

A Deep Dive into Public Access for Research Data

<https://doi.org/10.21949/1522407>

Leighton Christiansen <https://orcid.org/0000-0002-0543-4268>
Data Curator, National Transportation Library (NTL),
Chair, RD&T Public Access Implementation Working Group (PAIWG)
leighton.christiansen@dot.gov

Presented to: Federal Aviation Administration WJHTC/CAMI Library Services for Researchers Event
2021-05-12



U.S. Department of Transportation
Office of the Secretary of Transportation
Bureau of Transportation Statistics

Slide Title: Introduction to the U.S. DOT Public Access Policy & Resources

Speaker text:

Thank you to Mike and Anthony for inviting me to speak to you today.
Today, I will give a more in-depth exploration of the U.S. DOT Public Access Policy, its context, & steps we can take towards implementation of the Public Access.

But first, let's remind ourselves of the definition of "Public Access." Public Access is a very simple concept. By Public Access, we mean that the people of the United States (and the networked world) are able to discover and access U.S. government-funded research outputs, such as reports and datasets, while protecting privacy and security.

So Public Access is simply the practice of a more deliberate research transparency. In recent years, this concept has been applied to research data, as well as the research reports we have been sharing for decades.

Please keep this definition in mind as we move forward.
[Next slide]

[Slide text not presented orally:

Leighton Christiansen <https://orcid.org/0000-0002-0543-4268>
Data Curator, National Transportation Library

leighton.christiansen@dot.gov

Presented to: Federal Aviation Administration WJHTC/CAMI Library Services for
Researchers Event
2021-04-06]

NOTE: There are "Links to resources" slides at the end of the deck for those of you who would like to explore after the workshop. Further, each slide's "speaker notes" section has the complete text of my prepared remarks for each slide.

Contents

- 1. Review of April Presentation**
- 2. US DOT Public Access Plan Data Deep Dive**
- 3. Collecting & Identifying Research**
- 4. Data Management, Data Curation, & Data Science**

Slide Title: Contents

Speaker Notes:

The four broad themes I will speak on are:

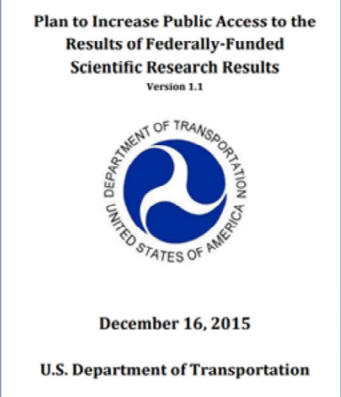
1. Review of April Presentation
2. US DOT Public Access Plan Data Deep Dive
3. Collecting & Identifying Research
4. Data Management, Data Curation, & Data Science

There will be a short pause at various points for questions, as well as some slides that respond to questions submitted in advance.

Let's get right to it.

[Next slide]

Review of 2021-04-05 Presentation



2021-04-05 Slides at:
<https://doi.org/10.21949/1522406>

- 1. How Public Access Came to Be**
 1. Open Science
 2. Federal Policies
- 2. What we Mean by Public Access**
 1. Public is aware of & can locate & analyze research outputs
 2. U.S. DOT Public Access Plan, December 2015:
<https://doi.org/10.21949/1503646>
- 3. Resources**
 1. U.S. DOT Public Access Plan Guidance Website:
<https://doi.org/10.21949/1503647>

3

Slide Title: Review of 2021-04-05 Presentation

Speaker notes:

In April, the presentation covered topics:

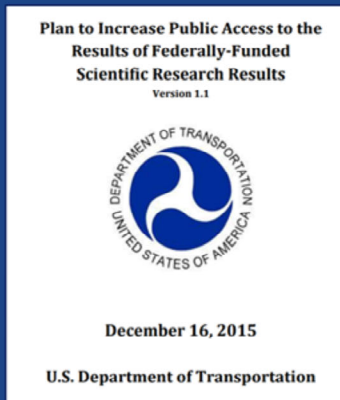
1. How Public Access Came to Be
 1. Open Science
 2. Federal Policies
2. What we Mean by Public Access
 1. Public is aware of & can locate & analyze research outputs
 2. U.S. DOT Public Access Plan: <https://doi.org/10.21949/1503646>
3. Resources
 1. U.S. DOT Public Access Plan Guidance Website:
<https://doi.org/10.21949/1503647>

You can access the slides handout, which includes my speaker notes at the DOI at the top of the slide: <https://doi.org/10.21949/1522406>

That means today we can do a deep dive into the Public Access Plan, especially as it concerns research data.

[Next slide]

Deep Dive:
Public Access Plan:
1. Background &
Purpose



U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

In December 2015, the U.S. DOT published its plan to affirm and enhance DOT's commitment to Public Access to Scientific Research results.

4

Slide Title: Deep Dive: Public Access Plan

Speaker notes: Consistent with the federal laws and policies discussed in the April presentation, in December 2015, the U.S. DOT published its public access plan to affirm and enhance DOT's commitment to Public Access to Scientific Research results, without charge to the maximum extent possible.

This plan went into effect on January 1, 2016, to ensure public access to Publications and Digital Data Sets arising from DOT-managed research and development programs. A further goal is to scale up the sharing of research data that had already been taking place in many DOT research units.

For the next several minutes, we are going to step through the various sections of the plan. I will highlight specific text, and I will answer questions you have submitted as I go.

I know most of you spent the month since I was here last, studying the plan in depth. However, you still may find it useful to have the plan open as we go. You can access the plan at the by using the DOI: <https://doi.org/10.21949/1503646>, which takes you to the plan's archival copy in NTL's Repository and Open Science Access Portal, or ROSA P.

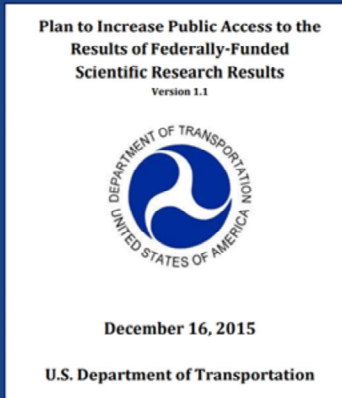
Alright, let's move on to Section 2.

[Next slide]

Slide text not read: U.S. DOT Public Access Plan:

<https://doi.org/10.21949/1503646>

Deep Dive: Public Access Plan: 2. Scope



U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

All Operating Administrations and Secretarial offices will adhere to the following directives:

- **Moving Ahead for Progress in the 21st Century Act: National Transportation Library (MAP-21; P.L. 112-141, July 6, 2012; 49 U.S.C. 6304).**
<https://www.govinfo.gov/content/pkg/BILLS-112hr4348enr/html/BILLS-112hr4348enr.htm>
- **OSTP Memorandum: Increasing Access to the Results of Federally Funded Scientific Research (February 22, 2013).** <https://rosap.ntl.bts.gov/view/dot/34953>
- **OMB Memorandum M-13-13: Open Data Policy - Managing Information as an Asset (May 9, 2013).** <https://rosap.ntl.bts.gov/view/dot/34954>

Slide Title: Deep Dive: Public Access Plan

Speaker notes: Section 2 defines the Scope of the Public Access Plan.

Section 2 starts:

All Operating Administrations and Secretarial offices will adhere to the following directives:

- **Moving Ahead for Progress in the 21st Century Act: National Transportation Library (MAP-21; P.L. 112-141, July 6, 2012; 49 U.S.C. 6304).**
<https://www.govinfo.gov/content/pkg/BILLS-112hr4348enr/html/BILLS-112hr4348enr.htm>
- **OSTP Memorandum: Increasing Access to the Results of Federally Funded Scientific Research (February 22, 2013).**
<https://rosap.ntl.bts.gov/view/dot/34953>
- **OMB Memorandum M-13-13: Open Data Policy - Managing Information as an Asset (May 9, 2013).** <https://rosap.ntl.bts.gov/view/dot/34954>

The important thing to note here is that all DOT operating administrations and secretarial offices are to adhere to the directives, which make up the backbone of the Public Access Plan. In other words, the plan is built around these laws and memos.

I have provided the link to each of these in the slide and in the slide notes. Of course with the updated plan, this list will need to be updated to include things such as the OPEN Government Data Act.

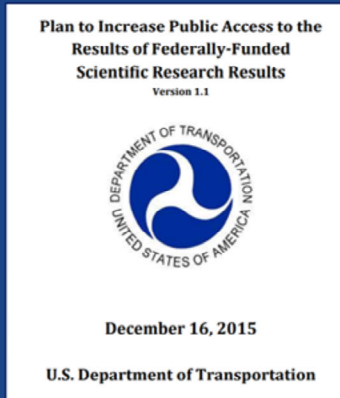
Section 2 also includes important definitions. So let us take some time with the definition of Digital Data Sets.

[Next slide]

Slide text not read: U.S. DOT Public Access Plan:

<https://doi.org/10.21949/1503646>

Deep Dive: Public Access Plan: 2. Definition of Digital Datasets



U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

“Digital Data Sets (sic):” For the purpose of this plan, will be defined as all scientific data collected through research projects funded, either fully or partially, by federal funds awarded through a DOT contract, grant or other agreement or collected by DOT employees. Such scientific data are the digitally recorded factual materials resulting from research that is necessary to validate research findings.

6

Slide Title: Deep Dive: Public Access Plan: 2. Definition of Digital Datasets

Speaker notes: Section 2 defines “Digital Data Sets (sic):” For the purpose of this plan, will be defined as all scientific data collected through research projects funded, either fully or partially, by federal funds awarded through a DOT contract, grant or other agreement or collected by DOT employees. Such scientific data are the digitally recorded factual materials resulting from research that is necessary to validate research findings.

Three things to call out here:

1. If a research project has 1 penny of DOT funding, it comes under the Public Access Plan; and,
2. The last sentence gives a partial answer to a very common question: “the factual materials...necessary to validate research findings.”
3. The definition of Public Access says that the public has the right to access “unless specifically precluded by privacy, confidentiality or National/Homeland security concerns; wherein, Public Access may be restricted to subsets of the Public based on the sensitivity of the Publication and/or Digital Data Set; and, Public Access may be controlled by Institutional Review Boards or other means or terms as necessary.”

We all know that lots data is collected that is not used in final research results.
And this plan specifically scopes the required data as that subset which
validates the research findings.
However, we are often asked to clarify...

[Next slide]

Slide text not read: U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

Question 1: What is, and is not, data? How much data are we supposed to share?

What is Digital Data?

Digitally recorded factual material commonly accepted in the scientific community as necessary to validate research findings.

What is NOT Digital Data?

- Laboratory notebooks;
- Preliminary analyses;
- Drafts of scientific papers;
- Plans for future research;
- Communications with colleagues; or,
- Physical objects, such as laboratory specimens.

See Holdren Memo page 5 at
<https://rosap.ntl.bts.gov/view/dot/34953>

How much Digital Data?

The subset of data collected that is necessary to validate research findings. The data to replicate and validate the research report.

NOT all of the Raw Digital Data, unless...

- Very unique event that would be hard to replicate;
- Obviously has long-term interest to transportation research:
 - Example: 100-Car Naturalistic Driving Study data
<https://vtnews.vt.edu/articles/2005/06/2005-834.html>

7

Slide Title: Question 1: What is, and is not, data? How much data are we supposed to share?

Speaker Notes: Here I have combined and paraphrased 2 questions from folks at FAA. One came after the April presentation, and one just this week.

Let's answer this question:

What is Digital Data? Digital data is the digitally recorded factual material commonly accepted in the scientific community as necessary to validate research findings. These are often known as measurements, observations, or responses. Most science today uses digital computing tools, whether to record sensor readings, traffic counts, aircraft altitude and speed, or survey answers. Most data today is captured digitally. Then digital computing tools are often used, in combination with human analysis, to show trends or suggest correlations. That digital data is then interpreted and presented to interested people in the form of reports or papers or articles or visualizations. The data needed to reproduce and replicate those research conclusions are digital data, according to federal memos, orders, and laws, as well as DOT policy.

What is NOT Digital Data?

- Laboratory notebooks;
- Preliminary analyses;
- Drafts of scientific papers;

- Plans for future research;
- Communications with colleagues; or,
- Physical objects, such as laboratory specimens.

I encourage folks to review the Holdren Memo, especially page 5, at <https://rosap.ntl.bts.gov/view/dot/34953>

Next, How much Digital Data do we need to preserve under the DOT Public Access Plan? The subset of data collected that is necessary to validate research findings. The data to replicate and validate the research report or other representations of findings.

This means, that NOT all of the Raw Digital Data needs to be preserved and shared...UNLESS...and this is a big unless, the data

- Records a very unique event that would be hard to replicate;
- Obviously have long-term interest to transportation research:
 - Example: 100-Car Naturalistic Driving Study data
<https://vtnews.vt.edu/articles/2005/06/2005-834.html>

Now, I will admit that the big unless is not recorded in the Public Access Plan, but you will likely find guidance around this issue in the future iteration of the implementation website.

Further, I am going to refer folks back to the definition of “Public Access” that allows access restrictions based on privacy and security concerns.

This means, there are a number of considerations to make when collecting data, and those need to be dealt with in the research planning stage, because the US Government policy is to default to sharing and openness.

Ok, this was a big deal. I will stop here for 1 clarify question or new question.
[Next Slide]

Slide Text: Question 1: What is, and is not, data? How much data are we supposed to share?

What is Digital Data?

Digitally recorded factual material commonly accepted in the scientific community as necessary to validate research findings.

What is NOT Digital Data?

- Laboratory notebooks;
- Preliminary analyses;
- Drafts of scientific papers;
- Plans for future research;
- Communications with colleagues; or,

- Physical objects, such as laboratory specimens.

See Holdren Memo page 5 at <https://rosap.ntl.bts.gov/view/dot/34953>

How much Digital Data?

The subset of data collected that is necessary to validate research findings. The data to replicate and validate the research report.

NOT all of the Raw Digital Data, unless...

- Very unique event that would be hard to replicate;
- Obviously has long-term interest to transportation research;
- Example: 100-Car Naturalistic Driving Study data
<https://vtnews.vt.edu/articles/2005/06/2005-834.html>

Audience Question Break

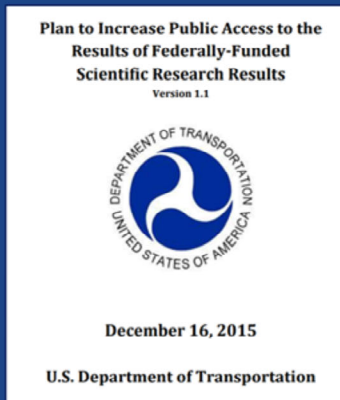


Slide Title: Audience Question Break

Let's take one question and then move on.

[Next slide]

Deep Dive: Public Access Plan: 3. Applicability



U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

This DOT Public Access Plan applies to the following individuals:

- **All DOT employees**, including full- and part-time employees; as well as support service contract employees, consultants and temporary and special government employees.
- **Awardees from non-DOT organizations** that publish Scientific Research material or compile Digital Data Sets resulting from research and development programs conducted **under a DOT grant, contract, or other agreement**.

9

Slide Title: Deep Dive: Public Access Plan: 3. Applicability

Speaker notes: Section 3, Applicability, is very simple, and short. You see most of it on the screen now.

This DOT Public Access Plan applies to the following individuals:

- ALL DOT Employees, no matter how employed or contracted; and,
- All AWARDEES from non-DOT organizations...under a DOT grant, contract, or other agreement.

That is it. The ONLY exception is DOT's Small Business Innovation Research (SBIR) program, as noted in the last sentence of Section 2, directly above.

Alright, let's move on to Section 4.
[Next slide]

Expanded slide note:

This DOT Public Access Plan applies to the following individuals:

- All DOT employees, including full- and part-time employees; as well as support service contract employees, consultants and temporary and special government employees.
- Awardees from non-DOT organizations that publish Scientific Research

material or compile Digital Data Sets resulting from research and development programs conducted under a DOT grant, contract, or other agreement. This includes but is not limited to states, localities, regulated parties, non-profit and volunteer organizations, contractors, cooperative agreement holders, grantees, cooperating federal agencies, intergovernmental organizations, universities and other educational institutions

Slide text not read: U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

Deep Dive: Public Access Plan: 4.2 Data Requirements

Useful Links

Evaluating Repositories:
<https://doi.org/10.21949/1520563>

DOT Conformant Repositories:
<https://doi.org/10.21949/1520566>

Creating DMPs:
<https://doi.org/10.21949/1520562>

ROSA P DMP Collection:
https://rosap.ntl.bts.gov/collection_pa_dmp

U.S. DOT Public Access Plan: <https://doi.org/10.21949/1503646>

There are 6 digital dataset requirements:

1. Stored and publicly accessible for search, retrieval, and analysis;
2. While protecting national/homeland security, individual privacy, and confidentiality.
3. DOT will allow the inclusion of **appropriate costs** for data management and access in funding proposals.
4. All digital datasets inventoried in the DOT Public Data Listing.
5. Researchers must comply with OMB's M-13-13 as well as DOT Order 1351.34.
6. All DOT-funded research proposals must include a **"Data Management Plan" (DMP)**:
 1. Including preservation information or justification for non-preservation;
 2. Including choice of repository that fits DOT specifications;
 3. To be reviewed and approved by OA funding research; and
 4. A sample DMP and guidance will be provided for researchers.

10

Slide Title: Deep Dive: Public Access Plan: 4. Requirements

Speaker notes:

Section 4.2 Data Requirements has 6 major requirements, as you can see on the slide. I am going to focus on two of these:

- Requirement 3: DOT will allow the inclusion of appropriate costs for data management and access in funding proposals.
 - DOT will pay for up to 5 years of storage, but researchers must choose and work with a repository before submitting a proposal, so that they can get a cost estimate. The Public Access Guidance website has help for making that choice. The links are listed in the side box <https://doi.org/10.21949/1520563> and <https://doi.org/10.21949/1520566>
- Requirement 6: Data Management Plans (DMPs) are required. The Plan goes into some detail on DMPs. The Guidance website has extensive help on DMPs including which sections to include, available at the link in the side box: <https://doi.org/10.21949/1520562>

Are you new to DMPs? NTL's ROSA P has an entire collection of DMPs submitted by research programs and projects you can review. Again, you will find the link in the side box https://rosap.ntl.bts.gov/collection_pa_dmp

Discussion of Appropriate Costs segues us into a questions submitted this week.

[Next slide]

Slide text not read:

U.S. DOT Public Access Plan: <https://doi.org/10.21949/1503646>

There are 6 digital dataset requirements:

1. Stored and publicly accessible for search, retrieval, and analysis;
2. While protecting national/homeland security, individual privacy, and confidentiality.
3. DOT will allow the inclusion of appropriate costs for data management and access in funding proposals.
4. All digital datasets inventoried in the DOT Public Data Listing.
5. Researchers must comply with OMB's M-13-13 as well as DOT Order 1351.34.
6. All DOT-funded research proposals must include a "Data Management Plan" (DMP):
 1. Including preservation information or justification for non-preservation;
 2. Including choice of repository that fits DOT specifications;
 3. To be reviewed and approved by OA funding research; and
 4. A sample DMP and guidance will be provided for researchers.

Useful Links

Evaluating Repositories: <https://doi.org/10.21949/1520563>

DOT Conformant Repositories: <https://doi.org/10.21949/1520566>

Creating DMPs: <https://doi.org/10.21949/1520562>

ROSA P DMP Collection: https://rosap.ntl.bts.gov/collection_pa_dmp

Extended Note: Full text of Section 4.2 Data Requirements is as follows:

This plan to the extent feasible and consistent with applicable law and policy; agency mission; resource constraints; U.S. national, homeland and economic security; and the objectives listed below, require digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding to be stored and publicly accessible for search, retrieval, and analysis. This plan requires that awardee(s) and/or the respective Operating Administration ensure Public Access to final research data, subject to the above restrictions and those imposed by data quality and the need to protect national/homeland security, individual privacy, and confidentiality. Ensuring

Public Access may include making such Digital Data Set available to the respective Operating Administration or the DOT for dissemination purposes in keeping with the Departmental Data Release Policy (DOT Order 1351.34). DOT will allow the inclusion of appropriate costs for data management and access in proposals for federal funding for Scientific Research. All Digital Data Sets subject to this plan will be inventoried in the DOT Public Data Listing, whether performed by intramural or extramural researchers.

DOT employees, grantees, contractors, and cooperative agreement awardees must comply with OMB's M-13-13 as well as DOT Order 1351.34 (Departmental Data Release Policy, adopted March 28, 2011). All DOT-funded research proposals, intramural and extramural, must include a supplementary document labeled "Data Management Plan" (DMP). In addition to providing long-term Digital Data Set preservation and storage location information, such DMPs may also discuss why long-term preservation and/or Public Access cannot be justified, if applicable.

In the DMP, researchers will propose their strategy(ies) to deposit Digital Data Sets resulting from DOT-funded Scientific Research in a repository that enables and allows for Public Access and sharing. Such proposed DMP will be reviewed and must be approved by each respective Operating Administration. A sample DMP will be provided as guidance to all extramural researchers.

Question 2. Is there a checklist or criteria associated with the value of a data set to be shared (would it be of value to anyone else) versus the cost associated with curating and making the data set accessible?

1. No. There is no checklist and there should not be.
2. It is not possible to decide how valuable research data is going to be in the future before research is done.
3. US Government and the Open Science movement, believe it is better to spend some resources on data preservation today than mourn lost data.
4. Over time curatorial staff and stakeholders, can decide about decommissioning data.
5. That decision is not to be made before research is even started.
6. OPEN Government Data Act: Section 3562(b) Open by default: we are required by law to plan to share data. <https://www.congress.gov/bill/115th-congress/house-bill/1770/text>

11

Slide Title: Question 2. Is there a checklist or criteria associated with the value of a data set to be shared (would it be of value to anyone else) versus the cost associated with curating and making the data set accessible?

Speaker Notes:

No. There is no checklist and there should not be. We don't know what we don't know, and we don't know what we will want to know. It is not possible to decide how valuable research data is going to be in the future before research is done. Based on this cautious approach, US Government, as well as global governments and research organizations have decided, as part of the Open Science movement, that spending some resources on data preservation today is better than mourning data that was not saved in the future. Therefore, Federal law requires making the public aware of data and sharing it. We don't know what new idea in research, or what change in the natural or economic environments will suddenly make "worthless" data the most useful data of the time. Only over time can curatorial staff, in conversation with stakeholders, make a decision about decommissioning data. That decision is not to be made before research is even started.

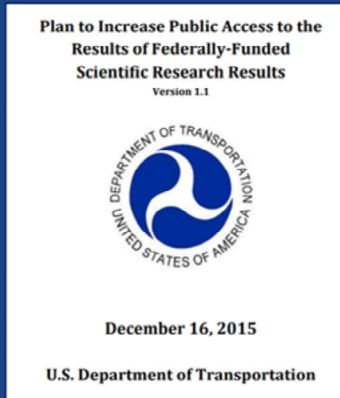
OPEN Government Data Act: Section 3562(b) Open by default: we are required by law to plan to share data, see <https://www.congress.gov/bill/115th-congress/house-bill/1770/text>

[Next Slide]

[Slide text not read]

1. No. There is no checklist and there should not be.
2. It is not possible to decide how valuable research data is going to be in the future before research is done.
3. US Government and the Open Science movement, believe it is better to spend some resources on data preservation today than mourn lost data.
4. Over time curatorial staff and stakeholders, can decide about decommissioning data.
5. That decision is not to be made before research is even started.
6. OPEN Government Data Act: Section 3562(b) Open by default: we are required by law to plan to share data. <https://www.congress.gov/bill/115th-congress/house-bill/1770/text>

Deep Dive: Public Access Plan: 6. Roles & Responsibilities



U.S. DOT Public Access Plan: <https://doi.org/10.21949/1503646>

The Assistant Secretary for Research and Technology:

- Will coordinate the implementation of this plan with OAs.

Heads of DOT OAs and Secretarial Offices:

- Will include the requirements of this plan as terms and conditions for grants, contracts, and other funding agreements

Awardees and Their Institutions:

- Ensure that sub-awardees, researchers and authors are aware of and comply with the DOT Public Access Plan.

Principal Investigators:

- Ensure that all rights under copyright are non-exclusively retained by DOT and that the terms and conditions of publication do not impair the obligation of the authors to comply with the DOT Public Access Plan.

Slide Title: Deep Dive: Public Access Plan: 6. Roles & Responsibilities

Speaker notes: Public access is team sport. As you can see, the Plan calls out 4 groups for actions, these include the Assistant Secretary for Research and Technology; Heads of Operating Administrations and Secretarial Offices; Awardees and research institutions; and, Principle Investigators. I have pulled out 1 require for each to include here, while the plan includes more.

I want to highlight that:

- The Assistant Secretary for Research and Technology: Will coordinate the implementation of this plan with OAs.
- Heads of DOT OAs and Secretarial Offices: Will include the requirements of this plan as terms and conditions for grants, contracts, and other funding agreements
- Awardees and Their Institutions: Will ensure that sub-awardees, researchers and authors are aware of and comply with the DOT Public Access Plan.
- Principal Investigators: Will ensure that all rights under copyright are non-exclusively retained by DOT and that the terms and conditions of publication do not impair the obligation of the authors to comply with the DOT Public

Access Plan.

Trainings like this can help us work together. But we also need DOT research offices to follow through, and check researchers for compliance.

I now want to pause for another question break.

[Next slide]

Slide text not read: U.S. DOT Public Access Plan:

<https://doi.org/10.21949/1503646>

Audience Question Break



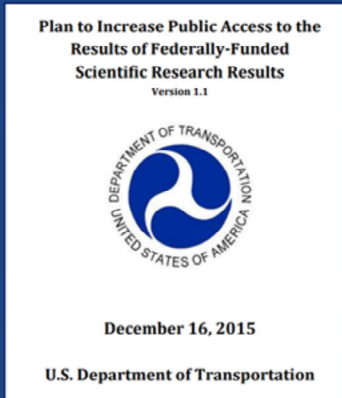
Let's pause for questions from Sections 3, 4, and 6.
I skipped section 5 because it is just a long list of policies and laws, like I showed on the timeline slide in the April presentation.

[PAUSE for a question]

Ok. The next section, Section 7. Implementation, is pretty dense, and we are going to spend some time on it. Let's go.

[Next Slide]

Deep Dive: Public Access Plan: 7. Implementation



U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

Most detailed section. We will focus on:

- **7.2 Before Research Begins**
- **7.3.2 Data Submission**
- **7.4.2 Data Management**
- **7.6.2 Data Preservation**

14

Slide Title: Deep Dive: Public Access Plan: 7. Implementation

Speaker notes: As I mentioned, Section 7. Implementation is a heavy lift, so we will spend some time. I have selected 4 areas to touch on. These are:

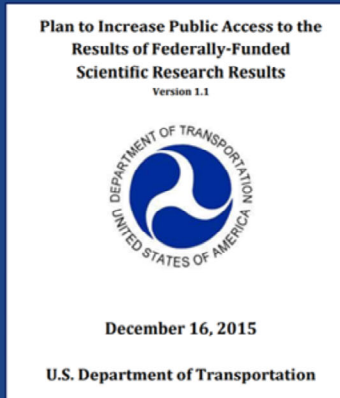
- 7.2 Before Research Begins
- 7.3.2 Data Submission
- 7.4.2 Data Management
- 7.6.2 Data Preservation

Alright, let's move on.

[Next slide]

Slide text not read: U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

Deep Dive: Public Access Plan: 7.2 Before Research Begins



U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

DOT will:

- Establish funding agreements requiring both the immediate grant of a comprehensive non-exclusive, paid-up, royalty-free copyright license to the DOT and the submission of any Publications to the NTL Digital Repository. (DOT DASH 2016-03 and 2016-05)
- Use digital object identifiers (DOI) to individually identify each Publication and Digital Data Set.
- Require all researchers to obtain and report his or her unique ORCID (Open Researcher and Contributor ID).
- Require researchers to include the appropriate funding agreement number(s) on all submissions of research results.

15

Slide Title: Deep Dive: Public Access Plan: 7.2 Before Research Begins

Speaker notes:

DOT will:

- Establish funding agreements requiring both the immediate grant of a comprehensive non-exclusive, paid-up, royalty-free copyright license to the DOT and the submission of any Publications to the NTL Digital Repository. (DOT DASH 2016-03 and 2016-05)
- Use digital object identifiers (DOI) to individually identify each Publication and Digital Data Set.
- Require all researchers to obtain and report his or her unique ORCID (Open Researcher and Contributor ID). <https://orcid.org/>
- Require researchers to include the appropriate funding agreement number(s) on all submissions of research results.

Now, DOT has completed all of these steps. Implementation has been spotty and is still very manual. But we are working on it.

Alright, let's move on to Section 2.

[Next slide]

Slide text not read: U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

[Extended Notes]

7.2 Before Research Begins, reads:

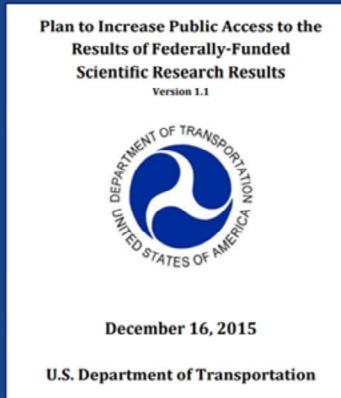
DOT will:

Establish new terms and conditions for all DOT funding agreements requiring both the immediate grant of a comprehensive non-exclusive, paid-up, royalty-free copyright license to the DOT and the submission of any Publications to the NTL Digital Repository. The copyright license must include “all rights under copyright,” including, but not limited to: Right to copy; Right to distribute; Right to prepare derivative works; Right to display; and Right to perform in public. Use digital object identifiers (DOI) to individually identify each Publication and Digital Data Set, to allow for correlation between associated Publications and supporting Digital Data Sets.

Require all researchers to obtain and report his or her unique ORCID (Open Researcher and Contributor ID) identification on submissions of research results to DOT and/or publishers.

Require researchers to include the appropriate funding agreement number(s) on all submissions of research results to DOT and/or publishers.

Deep Dive: Public Access Plan: 7.3.2 Data Submission



U.S. DOT Public Access Plan: <https://doi.org/10.21949/1503646>

- Intramural Research: both OMB's M-1313 and DOT Order 1351.34, Departmental Data Release Policy govern generation, management, and Public Access to digital research data.
- Intramural and Extramural Research: DOT will develop new, standardized requirements for the Data Management Plans (DMPs): at <https://doi.org/10.21949/1520562>

16

Slide Title: Deep Dive: Public Access Plan: 7.3.2 Data Submission

Speaker notes:

Again, DOT has completed the initial steps to implementation, as described in the plan as:

- Intramural Research: both OMB's M-1313 and DOT Order 1351.34, Departmental Data Release Policy govern generation, management, and Public Access to digital research data.
- Intramural and Extramural Research: DOT will develop new, standardized requirements for the Data Management Plans (DMPs): at <https://doi.org/10.21949/1520562>

However, we are far behind in implementation at the OA and Researcher levels.

[Next slide]

Slide text not read: U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

Deep Dive: Public Access Plan: 7.4.2 Data Management

Useful Links

Creating DMPs:
<https://doi.org/10.21949/1520562>

ROSA P DMP Collection:
https://rosap.ntl.bts.gov/collection_pa_dmp

Evaluating Repositories:
<https://doi.org/10.21949/1520563>

DOT Conformant Repositories:
<https://doi.org/10.21949/1520566>

U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

DMPs will include sections that:

1. Describe the data;
2. State standards and file formats;
3. Discuss access policies to protect PII and sensitive info disclosure;
4. State re-use policies; and,
5. State chosen repository, and preservation plan.

See Creating DMPs at <https://doi.org/10.21949/1520562>

Repositories should:

1. Meet essential metadata requirements;
2. Provide persistent identification of datasets; and
3. Provide long-term access.

See Evaluating Repositories at <https://doi.org/10.21949/1520563>

17

Slide Title: Deep Dive: Public Access Plan: 7.4.2 Data Management

Speaker notes: In Section 7.4.2, the DOT plan starts to put more emphasis on research organization and researcher actions. This includes:

Writing DMPs which include sections that:

1. Describe the data;
2. State standards and file formats;
3. Discuss access policies to protect PII and sensitive info disclosure;
4. State re-use policies; and,
5. State chosen repository, and preservation plan.

See Creating DMPs at <https://doi.org/10.21949/1520562>

And choosing data repositories which:

1. Meet essential metadata requirements;
2. Provide persistent identification of datasets; and
3. Provide long-term access.

See Evaluating Repositories at <https://doi.org/10.21949/1520563>

I have included the links to the Guidance webpages on the side box again.

Alright, let's move on to Section .

[Next slide]

Slide text not read: U.S. DOT Public Access Plan:

<https://doi.org/10.21949/1503646>

Useful Links

- Creating DMPs: <https://doi.org/10.21949/1520562>
- ROSA P DMP Collection: https://rosap.ntl.bts.gov/collection_pa_dmp
- Evaluating Repositories: <https://doi.org/10.21949/1520563>
- DOT Conformant Repositories: <https://doi.org/10.21949/1520566>

Question 3. What are the metadata requirements, standards, and resources available?

DOT Public Access Plan specifies Project Open Data metadata schema, now known as DCAT-US

<https://resources.data.gov/resources/dcat-us/>

Datasets should be accompanied by a .json metadata file. Templates can be found at <https://resources.data.gov/resources/podm-field-mapping/>

For other Federal data and metadata tools and training, go to Resources.data.gov at <https://resources.data.gov/>

Slide Title: Question 3. What are the metadata requirements, standards, and resources available?

Speaker Notes: Thank you for submitting this question, as it is easily answered.

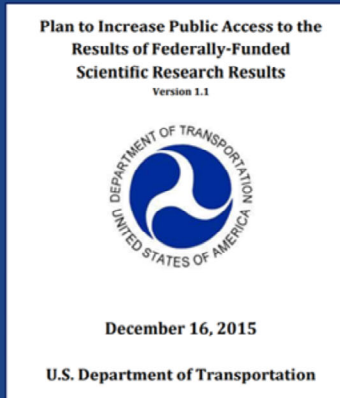
DOT Public Access Plan specifies Project Open Data metadata schema, now known as DCAT-US <https://resources.data.gov/resources/dcat-us/>

Datasets should be accompanied by a .json metadata file. Templates can be found at <https://resources.data.gov/resources/podm-field-mapping/>

For other Federal data and metadata tools and training, go to Resources.data.gov at <https://resources.data.gov/>

Lets move on to Access.
[Next Slide]

Deep Dive: Public Access Plan: 7.6.2 Data Preservation



U.S. DOT Public Access Plan: <https://doi.org/10.21949/1503646>

The DOT will:

- Expand NTL repository to meet trusted digital repository requirements;
- Engage in digital preservation networks;
- Ensure the permanent preservation and long-term accessibility of digital datasets by:
 - Adopting sound preservation standards and archival formats;
 - Developing practical backup, migration, and technology refreshing strategies;
 - Partnering with other appropriate archives;
 - Take into account the relative value of long-term preservation and access of Digital Data Sets against the associated cost and administrative burden.

19

Slide Title: Deep Dive: Public Access Plan: 7.6.2 Data Preservation

Speaker notes:

The DOT will:

- Expand NTL repository to meet trusted digital repository requirements;
- Engage in digital preservation networks;
- Ensure the permanent preservation and long-term accessibility of digital datasets by:
 - Adopting sound preservation standards and archival formats;
 - Developing practical backup, migration, and technology refreshing strategies;
 - Partnering with other appropriate archives;
 - Take into account the relative value of long-term preservation and access of Digital Data Sets against the associated cost and administrative burden.

These are ambitious goals.

We have not met our goal of trusted status yet, because we are not technologically ready, and we need to produce more robust documentation of policies and practices.

However, our 3-2-1 Backup strategy is an industry best practice: 3 copies of

an item; in at least 2 different systems or on 2 different media; and at least 1 copy stored in a totally different geographic region.

Other goals have started on, but mostly with only NTL resources, not with the entire commitment of the DOT, so progress is slow.

Our goals of long-term, or perpetual, preservation raise an important question, submitted by an FAA co-worker.

[Next slide]

Slide text not read: U.S. DOT Public Access Plan:

<https://doi.org/10.21949/1503646>

Question 4. What plans are in place for digital data to preserve it long term because all digital hardware does expire at some point?

The Public Access plan requires digital datasets be shared in open file formats, whenever possible, such as .csv. These open file formats are ubiquitous, long-lived, and less prone to software or hardware obsolescence issues. Other formats have to be monitored over time to watch for obsolescence. When possible, NTL will work with researchers to migrate data into open, preservation friendly formats at the time of submission to NTL. If that is not possible, NTL will, as needed, migrate data from old formats to new formats, using archival best practices.

Slide Title: Question 4. What plans are in place for digital data to preserve it long term because all digital hardware does expire at some point?

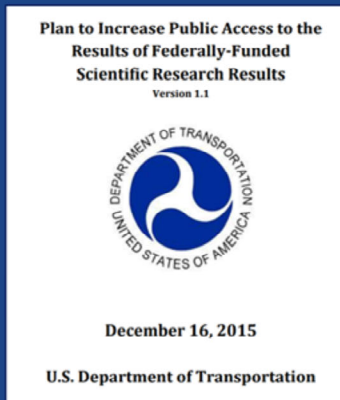
Speaker Notes:

The Public Access plan requires digital datasets be shared in open file formats, whenever possible, such as .csv. These open file formats are ubiquitous, long-lived, and less prone to software or hardware obsolescence issues. Other formats have to be monitored over time to watch for obsolescence. When possible, NTL will work with researchers to migrate data into open, preservation friendly formats at the time of submission to NTL. If that is not possible, NTL will, as needed, migrate data from old formats to new formats, using archival best practices.

Now before we leave the discussion of the implementation, I want to talk about the Public Access Implementation Working Group.

[Next Slide]

Public Access Implementation Working Group (PAIWG)



- Mission: Enable cross-modal collaboration to ensure the best possible public access to USDOT scientific research through implementation of the DOT Public Access Plan, common best practices, and shared resources.
- Scope:
 - USDOT Public Access Plan development, implementation, and compliance monitoring
 - Charters time-limited implementation task forces with modal and OST experts;
 - Reports Public Access Plan progress and obstacles to the RD&T Planning Team, including compliance monitoring; and
 - Coordinates U.S. DOT participation in U.S. Federal, domestic and international Public Access, Open Science, and Data Strategy efforts and activities.

21

Slide Title: Public Access Implementation Working Group (PAIWG)

Speaker notes:

The US DOT's Public Access Implementation Working Group harnesses the energy and talents of about 60 people to ensure the best possible public access to USDOT scientific research through implementation of the DOT Public Access Plan, common best practices, and shared resources.

You can see the mission and the scope on the screen and in the slides, so I won't read those to you.

Currently the PAIWG has task forces working on the updated language for the Public Access Plan, more robust research publication sharing and preservation, and data access and sharing implementation. Anyone in DOT can participate. Please let me know if you would like to.

Let's shift quickly to talking about timelines.
[Next slide]

[Slide text not presented orally:

- Mission: Enable cross-modal collaboration to ensure the best possible public access to USDOT scientific research through implementation of the DOT Public Access Plan, common best practices, and shared resources.

- Scope The Public Access Implementation Working Group (PAIWG):
 - Owns USDOT Public Access Plan development, implementation, and compliance monitoring across all categories of public access outputs, including consistent-facing communications and inputs to implementation support resources;
 - Charters time-limited implementation task forces with modal and OST experts;
 - Reports Public Access Plan progress and obstacles to the RD&T Planning Team, including Operating Administration compliance monitoring once the revised plan is implemented; and
 - Coordinates U.S. DOT participation in U.S. Federal, domestic and international Public Access, Open Science, and Data Strategy efforts and activities, and enables knowledge sharing of these activities with the Department.]

[Expanded background text:

In mid-December 2015, the U.S. DOT published its “Plan to Increase Public Access to the Results of Federally-Funded Scientific Research,” or “Public Access Plan.”

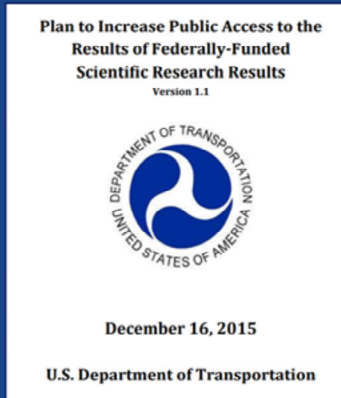
Plan language states that “Public Access” to Publications and Digital Data Sets, will mean:

- The Public is aware of the Digital Data Set holdings and/or the Digital Data Sets generated, fully or partially, through federally funded Scientific Research;
- The Public is able to download and analyze unclassified Publications and/or Digital Data Sets unless specifically precluded by privacy, confidentiality or National/Homeland security concerns;

For the next couple of years implementation of the plan was lead by an ad hoc group chaired by the DOT Office of the Assistant Secretary of Transportation for Research & Technology and the National Transportation Library. In 2018, the Public Access Implementation Working Group (PAIWG) was formally organized to harmonize public access and open science activities across DOT. Then in 2020, PAWIG became a working group of the DOT Research, Development & Technology Planning Team. There are now about 60 DOT employee engaged in the effort to fulfill the PAIWG mission of: enabling cross-

modal collaboration to ensure the best possible public access to USDOT scientific research through implementation of the DOT Public Access Plan, common best practices, and shared resources.]

Deep Dive: Public Access Plan: 8.1 Timeline



U.S. DOT Public Access Plan: <https://doi.org/10.21949/1503646>

- Submission of draft plan to OSTP/OMB/OIRA for review and approval – June 10, 2013;
- Finalize incorporation of OSTP/OMB/OIRA changes; initiate formal concurrence process within DOT – Not greater than six weeks following receipt of OSTP approval;
- Obtain required senior DOT Official signature for implementation – October 1, 2015;
- Begin internal DOT initiatives required for implementation of this Plan – May 1, 2015;
- Commence effective implementation – December 31, 2015;

22

Slide Title: Deep Dive: Public Access Plan: 8.1 Timeline

Speaker notes:

So let's look at the timeline for the Public Access Plan quickly.

- Submission of draft plan to OSTP/OMB/OIRA for review and approval – June 10, 2013;
- Finalize incorporation of OSTP/OMB/OIRA changes; initiate formal concurrence process within DOT – Not greater than six weeks following receipt of OSTP approval;
- Obtain required senior DOT Official signature for implementation – October 1, 2015;
- Begin internal DOT initiatives required for implementation of this Plan – May 1, 2015;
- Commence effective implementation – December 31, 2015;

The main 2 things I want to call out here, are:

1. that this plan began the rounds of the DOT in mid-2013; and,
2. this plan went into effect December 31, 2015.

We are now five and half years beyond that implementation date. Ideally, we should not still be doing introductory session on public access to DOT

research units.

But as I am always the first to point out in any meeting where I can get an ear, the DOT has not provided the resources needed for full socialization of the plan, much less the full implementation. That began to change for the better in 2020, and I hope that continues to change.

However, we still have a long way to go, which leads to another question submitted by FAA.

[Next slide]

Slide text not read: U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

Question 5. What is the time-line we have to provide the access to our output data from the time producing this data to providing access? For existing data sets, will we need to go back and make accessible?

FAA should have been providing Public Access to datasets generated from all research that began on or after January 1, 2016.

Realistically, I recommend FAA prepare to begin full implementation starting on January 1, 2022, or some date soon after.

No, there is no expectation that modes go back and try to make existing dataset, created since January 1, 2016, fully publicly accessible. Legacy work is low-return on investment.

However, FAA may choose to do so for specific, high-value, research data. NTL can help.

3

Slide Title: Question 5. What is the time-line we have to provide the access to our output data from the time producing this data to providing access? For existing data sets, will we need to go back and make accessible?

Speaker Notes:

FAA should have been providing Public Access to datasets generated from all research that began on or after January 1, 2016.

Realistically, I recommend FAA prepare to begin full implementation starting on January 1, 2022, or some date soon after. In my role as PAIWG chair, I will stand by this statement to DOT leadership, if you need me to.

No, there is no expectation that modes go back and try to make existing dataset, created since January 1, 2016, fully publicly accessible. Legacy work is low-return on investment. In my role as PAIWG chair, I will also stand by this statement to DOT leadership, if you need me to.

However, FAA may choose to do so for specific, high-value, research data. NTL can help.

We had a final question from FAA prior to the session.
[Next Slide]

Question 6: What about Software and Code?

Question: Besides data itself we also generate software code or configuration files for a software tool in our various research projects. Does the accessibility laws require making this code accessible? Can the executable code be shared without the actual source code? There is concern that if the source code is published, it could be altered and thus modified in a way not beneficial for the government because we wouldn't know the code was altered and gov't relied on the output as reported?

Response: Yes. The updated plan calls for the sharing of "research computer software," its documentation, and "source code." Of course this sharing is covered by the same cautions around national/organizational security as apply to publications, datasets, and other research outputs.

Government research results, for the most part, is in the public domain. Because of that, we produce things knowing that people will use them in ways we cannot predict. It is accepted risk.

4

Slide Title: Question 6: What about software and code?

Speaker Notes:

Question: Besides data itself we also generate software code or configuration files for a software tool in our various research projects. Does the accessibility laws require making this code accessible? Can the executable code be shared without the actual source code? There is concern that if the source code is published, it could be altered and thus modified in a way not beneficial for the government because we wouldn't know the code was altered and gov't relied on the output as reported?

Response: Yes. The updated plan calls for the sharing of "research computer software," its documentation, and "source code." Of course this sharing is covered by the same cautions around national/organizational security as apply to publications, datasets, and other research outputs.

Government research results, for the most part, is in the public domain. Because of that, we produce things knowing that people will use them in ways we cannot predict. It is accepted risk.

Alright, let's see if there are new questions on sections 7 and 8.

[Next Slide]

Audience Question Break



Let's pause for questions from Sections 7 and 8.

[PAUSE for a question]

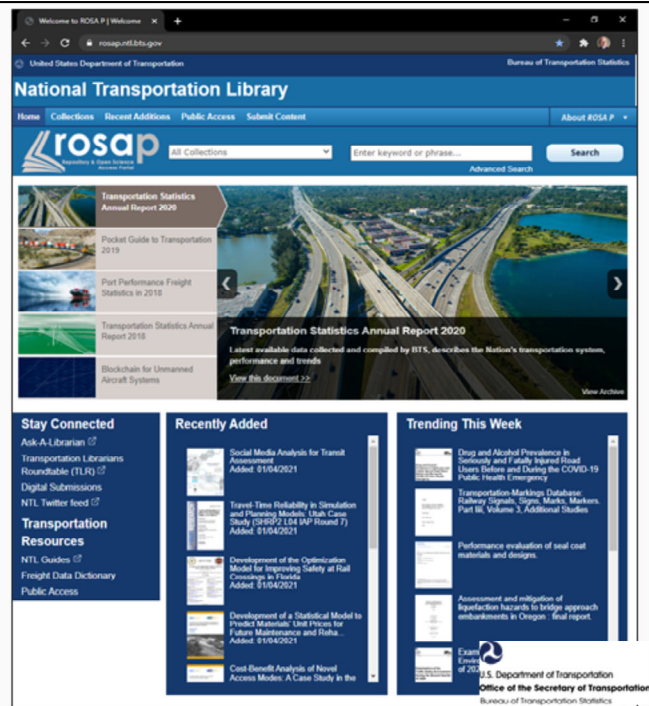
Ok. Now I want to share some implementation guidelines and tools to help you on your way.

[Next Slide]

Collecting & Identifying Research: Repository & Open Science Access Portal (ROSA P)

ROSA P is the National Transportation Library's *Repository and Open Science Access Portal*. The name ROSA P was chosen to honor the role public transportation played in the civil rights movement, along with one of the important figures, Rosa Parks.

Visit ROSA P at:
<https://rosap.ntl.bts.gov/welcome>



26

Slide Title: Repository & Open Science Access Portal (ROSA P)

Speaker notes:

The Repository & Open Science Access Portal, or ROSA P is operated by the National Transportation Library (NTL), and serves as the central repository for DOT research outputs, now numbering in the tens of thousands.

U.S.'s DOT 2015 Public Access Policy requires that a copy of all DOT-funded research reports, and a metadata record for all DOT-funded research datasets, be available to the public through the NTL repository.

ROSA P has many modal research collections, to serve as long-term archives of DOT-funded research, as a supplement to modal websites and databases. Remember 3-2-1 Backups? We can build general and special collections for FAA research, based on topic or facility, as you like. Mary Moulton show how collections worked in her presentation on May 5.

[Next slide]

[Slide text not presented orally:

ROSA P is the National Transportation Library's Repository and Open Science Access Portal. The name ROSA P was chosen to honor the role public transportation played in the civil rights movement, along with one of the important figures, Rosa Parks.

Visit ROSA P at: <https://rosap.ntl.bts.gov/welcome>]

[Expanded background text:

The Repository & Open Science Access Portal, or *ROSA P*, is managed and maintained by my coworkers and I at the National Transportation Library (NTL). The 2012 law, "Moving Ahead for Progress in the 21st Century Act (MAP-21)," requires that NTL's repository "Serve as the central repository for DOT research results and technical publications; and, Serve as the central clearinghouse for transportation data and information of the Federal Government."

Further, U.S.'s DOT 2015 Public Access Policy requires that a copy of all DOT-funded research reports, and a metadata record for all DOT-funded research datasets, be available to the public through the NTL repository.

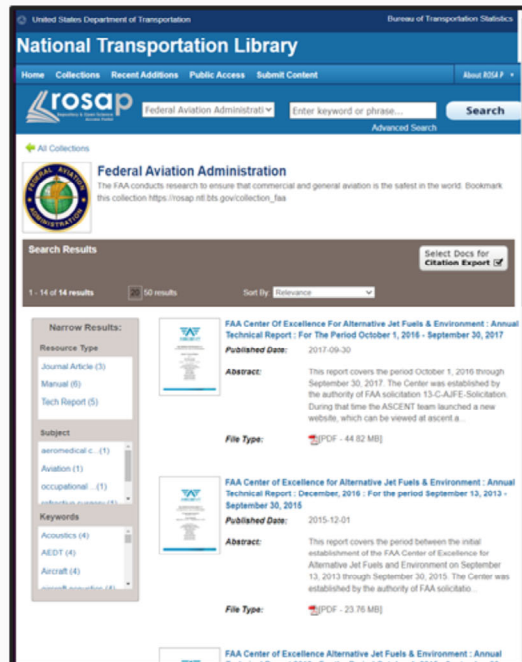
ROSA P is the repository and archive for tens of thousands of DOT-produced or DOT-funded research reports, going back, as we have seen, decades.

]

Collecting & Identifying Research: FAA's Proof of Concept ROSA P Collection

The FAA conducts research to ensure that commercial and general aviation is the safest in the world. Bookmark this collection https://rosap.ntl.bts.gov/collection_faa

Visit ROSA P at:
<https://rosap.ntl.bts.gov/welcome>



27

Slide Title: Repository & Open Science Access Portal (ROSA P)

Speaker notes:

The current FAA collection has just 14 items. But we have cataloged hundreds more in the general collection.

And remember, 1 item can belong in many collections.

[Next slide]

[Slide text not presented orally:

Visit ROSA P at: <https://rosap.ntl.bts.gov/welcome>]

Collecting & Identifying Research: Datasets and Data Packages

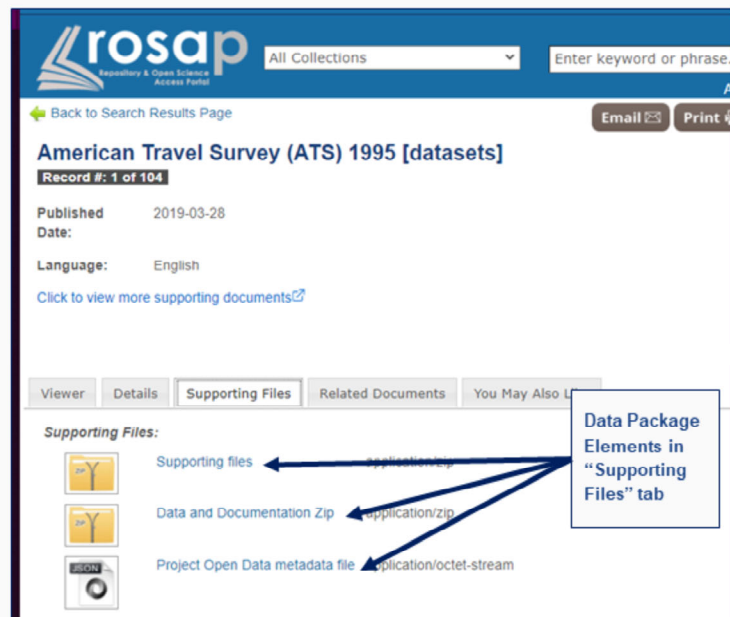
DOT-produced Research Data

DOT-funded Research Data

What is a "data package"?

- Dataset
- README.txt with data dictionary
- DCAT-US Metadata file
- Data management plan (DMP)
- Codes or scripts for analysis
- Supporting files and tables

See <https://doi.org/10.21949/1500456>



Slide Title: Collecting & Identifying Research: Datasets and Data Packages

Speaker Notes: So when we at NTL talk about fully documenting a research digital dataset, we talk about a "data package."

A data package is a robust collection of the dataset and the essential documentation needed to contextualize the data for any future re-user, including the original data collector.

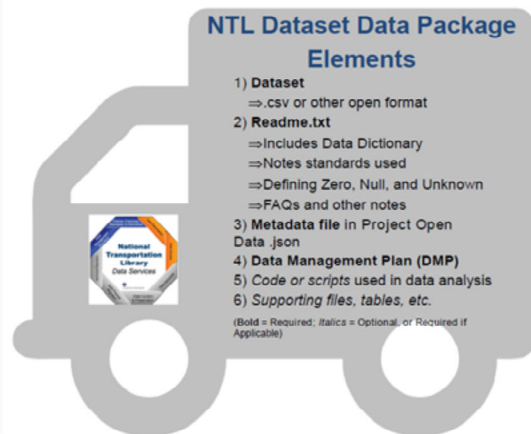
The elements of a good data package include:

- The Dataset
- A README.txt with data dictionary (we have a robust template for this)
- DCAT-US Metadata .json file
- A Data management plan (DMP)
- Codes or scripts for analysis
- Supporting files and tables

For more detail, please see the next slide.

[NEXT SLIDE]

Collecting & Identifying Research: Data Package Publications



Delivering Data Packages for Discovery, Analysis, and Preservation

Leighton Christiansen

<https://doi.org/10.21949/1500456>

Data Management Strategies for the National Transportation Data Archive: Dealing with Legacy Data

Jesse Long

<https://doi.org/10.21949/1506098>

29

Slide Title: Collecting & Identifying Research: Data Package Publications

Speaker Notes:

For those interested in more quick reading on data packages and their application, I suggest the work that NTL Data Curation Fellow Jesse Long, and I, have presented at the TRB Annual Meeting over the years.

These are:

Delivering Data Packages for Discovery, Analysis, and Preservation

Leighton Christiansen

<https://doi.org/10.21949/1500456>

Data Management Strategies for the National Transportation Data Archive: Dealing with Legacy Data

Jesse Long

<https://doi.org/10.21949/1506098>

Given the requirements of the Public Access Plan, NTL also expects to get data packages for dataset from outside researchers funded by the DOT.

[Next Slide]

Collecting & Identifying Research: External Datasets

DOT-funded Research Data held in third-party repository

- NTL Policy: We hold a local copy against loss, and link to external repository in metadata and ROSA P
 - DOT paid for it; DOT must hold a copy (if not too large)
- NTL librarians create "metadata data cover sheet" from external repository metadata
- NTL librarians relate datasets to reports, but this manual process and can be out of sync

The screenshot shows the ROSA P website with a search bar at the top. The main content area displays a record for 'Integrated Implementation of Innovative Intersection Designs [supporting datasets]'. It includes a 'Download' button and a 'Viewer' tab. A blue arrow points from the text 'NTL-created "data cover sheet" for Viewer Tab' to the 'Viewer' tab. Another blue arrow points from the text 'Includes DOI to external repository landing page' to the DOI link 'https://doi.org/10.5281/zenodo.3887540'.

Slide Title: Collecting & Identifying Research: External Datasets

Speaker Notes: The previous mention of data packages from external researchers leads me to this next topic on how we handle external datasets.

The US DOT Public Access Plan requires that the NTL repository hold a copy of every research report or technical report that is created from USDOT-produced or –funded research. And Mary talked about that last week.

With datasets, the Plan only requires that the NTL repository collect the metadata about datasets. This is useful for as us, as some large datasets are beyond our repository capacity at this moment. So NTL creates a metadata "cover sheet" record that points to the repository where the external dataset sits. We look for the data package elements there as well. And in the future, as we are better able to track compliance of all DOT researchers with Public Access, we will expect to see complete data packages in these repositories.

One other thing we do at NTL is to download a copy of the dataset and its documentation, to hold as a local copy, to guard against loss of the tax payer and DOT investment. Remember: 3-2-1 backup!

Now, how do we find all of this data spread across the web? If we are good data stewards, we use persistent identifiers, or in our case, Digital Object Identifiers (DOIs)

Let's look at those next.

[Next slide]

[Slide text not read]

DOT-funded Research Data held in third-party repository

- NTL Policy: We hold a local copy against loss, and link to external repository in metadata and ROSA P
 - DOT paid for it; DOT must hold a copy (if not too large)
- NTL librarians create “metadata data cover sheet” from external repository metadata
- NTL librarians relate datasets to reports, but this manual process and can be out of sync

Collecting & Identifying Research: Digital Object Identifiers (DOIs)

- Persistent and Unique identifier for any object that can be described in a computerized (digital) environment
- NTL DOIs lead to landing pages, they do NOT trigger downloads
- NTL registers DOIs with DataCite through contract with DOE OSTI
- NTL mints and supplies to any mode which asks for inclusion in publications and metadata

ROSA P records 2021-12-11: 51,655
ROSA P records with DOIs: 11,310
NTL-minted DOIs: 11,099
External DOIs: 211

U.S. Open Science Policy Perspectives & Transportation: Open Science in Transportation: Challenges and Opportunities in a COVID-19 Era
 Record #: 3 of 86
 Published Date: 2021-01-21
 Language: English

Details:
 Creators: Christiansen, Leighton L.
 Corporate Contributors: United States. Department of Transportation. National Transportation Library
 Subject/TRM Terms: [+]
 DOI: <https://doi.org/10.21949/1520725>
 Resource Type: Presentations

NTL-minted DOI. Could also be external DOI

Slide Title: Collecting & Identifying Research: Digital Object Identifiers (DOIs)

Speaker Notes: An industry best practice for identifying and tracking research outputs, is through the use of persistent identifiers or PIDs.

PIDs come in many flavors, to suite many different purposed. At NTL we use Digital Object Identifiers (DOIs) to uniquely and persistently identify publications and datasets.

Because DOIs work in a networked computing environment, the best semantic approach is to think of DOIs as “digital identifiers for objects.” These objects can be physical or virtual, as long as they can be represented in some way in a computerized environment. For example, the Eiffel Tower can have a DOI. That DOI would lead to a webpage that gives information about the Eiffel Tower.

So,

- Persistent and Unique identifier for any object that can be described in a computerized (digital) environment
- NTL DOIs lead to landing pages, they do NOT trigger downloads

- NTL registers DOIs with DataCite through contract with DOE OSTI
- NTL mints and supplies to any mode which asks for inclusion in publications and metadata

You will have noticed that NTL DOIs lead to landing pages. This is industry best practice, in case the dataset does get lost or deprecated, the public can still find the metadata about the data, which conforms to US law around government data.

Ok, so let us stop here and see if there are other questions.
[Next slide]

Audience Question Break



Let's pause for questions from tools and resources.

[PAUSE for a question]

Finally, I want to do a little more context setting around Data Management, Data Curation, & Data Science.

[Next Slide]

Data Management, Data
Curation, & Data
Science:
**Expected
Outcomes**



Image from UNESCO:
<https://unesdoc.unesco.org/ark:/48223/pf0000374837>

U.S. DOT Public Access Plan:
<https://doi.org/10.21949/1503646>

Expected Outcomes include:

- **Long-term access to, & preservation of, research;**
- **Enhanced scientific discovery and deployment; &**
- **Promotion of scientific & economic innovation.**

33

Slide Title: Data Management, Data Curation, & Data Science:: **Expected Outcomes**

Speaker notes:

So let us return to a question I posed in the April presentation: Why are we doing this? What do we hope to get out of Public Access and Open Science?

Keep in mind that the guidelines and requirements of Public Access and Open Science can be seen as the implementation of research data management best practices, along with the commitment to share data as global asset.

With that approach to data management and sharing, we expect:

- Long-term access to, & preservation of, research;
- Enhanced scientific discovery and deployment; &
- Promotion of scientific & economic innovation.

We think it does work. Data sharing was key to the development of the COVID-19 vaccines. And there are other global examples.

But I want to explore next how the practices of public access can benefit FAA

in an unexpected way.
[Next slide]

Data Management, Data Curation, & Data Science: Definitions



Image from UNESCO:
<https://unesdoc.unesco.org/ark:/48223/pf0000374837>

Data Management (DM):
deliberate planning, creation, storage,
access and preservation of data produced
from a given investigation^{1,2}

34

Slide Title: Data Management, Data Curation, & Data Science: Definitions (1)

Speaker notes:

Let's start with defining Data Management:

"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."

1: 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."

2: Source: Briney, Kristin. 2015. Data management for researchers: organize, maintain and share your data for research success. (6)

Now, let's take the next step.

[Next slide]

Data Management, Data Curation, & Data Science: Definitions



Image from UNESCO:
<https://unesdoc.unesco.org/ark:/48223/pf0000374837>

Data Management (DM):

deliberate planning, creation, storage, access and preservation of data produced from a given investigation

Data Curation (DC):

enables data discovery and retrieval, maintains data quality, adds value, and provides for re-use over time

35

Slide Title: Data Management, Data Curation, & Data Science: **Definitions** (2)

Speaker notes:

Now Data Curation:

“Data curation is the active and ongoing management of data through its lifecycle of interest and usefulness to scholarship, science, and education. Data curation enables data discovery and retrieval, maintains data quality, adds value, and provides for re-use over time through activities including authentication, archiving, management, preservation, and representation.”
And re-use in unexpected ways!

3: Source: Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. “Specialization in Data Curation,” 2013.

http://www.lis.illinois.edu/academics/programs/specializations/data_curation.

Finally....

[Next slide]

Data Management, Data Curation, & Data Science: Definitions



Image from UNESCO:
<https://unesdoc.unesco.org/ark:/48223/pf0000374837>

Data Management (DM):

deliberate planning, creation, storage, access and preservation of data produced from a given investigation

Data Curation (DC):

enables data discovery and retrieval, maintains data quality, adds value, and provides for re-use over time

Data Science (DS):

drawing useful conclusions from large and diverse data sets through exploration, prediction, and inference

36

Slide Title: Data Management, Data Curation, & Data Science: Definitions (3)

Speaker notes:

... Data Science:

“Data Science is about drawing useful conclusions from large and diverse datasets through exploration, prediction, and inference, using the skills and practices of statistics, information science, and computer programming.”

4: Based on: <http://www.inferentialthinking.com/chapter1/what-is-data-science.html>

Notice the definition said “diverse datasets.” We are saving data today because we do not know how it might be useful tomorrow.

As we talk about data management and data curation, you will likely recognize your own work touches on many of the actions described in these definitions. What may be missing from your current practice is seeing these individual actions as part of a holistic strategy for sharing and preserving data.

We can visualize this holistic strategy in the linked processes of

[Next slide]

Data Management, Data
Curation, & Data
Science:
Linked
Processes



Image from UNESCO:
<https://unesdoc.unesco.org/ark:/48223/pf0000374837>

DM is a **Necessary** element of **DC**

Data Management \in Data Curation

DC Enables robust **DS**

Data Curation \Rightarrow Data Science

37

Slide Title: Data Management, Data Curation, & Data Science: Linked Processes

Speaker notes:

Here I attempt to illustrate the interconnectedness of skills that we hope to harness to improve the transparency of federal data and statistics.

Data management is a necessary element of data curation. And to enable good data curation, it often means that we have to encourage researchers and data collections to think beyond a specific investigation or survey, and adapt good data management practices to meet future needs.

Good Data curation, in turn, enables broader, longitudinal Data Science. By preserving and adding value to data, data curation makes the task of data science more efficient and effective, as well as opening new output possibilities.

Let us go a step further.

[Next slide]

Data Management, Data
Curation, & Data
Science:
Linked
Processes



Image from UNESCO:
<https://unesdoc.unesco.org/ark:/48223/pf0000374837>

DM is a **Necessary** element of **DC**

Data Management \in Data Curation

DC Enables robust **DS**

Data Curation \Rightarrow Data Science

38

Slide Title: Data Management, Data Curation, & Data Science: Linked Processes

Speaker notes:

Here I attempt to illustrate the interconnectedness of skills that we hope to harness to improve the transparency of federal data and statistics.

Data management is a necessary element of data curation. And to enable good data curation, it often means that we have to encourage researchers and data collections to think beyond a specific investigation or survey, and adapt good data management practices to meet future needs.

Good Data curation, in turn, enables broader, longitudinal Data Science. By preserving and adding value to data, data curation makes the task of data science more efficient and effective, as well as opening new output possibilities.

Let us go a step further.

[Next slide]

Data Management, Data Curation, & Data Science Dependencies Model

Data Management \in Data Curation \Rightarrow Data Science

DM \in DC \Rightarrow DS

Slide Title: Data Management, Data Curation, & Data Science Dependencies Model

Speaker Notes: To visualize these dependencies altogether: DM is a *necessary element* of DC which *enables* DS. As far as I know, the above dependency model is original to me, beginning in July 2016, but I have yet to publish on it.

Soapboxing time: The reason I think that talking about these linked processes is that it helps to provide important context for public access policies, and laws that require federal research be more transparent: there is a hope that by opening these datasets more broadly to the research, business, scientific, policy, and public communities, that new discoveries can be made by current and future data scientists.

So we as data collectors of today have a responsibility to the data users who follow us. And those data users may even be us!! [Wait we can help our future selves? Amazing!!]

That responsibility can be summed up as we should use the best resources and practices at our disposal to steward those data and statistics into the future, for as long as they will be of interest. And given the nature of many federal transportation research projects, the period of interest may extend decades or longer.

So let's wrap up.
[Next Slide]

Review

- 1. Review of April Presentation**
- 2. US DOT Public Access Plan Data Deep Dive**
- 3. Collecting & Identifying Research**
- 4. Data Management, Data Curation, & Data Science**



Slide Title: Contents

Speaker Notes:

Today we:

1. Reviewed the high points of the April Presentation
2. Took a deep dive into the US DOT Public Access Plan around datasets
3. Introduced tools for Collecting & Identifying Research
4. Got into theory with Data Management, Data Curation, & Data Science

[Next slide]

Resources:
Contact Info

Leighton Christiansen
Data Curator,
National Transportation Library
leighton.christiansen@dot.gov
202-366-2759

41

Slide Title: Resources: Contact Info

In conclusion,

The themes I spoke on today were:

1. How Public Access Came to Be
2. What we Mean by Public Access
3. Resources

There are many more resources on the Supplemental Slides of this deck, which will be shared with you after the presentation.

You can always reach out to me with other question.

You may see and email from Mike and Anthony ahead of the May 12 presentation, asking for your questions ahead of time, so that we can target that presentation.

[Next slide]

Thank you!

Leighton Christiansen <https://orcid.org/0000-0002-0543-4268>
Data Curator, National Transportation Library (NTL),
Bureau of Transportation Statistics (BTS), Office of the Assistant
Secretary for Research and Technology (OST-R); U.S. Department of
Transportation (U.S. DOT)
leighton.christiansen@dot.gov



U.S. Department of Transportation
Office of the Secretary of Transportation
Bureau of Transportation Statistics

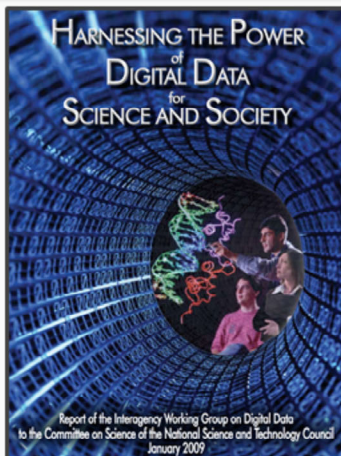
Supplemental Slides

The following Supplemental Slides include further resources for those interested.

Slide Title: Supplemental Slides

Slide text: The following Supplemental Slides include further resources for those interested.

Opening U.S. Government-Funded Science: Practices



Guiding Principles

- Science is global and thrives in the digital dimensions;
- Digital scientific data are national and global assets;
- Not all digital scientific data need to be preserved and not all preserved data need to be preserved indefinitely;
- Communities of practice are an essential feature of the digital landscape;
- Preservation of digital scientific data is both a government and private sector responsibility and benefits society as a whole;
- Long-term preservation, access, and interoperability require management of the full data life cycle; and
- Dynamic strategies are required

<https://www.nitrd.gov/Publications/PublicationDetail.aspx?pubid=25>



U.S. Department of Transportation
Office of the Secretary of Transportation
Bureau of Transportation Statistics

44

Slide Title: Opening U.S. Government-Funded Science: Practices

Speaker notes: The impacts of digital technology on science are fundamental, and were summarized by the Interagency Working Group on Digital Data with their 2009 report “Harnessing the Power of Digital Data for Science and Society.”

The authors note that science will now be conducted in a “fully digital world” and that data is “an endless fuel for creativity.”

The authors list seven guiding principles for the new research reality. Among these are:

- Digital scientific data are national and global assets;
- Communities of practice are an essential feature of the digital landscape;
- Preservation of digital scientific data is both a government and private sector responsibility and benefits society as a whole.

The evolution in research and digital data made the new U.S. policies necessary in order for U.S.-funded researchers to keep pace.

However, new Open Science policies and practices are only fully realized if they are implemented. Let us look at two groups engaged in implementation.

[Next slide]

[Text of slide not presented orally:

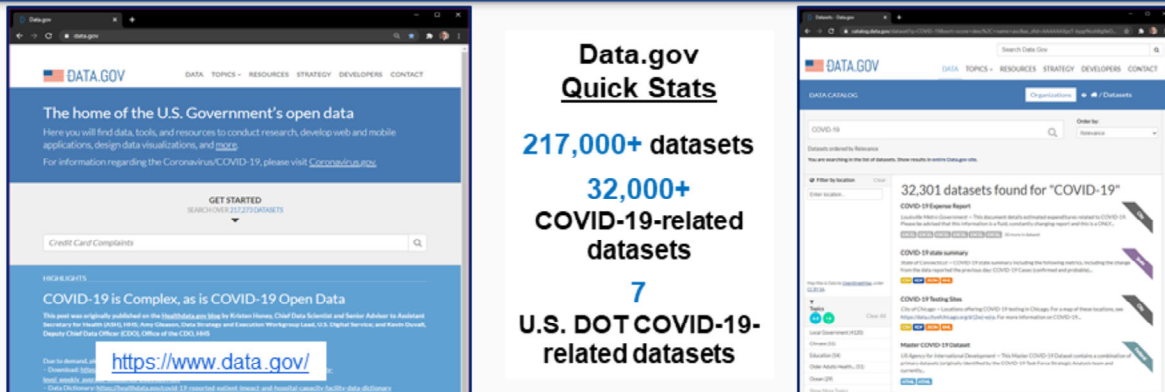
Full list of Guiding Principles:

- Science is global and thrives in the digital dimensions;
- Digital scientific data are national and global assets;
- Not all digital scientific data need to be preserved and not all preserved data need to be preserved indefinitely;
- Communities of practice are an essential feature of the digital landscape;
- Preservation of digital scientific data is both a government and private sector responsibility and benefits society as a whole;
- Long-term preservation, access, and interoperability require management of the full data life cycle; and
- Dynamic strategies are required

URL to “Harnessing the Power of Digital Data” report:

<https://www.nitrd.gov/Publications/PublicationDetail.aspx?pubid=25>]

Opening U.S. Government-Funded Science: Technology: Data.gov



Want just the DOT data in data.gov?
<https://catalog.data.gov/organization/dot-gov>

U.S. Department of Transportation
Office of the Secretary of Transportation
Bureau of Transportation Statistics

45

Slide Title: Opening U.S. Government-Funded Science: Technology: Data.gov

Speakers notes;

As U.S. agencies open their data, we have seen a number of new technologies deployed, including data portals such as DATA.gov, in May 2009. Data.gov harvests its information from agency data inventories, giving the public a “one-stop federal-shop.”

As of January 2021, it indexed more than 217,000 datasets.

As of January 6, data.gov is linking to more than 32,000 COVID-19-related datasets.

- Of these, more than 23,000 were federal government data.

A couple of notes about Data.gov’s holdings:

- Not all 217,000 records link to machine-readable datasets.
 - Some of the “datasets” are PDFs of data tables
- Not all data indexed by Data.gov is publicly accessible.
 - Data’s metadata is discoverable, but access may be limited for privacy or security concerns.

And if you want to search just the U.S. DOT datasets in data.gov, you can use this specific link <https://catalog.data.gov/organization/dot-gov>

Speaking of DOT data, let’s now take a quick look at DOT’s data inventory

[Next slide]

[Expanded background text:

- As U.S. agencies, following policy leads we just reviewed, have been moving forward towards data openness, and we have seen a number of new technologies deployed, including data portals such as DATA.gov. Data.gov was launched in May 2009 in order to increase public access to data across all federal agencies. Data.gov harvests its information from all governmental agency enterprise data inventories, giving the public a “one-stop shop” for government data. As of January 2021, Data.gov lists more than 217,000 datasets.
- As of January 6, 2021, data.gov is linking to more than 32,000 COVID-19-related datasets.
 - Of these, more than 23,000 were federal government data, nearly 5,000 from U.S. states, and more than 2300 from city governments .
 - 7 of the U.S. DOT datasets have data related to COVID-19, coronavirus, or pandemic, by search term. There are likely many more, but they may not have metadata that indicates data during the COVID-19 period.
- A couple of notes about Data.gov’s holdings:
 - Not all 217,000 records link to machine-readable datasets.
 - Some of the “datasets” are PDFs of data tables
 - Not all data indexed by Data.gov is publicly accessible.
 - U.S. law requires that citizens are able to discover the metadata about federal data. However, access may be limited for personal and business privacy or national security concerns.
 - And if you want to search just the U.S. DOT datasets in data.gov, you can use this specific link <https://catalog.data.gov/organization/dot-gov>

As I mentioned, Data.gov harvests metadata and indexes federal data inventories. So let’s now take a quick look at DOT’s data.transportation.gov.]

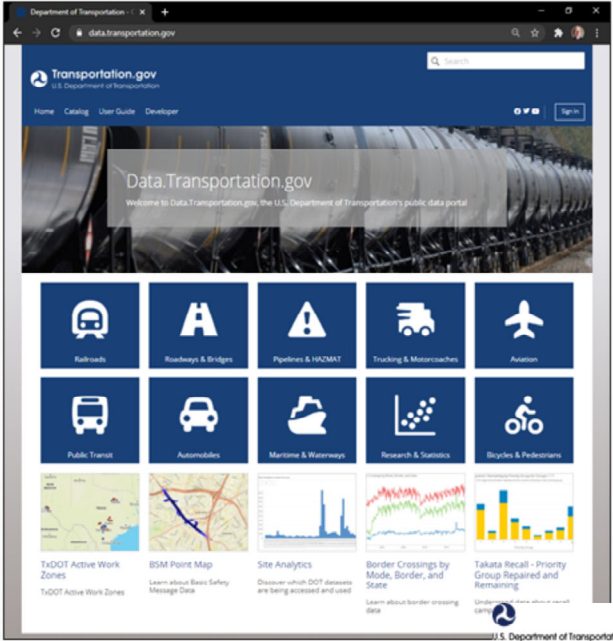
U.S. DOT's Open Data

Data.transportation.gov

Highlights:

- 4000+ datasets
- All transport modes
- Visualization tools
- Data management best practices:
 - Machine-readable datasets and subsets
 - Open formats
 - API access

<https://data.transportation.gov/>



Slide Title: U.S. DOT's Open Data: Data.transportation.gov

Speaker notes:

Data.gov harvests metadata from DOT's DATA.TRANSPORTATION.gov, a data catalog, warehouse, and visualization suite.

Most of the 4000-plus datasets are available for public download.

Some of the highlights include:

- All transportation modes are represented
- Data visualization tools are built in

We should now visit DOT's open scientific research report portal.

[Next slide]

[Slide text presented orally:

Be sure to visit <https://data.transportation.gov>

List of data.transportation.gov highlights

Highlights:

- 4000+ datasets
- All transport modes

- Visualization tools
- Data management best practices:
 - Machine-readable datasets and subsets
 - Open formats
 - API access]

[Expanded background text:

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. U.S. DOT launched its first data inventory in September 2010. Data.transportation.gov now utilizes the data visualization platform Socrata. As of January 2021, Data.transportation.gov contains records of more than 4000 datasets. Most of these are available to the public for download.

Some of the highlights of data.transportation.gov include:

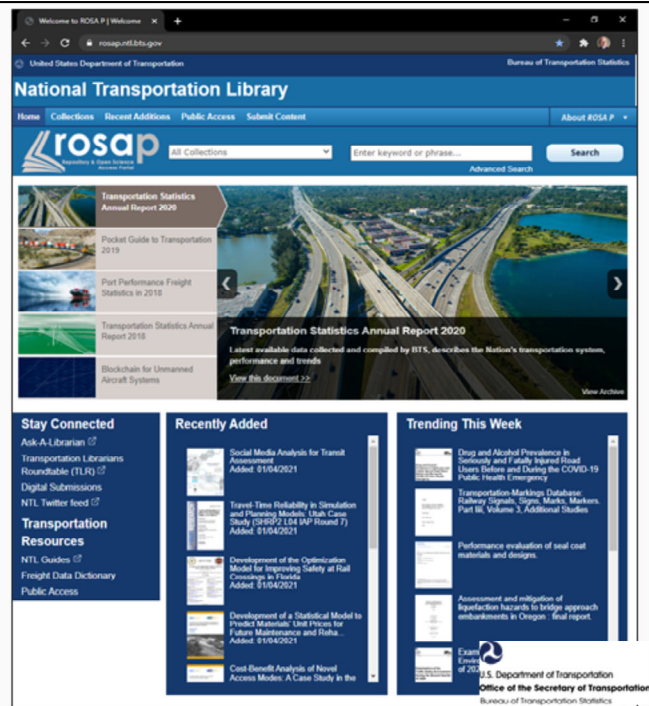
- There are currently over 4000 datasets
- All transportation modes are represented
- Data visualization tools are built into the user interface
- The new system meets several data management best practices including:
 - Allowing users to download Machine-readable datasets and subsets
 - Downloaded data is in open formats
 - The interface also allows for API access

Be sure to visit <https://data.transportation.gov>]

Repository & Open Science Access Portal (ROSA P)

ROSA P is the National Transportation Library's *Repository and Open Science Access Portal*. The name ROSA P was chosen to honor the role public transportation played in the civil rights movement, along with one of the important figures, Rosa Parks.

Visit ROSA P at:
<https://rosap.ntl.bts.gov/welcome>



47

Slide Title: Repository & Open Science Access Portal (ROSA P)

Speaker notes:

The Repository & Open Science Access Portal, or ROSA P is operated by the National Transportation Library (NTL), and serves as the central repository for DOT research outputs, now numbering in the tens of thousands.

And there are already 6 research outputs that deal specifically with COVID-19.

But what if you up-to-date info on how COVID is impacting transportation?

[Next slide]

[Slide text not presented orally:

ROSA P is the National Transportation Library's Repository and Open Science Access Portal. The name ROSA P was chosen to honor the role public transportation played in the civil rights movement, along with one of the important figures, Rosa Parks.

Visit ROSA P at: <https://rosap.ntl.bts.gov/welcome>]


[Expanded background text:

The Repository & Open Science Access Portal, or *ROSA P*. is managed and maintained by my coworkers and I at the National Transportation Library (NTL). The 2012 law, “Moving Ahead for Progress in the 21st Century Act (MAP-21),” requires that NTL’s repository “Serve as the central repository for DOT research results and technical publications; and, Serve as the central clearinghouse for transportation data and information of the Federal Government.”

Further, U.S.’s DOT 2015 Public Access Policy requires that a copy of all DOT-funded research reports, and a metadata record for all DOT-funded research datasets, be available to the public through the NTL repository.

ROSA P is the repository and archive for tens of thousands of DOT-produced or DOT-funded research reports, going back, as we have seen, decades. And there are already about 6 reports, statistical publications, or other research outputs that deal specifically with COVID-19.]


COVID-19 Transportation Statistics from BTS




<https://www.bts.dot.gov/covid-19>

COVID-19 Related Statistics

- Daily Travel During the COVID-19 Public Health Emergency
- Mobility Over Time by State and By Trip Distance
- The Week in Transportation: Selected Measures During COVID-19
- Monthly Transportation Statistics
- County Transportation Profiles
- Daily Vehicle Travel
- Effects of COVID-19 On Travel Behavior
- Effects of COVID-19 On Travel Behavior by Income Groups
- Effects of COVID-19 On Bikeshare and E-Scooter Operations
- Docked Bikeshare Ridership: COVID-19 Effects
- Ferry Operators Status
- Ferry Routes for Top Ten Operators





48

Slide Title: COVID-19 Transportation Statistics from BTS

Speaker notes: The Bureau of Transportation Statistics (BTS) publishes stats on aviation, freight activity, and transportation economics, as well the effects of COVID-19 on travel and shipping, including:

- Travel Behavior by Income Groups
- Bikeshare and E-Scooter Operations

Let us next turn our attention to federal Open Science resources.

[Next slide]

[Slide text not presented orally:

COVID-19 Related Statistics:

- Daily Travel During the COVID-19 Public Health Emergency
- Mobility Over Time by State and By Trip Distance
- The Week in Transportation: Selected Measures During COVID-19
- Monthly Transportation Statistics
- County Transportation Profiles
- Daily Vehicle Travel
- Effects of COVID-19 On Travel Behavior

- Effects of COVID-19 On Travel Behavior by Income Groups
- Effects of COVID-19 On Bikeshare and E-Scooter Operations
- Docked Bikeshare Ridership: COVID-19 Effects
- Ferry Operators Status
- Ferry Routes for Top Ten Operators

<https://www.bts.dot.gov/covid-19>]

[Expanded background text:

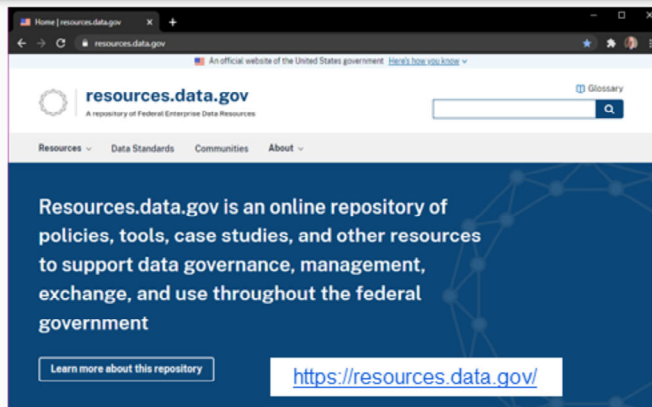
The Bureau of Transportation Statistics (BTS) is one of the 13 principle federal statistical agencies. BTS is the preeminent source of statistics on commercial aviation, multimodal freight activity, and transportation economics, and provides context to decision makers and the public for understanding statistics on transportation.

Responding to interest in the most recent coronavirus-related data, BTS has created web pages of transportation statistics allowing comparison of pre-COVID-19 and current numbers for passenger travel and freight shipments.

These pages present a wide range of data on all transportation modes from various sources. Some of these pages are update daily, weekly, or as data becomes available from the numerous providers BTS works with. To visit the COVID-19 related pages, go to <https://www.bts.dot.gov/covid-19>

Among the list of regularly updated pages of COVID-19 Related Statistics, are:
(see list above)]

Opening U.S. Government-Funded Science: Resources.data.gov



Some Available Resources:

- DCAT-US Schema v1.1 (Project Open Data Metadata Schema)
- Principles of Open Government Data
- Data Ethics Framework
- Geoportal Server
- JSON Validator
- Digital Analytics Program (DAP)
- Improving Agency Data Skills Playbook
- Case studies & examples

Slide Title: Opening U.S. Government-Funded Science: Resources.data.gov

Speaker notes: Resources.data.gov is a repository of policies, tools, case studies, and resources to support data governance, management, and use throughout the U.S. government.

Some of the available resources include:

- The DCAT-US Schema v1.1
- The Data Ethics Framework
- And Case studies & examples

As we near the end of this trip, let's turn to the challenges we face around open science.

[Next slide]

[Slide text not presented orally:

Some of the available resources include:

- The DCAT-US Schema v1.1 (Project Open Data Metadata Schema)
- The Principles of Open Government Data
- The Data Ethics Framework
- A Geoportal Server
- A JSON Validator

- Digital Analytics Program (DAP) for measuring use metrics
- An Improving Agency Data Skills Playbook
- And Case studies & examples

<https://resources.data.gov/>]

Science.gov

- Interagency federated search
- Focused COVID-19 search
- Results include:
 - Journal articles
 - Technical reports
 - Datasets
 - Conference papers
 - Videos
 - Audio files
 - Images

<https://www.science.gov/>



Science.gov Alliance Members

- Department of Agriculture (USDA, Forest Service)
- Department of Commerce (NTIS, NIST)
- Department of Defense
- Department of Education
- Department of Energy
- Department of Health and Human Services (NIH)
- Department of Homeland Security
- Department of Transportation
- Environmental Protection Agency
- Government Publishing Office
- National Aeronautics and Space Administration
- National Science Foundation

[Click here for the Science.gov COVID-19 search results.](#)



U.S. Department of Transportation
Office of the Secretary of Transportation
Bureau of Transportation Statistics

50

Slide Title: Science.gov

Speaker notes:

Science.gov is an interagency initiative providing a gateway to U.S. government science information. As a federated search interface, Science.gov offers free access to R&D results, as well as scientific and technical information (STI) from a long list of federal agencies. This includes journal articles, technical reports, conference papers, videos, audio files, images, and other multimedia, scientific and technical data sets and collections.

Science.gov has also added a directed search for federally-funded COVID-19 research.

All U.S. research outputs – report, dataset, software code, etc. – are tied to research projects. Let us take a quick look at the U.S. DOT research project database.

[Next slide]

[Slide text not presented orally:

Science.gov URL: <https://www.science.gov/>

List of all agencies that are part of Science.gov alliance:

- Department of Agriculture (USDA, Forest Service)
- Department of Commerce (NTIS, NIST)
- Department of Defense
- Department of Education
- Department of Energy
- Department of Health and Human Services (NIH)
- Department of Homeland Security
- Department of Transportation
- Environmental Protection Agency
- Government Publishing Office
- National Aeronautics and Space Administration
- National Science Foundation

Click here for the Science.gov COVID-19 search results:

[https://www.science.gov/scigov/desktop/en/service/link/runSearch/fullRecord:
%22Coronavirus%22%20OR%20%222019-
nCoV%22%20OR%20%222019nCoV%22%20OR%20%22COVID-
19%22%20OR%20%22SARS-CoV-2%22](https://www.science.gov/scigov/desktop/en/service/link/runSearch/fullRecord:%22Coronavirus%22%20OR%20%222019-nCoV%22%20OR%20%222019nCoV%22%20OR%20%22COVID-19%22%20OR%20%22SARS-CoV-2%22)

Detailed background:

Science.gov is an interagency initiative providing a gateway to U.S. government science information. As a federated search interface, Science.gov offers free access to R&D results, as well as scientific and technical information from a long list of federal agencies. This includes journal articles, technical reports, conference papers, videos, audio files, images, and other multimedia, scientific and technical data sets and collections.

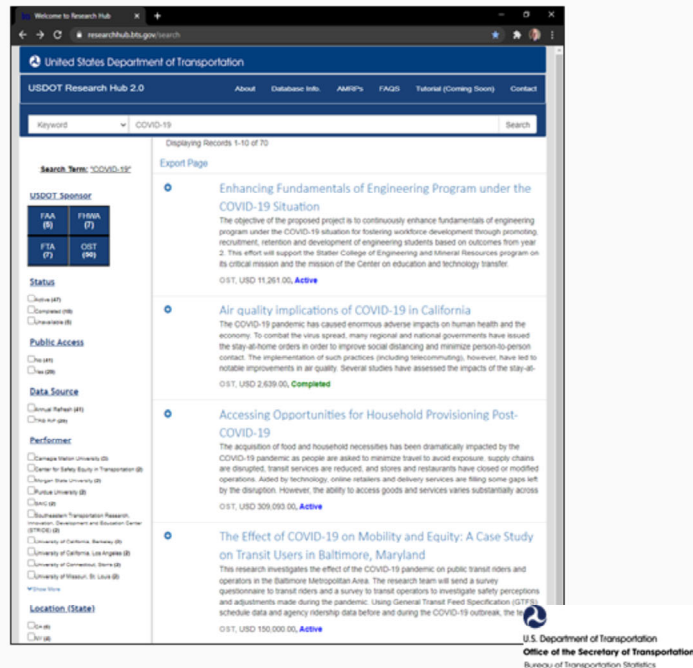
Science.gov has also added a directed search for federally-funded COVID-19 research.

Most U.S. research outputs – reports, datasets, software code, etc. – are tied to research projects. Let us take a quick look at the U.S. DOT research project database.

U.S. DOT Research Hub

Research Hub is a publicly accessible database of USDOT-sponsored research, development, and technology project records.

<https://researchhub.bts.gov/search>



Slide Title: Research Hub

Speaker notes:

Another tool in the US DOT Open Science toolbox is Research Hub.

The USDOT Research Hub is a publicly accessible 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 USDOT's Operating Administrations, providing a comprehensive account of the Department's research portfolio at the project level.

A quick search of Research Hub shows a number of active COVID-19 related projects, and at least 1 that is already completed.

For the very latest on COVID-19 impacts on transportation, however, we should turn to the Bureau of Transportation Statistics.

[Next slide]

[Slide Text of slide not presented orally:


Research Hub is a publicly accessible database of USDOT-sponsored research, development, and technology project records.

<https://researchhub.bts.gov/search>]

ITS JPO CodeHub

ITS CodeHub promotes a reuse-first mentality and aims to support the discovery of open source code by putting it directly into the hands of developers to customize, transform, expand, and improve, as trends evolve and needs change

<https://its.dot.gov/code/>




Purpose
Empower innovation through code reuse, collaboration, and continuous improvement in the open

Capabilities

- Discover projects and modules
- Evaluate code health for reuse
- Connect to developers and other re-users
- Analyze development trends

Community
Grassroots, collaborative development of open-source ITS software



U.S. Department of Transportation
 Office of the Secretary of Transportation
 Bureau of Transportation Statistics

52

Slide Title: ITS JPO CodeHub

Speaker notes:

Of course, an important aspect of Open Science is also the sharing and opening of computer and software code. To that end the U.S. Department of Transportation (U.S. DOT) Intelligent Transportation Systems (ITS) Joint Program Office's (JPO) launched CodeHub for its intelligent transportation systems projects.

ITS CodeHub is the source code management system. It is a resource for the ITS community to discover open source code, software, and more.

ITS CodeHub promotes a reuse-first mentality and aims to support the discovery of open source code by putting it directly into the hands of developers to customize, transform, expand, and improve, as trends evolve and needs change.

This approach has the benefits of lowering costs, increasing interoperability and transparency, and accelerating the path to high-quality software deployment—collectively advancing our nation's transportation system.

The Purpose

Empower innovation through code reuse, collaboration, and continuous improvement in the open...

ITS CodeHub's primary objectives are to:

- Source open source code.
- Encourage code reuse.
- Foster open-source development.

The Capabilities

More than just a catalogue of software development projects...

ITS CodeHub goes beyond cataloging U.S. DOT-funded software development projects. ITS CodeHub offers the transportation community the following capabilities:

- Discover projects and modules already built within the U.S. DOT and across the open-source community.
- Evaluate code health, statistics, dependencies, and compatibility to reuse in projects.
- Connect to developers and others who have reused and extended code.
- Analyze development trends and statistics to understand evolving software development needs.

The Community

The community plays a pivotal role in the development of open-source products—from discovery, to quality assurance, to coding and development, to adoption and integration.

ITS CodeHub fosters a community for the grassroots, collaborative development of open-source ITS software among the U.S. DOT, state and local agencies, researchers, and companies.

Contact Information

For more information about the ITS CodeHub, please contact the ITS JPO Data Program support team at: data.itsjpo@dot.gov.

It is hoped that by the end of 2021, CodeHub will expand to be a service available to all DOT offices and their funded researchers.

[Next slide]

[Slide Text of slide not presented orally:

ITS CodeHub promotes a reuse-first mentality and aims to support the discovery of open source code by putting it directly into the hands of developers to customize, transform, expand, and improve, as trends evolve

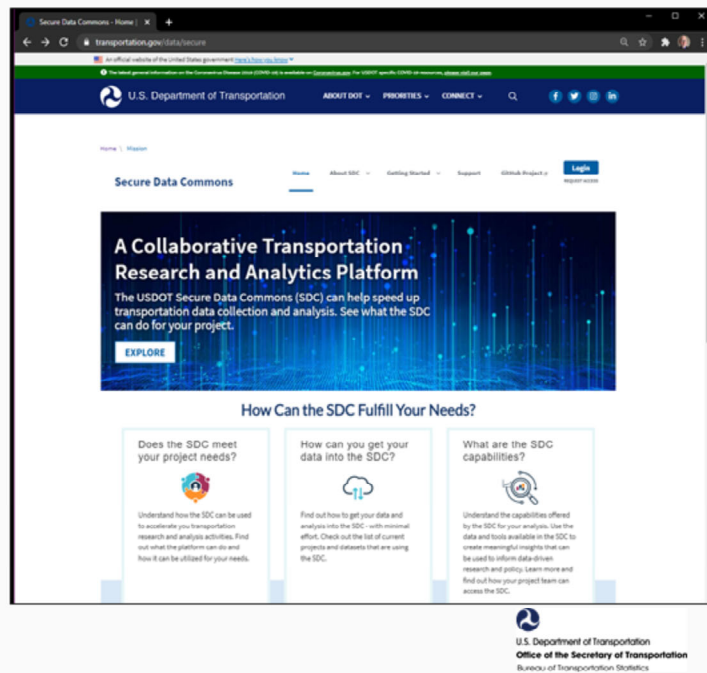
and needs change

<https://its.dot.gov/code/>]

U.S. DOT Secure Data Commons

The USDOT Secure Data Commons (SDC) can help speed up transportation data collection and analysis.

<https://www.transportation.gov/data/secure>



Slide Title: U.S. DOT Secure Data Commons

Speaker notes:

While we are moving to make our research and other data open, we also must be aware of the needs to protect privacy, provide security, and behave ethically. To help protect sensitive data, but also allowing for its analysis, the ITS JPO developed the Secure Data Commons (SDC). SDC has now become a DOT-wide shared service.

The Secure Data Commons (SDC) is a cloud-based analytics platform that enables traffic engineers, researchers, and data scientists to access transportation-related datasets. The U.S. Department of Transportation (USDOT) created the SDC to provide a secure platform for sharing and collaborating on research, tools, algorithms, and analysis involving moderate sensitivity level (PII & CBI) datasets using commercially available tools, without needing to install tools or software locally.

The SDC offers a common platform for innovative data analysis and sharing of results that cuts across the Department's data silos.

[Next slide]


[Slide text not presented orally:

The USDOT Secure Data Commons (SDC) can help speed up transportation data collection and analysis.

<https://www.transportation.gov/data/secure>

]

NCHRP Report 936



NCHRP 936: A Guide to Ensure Access to the Results of Federally Funded Transportation Research

- Report Link:
<http://www.trb.org/main/blurbs/180230.aspx>
 - Project NCHRP 20-110:
<https://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=4062>
 - Designed to help DOT-funded researchers improve data management and data sharing
 - Already a little out of date because of things like Federal Data Strategy that came about while report in publication limbo
 - National Transportation Library planning series of video trainings

54

Slide Title: NCHRP Report 936

Speaker notes:

The National Transportation Library staff has been leading U.S. DOT efforts to affect transportation research culture change towards public access and open science since 2014. We do this by offering training, creating guidelines, and consulting on the creation of new resources.

One example is NCHRP Report 936.

In late January, the Transportation Research Board (TRB) published the pre-print version of National Cooperative Highway Research Program (NCHRP) Report 936: A Guide to Ensure Access to the Results of Federally Funded Transportation Research.

I was one of the original authors of the research needs statement in 2014, and I and other NTL staff served on the research project panel, and helped to review and edit the report as it became finalized.

You can access the Report at <http://www.trb.org/Main/Blurbs/180230.aspx>

- This report is the final output of NCHRP Project NCHRP 20-110:
<https://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=4062>

- The Guide is designed to help DOT-funded researchers improve data management and data sharing
- The Guide, which experienced some publication delays, is already a little out of date because of things like Federal Data Strategy that came about while report in publication limbo
- The National Transportation Library planning series of video trainings

Once we get DOT-funded researchers all managing their data well, the next questions is where to preserve and share that data.

Links to resources

United States. Government Publishing Office (GPO). (2021). Washington, D.C. <https://www.gpo.gov/>

United States. Government Publishing Office (GPO). (2021). *GPO Style Manual*. Washington, D.C. <https://www.govinfo.gov/content/pkg/GPO-STYLEMANUAL-2016/pdf/GPO-STYLEMANUAL-2016.pdf>

United States. National Technical Information Service (NTIS). (2021). Washington, D.C. <https://www.ntis.gov/>

United States. National Technical Information Service (NTIS). (2021). *National Technical Reports Library (NTRL)*. Washington, D.C. <https://ntrl.ntis.gov/NTRL/>

United State. Department of Transportation. Federal Highway Administration. (1976). *Freeway Data for Incident and Nonincident Conditions – Vol. 1: Traffic Data Sets from Widely Spaced Detectors*. Washington, D.C. <https://doi.org/10.21949/1520658>

United State. Department of Transportation. Federal Highway Administration. (1977). *Freeway Data for Incident and Nonincident Conditions – Vol. 2: Traffic Data Sets from Closely Spaced Detectors*. Washington, D.C. <https://doi.org/10.21949/1520659>

United State. Department of Transportation. Federal Highway Administration. (1977). *Freeway Data for Incident and Nonincident Conditions – Vol. 3: FORTRAN Program Documentation for Analyzing Individual Data Sets*. Washington, D.C. <https://doi.org/10.21949/1520660>



U.S. Department of Transportation
Office of the Secretary of Transportation
Bureau of Transportation Statistics

55

Slide Title: Links to resources

Slide text:

United States. Government Publishing Office (GPO). (2021). Washington, D.C. <https://www.gpo.gov/>

United States. Government Publishing Office (GPO). (2021). *GPO Style Manual*. Washington, D.C. <https://www.govinfo.gov/content/pkg/GPO-STYLEMANUAL-2016/pdf/GPO-STYLEMANUAL-2016.pdf>

United States. National Technical Information Service (NTIS). (2021). Washington, D.C. <https://www.ntis.gov/>

United States. National Technical Information Service (NTIS). (2021). *National Technical Reports Library (NTRL)*. Washington, D.C. <https://ntrl.ntis.gov/NTRL/>

United State. Department of Transportation. Federal Highway Administration. (1976). *Freeway Data for Incident and Nonincident Conditions – Vol. 1: Traffic Data Sets from Widely Spaced Detectors*. Washington, D.C. <https://doi.org/10.21949/1520658>

United State. Department of Transportation. Federal Highway Administration. (1977). *Freeway Data for Incident and Nonincident Conditions – Vol. 2: Traffic Data Sets from Closely Spaced Detectors*. Washington, D.C. <https://doi.org/10.21949/1520659>

United State. Department of Transportation. Federal Highway Administration. (1977). *Freeway Data for Incident and Nonincident Conditions – Vol. 3: FORTRAN Program Documentation for Analyzing Individual Data Sets*. Washington, D.C. <https://doi.org/10.21949/1520660>

Links to resources (continued)

United States. Office of Management and Budget (OMB). (2005). Memo M-06-02, "Improving Public Access to and Dissemination of Government Information and Using the Federal Enterprise Architecture Data Reference Model."

<https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2006/m06-02.pdf>

United States. White House. (2009). "Transparency and Open Government." [Memorandum].

<https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>

United States. Office of Management and Budget (OMB). (2009) Memo M-10-06, "Open Government Directive."

<https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2010/m10-06.pdf>

United States. White House. Office of Science and Technology Policy (OSTP). (2013). "Increasing Access to the Results of Federally Funded Scientific Research." [Memorandum.]

https://web.archive.org/web/20130308142014/https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

United States. Office of Management and Budget (OMB). (2013). Memo M-13-13, "Open Data Policy – Managing Information as an Asset."

<https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2013/m-13-13.pdf>

United States. White House. (2013). Executive Order 13642, "Making

Open and Machine Readable the New Default for Government Information." <https://www.govinfo.gov/content/pkg/FR-2013-05-14/pdf/2013-11533.pdf>



U.S. Department of Transportation
Office of the Secretary of Transportation
Bureau of Transportation Statistics

56

Slide Title: Links to resources

Slide text:

United States. Office of Management and Budget (OMB). (2005). Memo M-06-02, "Improving Public Access to and Dissemination of Government Information and Using the Federal Enterprise Architecture Data Reference Model."

<https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2006/m06-02.pdf>

United States. White House. (2009). "Transparency and Open Government."

[Memorandum]. <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>

United States. Office of Management and Budget (OMB). (2009) Memo M-10-06, "Open Government Directive."

<https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2010/m10-06.pdf>

United States. White House. Office of Science and Technology Policy (OSTP).

(2013). "Increasing Access to the Results of Federally Funded Scientific Research." [Memorandum.]

https://web.archive.org/web/20130308142014/https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

United States. Office of Management and Budget (OMB). (2013). Memo M-13-13, "Open Data Policy – Managing Information as an Asset."

[https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2013/m-13-](https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2013/m-13-13.pdf)

13.pdf

United States. White House. (2013). Executive Order 13642, "Making Open and Machine Readable the New Default for Government Information."

<https://www.govinfo.gov/content/pkg/FR-2013-05-14/pdf/2013-11533.pdf>

Links to resources (continued)

United States. House of Representatives. (2017). H.R. 4174, "Foundations for Evidence-Based Policymaking Act of 2018." <https://www.congress.gov/bill/115th-congress/house-bill/4174>

United States. Senate. (2017). S.760 - "Open, Public, Electronic, and Necessary Government Data Act." <https://www.congress.gov/bill/115th-congress/senate-bill/760/text>

United States. (2019). Public Law No. 115-435 "Foundations for Evidence-Based Policymaking Act of 2018." <https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf>

United States. National Science and Technology Council. Interagency Working Group on Digital Data. (2009). *Harnessing the Power of Digital Data for Science and Society*. Washington, D.C. <https://www.nitrd.gov/Publications/PublicationDetail.aspx?pubid=25>

White House. Office of Science and Technology Policy (OSTP). (2020). <https://www.whitehouse.gov/ostp/>

United States. Department of Transportation. (2015). "Plan to Increase Public Access to the Results of Federally-Funded Scientific Research." <https://doi.org/10.21949/1520559>

United States. General Services Administration. Technology Transformation Services. (2021). "Data.gov." <https://www.data.gov/>

United States. General Services Administration. Technology Transformation Services. (2021). "Department of Transportation Data Catalog." <https://catalog.data.gov/organization/dot-gov>

United States. Department of Transportation. (2021). "Data.transportation.gov." <https://data.transportation.gov>



U.S. Department of Transportation
Office of the Secretary of Transportation
Bureau of Transportation Statistics

57

Slide Title: Links to resources

Slide text:

United States. House of Representatives. (2017). H.R. 4174, "Foundations for Evidence-Based Policymaking Act of 2018." <https://www.congress.gov/bill/115th-congress/house-bill/4174>

United States. Senate. (2017). S.760 - "Open, Public, Electronic, and Necessary Government Data Act." <https://www.congress.gov/bill/115th-congress/senate-bill/760/text>

United States. (2019). Public Law No. 115-435 "Foundations for Evidence-Based Policymaking Act of 2018." <https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf>

United States. National Science and Technology Council. Interagency Working Group on Digital Data. (2009). *Harnessing the Power of Digital Data for Science and Society*. Washington, D.C.

<https://www.nitrd.gov/Publications/PublicationDetail.aspx?pubid=25>

White House. Office of Science and Technology Policy (OSTP). (2020). <https://www.whitehouse.gov/ostp/>

United States. Department of Transportation. (2015). "Plan to Increase Public Access to the Results of Federally-Funded Scientific Research." <https://doi.org/10.21949/1520559>

United States. General Services Administration. Technology Transformation Services. (2021). "Data.gov." <https://www.data.gov/>

United States. General Services Administration. Technology Transformation Services.
(2021). "Department of Transportation Data Catalog."

<https://catalog.data.gov/organization/dot-gov>

United States. Department of Transportation. (2021). "Data.transportation.gov."

<https://data.transportation.gov>

Links to resources (continued)

United States. Department of Transportation. Bureau of Transportation Statistics. National Transportation Library. (2021). "Repository & Open Science Access Portal (ROSA P)." <https://doi.org/10.21949/1398953>

United States. Department of Energy. Office of Scientific and Technical Information. (2021). "Science.gov." <https://www.science.gov/>

United States. Department of Transportation. Office of the Assistant Secretary for Research and Technology. (2021). "Research Hub." <https://researchhub.bts.gov/search>

United States. Department of Transportation. Bureau of Transportation Statistics. (2021). "COVID-19 Related Transportation Statistics." <https://www.bts.dot.gov/covid-19>

United States. General Services Administration. Technology Transformation Services. (2021). "resources.data.gov." <https://resources.data.gov/>

United States. Department of Health and Human Services. National Institutes of Health. (2020). "The National Library of Medicine expands access to coronavirus literature through PubMed Central."

<https://www.nih.gov/news-events/news-releases/national-library-medicine-expands-access-coronavirus-literature-through-pubmed-central>



U.S. Department of Transportation
Office of the Secretary of Transportation
Bureau of Transportation Statistics

58

Slide Title: Links to resources

Slide text:

United States. Department of Transportation. Bureau of Transportation Statistics. National Transportation Library. (2021). "Repository & Open Science Access Portal (ROSA P)." <https://doi.org/10.21949/1398953>

United States. Department of Energy. Office of Scientific and Technical Information. (2021). "Science.gov." <https://www.science.gov/>

United States. Department of Transportation. Office of the Assistant Secretary for Research and Technology. (2021). "Research Hub."

<https://researchhub.bts.gov/search>

United States. Department of Transportation. Bureau of Transportation Statistics. (2021). "COVID-19 Related Transportation Statistics." <https://www.bts.dot.gov/covid-19>

United States. General Services Administration. Technology Transformation Services. (2021). "resources.data.gov." <https://resources.data.gov/>

United States. Department of Health and Human Services. National Institutes of Health. (2020). "The National Library of Medicine expands access to coronavirus literature through PubMed Central." <https://www.nih.gov/news-events/news-releases/national-library-medicine-expands-access-coronavirus-literature-through-pubmed-central>