Abstract

The legislative mandate for the National Transportation Library $(NTL)^1$ includes direction to serve as the central repository for transportation information and a portal to federal transportation data.

This mandate means that NTL's Repository and Open Science Access Portal (ROSA P) is the permanent repository for publications in the USDOT Public Access Plan², a response to the White House Office of Science and Technology Policy memorandum "Increasing Access to the Results of Federally Funded Scientific Research³." ROSA P will also preserve and disseminate data and publications from the Bureau of Transportation Statistics, USDOT, and the US transportation community.

The USDOT Public Access Plan further requires active management of data throughout the research lifecycle. Data management is deliberate planning, creation, storage, access and preservation of data.

The NTL, in supporting public access, now offers a growing suite of data services, including implementing data management planning for federally funded transportation research.

This poster highlights some tools and updates made to existing guidance to help intramural and extramural data creators employ better data management practices and comply with the USDOT Public Access Plan. The poster also includes leading edge policies and recommendations authored or employed by NTL in order to maintain pace with other state-ofthe-practice data repositories.



Author Information

Leighton L. Christiansen Nicole Strayhorn http://orcid.org/0000-0002-0543-4268 http://orcid.org/0000-0001-9328-4556

Intern, NTL Data Curator, NTL ncstrayhorn@gmail.com

leighton.christiansen@dot.gov

Amanda J. Wilson

Director, NTL Intern, NTL

amanda.wilson@dot.gov

http://orcid.org/0000-0001-6580-2328 http://orcid.org/0000-0001-7768-869X

Nikki Olsen

nolsen1226@gmail.com

Defining Data Management & Playing Bingo!

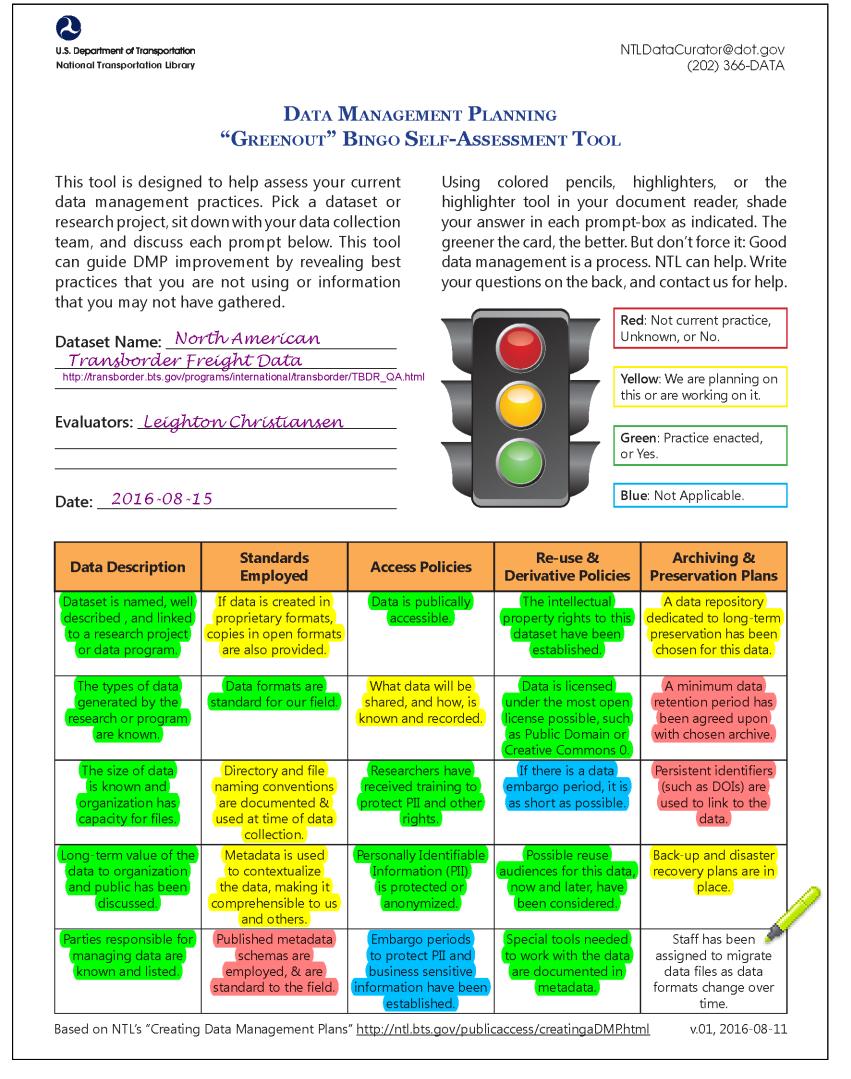
BTS Road Shows

The National Transportation Library (NTL) is part of USDOT's Bureau of Transportation Statistics (BTS)⁴. BTS was created in 1992 to administer data collection, analysis, and reporting of transportation statistical information.

BTS data collections include traffic, passenger flow, employment, financial condition, and on-time performance of commercial aviation; the Commodity Flow Survey; transborder movement of freight by mode of transportation; a census of ferry operations, precursor safety data for transit operations, and data on near misses and equipmen failures in offshore operations.

To date, BTS data practices have focused on making data accessible, particularly when complying with the current Administration's Open Data memorandum. Planning for the long-term preservation and re-use of BTS data has been uneven, as BTS data programs' focus has been on creation, compilation, and dissemination. In adopting and employing modern data dissemination practices and visualization tools, there is a greater understanding of, and desire for, improved practices to ensure preservation and reuse of BTS data for the long-term.

NTL's new Data Curator has focused on education, training, and consultation on data management best practices, as well as developed and presented an overview of the topic to each BTS office, called "Data Management Roadshows." One tool NTL has created for the sessions is the "Data Management Planning 'Greenout' Bingo Self-Assessment Tool." By honestly assessing each prompt, users create a visual representation of data management strengths and weaknesses around a specific dataset.



During the hour-long Roadshows, the Data Curator walks the audience through using the Bingo tool, using a dataset they produce as an example.

NTL's Data Curator can then help data creators implement missing data management practices, thus improving the preservation and reuse potential of these publicly accessible datasets.

USDOT & NTL Public Access Links

USDOT Public Access Plan:

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

USDOT Open Government site: https://www.transportation.gov/mission/open/open-government

> NTL Public Access Guidance site: http://ntl.bts.gov/publicaccess/

NTL "Creating DMPs" Guidance:

http://ntl.bts.gov/publicaccess/creatingaDMP.html

NTL Public Access Plan FAQs: http://ntl.bts.gov/publicaccess/FAQs.html

> **NTL Digital Repository:** http://ntlsearch.bts.gov/

USDOT Research Hub:

http://ntlsearch.bts.gov/researchhub/index.do

NTL Sufficiency Checklist for DMP Evaluation

USDOT's Public Access Plan requires DMPs submitted to address the following areas:

- Data Description
- Standards Used; Access Policies;
- Re-Use, Redistribution, and Derivative Products Policies; and,

Archiving and Preservation Plans.

USDOT and NTL staff created guidance pages for intramural and extramural data creators, which include prompts for the types of information that could be supplied in each section listed above. These include, among many others:

- List in what format(s) the data will be collected, and whether
 - they are open or proprietary. Describe what data will be publicly shared, and how data files
- will be shared.

Using the information prompts from these guidance pages, NTL staff has created a "DMP Sufficiency Checklist" in order to aid USDOT staff in the assessment of submitted data management plans. Example pages of the checklist are shown in the images.

The DMP discusses the period of time data will be collected and The DMP describes the relationship between the new data collected for this effort and any existing data also used.. 2.07 The DMP lists potential users of the data. The DMP discusses the potential value the data have over the long-term for not only U.S. DOT, but also for the public. If the DMP contains a request permission to not make the data publicly accessible, it explains the rationale for lack of public 2.10 The DMP indicates the party responsible for managing the data. The DMP describes how project leads will check for adherence to this data management plan.

Total of checked boxes for each column, out of 11: 6 1 2 2 • Did a majority of the prompts rate "Explained Fully"? is responsible for managing data, and how the data will be managed?

Re-Use, Redistribution, and Derivative Products Policies 5.01 The DMP names the party who has the right to manage the data. The DMP indicates who holds the intellectual property rights to the The DMP lists any copyrights to the data, and indicates who owns 5.04 The DMP discusses any rights be transferred to a data archive. 5.05 The DMP describes how the data will be licensed for reuse. Total of checked boxes for each column, out of 5: 4 0 1

The "DMP Sufficiency Checklist" is designed to assist evaluators assessing the sufficiency of the DMPs required of research projects funded by USDOT. The checklist provides a quantitative measure to guide what is essentially a qualitative decision. Further, the check list helps to identify the precise areas where researchers should provide more data management information.

A sufficiently detailed DMP, like a well-developed research methodology, is one vital component of a research proposal for extramural researchers and the research project plan for intramural researchers.

Data Citation Recommendations

FORCE11 in 2014.

Properly citing datasets is a hot topic wherever public and open access are discussed. Many data creators feel datasets should be given the same weight as publications, and would like to receive credit for the work they do. NTL staff share the view that datasets should be cited in the same manner as other information resources. As USDOT is implementing a new dataset management system, NTL felt it was important to set a data citation recommendation before its launch.

Other federal scientific and technical information (STI) agencies with which NTL partners share the same desire to promulgate a data citation standard for datasets produced by their agencies as part of a data management program.

Work by U.S. DOT Office:

Corporate Author. (Release Date/Year of publication, Update Frequency or Revision, Modified Date). Title of dataset. Version number/Edition number. Subset information. [Data file type]. Archive/Source. Accessed date from Persistent

U.S. Department of Transportation, Federal Highway Administration. (2015). Public Road and Street Mileage by Functional System(a) 1990-2013. Table 1-5: U.S. [statistical table]. National Transportation Statistics. Accessed 2016-07-15 from http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/ national_transportation_statistics/html/table_01_05.html

This is the date the dataset

ongoing updates that may

**Optional. Please provide if | constitute a new version or | Format 2: Updated Month Year

If you use only a subset of

the full dataset, you can

create a "micro-citation"

elling readers precisely

| book, and allows for greater |

users. Please be as granular

reproducibility for other

occur daily, weekly, monthly, or annually, but may not Ex. 1: (2015, Updated daily).

edition of the data set. Ex. 2: (2014, Updated 2016-07).

which specific data you | Ex. 1: 2003-10-15 to 2003-11-18.

Citation Component Explanation

Usage Tips & Examples

Ex. 2: Taggert, Robert, and Tavian Williams

Ex. 2: (2015-10). (Use for: October 2015

7x. 3: (2013-05-26). (*Use for:* May 26, 2013).

Ex. 1: National Census of Ferry Operators (NCFO).

Ex. 3: State Transportation Statistics.

itilized. This is similar to Ex. 2: National Census of Ferry Operators. Terminal Informatio

citing a specific page from a | subset (Terminal; Year 2010; All Items; All Variables).

Ex. 2: Airline Delays, Cancellations and Tarmac Times.

Ex. 3: Stroizer, Tom, Peter Patterson, Kelly Rivers, et al.

x. 2: U.S. Department of Energy and U.S. Geological Survey. x. 3: U.S. Department of Transportation, Bureau of Transporta-

ORCID iDs should be included after each author or principle inve

rmat: http://orcid.org/0000-0002-0543-4268
. 1: Christiansen, Leighton L http://orcid.org/0000-0002-0543-4268

543-4268, Amanda J. Wilson http://orcid.org/0000-0001

Although optional, the update frequency or modification date

Work by named author:

Author(s)/Principle Investigator(s) ORCID iDs. (Release Date/Year of publication, date Frequency or Revision/Modified Date). Title of dataset, Version number/ Edition number. [Data file type]. Archive/Source. Accessed date from Persistent identifier or URL

Moore, Jason K. http://orcid.org/xxxx-xxxx-xxxx-xxxx, Koojiman, J. D. G. http:// orcid.org/xxxx-xxxx-xxxx-xxxx, & Schwab, Arend L. http://orcid.org/xxxx-xxxx-xxxxxxxx . (2015-06-23). Delft Instrumented Bicycle Data and Videos. [dataset]. Zenodo. Accessed 2016-07- 19 from http://doi.org/10.5281/zenodo.18862

lataset was created or

locate the dataset.

e for the preservation of Ex. 2: National Transportation Data Archive

x. 1: 2013-05-26 from

as this may cause link to break

Ex. 1: http://dx.doi.org/10.5284/1000389

ble for the distribution of Ex. 2: Dryad Digital Repository.

the data. It could also be the publisher of the data.

Ex. 4: In Transportation Statistics Annual Report December

persistent identifier is not provided, you may have to use the

xisting URL. If using URL, it should include the full path of the

Do **not** add a period or other punctuation after the DOI or URL

Ex. 3: http://www.its.dot.gov/communications/image_gallery/

Format: Digital Object Identifier (DOI) (long format)

Work by U.S. DOT Office, with subset information:

Corporate Author. (Release Date/Year of publication, Update Frequency or Revision/Modified Date). Title of dataset. Version number/Edition number. Subset information. [Data file type]. Archive/Source. Accessed date from Persistent identifier

U.S. Census Bureau. (2015). 2010-2014 American Community Survey 5-Year Estimates. Subset: DP05 - ACS Demographic and Housing Estimates. [statistical table]. American FactFinder. Accessed 2016-07-27 from http://factfinder.census.gov/ faces/table-services/jsf/pages/productview.xhtm-l?pid=ACS 14 5YR DP05&src=pt

Digital Repository Landing Page Best Practices

While individual digital repositories may each have unique characteristics, a certain amount of standardized information can help users find information and data more easily across platforms. FORCE11, a collaborative effort comprised of librarians, scholars, archivists, publishers, and research funders, released its recommendations of best practices for digital repository landing pages in 2015¹⁷. In order to improve knowledge creation and sharing, FORCE11 recommends that persistent identifiers (PIDs) (e.g. digital object identifiers (DOIs), archival resource keys

(ARKs), etc.) should resolve to landing pages rather than directly to data or information objects themselves.

Metadata and data may have different lifespans, and as metadata (structured or unstructured) is likely to survive data given the possibility of deaccessioning it is important that the persisten identifier lead to something tangible and continually accessible.

Furthermore, there may be times where cited data is not legally available to researchers due to restrictive licenses, confidentiality, or privacy issues; in these circumstances, researchers may still be able to glean sought after information from the

for access. Lastly, if multiple encodings or versions of a dataset exist, a common landing page provides an independent access point for researchers to reach each instantiation.

NTL's Repository and Open Science Access Portal (ROSA P) will follow the FORCE11 guidelines for repository landing pages.

In practice, this will mean that the "Details" tab for each item in ROSA P will contain, at a minimum, the following information, when available or appropriate (see image):

Bureau of Transportation Statistics (BTS) Transportation Services Index (.CSV - 1.01 MB) Download Transportation Services Index Transportation Services Index provides the latest high-level Bureau of Transportation Statistics data in an easy to digest format. N/A [NOTE: Will only display if values present.] N/A [NOTE: Will only display if values present.] U.S. Department of Transportation, Bureau of Transportation Statistics metadata file, or learn how to qualify U.S. Department of Transportation September 2016 Geographic Coverage: United States, Canada, and Mexico CC0 1.0 (Public Domain) Digital Object Identifier: http://dx.doi.org/10.21949/bts_trans_services_index_201609/12345 Recommended Citation: U.S. Department of Transportation, Bureau of Transportation Statistic (2000 to 2016). Transportation Services Index - Monthly Data 2000-Present. [statistical table]. ROSA P. Accessed 2016-11-20 from http://dx.doi.org/10.21949/bts_trans_services_index_201609/12345 NotePad++, Microsoft Excel, or other text or comma separated value file Page Viewed ____ Times
Downloaded ____ Times d14a028c2a3a2bc9476102bb288234c415a2b01f828ea62ac5b3e42f

Persistence Statement

FORCE11 also recommends the inclusion of a persistence statement, which should include the persistence policies of the repository. For example, the statement for ROSA P might state:

The National Transportation Library is committed to maintaining persistent identifiers in *Repository and* Open Science Access Portal (ROSA P) so that they will continue to resolve to a landing page providing metadata describing data and information, including elements of stewardship, provenance, and availability. The National Transportation Library's plan for organizational persistence and succession can be located in its "Digital Curation Policy," which can be accessed on the ROSA P homepage.

Future Actions

 Work with Bureau of Transportation Statistics (BTS) offices to create Data Management Plans (DMPs) for each dataset created by BTS.

 Help BTS and USDOT data creators obtain ORCID iDs. Integrate ORCID iDs into repository search. Train USDOT staff on use of DMP Sufficiency Evaluation Tool (in-person and online module formats).

 Include DMPs in required data package for NTL's digital repository, to enable DMP searching and analysis. Work with federal repositories towards trusted repository status.

 Complete migration to a cloud-based digital repository solution, using CDC's Public Access Platform.

Christiansen, Leighton L http://orcid.org/0000-0002-0543-4268, Amanda J. Wilson http://orcid.org/0000-0001-6580-2328, Nicole Strayhorn http://orcid.org/0000-0001-9328-4556, & Nikki Olsen http://orcid.org/0000-0001-7768-869X. (2017). "Opening Transportation Data for Innovation: Getting Our Public Access Bits in a Row." Transportation Research Board 96th Annual Meeting. Washington, D.C., USA.

Acknowledgements The authors would like to thank Alpha Wingfield for design assistance.

Image Credits: FatCow. (2014). Farm-Fresh highlighter.png. CC BY 3.0. Wikimedia Commons. Accessed 202 -08-11 from https://commons.wikimedia.org/wiki/File:Farm-Fresh_highlighter.png

Citations

1) National Transportation Library. (2016). Homepage. National Transportation Library. Accessed 2016-12-12 U.S. Department of Transportation. (2015). Official DOT Public Access Plan v1.1. Transportation.gc /Accessed 2016-12-12 from: https://www.transportation.gov/mission/open/official-dot-public-access-plan-v cientific Research. [memorandum]. Whitehouse.gov. Accessed 2016-12-12 from: https:// www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf ransportation Library. Accessed 2016-12-12 from: http://ntl.bts.gov/publicaccess/ 6) National Transportation Library. (2016). Creating Data Management Plans for Extramural Research. Nationa

Transportation Library. Accessed 2016-12-12 from: http://ntl.bts.gov/publicacces.

7) Dryad. (2016). Dryad Digital Repository. Accessed 2016-12-12 from: http://www.datadryad.org/ 8) The President & Fellows of Harvard College. (2016). Harvard Dataverse. Accessed 2016-12-12 from: https:

9) European Organization for Nuclear Research (CERN). (2016). Zenodo. Accessed 2016-12-12 from: https:// 10) Regents of the University of Michigan. (2016). ICPSR [Inter-university Consortium for Political and Socia Research]. Accessed 2016-12-12 from: http://www.icpsr.umich.edu/icpsrweb/

11) Digital Curation Centre. (2004-2016). DCC. Accessed 2016-12-12 from: http://www.dcc.ac.uk/ www.chicagomanualofstyle.org/home.html 13) The Modern Language Association. (2016). The MLA Style Center. Accessed 2016-12-12 from: https://

14) American Psychological Association. (2016). APA Style Central. Accessed 2016-12-12 from: http:// 15) U.S. Government Printing Office. (2008). U.S. Government Printing Office Style Manual. Accessed 2016-12 12 from: https://www.gpo.gov/fdsys/search/pagedetails.action?granuleId=&packageId=GPO-STYLEMANUAI

Science 1:e1. Accessed 2016-12-12 from: https://doi.org/10.7717/peerj-cs.1

16) Data Citation Synthesis Group. (2014). Joint Declaration of Data Citation Principles – Final. Accessed 2016-12 -12 from: https://www.force11.org/group/joint-declaration-data-citation-principles-final 17) Starr J, Castro E, Crosas M, Dumontier M, Downs RR, Duerr R, Haak LL, Haendel M, Herman I, Hodson S, Hourclé J, Kratz JE, Lin J, Nielsen LH, Nurnberger A, Proell S, Rauber A, Sacchi S, Smith A, Taylor M, Clark T. (2015) Achieving Human and Machine Accessibility of Cited Data in Scholarly Publications. PeerJ Computer

Discussion and Debate Usage Tips & Examples As we have discussed these recommendations with other interested

While an international standard has not yet been set, NTL consulted existing recom-

mendations from data repositories and data curation organizations (i.e., Dryad⁷, Har-

vard Dataverse⁸, Zenodo⁹, ICPSR¹⁰, Digital Curation Centre¹¹), style guides (i.e., Chica-

go¹², MLA¹³, and APA¹⁴), as well as the required citation information for bibliographic

mendations follow the *Joint Declaration of Data Citation Principles* ¹⁶ released by

Open Science Access Portal (ROSA P). The results are as follows:

items in the U.S. Government Printing Office Style Manual, 2008¹⁵. Further, our recom-

These guidelines will be used by NTL staff to create a "Recommended Citation" for each

publication and dataset in NTL's digital repository, soon to be rebranded Repository and

parties, there has been some concern that the resulting citations may be too long and that some journals have limits on the length of the individual citations they are willing to print.

In response, NTL's perspective is that in order to achieve the greatest reproducibility, specificity, and interoperability, as recommended by the sources we cite, the level of detail in our landing page citations is nec-

Second, it is much more difficult to locate information that has been left out of a citation, than it is to edit a recommended citation to fit a particular journal's length requirement. Further, no technology exists at present to automatically reproduce the varying journal citation styles for citation.

NTL offers our recommended citations for BTS and USDOT datasets as a reflection of current best practices and recommendations to support scientific reproducibility in a digital world.

As well, NTL's aim is that the data citation recommendations will allow users to more fully benefit from current and future networked electronic search, discovery, and retrieval tools.