

Expanding the U.S. Federal Data Curation Community: Year 01 at the National Transportation Library

Montreal, QC, CA, September 19-21, 2017
Research Data Alliance 10th Plenary

Abstract

At the RDA Eighth Plenary Meeting, staff from the National Transportation Library (NTL), part of the Bureau of Transportation Statistics (BTS), introduced the research data community to the efforts NTL is making to open data created by the U.S. Department of Transportation (USDOT), its contract researchers, and BTS to researchers and the public.

Over the past year NTL's work has expanded far beyond the walls of the USDOT. NTL's data curation and digital library staff are engaged in federal and national efforts to improve the quality of U.S. government data as it is created, as well as the preservation of these datasets over their lifecycles. Our goal is to improve the quality of government research and statistical data, allowing our data users – from other federal agencies to the citizen data scientist – to make well-informed and robust decisions.

This poster will highlight NTL and partner activities around:

- Data curation community-building across federal agencies;
- Federal data management and curation reskilling, including curriculum development;
- Trusted repository certification among peer federal agencies;
- Creating a network of academic research repositories which are conformant with the USDOT public access plan to ease the researcher's decision making around choosing a repository for their data;
- NTL data package requirements for federally-funded public access data to be deposited in the NTL repository; &
- Challenges we face as we mature and expand NTL's data sharing infrastructure and practices.

DM/DC Community Building



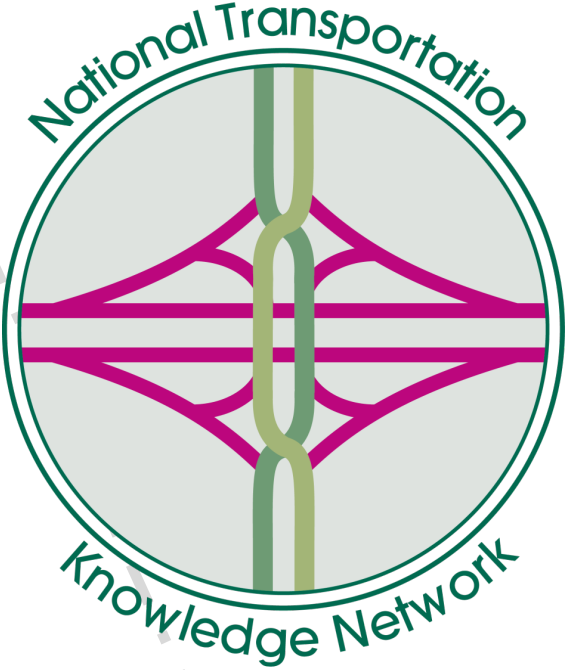
The NTL Data Curator proposed and co-chairs CENDI Data Curation Discussion Group (DCDG)

Purpose: Organize agency hands-on-data management and curation staff to:

- Collaborate across agencies;
- Employ data curation best practices, tools, & workflows;
- Promote efficiencies and consistency;
- Work through challenges;
- Avoid "reinvention"

Growing membership from:

- Dept. of Energy (DOE) OSTI
- Dept. of Defense (DOD)
- Environmental Protection Agency (EPA)
- National Aeronautics & Space Administration (NASA)
- National Institute of Standards & Technology (NIST)
- National Library of Medicine (NIH)
- National Science Foundation (NSF)
- National Transportation Library (DOT)
- U.S. Agency for International Development (USAID)
- U.S. Dept. of Agriculture (USDA)
- U.S. Government Publishing Office (GPO)
- Research Data Alliance (RDA)
- And others (See Leighton to join!)



NTL coordinates the National Transportation Knowledge Network (NTKN), which leads, supports, and guides the nation's transportation knowledge networks in their efforts to organize, share, and preserve transportation information, data, and knowledge for their stakeholders in the transportation community.

Regional TKN members come from state DOTs, university research centers, federal agencies, industry, and other transportation organizations.

DM/DC Reskilling

For Academic Researchers



The NTL Data Curator co-chairs the Transportation Research Board (TRB) Library and Information Science for Transportation (LIST) committee. At the TRB 2017 Annual Meeting, LIST organized a "Public Access Speed Dating" workshop. The workshop allowed researchers and staff from University Transportation Centers (UTCs) to ask questions of federal Public Access experts. UTCs do a large volume of contracted research for the U.S. Department of Transportation. Helping UTC researchers and staff improve their data management practices is a major outreach effort of the NTL. These efforts should lead to improved long-term data curation.

For State DOT Researchers



The NTL Data Curator authored the proposal for, and sits on the panel for, NCHRP 20-110 Project Panel A Guide to Ensure Access to the Publications and Data of Federally Funded Transportation-Related Research.

The objective of this research is to develop guidance and activities to assist state DOTs, other public agencies, and transportation research organizations to efficiently and effectively ensure access to the results of federally funded transportation-related research.

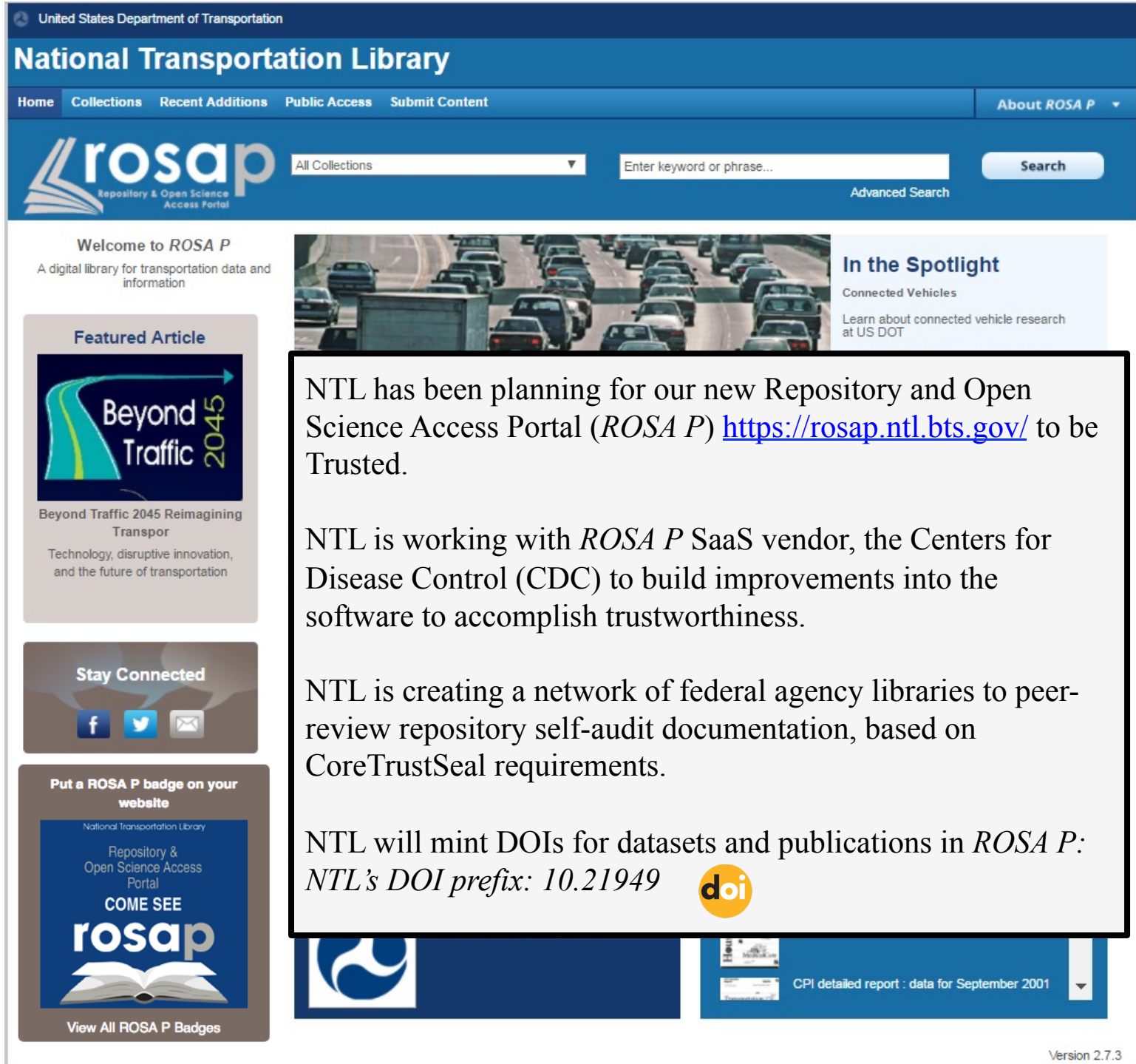
For Federal Employees



The NTL Data Curator and CENDI Data Curation Discussion Group members are co-authoring a Data Management/Data Curation curriculum for federal data workers. Many federal employees have seen their jobs change to include data management and data curation, but very few have had the opportunity for further training. The curriculum modules will allow federal data workers to improve skills on the job, and improve long-term preservation of federal data.

Trusted Repository Certification

NTL's Digital Librarian leads CENDI's Trusted Repository Working Group, which evaluated standards and guidelines for Trusted Repositories, as well as organizing well-attended series of webinars featuring international thought leaders.



Transportation Research Data Repository Network

Proposal: Create a network of USDOT Public Access conformant and trustworthy transportation repositories.

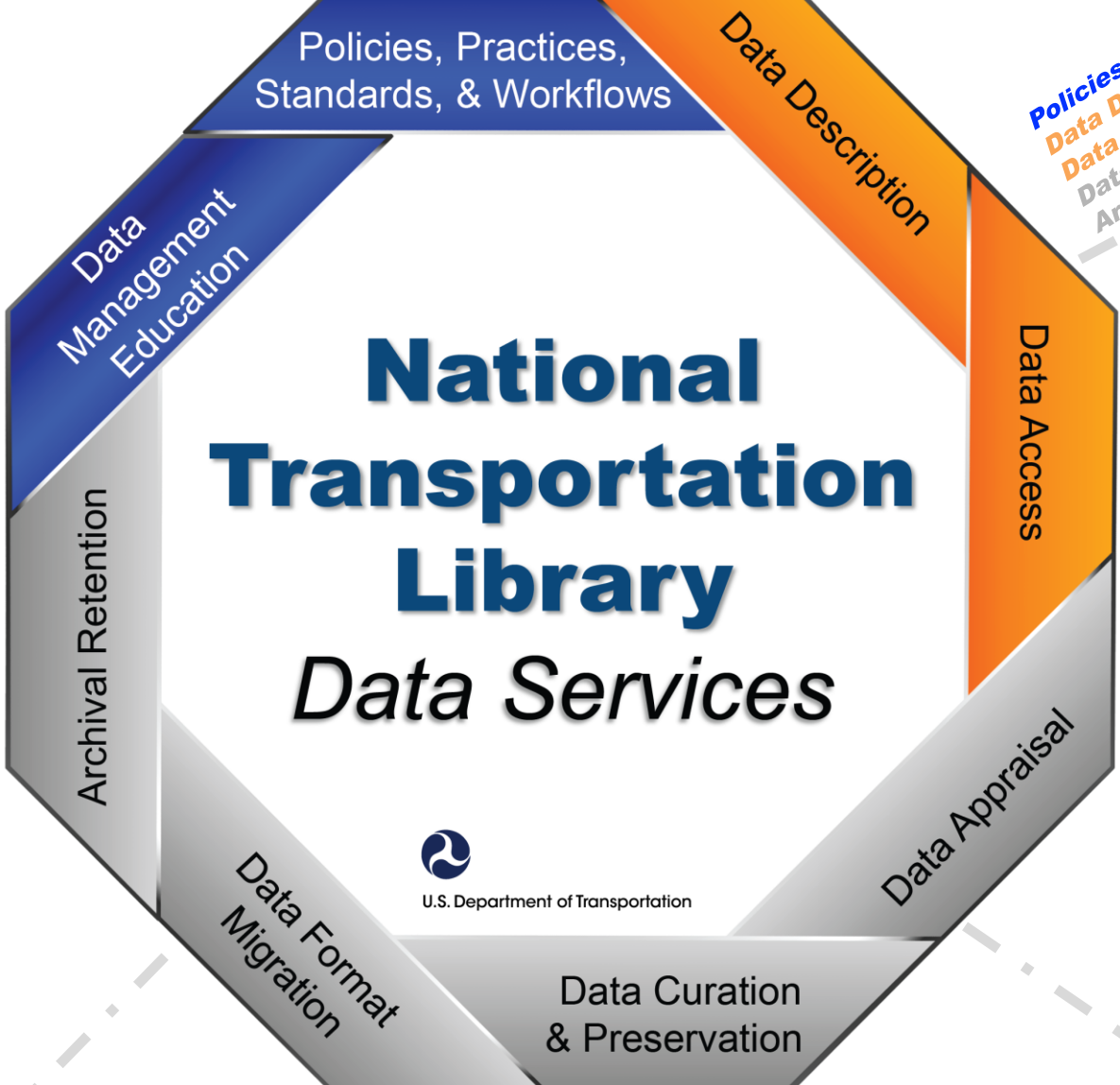
USDOT Public Access Policy: <https://ntl.bts.gov/publicaccess/>
CoreTrustSeal Data Repository Requirements: <https://www.coretrustseal.org/why-certification/requirements/>

The development of such a network of repositories, with substantial transportation data collections, preserving the output of transportation researchers in conformance with USDOT Public Access rules and recommendations, would be a major step forward in transportation data preservation.

First step: NTL seeks funding for 2018 meeting of interested repositories.

Interested? Contact Leighton Christiansen at leighton.christiansen@dot.gov or add a sticky note or business card below:

your name here



Blue: Maturing
Orange: Early stages
Silver: Future Actions

NTL Repository Data Packages

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

NTL Dataset Data Package Guidelines

Dataset .csv or other open format (whenever possible, or complete documentation of proprietary format)

Readme.txt Includes Data Dictionary; Notes standards used; Defines Zero, Null, & Unknown; includes FAQs and other notes

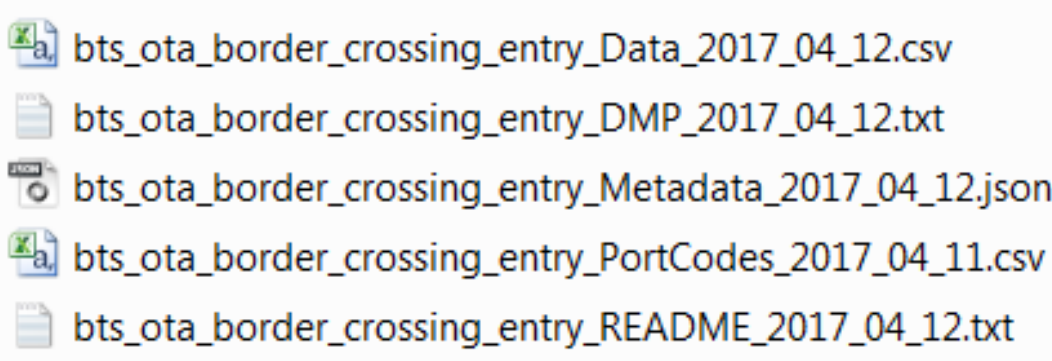
Metadata XML file in Project Open Data .json and/or other schema

Data Management Plan (DMP) in .txt for future machine readability & .pdf for human readability [Is XML next?]

Code or scripts used in data analysis

Supporting files, tables, etc.

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



NTL Recommended practice: Files names should be human readable, consistently formatted, and versioned with a date stamp in ISO 8601.

Adapted from:
Strasser, Carly. Research Data Management: A Primer Publication of the National Information Standards Organization. Baltimore, MD: NISO, 2015.
http://www.niso.org/apps/group_public/download.php/15375/PrimerRDM-2015-0727.pdf
Briney, Kristin. 2015. Data management for researchers: organize, maintain and share your data for research success.

Our Challenges

- Small staff
- Lack of standards usage at federal level
 - Desired Standards:
 - ISO 8601 for Dates and Times
 - ISO 3166-2 for Place names
- Legacy workflows
- Public Access Pushback
- Bad spreadsheet practices
- Lack of ORCID use
- Poorly managed and documented legacy data
- Cloud security

Your Suggestions

Please use sticky notes to leave suggestions or comments. Thanks!

your comments here

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

U.S. Department of Transportation
National Transportation Library

Recommended Citation
Christiansen, Leighton L. <http://orcid.org/0000-0002-0543-4268> (2017).
"Expanding the U.S. Federal Data Curation Community: Year 01 at the National Transportation Library." Research Data Alliance Tenth Plenary. Montreal, Quebec, Canada.