

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

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**



Review of 2021-04-05 Presentation

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

**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>

Deep Dive:

Public Access Plan:

1. Background & Purpose

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

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.

Deep Dive:

Public Access Plan:

2. Scope

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

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>

Deep Dive:

Public Access Plan:

2. Definition of Digital Datasets

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

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.

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



Deep Dive:

Public Access Plan:

3. Applicability

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

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.

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.**

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>

Deep Dive:

Public Access Plan:

6. Roles & Responsibilities

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

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.

Audience Question Break



Deep Dive:

Public Access Plan:

7. Implementation

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

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**

Deep Dive:

Public Access Plan:

7.2 Before Research Begins

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

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.**

Deep Dive:

Public Access Plan:

7.3.2 Data Submission

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

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>

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:

- 6. Meet essential metadata requirements;**
- 7. Provide persistent identification of datasets; and**
- 8. Provide long-term access.**

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

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-fie>

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



Deep Dive:

Public Access Plan:

7.6.2 Data Preservation

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

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.

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.



Public Access Implementation Working Group (PAIWG)

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

- **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.

Deep Dive:

Public Access Plan:

8.1 Timeline

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

Version 1.1



December 16, 2015

U.S. Department of Transportation

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;**

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.

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.

Audience Question Break



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>

The screenshot displays the National Transportation Library (NTL) website. At the top, the browser address bar shows rosap.ntl.bts.gov. The page header includes the United States Department of Transportation logo and the Bureau of Transportation Statistics. The main navigation bar contains links for Home, Collections, Recent Additions, Public Access, and Submit Content. A search bar is prominently displayed with the text "Enter keyword or phrase..." and a "Search" button. Below the search bar, there is a large banner for the "Transportation Statistics Annual Report 2020" with a "View Archive" link. To the left of the banner, a vertical list of documents is shown, including "Pocket Guide to Transportation 2019", "Port Performance Freight Statistics in 2018", "Transportation Statistics Annual Report 2018", and "Blockchain for Unmanned Aircraft Systems". Below the banner, there are three columns of content: "Stay Connected" with links to "Ask-A-Librarian", "Transportation Librarians Roundtable (TLR)", and "Digital Submissions"; "Transportation Resources" with links to "NTL Guides" and "Freight Data Dictionary Public Access"; "Recently Added" with a list of documents such as "Social Media Analysis for Transit Assessment" and "Travel-Time Reliability in Simulation and Planning Models"; and "Trending This Week" with documents like "Drug and Alcohol Prevalence in Seriously and Fatally Injured Road Users" and "Performance evaluation of seal coat materials". The footer of the page includes the U.S. Department of Transportation logo and the text "Office of the Secretary of Transportation Bureau of Transportation Statistics".

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>

The screenshot displays the National Transportation Library (NTL) website interface. At the top, it identifies the United States Department of Transportation and the Bureau of Transportation Statistics. The main header is the "National Transportation Library" logo. Below this is a navigation bar with links for Home, Collections, Recent Additions, Public Access, and Submit Content, along with an "About ROSA P" dropdown. A search bar is prominently featured, with a dropdown menu currently set to "Federal Aviation Administration" and a search button labeled "Search". Below the search bar, there are options for "Advanced Search" and "All Collections".

The main content area features the "Federal Aviation Administration" logo and a brief description: "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".

Below this is a "Search Results" section. It includes a "Select Docs for Citation Export" button with a checkmark. The results are displayed as "1 - 14 of 14 results" with a "50 results" filter and a "Sort By" dropdown set to "Relevance".

On the left side of the search results, there is a "Narrow Results:" sidebar with three sections: "Resource Type" (Journal Article (3), Manual (6), Tech Report (5)), "Subject" (aeromedical c... (1), Aviation (1), occupational... (1), reflective curv... (1)), and "Keywords" (Acoustics (4), AEDT (4), Aircraft (4), aircraft acoustics (4)).

The search results list two technical reports from the FAA Center of Excellence for Alternative Jet Fuels & Environment:

- Technical Report : For The Period October 1, 2016 - September 30, 2017**
Published Date: 2017-09-30
Abstract: This report covers the period October 1, 2016 through September 30, 2017. The Center was established by the authority of FAA solicitation 13-C-AJFE-Solicitation. During that time the ASCENT team launched a new website, which can be viewed at ascent.a...
File Type: PDF - 44.82 MB
- Technical Report : December, 2016 : For the period September 13, 2013 - September 30, 2015**
Published Date: 2015-12-01
Abstract: This report covers the period between the initial establishment of the FAA Center of Excellence for Alternative Jet Fuels and Environment on September 13, 2013 through September 30, 2015. The Center was established by the authority of FAA sollicitatio...
File Type: PDF - 23.76 MB

At the bottom of the page, a partial result for "FAA Center of Excellence Alternative Jet Fuels & Environment : Annual" is visible.

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>

rosap
Repository & Open Science
Access Portal

All Collections

Enter keyword or phrase.

Back to Search Results Page

Email Print

American Travel Survey (ATS) 1995 [datasets]

Record #: 1 of 104

Published Date: 2019-03-28

Language: English

[Click to view more supporting documents](#)

Viewer Details **Supporting Files** Related Documents You May Also Like

Supporting Files:

- Supporting files application/pdf/zip
- Data and Documentation Zip application/zip
- Project Open Data metadata file application/octet-stream

Data Package Elements in “Supporting Files” tab



Collecting & Identifying Research: Data Package Publications

NTL Dataset Data Package Elements

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

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



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>

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

rosap
Repository & Open Science
Access Portal

All Collections [v] Enter keyword or phrase... [Search]

Advanced Search

Back to Search Results Page [Email] [Print] [Share] [Export Citations]

Integrated Implementation of Innovative Intersection Designs [supporting datasets]

Record #: 1 of 104

Published Date: 2020-06-10
Language: English

[PDF-314.69 KB] [Download]

Viewer | Details | Supporting Files | Related Documents | You May Also Like

1 of 3 Automatic Zoom

Integrated Implementation of Innovative Intersection Designs [supporting datasets]

Dataset available at: <http://doi.org/10.5281/zenodo.3887540>

(This dataset supports research report: **Integrated Implementation of Innovative Intersection Designs**.)

This U.S. Department of Transportation-funded dataset is preserved by Southeastern Transportation Research, Innovation, Development and Education Center (STRIDE) in Zenodo, and is available at: <http://doi.org/10.5281/zenodo.3887540>

The related final report **Integrated Implementation of Innovative Intersection Designs**, will be available from the National Transportation Library's digital repository, the Repository & Open Science Access Portal (ROSA P <https://doi.org/10.21949/1398953>).

Metadata from the repository record:

Description: These files belong to a research project on evaluating different bicycle and pedestrian crossing options at Continuous Flow Intersections (CFI). These are the simulation models for different CFI crossing alternatives for pedestrians and bicycles. Three parameters were varied to create the scenario: a) number of displaced left turn legs (4 vs. 2) b) right turn signalization (signalized vs. unsignalized) c) crossing types (traditional vs. midblock vs. offset). In addition to the CFI models, equivalent standard intersection models are also included here. The models were created using VISSIM 10 (64 bit) software. Users may need to show the VISSIM file to the corresponding .rbc files for each signal controller.

NTL-created
“data cover
sheet” for
Viewer Tab

Includes DOI to
external
repository
landing page

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

rosap
Repository & Open Science Access Portal

All Collections [v] Enter keyword or phrase

Back to Search Results Page [Email] [Print]

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

Viewer [Details] Supporting Files [R]

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

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

NTL-minted DOI. Could also be external DOI

Audience Question Break



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.**

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}

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 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

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

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

Data Management, Data Curation, & Data Science Dependencies Model

Data Management \in Data Curation \Rightarrow Data Science

DM \in DC \Rightarrow DS



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**



Resources:
Contact Info

**Leighton Christiansen
Data Curator,
National Transportation Library**

**leighton.christiansen@dot.gov
202-366-2759**

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



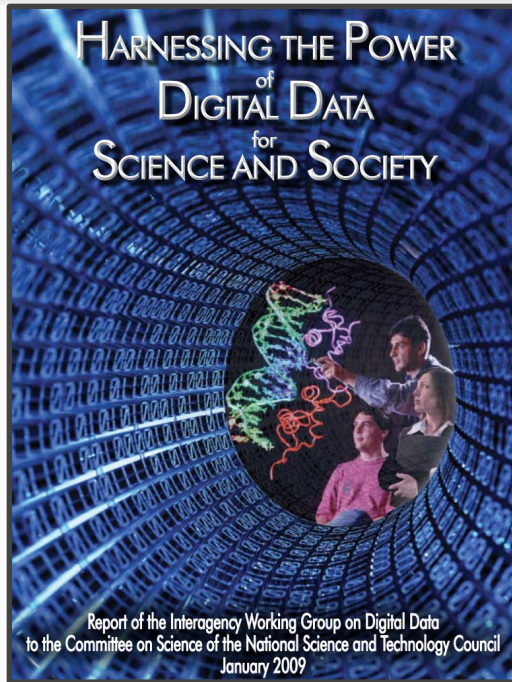
U.S. Department of Transportation
Office of the Secretary of Transportation
Bureau of Transportation Statistics

Supplemental Slides

The following Supplemental Slides include further resources for those interested.



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

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

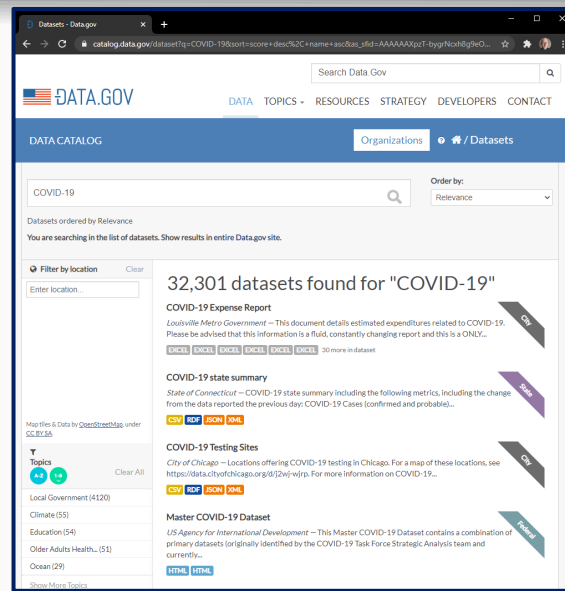


Data.gov
Quick Stats

217,000+ datasets

32,000+
COVID-19-related
datasets

7
U.S. DOT COVID-19-
related datasets



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

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/>

The screenshot shows the homepage of Data.Transportation.gov. At the top, there is a navigation bar with the logo, search bar, and links for Home, Catalog, User Guide, and Developer. Below the navigation bar is a large banner image of a train with the text "Data.Transportation.gov" and "Welcome to Data.Transportation.gov, the U.S. Department of Transportation's public data portal". The main content area features a grid of 10 icons representing different transport modes: Railroads, Roadways & Bridges, Pipelines & HAZMAT, Trucking & Motorcoaches, Aviation, Public Transit, Automobiles, Maritime & Waterways, Research & Statistics, and Bicycles & Pedestrians. Below the icons are five featured datasets: TxDOT Active Work Zones (map), BSM Point Map (map), Site Analytics (bar chart), Border Crossings by Mode, Border, and State (line chart), and Takata Recall - Priority Group Repaired and Remaining (bar chart).

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>

The screenshot shows the ROSA P website interface. At the top, the browser address bar displays "rosap.ntl.bts.gov". The page header includes "United States Department of Transportation" and "Bureau of Transportation Statistics". The main title is "National Transportation Library". Below the header, there are navigation links: "Home", "Collections", "Recent Additions", "Public Access", "Submit Content", and "About ROSA P". A search bar is present with the text "rosap Repository & Open Science Access Portal" and a search button. The main content area features a large banner for the "Transportation Statistics Annual Report 2020" with a description: "Latest available data collected and compiled by BTS, describes the Nation's transportation system, performance and trends." Below the banner, there are several featured items in a "Recently Added" section, including "Social Media Analysis for Transit Assessment", "Travel-Time Reliability in Simulation and Planning Models: Utah Case Study (SHRP2 L04 IAP Round 7)", "Development of the Optimization Model for Improving Safety at Rail Crossings in Florida", "Development of a Statistical Model to Predict Materials' Unit Prices for Future Maintenance and Reha...", and "Cost-Benefit Analysis of Novel Access Modes: A Case Study in the". A "Trending This Week" section is also visible, featuring "Drug and Alcohol Prevalence in Seriously and Fatally Injured Road Users Before and During the COVID-19 Public Health Emergency", "Transportation-Markings Database: Railway Signals, Signs, Marks, Markers. Part III, Volume 3, Additional Studies", "Performance evaluation of seal coat materials and designs.", and "Assessment and mitigation of liquefaction hazards to bridge approach embankments in Oregon: final report." The footer includes the U.S. Department of Transportation logo and the text "Office of the Secretary of Transportation Bureau of Transportation Statistics".

COVID-19 Transportation Statistics from BTS

Find the latest Coronavirus-related transportation statistics on the [BTS Covid-19 landing page](#)

United States Department of Transportation

Ask-A-Librarian # | A-Z Index

Bureau of Transportation Statistics

Search BTS site

Topics and Geography | Statistical Products and Data | National Transportation Library | Newsroom | About BTS

Home

Catalog of COVID-19 Related Transportation Statistics

Daily Travel (National, State, and County)

The Week in Transportation (National)

COVID-19-Related Data Spotlights

Bikes and E-scooters

Ferry Operations

COVID-19 Related Transportation Statistics

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, and BTS will add more measures as they become available.

DAILY TRAVEL DURING THE COVID-19 PUBLIC HEALTH EMERGENCY

- **What it shows:** The percent and number of people staying home and not staying home each day, and the number of trips taken each day for 10 different distance groupings.
- **How recent:** One-week lag
- **Update Frequency:** Weekly on Monday
- **Geography level:** National, State, County

MOBILITY OVER TIME BY STATE AND BY TRIP DISTANCE

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

Daily Travel during the COVID-19 Public Health Emergency

Find the latest Coronavirus-related transportation statistics on the [BTS Covid-19 landing page](#)

United States Department of Transportation

Bureau of Transportation Statistics

Home > Daily Travel (National, State, and County)

Catalog of COVID-19 Related Transportation Statistics

Daily Travel (National, State, and County)

The Week in Transportation (National)

COVID-19-Related Data Spotlights

Bikes and E-scooters

Ferry Operations

Daily Travel during the COVID-19 Public Health Emergency

How many people are staying at home during the COVID-19 pandemic? How far are people traveling when they don't stay home? Which states and counties have more people taking trips? Start exploring our Daily Travel data to answer those questions.

In which States and Counties are people staying at home? Which ones show the most activity?

Dive into the map below to see what percentage of the population is staying at home in your state or county. You can also use the *Select a Metric* drop-down to see state or county-level measures for the average number of daily trips people are taking and more.

Map of Activity by State or County

Average Percent of People Staying at Home per Day

Select a Month
November 2020

Select a Geographic Level
 State
 County

Select a Metric
Percent of People Staying at Home

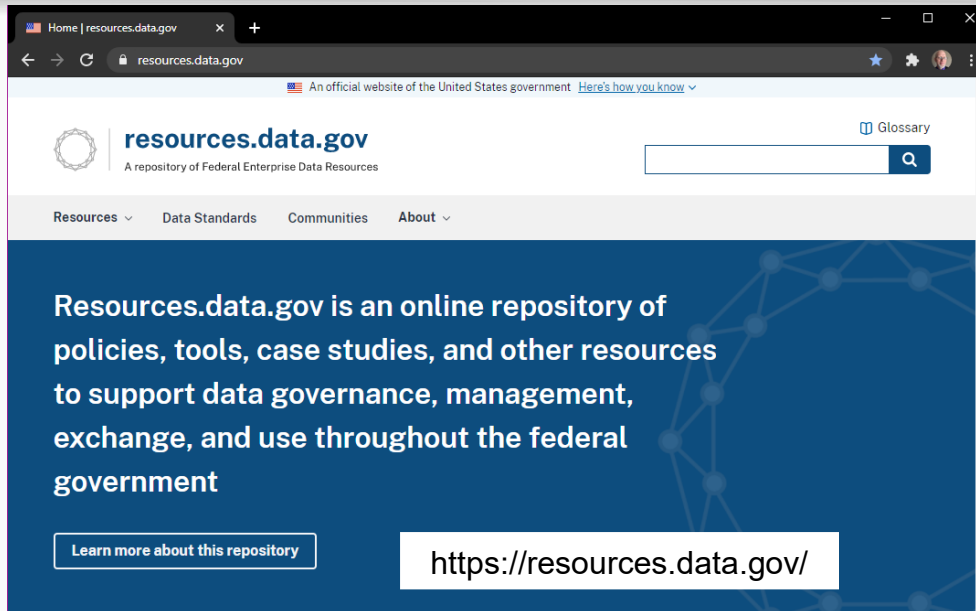
← → ↺ ↻ ⌂ Share Download Full Screen

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 by Income Groups
- 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



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



The screenshot shows a web browser window displaying the homepage of Resources.data.gov. The browser's address bar shows the URL resources.data.gov. The page header includes the site logo, the text "resources.data.gov A repository of Federal Enterprise Data Resources", a search bar, and a "Glossary" link. A navigation menu contains "Resources", "Data Standards", "Communities", and "About". The main content area features a blue background with a network diagram and the text: "Resources.data.gov is an online repository of policies, tools, case studies, and other resources to support data governance, management, exchange, and use throughout the federal government". Below this text is a button labeled "Learn more about this repository" and a white box containing the URL "https://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



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
Your Gateway to U.S. Federal Science

Home About STEM Opportunities Translate

Science.gov searches over 60 databases and over 2,200 scientific websites to provide users with access to more than 200 million pages of authoritative federal science information including research and development results.

New:
Find federal research on Coronavirus (COVID-19)

Find out how the COVID-19 search works >
For the latest public health information about COVID-19, visit the CDC >
For information about the U.S. Government's response, visit USA.gov >

Enter Search Terms

[Advanced Search](#)

[U.S. Federal Science Agencies' Public Access Plans](#)
[How To Submit Research Papers to Funding Agencies](#)
[Inventory of Agency APIs and Other Services for Public Access Collections](#)

Science.gov is governed by the interagency *Science.gov Alliance*. Participating agencies are:

USDA USGS NIH U.S. National Library of Medicine

USGS USGS USGS GAO NASA NIST

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. DOT Research Hub

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

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

The screenshot shows the USDOT Research Hub 2.0 search results for the keyword "COVID-19". The page displays a list of four research records, each with a title, a brief description, and funding information. The left sidebar contains various filters for refining the search results.

United States Department of Transportation
USDOT Research Hub 2.0

Search Term: "COVID-19"

Displaying Records 1-10 of 70

Export Page

- Enhancing Fundamentals of Engineering Program under the COVID-19 Situation**
The objective of the proposed project is to continuously enhance fundamentals of engineering program under the COVID-19 situation for fostering workforce development through promoting, recruitment, retention and development of engineering students based on outcomes from year 2. This effort will support the Statler College of Engineering and Mineral Resources program on its critical mission and the mission of the Center on education and technology transfer.
OST, USD 11,261.00, **Active**
- Air quality implications of COVID-19 in California**
The COVID-19 pandemic has caused enormous adverse impacts on human health and the economy. To combat the virus spread, many regional and national governments have issued the stay-at-home orders in order to improve social distancing and minimize person-to-person contact. The implementation of such practices (including telecommuting), however, have led to notable improvements in air quality. Several studies have assessed the impacts of the stay-at-home orders on air quality.
OST, USD 2,639.00, **Completed**
- Accessing Opportunities for Household Provisioning Post-COVID-19**
The acquisition of food and household necessities has been dramatically impacted by the COVID-19 pandemic as people are asked to minimize travel to avoid exposure, supply chains are disrupted, transit services are reduced, and stores and restaurants have closed or modified operations. Aided by technology, online retailers and delivery services are filling some gaps left by the disruption. However, the ability to access goods and services varies substantially across geographic areas.
OST, USD 309,093.00, **Active**
- The Effect of COVID-19 on Mobility and Equity: A Case Study on Transit Users in Baltimore, Maryland**
This research investigates the effect of the COVID-19 pandemic on public transit riders and operators in the Baltimore Metropolitan Area. The research team will send a survey questionnaire to transit riders and a survey to transit operators to investigate safety perceptions and adjustments made during the pandemic. Using General Transit Feed Specification (GTFS) schedule data and agency ridership data before and during the COVID-19 outbreak, the team will analyze the impact of the pandemic on transit users and operators.
OST, USD 150,000.00, **Active**

USDOT Sponsor

FAA (5)	FHWA (7)
FTA (7)	OST (50)

Status

- Active (47)
- Completed (18)
- Unavailable (5)

Public Access

- No (41)
- Yes (29)

Data Source

- Annual Refresh (41)
- TRB RIP (29)

Performer

- Carnegie Mellon University (3)
- Center for Safety Equity in Transportation (2)
- Morgan State University (2)
- Purdue University (2)
- SAIC (2)
- Southeastern Transportation Research, Innovation, Development and Education Center (STRIDE) (2)
- University of California, Berkeley (2)
- University of California, Los Angeles (2)
- University of Connecticut, Storrs (2)
- University of Missouri, St. Louis (2)

Location (State)

- CA (6)
- NY (4)

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

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

The screenshot shows the ITS CodeHub website homepage. At the top, there is a navigation bar with links for Home, About, Explore, and Resources. Below this is a large header section with the title "Explore ITS CodeHub" and a subtitle "The U.S. Department of Transportation's portal for open-source, ITS source code." A search bar is prominently displayed. The main content area is divided into several sections: "Repository Spotlight" featuring three featured repositories (carma-platform, jpo-ode, and carma-cloud), "Popular Categories" with icons for Connected Vehicles, Road Weather Management, Roadway Maintenance, Vehicle Traffic Control, Traffic Incident Management, and Traffic Signal, and "Open Source Community" which includes a call to action to "Join the ITS Open-Source Community" and "Contribute Repository", as well as sections for "ITS Repositories" and "Repository Metrics".

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. 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>

The screenshot shows the homepage of the U.S. DOT Secure Data Commons. The browser address bar displays [transportation.gov/data/secure](https://www.transportation.gov/data/secure). The page features a dark blue header with the U.S. Department of Transportation logo and navigation links for ABOUT DOT, PRIORITIES, and CONNECT. A search bar and social media icons are also present. Below the header, a navigation menu includes Home, About SDC, Getting Started, Support, GitHub Project, and a Login button. The main content area is dominated by a large blue graphic with the text: "A Collaborative Transportation Research and Analytics Platform". Below this, a sub-headline reads: "The USDOT Secure Data Commons (SDC) can help speed up transportation data collection and analysis. See what the SDC can do for your project." An "EXPLORE" button is positioned at the bottom left of the graphic. Underneath the graphic, a section titled "How Can the SDC Fulfill Your Needs?" contains three columns of information, each with an icon and a brief description:

- Does the SDC meet your project needs?** (Icon: People and gear) Understand how the SDC can be used to accelerate your transportation research and analysis activities. Find out what the platform can do and how it can be utilized for your needs.
- How can you get your data into the SDC?** (Icon: Cloud and upload arrow) Find out how to get your data and analysis into the SDC - with minimal effort. Check out the list of current projects and datasets that are using the SDC.
- What are the SDC capabilities?** (Icon: Magnifying glass over data points) Understand the capabilities offered by the SDC for your analysis. Use the data and tools available in the SDC to create meaningful insights that can be used to inform data-driven research and policy. Learn more and find out how your project team can access the SDC.



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.aspx>
 - 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

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-STYL>

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>



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>



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-ex>

