, Frequently Asked Questions, and other useful guidance are all available at our guidance website: <http://ntl.bts.gov/publicaccess/>

The current pages include:

- The Plan
- Information for researchers, especially on how to write DMPs
- Info for DMP evaluators
- Information for and on repositories
- FAQs, especially with answers on publishing, public access, and copyright issues; and,
- Training Resources, which are always under development

Links and resources mentioned in this webinar can all be found there.


[NEXT SLIDE]

U.S. DOT Public Access Policy

**Plan to Increase Public Access to the Results of Federally-Funded Scientific Research
(2015-12-15)**

<https://www.transportation.gov/mission/open/official-dot-public-access-plan-v11>

Plan to Increase Public Access to the
Results of Federally-Funded
Scientific Research Results
Version 1.1



December 16, 2015
U.S. Department of Transportation

- **Affirm and enhance DOT’s commitment to Public Access to Scientific Research results, including digitally formatted scientific data without charge to the maximum extent possible.**
- Support governance of and best practices for managing Public Access to peer-reviewed Publications and Digital Data Sets across DOT.
- **Ensure continuous access to and reliable preservation of DOT-funded Publications and Digital Data Sets for research, development and education purposes, within available resources.**
- Preserve and increase the use of Scientific Research results to enhance scientific discovery and deployment of research results.
- Enhance the use of Scientific Research results to promote innovation and economic competitiveness.
- Affirm DOT’s support for the reproducibility of Scientific Research results.
- Make DOT’s research portfolio available to the public at the project level.

8

[Leighton] Charles, I know that you were part of the team that crafted the DOT Public Access Policy. Could you please take a few minutes to review the plan?

[Charles] Sure. The U.S. DOT “Plan to Increase Public Access to the Results of Federally-Funded Scientific Research (2015-12-15)” (also known as the U.S. DOT Public Access Policy), is DOT’s response to a 2013 memorandum from the White House Office of Science and Technology Policy (OSTP). Through this memorandum, OSTP directed all Executive Departments with more than \$100 million in yearly research and development expenditures to prepare a plan for improving the public’s access to the results of federally funded research. As U.S. DOT, the various modal research offices spend more than \$1 billion dollars per year on transportation-related research.


The Public Access Policy sets out to:

- **Affirm and enhance DOT’s commitment to Public Access to Scientific Research results, including digitally formatted scientific data without charge to the maximum extent possible.**
- Support governance of and best practices for managing Public Access to peer-reviewed Publications and Digital Data Sets across DOT.
- **Ensure continuous access to and reliable preservation of DOT-funded Publications and Digital Data Sets for research, development and education purposes, within available resources.**
- Preserve and increase the use of Scientific Research results to enhance scientific discovery and deployment of research results.
- Enhance the use of Scientific Research results to promote innovation and economic competitiveness.
- Affirm DOT’s support for the reproducibility of Scientific Research results.
- Make DOT’s research portfolio available to the public at the project level.

[NEXT SLIDE]

U.S. DOT Public Access Policy from 30,000 feet

Plan to Increase Public Access to the
Results of Federally-Funded
Scientific Research Results
Version 1.1



December 16, 2015
U.S. Department of Transportation

- DOT-Funded Scientific Research
 - Not goods, services, or construction
- Three Components
 - Publications
 - Digital Datasets
 - Research Project Record
- Links to Final Digital Datasets
- Data Management Plan (DMP)
 - Explain case for long-term preservation and sharing; **OR**,
 - Justify case against long-term preservation and sharing
- Include preservation costs in proposals

9

[Charles] Let's take a quick 30,000 foot view of the plan. First, the plan is designed to apply only to scientific research results. While we have a broad definition of what constitutes "scientific research", not all government contracts, grants and other funding agreements are going to fall under the plan's requirements. Government contracts or grants for goods, services, or building things, for instance, are not going to fall under the plan's requirements.

As we mentioned before, the plan is directed to two primary components of scientific research – the written deliverables of a funding agreement (technical reports, peer-reviewed journal articles, etc.) and the final digital dataset that support the conclusions and results of the research. These are the items to which we are responsible for ensuring the Public has access. There is a third component of the plan – the concept of a Research Project Record. We have tried to utilize updated existing databases of DOT research to provide a project level research record in order to track our research on the most basic level possible. Internally, we also use it to ensure our funded researchers are complying with the plan's requirements.

With regard to the written deliverables, funded researchers are required to submit a copy of their materials to the Departmental Administration (i.e., FHWA, FAA, NHTSA, FRA, etc.) that funded their research, as well as to the National Transportation Library as Leighton will explain in just a minute. But, we are here today to focus on the other part of the plan – the digital datasets.

As with the written deliverables/publications, all DOT-funded scientific researchers are required to ensure public access to their digitally-formatted final research dataset(s). The Public Access plan does not require that they submit the actual dataset to DOT. Instead, we have asked that

they provide a link to an acceptable repository where access to the dataset can be preserved for a long time ... at least for some datasets.

As a part of ensuring the public's access to the preserved digital dataset, DOT-funded scientific researchers are also required to submit a Data Management Plan (DMP) for DOT review and approval prior to beginning their research. In addition to providing the DOT the long-term preservation and storage location information, such a DMP may explain why the long-term preservation and/or public access is not justified.

[Leighton] Wait! You are saying researchers can chose not to preserve their data?

[Charles] That's right, we leave it to the researchers to determine whether there is long-term value to the preservation of their to-be-collected dataset. They tell us that in their DMP. Frankly, some data has no value beyond the research for which it is collected. We wanted to offer the option not to preserve the data that falls into that camp. For those datasets that do have longer-term value (i.e., they can be used for additional research, they can be used to reproduce and confirm important research results, or they can be combined with other data for new research, etc.), the plan allows for researchers to include the long-term preservation costs in their proposals. The only requirements for allowable data storage repositories are that the digital datasets be publicly accessible at no cost to the extent possible.

And I understand that we have some preferred characteristics.

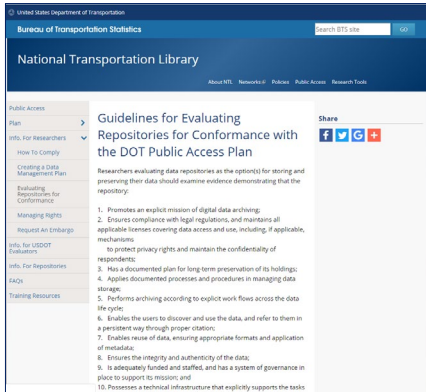
[Leighton] Indeed we do Charles. But before that, we had a recent question about the time frame in which university-based researchers need to be compliant with the Public Access Plan. Could review that too, please?

[Charles] Sure. As Section 2.0, Scope, states, in part, "Any new intramural program, as well as any award, modification to an existing award or extension of an existing agreement for extramural research made on or after implementation of this plan will be subject to this plan." As the plan was published on December 15, 2015, all newly funded, or modified agreement, research, begun after December 15, 2015, is subject to the this plan. As you know, the DOT has been working with researchers since December 2015 to help them come into compliance with the Public Access Plan.

[Leighton] That is right. And one measure of compliance is the preservation of digital datasets in a repository that conforms with the plan. So let us talk about repository characteristics next.

[NEXT SLIDE]

Repository Characteristics Review



The screenshot shows the 'National Transportation Library' website. The main content area is titled 'Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan'. It includes a list of 10 characteristics for evaluating data repositories. The URL <https://ntl.bts.gov/public-access/guidelines-evaluating-repositories> is displayed at the bottom of the page.

1. Explicit mission of digital data archiving
2. Protect privacy rights and maintain the confidentiality of research subjects
3. Enables the users to discover and use the data, and refer to them with a persistent identifier
4. Ensures the integrity and authenticity of the data

[Leighton] The DOT Public Access Guidance website has a page called: “Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan.” On that page we encourage researchers evaluating data repositories to look at 10 characteristics of good data repositories, based on the Data Seal of Approval.

[Charles] Why don’t you give us the top four or five for now, since they are all documented.

[Leighton] That is a good idea, although difficult. To highlight just four, I would choose these, in no particular order:

- Promotes an explicit mission of digital data archiving;
- Ensures compliance with legal regulations, and maintains all applicable licenses covering data access and use, including, if applicable, mechanisms to protect privacy rights and maintain the confidentiality of respondents;
- Enables the users to discover and use the data, and refer to them in a persistent way through proper citation; and,
- Ensures the integrity and authenticity of the data.

[Charles] Are there others ways to tell if a repository is conformant with the DOT plan?

[Leighton] Yes. For example, if a repository is certified as “trustworthy” under the Data Seal of Approval, the Core Trust Seal, or the ISO 16363 standards, that repository would be conformant with the DOT Public Access plan.

A good place to start is close to your own organization. If you are a researcher at a university, go talk to the university library or institutional repository. Many universities and colleges have invested in repositories because they see long-term value in preserving, managing, and sharing research undertaken within their walls. You might also try your state library as many states and state agencies are also moving to manage and preserve data as strategic, business, governmental, or public asset.

If your institution does not have repository resources, you might look to larger institutions or consortia for assistance.

In April 2019, the U.S. DOT hosted a workshop entitled “Building a National Transportation Data Preservation Network.” Starting with a small pool of academic and state transportation researchers, Charles, myself, and other DOT staff took the first steps to building a national transportation research consortium. We will be expanding the stakeholder pool for future meetings. If you are interested in taking part, please contact me.

So, Charles, a little earlier you mentioned researchers needed to submit reports and dataset links. Would this be a good time to review those?

[Charles] Yes, let us go to those next.

[NEXT SLIDE]

Slide 11

Submitting Final Reports and Final Datasets

11

Transition slide Title: Submitting Final Reports and Final Datasets

[NEXT SLIDE]

Submitting Final Reports and Final Datasets

<https://ntl.bts.gov/public-access/how-comply>

10. Send one email to Research Hub, NTL, and TRB:

- Final Report URL(s) or PDFs;
- URL(s) to final datasets and descriptive metadata;
- Funding agreement number;
- The Research Hub Display ID;
- ORCID(s) (unique researcher IDs) for all author(s); and,
- Any other documented project outputs or outcomes.

Dataset Differences

- Final datasets used to draw conclusions in report, NOT all raw data gathered;
- URLs to datasets, NOT electronic copies
 - Choose a conformant data repository
<https://ntl.bts.gov/public-access/data-repositories-conformant-dot-public-access-plan>
- Dataset & documentation

12

[Charles] I know we had a recent question about submitting final reports and datasets. Leighton could you go over the submission process for our viewers?

[Leighton] Glad to Charles.

The basic information is all available on our “How to Comply” page at <https://ntl.bts.gov/public-access/how-comply>

Step 10. asks researcher to send a single email to USDOT Research Hub at Research.Hub@dot.gov, the National Transportation Library at NTLDigitalSubmissions@dot.gov, and the Transportation Research Board TRID database at TRIS-TRB@nas.edu. Further, researchers are asked to include the following information:

- Final Report URL(s) or PDFs for any resulting publications;
- URL(s) to and associated descriptive metadata for any final datasets and the arising from the research project;
- The funding agreement number of the project;
- The RH Display ID for the project;
- ORCID(s) (unique researcher IDs) for all publication author(s); and,
- Any documented project outputs or outcomes resulting from the research project.

As you can see the we have tried to keep the data submission process as close to the same as the report submission process as possible.

[Charles] That is great, but aren't reports and datasets different assets, with different needs?

[Leighton] That is a good point Charles, there are some differences and I would like to call out a few here.

-- First Public Access plan Section 4.2 Data, requires only that awardees ensure DOT access to the final datasets, NOT all raw data collected. The Plan defines “scientific data,” in part, as “the digitally recorded factual materials resulting from research that is necessary to validate research findings.” In most research projects far more data is collected than is analyzed for final conclusions. DOT requires access to the final dataset so that findings can be replicated and verified.

-- While the public access plan and existing funding documents require that the U.S. DOT receive a copy of each final report and other project output, Section 4.3 Research Project Reports only requires that DOT is supplied a link to digital data sets, we do not require an electronic copy of the dataset.

--- I should add that it is best if the URL to the dataset leads to a data repository which is conformant with the Plan. Researchers can find information about conformant repositories, and how to evaluate a repository for conformance, by visiting our web page at <https://ntl.bts.gov/public-access/data-repositories-conformant-dot-public-access-plan> .

[Charles] The last point listed on the slide is “Dataset and documentation.” Can you tell us what you mean by that?

[Leighton] Yes I can, but that requires a bit more space, so we should go to the next slide.

[NEXT SLIDE]

Documenting Data

Data Package Guidelines

- **Dataset**
 - .csv or other open format
- **Readme.txt**
 - Includes Data Dictionary
 - Notes standards used
 - Defining Zero, Null, & Unknown
 - FAQs and other notes
- **Metadata file** in Project Open Data .json and/or other schema
- **Data Management Plan (DMP)**
- *Code or scripts* used in data analysis
- *Supporting files, tables, etc..*

(Bold = Required; *Italics* = Optional, or Required if Applicable)

A “Data Package” is the dataset, the DMP, and all other documentation needed to contextualize the dataset for any and all users.

```
bts_american_travel_survey_1995_DATA_20190328.csv
bts_american_travel_survey_1995_DMP_20190328.txt
bts_american_travel_survey_1995_Metadata_20190328.json
bts_american_travel_survey_1995_README_20190328.txt
bts_american_travel_survey_1995_DATA_DICTIONARY_20190328.txt
bts_american_travel_survey_1995_SUPPORTING_coded_values_20190328.csv
```

American Travel Survey (ATS) 1995
<https://doi.org/10.21949/1503648>

13

[Charles] Having just reviewed the Public Access Plan, I don't remember the plan call on researchers to provide specific dataset documentation, aside from a common core metadata file. Do you have some tips on how researchers can document their data to improve shareability?

[Leighton] Indeed I do. To be absolutely clear, the Public access plan does not spell out a specific level of documentation. But as we have been working on making our own data sharable in alignment with public access, and doing research on data management, we have come across some best practices that we have adopted ourselves. I want to take the opportunity to pass those practices along here.

In order to meet the spirit of public access and data sharing, data should be accompanied by a few pieces of documentation. The purpose of the documentation is to contextualize the data for future users. That is, supply enough information about the data that future users, whether that is you, someone on your research team, or a user not yet known or born, will be able to decide if the data is fit for the use they have in mind, and be able to draw valid conclusions from it. To do this well, some people who teach about data management refer to a “data package.”

What is a “data package”? A data package is the dataset, the data management plan, and all other documentation needed to contextualize the dataset for any and all users.

We feel that our data packages help to make the data as open as possible, by giving the naive user everything they will need to understand the data. They should be able to tell if the data is fit for the use they have in mind. And each variable should be well defined, so there is no ambiguity.

So what are the elements of a data package?

The elements of a data package are:

- The dataset (which is required); preferably in an open, non-proprietary format;
- A readme text that includes:
 - the data dictionary;
 - notes on any standards used;
 - definitions of null values, zeros, and unknown, or empty cells;
 - a description of the data;
 - contact information;
 - and other notes and FAQs.
- A machine readable .json metadata file written in the U.S. government standard known as Project Open Data Metadata Schema Version 1.1 (which is based on the international standard known as DCAT (or Data Catalog Vocabulary))
- And a data management plan (DMP)

Optional elements include:

- Code, or scripts used in data analysis; and -- Supporting files and tables, such as imputed value tables

At NTL we like to adopt best practices from others and give credit where it is due. Our data package guidelines are adopted from Kristin Briney's book "Data Management for Researchers."

If you would like to look at an example of data package from a Bureau of Transportation Statistics dataset, you can go to the "American Travel Survey (ATS) 1995" page at <https://doi.org/10.21949/1503648>

The documentation is quite robust, because the survey datasets are complex.

[Charles] That is an impressive amount of documentation. Will all repositories be looking for the same level of documentation?

[Leighton] That is an excellent question Charles. The answer is: it depends on the repository. But all will require or strongly suggest all or some elements of the data package described above.

Without documentation, data ceases to become sharable. As usual, that is a longer answer from me than was strictly needed. Let us move on to the next topic.

[NEXT SLIDE]

Slide 14

Benefits of Data Management

14

Transition slide Title: Benefits of Data Management

[NEXT SLIDE]

Data Management Definitions

“In the context of research and scholarship, ‘**Data Management**’ refers to the **storage, access and preservation** of data produced from a given investigation. Data management practices cover the **entire lifecycle of the data**, from planning the investigation to conducting it, and from backing up data as it is created and used; to long term preservation of data deliverables after the research investigation has concluded.”

- University Library, Texas A&M <http://guides.library.tamu.edu/DataManagement>

“Data management is the **compilation of many small practices that make your data** easier to find, easier to understand, less likely to be lost, and **more likely to be usable during a project or ten years later.**”

- Kristin Briney. 2015. Data management for researchers: organize, maintain and share your data for research success.

15

[Charles] So Public Access Plan section 4.2 Data, requires all DOT-funded research data be accompanied by one element of the data packages you mentioned: a Data Management Plan, or DMP. What are the benefits of planning to manage research data before the research project even starts?

[Leighton] That is an excellent question. Let me start with a definition before I talk about the benefits.

The University Library at Texas A&M defines data management in this way: “In the context of research and scholarship, “Data Management” refers to the storage, access and preservation of data produced from a given investigation. Data management practices cover the entire lifecycle of the data, from planning the investigation to conducting it, and from backing up data as it is created and used; to long term preservation of data deliverables after the research investigation has concluded.”

[Source: University Library, Texas A&M University. “Data Management Defined - Research Data Management - Guides at Texas A&M University.” Research Data Management, October 1, 2013. <http://guides.library.tamu.edu/DataManagement>]

Or to borrow a plain language definition from Kristin Briney, (page 7) “Data management is the compilation of many small practices that make your data easier to find, easier to understand, less likely to be lost, and more likely to be usable during a project or ten years later.”

Briney, Kristin. 2015. Data management for researchers: organize, maintain and share your data for research success.

I want to emphasize a major point that both definitions make about data management: Data management is about the entire life of the data, not just a form you fill out during proposals, or an added extra you throw on at the end. Data management is a group of practices and tools meant to help you improve your research methods and outcomes.

So let us keep those in mind as we talk about data management over the next few minutes.

[NEXT SLIDE]

Benefits of Managing Data

- Extend data's useful life
- Plan for software and hardware
- Plan for storage size & cost
- Save time
 - Documented file storage paths
 - Documented file naming convention and version control
 - Documented data roles
 - Data access levels
- Find data for follow-on research
- Backup and disaster recovery plan
 - 3-2-1 Strategy
- Increase funder confidence
- Improve data sharing culture

Citation: Vines, Timothy H., Arianne YK Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean-Sébastien Moore, Sébastien Renaut, and Diana J. Rennison. "The availability of research data declines rapidly with article age." *Current biology* 24, no. 1 (2014): 94-97. <http://doi.org/10.1016/j.cub.2013.11.014>

16

[Leighton] Now let us talk about the benefits.

Over the past couple of decades, research data collection across nearly all disciplines has become digital. Hard copies of research on paper are becoming a thing of the past. And that move from analog paper datasets to digital datasets actually puts data at greater risk. Modern paper is very archival, meaning that barring fire or flood, paper is shelf stable for decades or longer. At a past position, I analyzed and cataloged a transportation engineer's field notebook, with traffic counts, from 80 years ago. The quirks of handwriting aside, the data was still legible and the paper was in pretty good shape.

Digital data is far more fragile. As digital data is stored as a string of magnetized bits, small mishaps can lead to the loss of data. If data is NOT backed up, the change in polarity of single bit, the loss of a flash drive, a scratch or broken read/write disc, or a stolen laptop can mean all data is irretrievably lost.

In 2014, Timothy Vines and his co-authors wrote a paper for *Current Biology* called "The availability of research data declines rapidly with article age." In their study of 516 journal papers, they concluded that "the odds of a data set being reported as extant declined by 17 percent per year." This means that in 5 to 7 years after publication, authors could no longer locate or produce the dataset to back up their research when requested. Further, for some authors, data was lost within the first year after publication. This data loss has a huge negative impact on follow-on research, research replication and reproducibility, and future data sharing and novel re-use. (For those interested in the Vines paper, I have included the citation at the bottom of the slide and in the webinar transcript.)

The hope of the scientific research community is that robust data management should help us extend the useful life of digital data beyond its current 1- to 5-year life expectancy.

[Charles] Wow, that rate of data loss is a bit shocking. So if extending data's useful life is the a main benefit of data management, what are some others?

[Leighton] There are a number of benefits of good data management. I will briefly highlight just a few more.

- Writing a data management plan can help you plan for research project software and hardware needs, and help you to be sure to include those in your project proposal budget.

- The same is true of planning for data storage. Through data management planning, you can estimate your project storage costs, as well as your long-term preservation or repository storage, and you can include those costs in your proposal. Remember, research data repository storage costs are allowable proposal costs for DOT-funded projects. But you need to include those at the beginning.

- There is a data management rule of thumb that each 1 minute of planning saves 10 minutes of headaches later. I don't know if it is that quantifiable, but there are major time savings to be had through planning. Some ways you can save time through data management planning include:

- By having a data management plan that is shared with the research team, the entire team will know the location where data is stored; how data files are named and version controlled; and who has which roles in relation to the data; and who has access to various layers of data. This is important if you are collecting or re-using data that has sensitive information. Documenting all of these aspect of your data collection and management in a single location makes it much easier to manage changes in research staff and helps to reduce project knowledge loss, as well as on-boarding time.

- One good reason to plan your data management for your current project is that it can make it easier for you to prepare for the follow-on project. As we saw with paper from Vines, data loss can begin in the first year after publication. And as there is sometimes a lag between project funding, you don't want to take the chance that you will lose your data before you start phase 2. No funder will want to pay for repeating a data collection because past data was lost through poor data management.

- It is best practice to plan your data backup strategy BEFORE you begin to collect data. Each new data collection should be backed up as soon as possible, to prevent loss. A robust backup plan follows the 3-2-1 rule of thumb: 3 copies of the data; on at least 2 different kinds of media (disk drive, portable drive, local server, cloud storage, etc.); with 1 backup copy held in a geographically distinct and distant region from your own, in case of natural disaster, or regional loss of power.

- Data management plans are now required by all government scientific research funders, including U.S. DOT. A robust and well thought out DMP gives funders confidence that researchers will be good stewards of the data. This can in turn, lead to future funding opportunities. Conversely, bad data management planning, or actual funded-data loss, can mean that future funding awards go to other researchers, who manage their data better.

- Digital data sharing, when paired with computational data search, visualization, and analysis, can lead to previously gathered data being used in unexpected ways, and leading to unexpected discoveries. By managing, documenting, and sharing your data, you may share in the credit for future discoveries. Further, your research may benefit by using shared, third-party data. Data

sharing is not new to science and research. What is new, is that rather than sharing hard copies of data, we are now sharing digital datasets. And as we discussed before, digital datasets are fragile things, which need new levels of care.

There are more benefits to data management. But lets stop with these and move on to writing a data management plan.

[NEXT SLIDE]

[Citation] Vines, Timothy H., Arianne YK Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean-Sébastien Moore, Sébastien Renaut, and Diana J. Rennison. "The availability of research data declines rapidly with article age." *Current biology* 24, no. 1 (2014): 94-97.
<http://doi.org/10.1016/j.cub.2013.11.014>

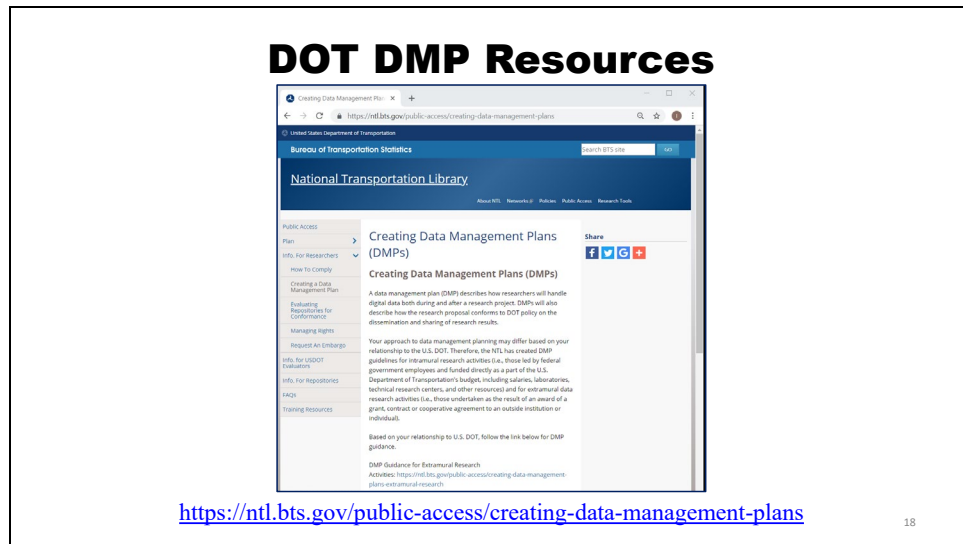
Slide 17

Writing Data Management Plans

17

Transition slide Title: Writing Data Management Plans

[NEXT SLIDE]



[Charles] As I recall, the DOT Public Access website has some guidance pages on writing data management plans.

[Leighton] That is correct Charles. If folks navigate to <https://ntl.bts.gov/public-access/creating-data-management-plans> they will find the information we are about to review

[NEXT SLIDE]

DOT DMP Overview

A data management plan (DMP) describes how researchers will handle digital data both during and after a research project. DMPs will describe how the research proposal conforms to DOT policy on the dissemination and sharing of research results. Each plan should include a 2-3 page narrative description covering:

- 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.

DOT-funded research projects are expected to be conducted pursuant to the approved DMP. A DMP may evolve as the research project evolves and should be reviewed for possible revision whenever a data management procedure is changed.

<https://ntl.bts.gov/public-access/creating-data-management-plans-extramural-research>

19

[Charles] Leighton, could you please give our viewers an overview of DOT DMP requirements?

[Leighton] Happy to Charles.

As the guidance page says, “A data management plan (DMP) describes how researchers will handle digital data both during and after a research project. DMPs will describe how the research proposal conforms to DOT policy on the dissemination and sharing of research results. Each plan should include a 2-3 page narrative description covering:

- 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.

DOT-funded research projects are expected to be conducted pursuant to the approved DMP. A DMP may evolve as the research project evolves and should be reviewed for possible revision whenever a data management procedure is changed.”

As you can see, our original definition of a DMP is pretty basic. We use these same sections when we write data management plans for datasets coming from the Bureau of Transportation Statistics, and we usually come to 3 pages or so.

Now let us take a deeper look at the sections of a data management plan.

DOT DMP Sections

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

<https://nfl.bts.gov/public-access/creating-data-management-plans-extramural-research>

20

[Leighton] Our DMP sections are based on and borrowed from other federal research bodies and common practice. In fact, the DOT DMP sections look very much like the National Science Foundation DMP sections as they existed in 2015, when we wrote our plan.

My suggestion is that you adopt the same 5, in order to help maintain consistency and help your DOT program managers review your DMP.

Now I want to take a deeper look at the guidance we provide for the sections of the data management plan. Charles would you like to pick a section?

[Charles] Well, how about we start at the top and look at the Data Description section?

[Leighton] Sounds good.

[NEXT SLIDE]

DOT DMP Section: Data Description

Section Description	Helpful Prompts
<ul style="list-style-type: none">• Provide a description of the data that you will be gathering in the course of your project.• Address the nature, scope, and scale of the data that will 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.• Discuss value of the data over the long-term.	<ol style="list-style-type: none">1. Name the data, data collection project, or data producing program.2. Describe the purpose of the research.3. Describe the data that will be generated in terms of nature and scale (e.g., numerical data, image data, text sequences, video, audio, database, modeling data, source code, etc.).4. 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; instrument generated digital data output such as images and video; etc.).5. Discuss the period of time data will be collected and frequency of update.6. If using existing data, describe the relationship between the data you are collecting and existing data.7. List potential users of the data.8. Discuss the potential value of the data have over the long-term for not only your institution, but also for the public.9. If you request permission not to make data publicly accessible, explain rationale for lack of public access.10. Indicate the party responsible for managing the data.11. Describe how you will check for adherence to this data management plan

<https://ntl.bts.gov/public-access/creating-data-management-plans-extramural-research>

21

[Leighton] The description of the Data Description section is pretty basic. In this section the data management plan author or authors should:

- Provide a description of the data that you will be gathering in the course of your project.
- Address the nature, scope, and scale of the data that will 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.
- Discuss value of the data over the long-term.

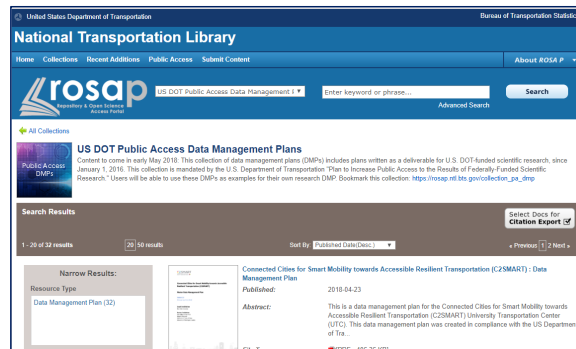
After the short description for each section, there are a series of Helpful Prompts, which will help the researchers create a narrative for their data management plan. If each prompt that applies is answered with a sentence or two (or more if needed), a DMP should emerge.

Again, the sections, their descriptions, and the prompts are all found in the Public Access Plan guidance website. In this case, on the “Creating Data Management Plans for Extramural Research” page. <https://ntl.bts.gov/public-access/creating-data-management-plans-extramural-research>

[Charles] That seems straight forward enough. But do we have any examples that we can share?

[Leighton] I am happy you asked that question Charles, because yes we do. [NEXT SLIDE]

Submitted U.S. DOT Public Access Data Management Plans



[Leighton] DOT Public Access Plan Section 4.3, Research Project Records, calls on the National Transportation Library to provide a searchable database of plans submitted under the plan. We have fulfilled this requirement by creating a special collection of U.S. DOT Public Access Data Management Plan in the National Transportation Library's Digital Repository, the Repository & Open Science Access Portal.

Researchers can look through a couple of dozen DMPs which were submitted and approved, for examples and ideas.

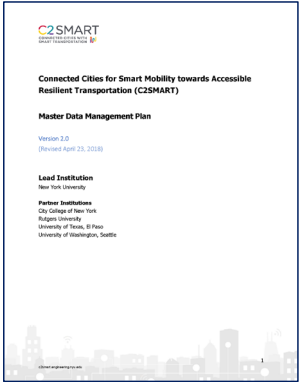
[Charles] That is great Leighton. I know that we had a recent request to talk about commonalities and lessons learned from the DMPs submitted so far. Do you have any thing you would like to share?

[Leighton] Indeed I do. Let us go to the next slide and talk about some lessons learned.

[NEXT SLIDE]

Lessons from Submitted DMPS

- Author/Creator missing from sections
- DMPS can stand alone
 - **Solution:** Add organizational information; branding
- Single research organization; many DMPS; lots of duplicated information
 - **Solutions:** Master DMP for common responses
 - Project level DMP for project-specific info, link to Master
- DOT Narrative DMP not as detailed as needs to be for researchers
 - **Solution:** Robust project DMPS



<http://c2smart.engineering.nyu.edu/c2smartpublications/#1534358159826-cb5e9a6e-27fc>

<https://rosap.ntl.bts.gov/view/dot/36367>

23

[Leighton] One of the first things we learned as we were reviewing submitted data management plans, is that we did not specify in our DMP requirements that the authors or organization creating the DMP name themselves, or supply us with contact information.

While this may not seem important if a DMP is part of larger proposal package, it becomes a problem when DMPS are seen out of context of other information.

For example, there was a day when I reviewing DMPS from several University Transportation Centers at the same time. The authors had done a great job filling out the sections that we described. But we had not included a section Zero: DMP author name and contact information. I started to lose track of which DMPS belong to which UTC. And as a DMP can stand alone from the other proposal documents, it should be treated an item that can stand alone. Think about the DMP being included in a data package; the DMP can help to link the data back to the organization that created the data.

In conversations with a few DMP authors at a couple of UTC, we came up with the idea of putting a cover page on the DMPS, and using a little branding, such as organizational letterhead. On the slide you can see the cover of the Connected Cities for Smart Mobility towards Accessible Resilient Transportation (C2SMART) UTC master DMP. I really like this because it even includes version control and a date. The second page give a description of the UTC and the research conducted. The DMP sections inside are very straight forward and align well with the DOT suggested sections.

When we first began socializing data management plans to transportation researchers, we heard some concern about time spent with in a research organization entering the same information over and over again with each DMP. In response to that we suggested that each organization create a master DMP. This DMP would be especially useful where the organization was using a

single solution for its data repository, or needed to report on its organization's IT infrastructure. Then, if the organization posted its master DMP on the web, researchers could reference the master via URL as they wrote their project level DMPs. In a connected digital environment, this is a much more efficient way for folks to create DMPs. I am happy to report, that I looked on the C2SMART UTC website, and was able to locate their master DMP! That is a great best practice, and I hope others will adopt it.

[Charles] That sounds like a good outcome from our experience. But what about the project level DMPs you mentioned?

[Leighton] Well, we don't have any specific guidance written for this yet, but we should think about it this year. Ideally, the project DMP would start with a statement that says something like "The research project complies with our organization's master data management plan, located at this URL:..... Data management activities unique to the research project described here are spelled out below." Or something like that. Then the project DMP would include things like project staff; their roles and access levels; the specific types and sizes of data from this project; file storage and naming conventions to be used; and any other deviations from or additions to the master DMP. This would allow researchers to focus on and plan for their data specific data management needs.

Now I haven't seen any project level DMPs yet, but I do hope we will start to get some and we will get feedback from researchers on this method.

[Charles] Now the last lesson you mention is that the narrative 2 to 3 page DMP DOT requires is not detailed enough. What do you mean?

[Leighton] Well, this is something I have been think about for the past three years reviewing submitted DMPs, engaging in data-centric professional organizations and conferences, and writing DMPs for data that I am curating at the National Transportation Library. The 2 to 3 page DMP we ask for is fairly high level. But for a DMP to actually be useful to a research team, it needs to be more of a project management document. The project DMP should record those things I described a minute ago: project staff; their roles and access levels; the specific types and sizes of data from this project; file storage and naming conventions to be used; and any other deviations from or additions to the master DMP. This type of DMP is then a useful knowledge bank for the research team. If it is posted on a networked drive or private intranet, the team, organization leadership, IT staff, can refer to the project DMP any time they have questions, or need to know who to contact about specific issues. The project DMP would serve as a useful on-boarding reference tool as research staff changes. And it would be easy to update and version the project DMP as needed. In turn, this should provide better managed data, and help to increase the shelf live of data.

[Charles] I can see the advantages of what you just described. Do we have any examples of this type from transportation researchers yet.

[Leighton] No. But if an organization or researcher wanted to contact me, I would be happy to help them implement such a DMP. It would be nice to track any actual process improvements.

[Charles] Speaking of implementing DMPs, I believe that is our next topic.

[NEXT SLIDE]

Slide 24

Implementing Data Management Plans

24

Transition slide Title: Implementing Data Management Plans

[NEXT SLIDE]

Implementing DMPs

- DMPs are Living Documents
 - Quarterly reviews
 - Record changes
 - Changes in research plan are expected
 - Version Control
 - Submitted updated DMPs to DOT
- Plan to manage and share data from the beginning
 - Many small practices
 - Data more useful to you
 - Longer shelf life for data
 - Data more easily shared

25

[Leighton] The last thing I want to talk about is implementing data management plans. We don't have any guidance on this yet, but we should create some soon.

The major point is that a DMP should be a living document. Even the best planned research project comes up against the unexpected. Key research personnel win the lottery; a sensor manufacturer goes out of business; new technology comes online; your IT department puts all your files on the cloud; laws and regulations change. We live in a dynamic world, and there is no penalty to responding to changed circumstances. DMPs should change as needed.

A good rule of thumb is to review a DMP at least quarterly. Record any changes in staff, data collection or storage, or policies that might affect the data. Update your DMP with new date and version information, if needed. Submit that updated DMP to your DOT research point of contact.

[Charles] "Living Documents:" That is a good way to think about DMPs. Do you have other tips for implementing DMPs?

[Leighton] Yes, the last point I want to make is to make is that under the Public Access Plan we are encouraging researchers to plan to manage and share their data from the beginning of the research process. That means taking the time to think through the data management planning process and to create a DMP that is useful to you and your research team. This also means think through how you share the data after the project. Keeping up with data documentation during the research process, such as writing the data dictionary as you name your variables, will be much more time efficient, and can save your team misunderstandings later. Then at the end of the project, the data dictionary will just need a quick review and spell check, and it will be ready to go into the data package.

Also, think back to the definition of data management we gave from Kristin Briney: “Data management is the compilation of many small practices that make your data easier to find, easier to understand, less likely to be lost, and more likely to be usable during a project or ten years later.” Take the small practices to heart. Start with one or two to add to your research process now. Such as recording file naming conventions and staff access levels to data. Once you are comfortable with those in your research workflow, add a couple more small practices. Over time these should help improve research and data sharing outcomes.

[NEXT SLIDE]

Topics Covered

- Opening U.S. Government-Funded Research Data
- U.S. DOT Public Access Review
- Submitting Final Reports and Final Datasets
- Benefits of Data Management
- Writing Data Management Plans
- Implementing Data Management Plans

Questions or feedback? Email: public.access@dot.gov

26

[Charles] In this webinar, we covered the following topics:

- Opening U.S. Government-Funded Research Data
- U.S. DOT Public Access Review
- Submitting Final Reports and Final Datasets
- Benefits of Data Management
- Writing Data Management Plans
- Implementing Data Management Plans

[Leighton] If you have further questions or would like to give us feedback on this webinar, please send an email to public.access@dot.gov and one of the team will get in contact with you.

Thank you Charles, this has been a lot of fun.

[Charles] Thank you Leighton. And thank you to our viewers.

[NEXT SLIDE]

Resources & Works Cited

- United States Department of Transportation. Plan to Increase Public Access to the Results of Federally-Funded Scientific Research. 2015. <https://www.transportation.gov/mission/open/official-dot-public-access-plan-v11>
- United States Department of Transportation, National Transportation Library. DOT Public Access [guidance website]. 2015. <https://ntl.bts.gov/public-access>
- United States Department of Transportation, National Transportation Library. Repository & Open Science Access Portal. 2017. <https://rosap.ntl.bts.gov>
- University Library, Texas A&M University. "Data Management Defined - Research Data Management - Guides at Texas A&M University." Research Data Management, October 1, 2013. <http://guides.library.tamu.edu/DataManagement>
- Briney, Kristin. 2015. Data Management for Researchers: Organize, Maintain and Share your Data for Research Success.
- Vines, Timothy H., Arianne YK Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean-Sébastien Moore, Sébastien Renault, and Diana J. Rennison. "The availability of research data declines rapidly with article age." Current biology 24, no. 1 (2014): 94-97. <http://doi.org/10.1016/j.cub.2013.11.014>
- Connected Cities for Smart Mobility towards Accessible Resilient Transportation (C2SMART). 2018. Master Data Management Plan. <http://c2smart.engineering.nyu.edu/c2smartpublications/#1534358159826-cb5e9a6e-27fc>

27

[Leighton] On this slide and in the recording transcript you will find a list of resources we discussed in this webinar.

These include:

- United States Department of Transportation. Plan to Increase Public Access to the Results of Federally-Funded Scientific Research. 2015. <https://www.transportation.gov/mission/open/official-dot-public-access-plan-v11>
- United States Department of Transportation, National Transportation Library. DOT Public Access [guidance website]. 2015. <https://ntl.bts.gov/public-access>
- United States Department of Transportation, National Transportation Library. Repository & Open Science Access Portal. 2017. <https://rosap.ntl.bts.gov>
- University Library, Texas A&M University. "Data Management Defined - Research Data Management - Guides at Texas A&M University." Research Data Management, October 1, 2013. <http://guides.library.tamu.edu/DataManagement>
- Briney, Kristin. 2015. Data Management for Researchers: Organize, Maintain and Share your Data for Research Success.
- Vines, Timothy H., Arianne YK Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean-Sébastien Moore, Sébastien Renault, and Diana J. Rennison. "The availability of research data declines rapidly with article age." Current biology 24, no. 1 (2014): 94-97. <http://doi.org/10.1016/j.cub.2013.11.014>
- Connected Cities for Smart Mobility towards Accessible Resilient Transportation (C2SMART). 2018. Master Data Management Plan. <http://c2smart.engineering.nyu.edu/c2smartpublications/#1534358159826-cb5e9a6e-27fc>



Thank you for watching this Special Transportation Librarians Roundtable.
To provide feedback or pose questions related to this TLR, please send an email to public.access@dot.gov
To watch this recording again, or see other webinars on this and other topics, please the TLR Archive at https://rosap.ntl.bts.gov/collection_tlr