December 12, 2019

National Transportation Library End-of-Year Update 2019

https://transportation.libguides.com/TLR

Audio call-in number: 877-336-1274; Access code: 5759713

National Transportation Library

End of Year Update 2019

Mary Moulton, Digital Librarian
Jesse Long, Fellow
Leighton Christiansen, Data Curator
Xin Wang, Systems Librarian
Vinod Koduri, Senior Web Engineer
Lisa Curtin, STIPDG Intern

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

General NTL Update

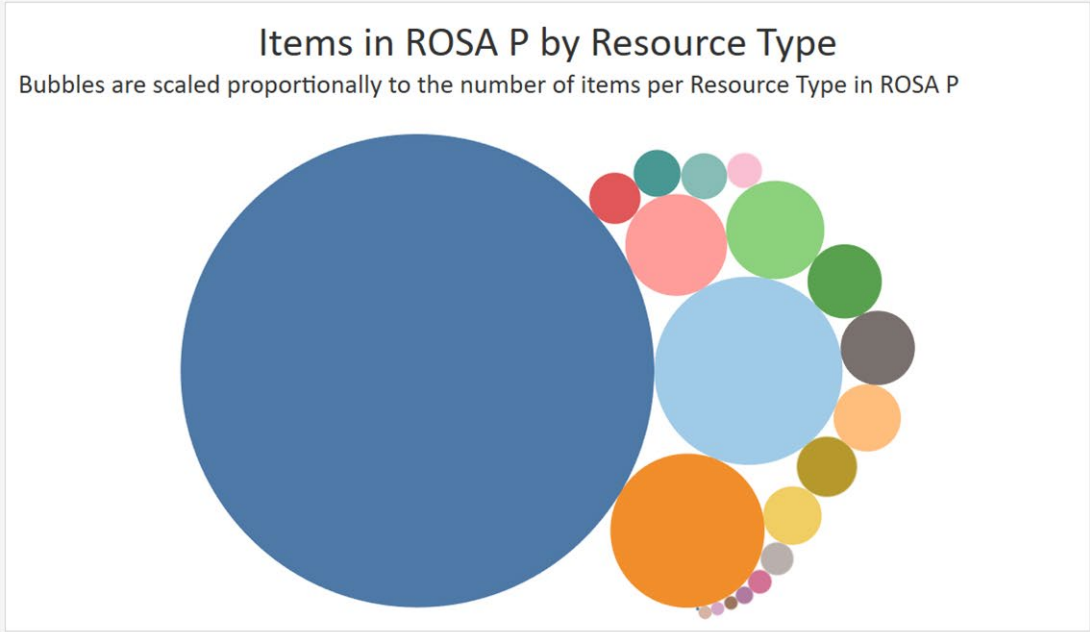
Mary Moulton, Digital Librarian

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

ROSA P: New Collections

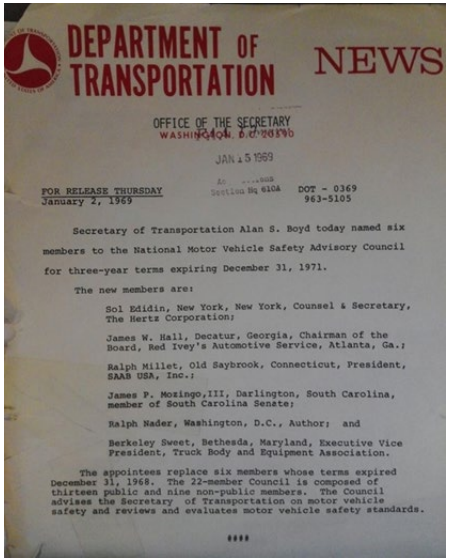
- NTL Digital Special Collections
  - Advisory Circulars
  - CAA and FAA Reports
  - Civil Aeronautics Manuals
  - Civil Aeronautics Regulations
  - Federal Aviation Regulations
  - Historic CAB/DOT Orders
  - Investigations of Railroad Accidents
  - Investigations of Aircraft Accidents
  - National Conferences on Street and Highway Safety
  - Papers by Fairbank, Turner & Macdonald
  - US Coastguard Circulars
- US DOT modal research
  - Federal Motor Carrier Safety Administration (FMCSA)
  - Federal Highway Administration (FHWA)
  - Federal Railroad Administration (FRA)
  - Pipeline and Hazardous Material Safety Administration Office of Pipeline Safety
  - Federal Transit Administration (FTA)
  - National Highway Traffic Safety Administration Vehicle Safety Research
  - National Highway Traffic Safety Administration Behavioral Safety Research

ROSA P: Metrics & Actions



Digitization Projects

- Public Roads
- USACE Port Series
- US DOT Newsletters



DEPARTMENT OF TRANSPORTATION NEWS

OFFICE OF THE SECRETARY
WASHINGTON, D.C. 20590

JAN 15 1969

FOR RELEASE THURSDAY
January 2, 1969

DOT - 0369
963-5105

Secretary of Transportation Alan S. Boyd today named six members to the National Motor Vehicle Safety Advisory Council for three-year terms expiring December 31, 1971.

The new members are:

Sol Edidin, New York, New York, Counsel & Secretary, The Hertz Corporation;

James W. Hall, Decatur, Georgia, Chairman of the Board, Red Ivey's Automotive Service, Atlanta, Ga.;

Ralph Millet, Old Saybrook, Connecticut, President, SAAB USA, Inc.;

James P. Mozingo,III, Darlington, South Carolina, member of South Carolina Senate;

Ralph Nader, Washington, D.C., Author; and

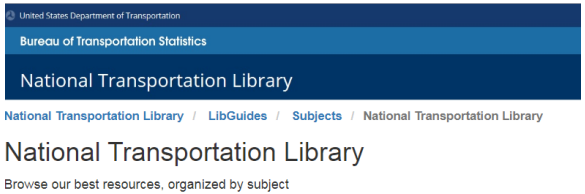
Berkeley Sweet, Bethesda, Maryland, Executive Vice President, Truck Body and Equipment Association.

The appointees replace six members whose terms expired December 31, 1968. The 22-member Council is composed of thirteen public and nine non-public members. The Council advises the Secretary of Transportation on motor vehicle safety and reviews and evaluates motor vehicle safety standards.

####

NTL Subject Guides

- Accessibility
  - Checklist, training resources, NTL policy
- Persistent Identifiers
  - Recommended practices, glossary, references
- Research Tools
  - Transportation information sources, with focus on government and public domain
- ROSA P
  - Guide to search and navigation
- Transportation Librarians Roundtable



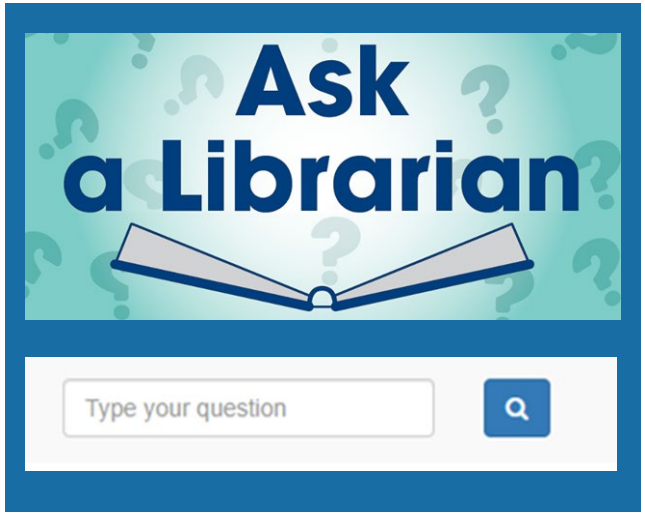
https://transportation.libguides.com/NTLsubjects

Ask-A-Librarian

Using LibApps virtual tools, NTL Reference Librarians provide:

- Place of entry to folks, from congress to the public, that have questions of DOT
- Basic research help for BTS products
- ROSA P assistance
- Referrals to US DOT OAs when appropriate
- Curated selection of answers to FAQs
- Understanding of what is on the public's mind

https://transportation.libanswers.com/



NTL @ TRB 2020

- **US DOT OST-R/NTL/BTS Booth**
  - Exhibit Hall 835
- **Section 508 Compliance and Document Accessibility: How to Make Your Reports Accessible for Everybody**
  - Workshop 1008
  - Sunday, January 12, 9:00 AM – noon
  - Convention Center, 147B
- **Research Data Management for State DOTs**
  - Workshop 1770
  - Thursday, January 16, 8:00 AM – noon
  - Convention Center, 103B
- **National Transportation Data Preservation Network (NTDPN) Workshop 2**
  - Thursday, January 16, 1:00 PM – 2:30 PM
  - Convention Center, 305, or remote participation
- **Moving Forward and Advancing Society with Transportation Libraries, Information, and Data**
  - Poster Session 1486
  - Tuesday, January 14, 10:15 AM – noon
  - Convention Center, Hall A

Data Team Update

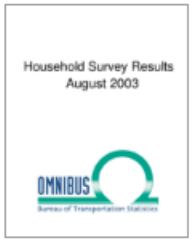
Jesse Long, Fellow
Leighton Christiansen, Data Curator

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

Omnibus Survey Program: 2000-2009

- **Goal:** Create Data Packages for each time the survey was conducted.
  - Omnibus Household Surveys (OHS) = 35
  - Omnibus Targeted Surveys = 7

Starting Point: Final Reports in ROSAP



Household Survey Results August 2003

Published Date: 2003-08-01

Abstract: The Bureau of Transportation Statistics (BTS) is conducting a series of monthly surveys to monitor expectations of and satisfaction with the transportation system and to gather event, issue, and mode-specific information. The surveys will serve as an...

File Type: [PDF - 478.96 KB]

Working with Legacy Data:

Data Package

Key Elements:

1. Dataset
2. Readme.txt
3. Metadata file
4. Data Management Plan

Optional:

5. Codes or Scripts
6. Supporting files, tables, etc.

	A	B	C	D	E
1	Date	Data	Documentation	Supporting Files Found	Data Package Complete
2	2000-08	Y	Y	Summary Tables (also in documentation doc)	Yes
3	2000-09	Y	Y	Summary Tables (also in documentation doc)	Yes
4	2000-10	Y	Y	Summary Tables (also in documentation doc)	Yes
5	2000-11	Y	Y	Summary Tables (also in documentation doc)	Yes
6	2000-12	Y	Y	Summary Tables (also in documentation doc)	Yes
7	2001-01	Y	Y	Summary Tables (also in documentation doc), Data Collection Report	Yes
8	2001-02	Y	Y	Summary Tables (also in documentation doc), Data Collection Report	Yes
9	2001-03	Y	Y	Summary Tables (also in documentation doc), Data Collection Report	Yes
10					
11	2001-07	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
12	2001-08	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
13	2001-10	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
14	2001-11	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
15	2001-12	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
16	2002-01	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
17	2002-02	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
18	2002-03	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
19	2002-04	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
20	2002-05	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
21	2002-06	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
22	2002-07	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
23	2002-08	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
24	2002-09	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
25	2002-10	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
26	2002-12	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
27	2003-02	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
28	2003-04	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
29	2003-06	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
30	2003-08	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
31	2003-10	Y	Y	data dictionary (.xls), survey table results (.xls, .pdf, .doc, .txt), SASLabels (.txt), SASFormat Library (.txt)	Yes
32					
33	2004-12	N	N	Nothing found	
34	2005-10	Y	N	data dictionary (.xls), survey table results (.txt), SASLabels (.txt), SASFormat Library (.txt), List of Variables to BTS (.xls)	
35	2006-11	Y	Y	data dictionary (.xls), survey results/frequency tables-also in stats format (.txt, .xls), SASFormat Library (.txt), SAS Labels (.txt), BTS Special Report for OHS 2006-2007 data (.pdf)	
36	2007-11	N	Y	data dictionary (.doc), data collection plan (.doc), stats for response rates (.xls), BTS Special Report for OHS 2006-2007 data (.pdf), BTS Special Report for OHS 2007-2008 data (.pdf)	
37	2008-11	Y?	Y	report tables (.xls), data dictionary (.xls), BTS Special Report for OHS 2007-2008 data (.pdf), BTS Special Report for OHS 2008 data (.pdf)	
38	2009-10	Y	Y	Survey results/frequency tables for both NAT and MSA data (.xls)	Yes

Omnibus Household Surveys (OHS) Inventory:

Common Issues:

Inconsistent or unknown naming structure

+ Hard to find files
+ Basic names
+ Repeating names
+ Unknown contents

Various file locations

+ Found new data at various stages of the project and had to back track

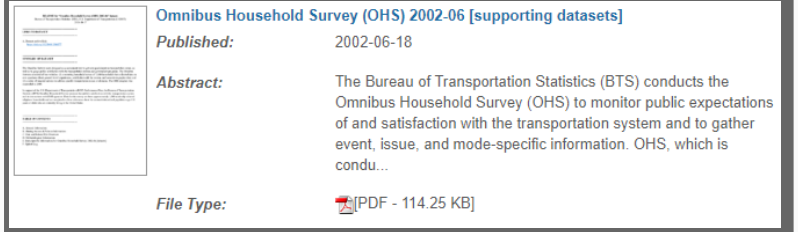
Limited Documentation

+ Creates gaps that make understanding and preserving the data difficult

Progress:

National Transportation Data Archive (NTDA)
https://doi.org/10.21949/1504517

+ 30 uploaded
+ 6 in-progress
+ 5 missing information



Omnibus Household Survey (OHS) 2002-06 [supporting datasets]

Published: 2002-06-18

Abstract: The Bureau of Transportation Statistics (BTS) conducts the Omnibus Household Survey (OHS) to monitor public expectations of and satisfaction with the transportation system and to gather event, issue, and mode-specific information. OHS, which is condu...

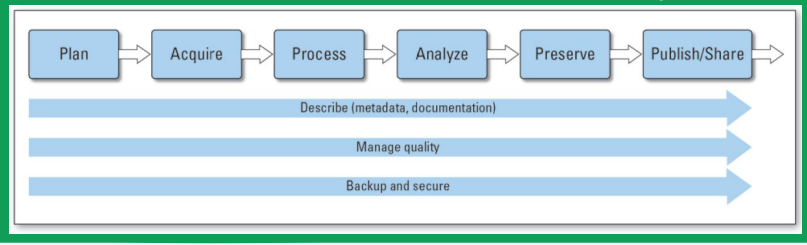
File Type: [PDF - 114.25 KB]

Looking Ahead:

Preventing issues identified by legacy data through:

+ Data Management and Sharing Plans
+ Embedded Data Curators

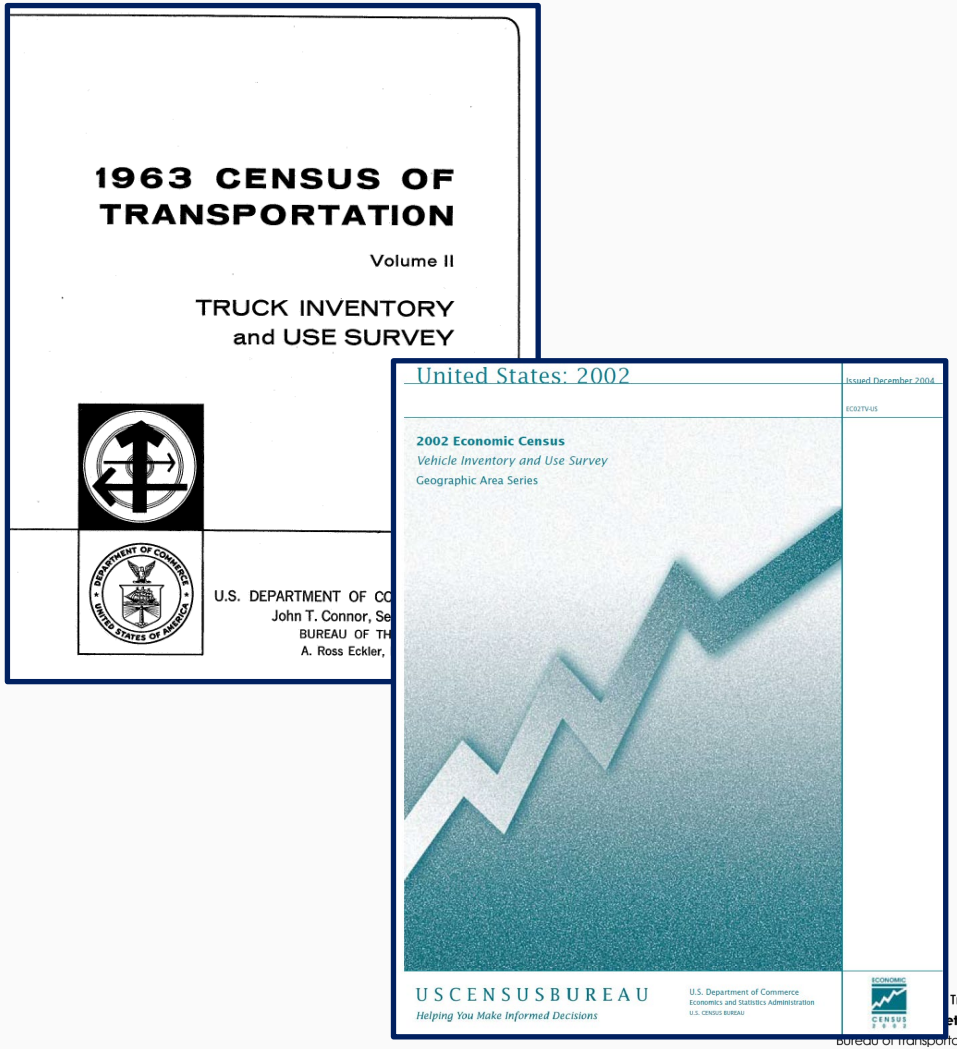
USGS Data Lifecycle Model



Vehicle Inventory & Use Survey (VIUS)

Historic Census Surveys halted in 2002

https://www.census.gov/library.html



Transportation
etary of Transportation
Bureau of Transportation Statistics

VIUS Print Data: PDF Prisoners

1963, Figure 1

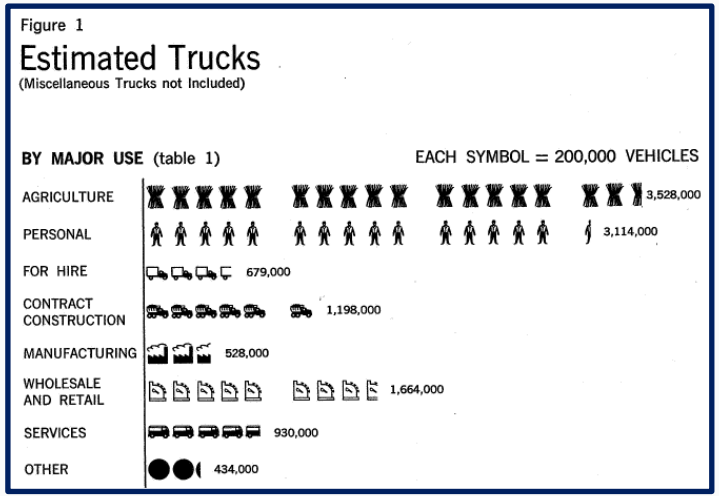


Figure 1

Estimated Trucks

(Miscellaneous Trucks not Included)

BY MAJOR USE (table 1) EACH SYMBOL = 200,000 VEHICLES

AGRICULTURE	3,528,000
PERSONAL	3,114,000
FOR HIRE	679,000
CONTRACT CONSTRUCTION	1,198,000
MANUFACTURING	528,000
WHOLESALE AND RETAIL	1,664,000
SERVICES	930,000
OTHER	434,000

2002, Table 1a

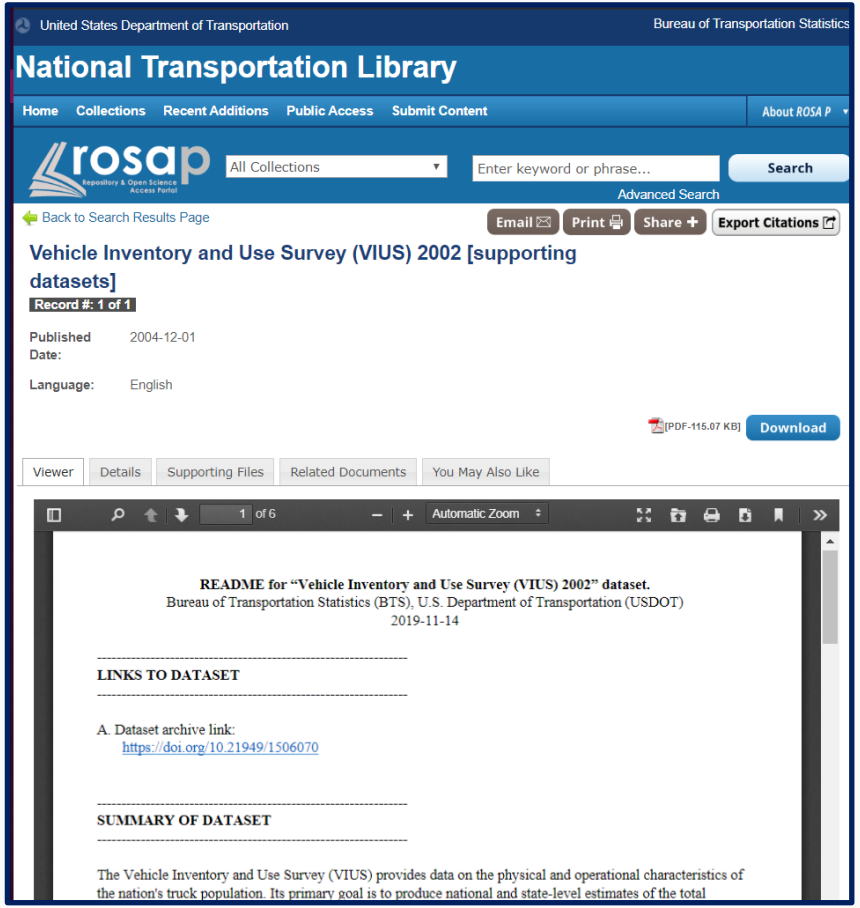
Table 1a. **Trucks—Comparative Summary: 2002**

[Estimates are shown as percents and are based on data from the 2002 and 1997 Vehicle Inventory a estimates may not be additive]

Vehicular and operational characteristics	2002	1997	1992	1987
Total	**100.0**	**100.0**	**100.0**	**100.0**
BUSINESS[1]				
Agriculture, forestry, fishing, or hunting	2.6	5.0	6.4	8.5
Mining	.2	.3	.4	.5
Construction	5.3	8.3	8.4	10.0
Manufacturing	.9	1.0	1.3	1.3
Wholesale and retail trade	2.7	4.8	5.2	5.6
For-hire transportation or warehousing	1.5	1.5	1.5	2.2
Utilities and all services	5.2	6.7	6.2	5.5
Personal transportation	76.7	70.0	68.3	65.7
Other, not reported, and not applicable	4.9	2.4	2.2	.6

VIUS 2002 Data

Now in Machine-Readable Format!!
With “Documentation”
(Amaze your friends!)



https://doi.org/10.21949/1506070

Public Access Updates

GAO Report

FEDERAL RESEARCH: Additional Actions Needed to Improve Public Access to Research Results

"implement mechanism to ensure researcher compliance"

https://www.gao.gov/products/GAO-20-81

PA Plan Review

- Published in December 2015
- Planned Review
- Reflect:
  - ~ GAO report;
  - ~ Evidence-Based Policy Act;
  - ~ Lessons learned

Feed Back?
public.access@dot.gov

National Transportation Data Preservation Network (NTDPN):

"Workshop 2" at TRB, January 16

https://doi.org/10.21949/1506118

Systems Team Update

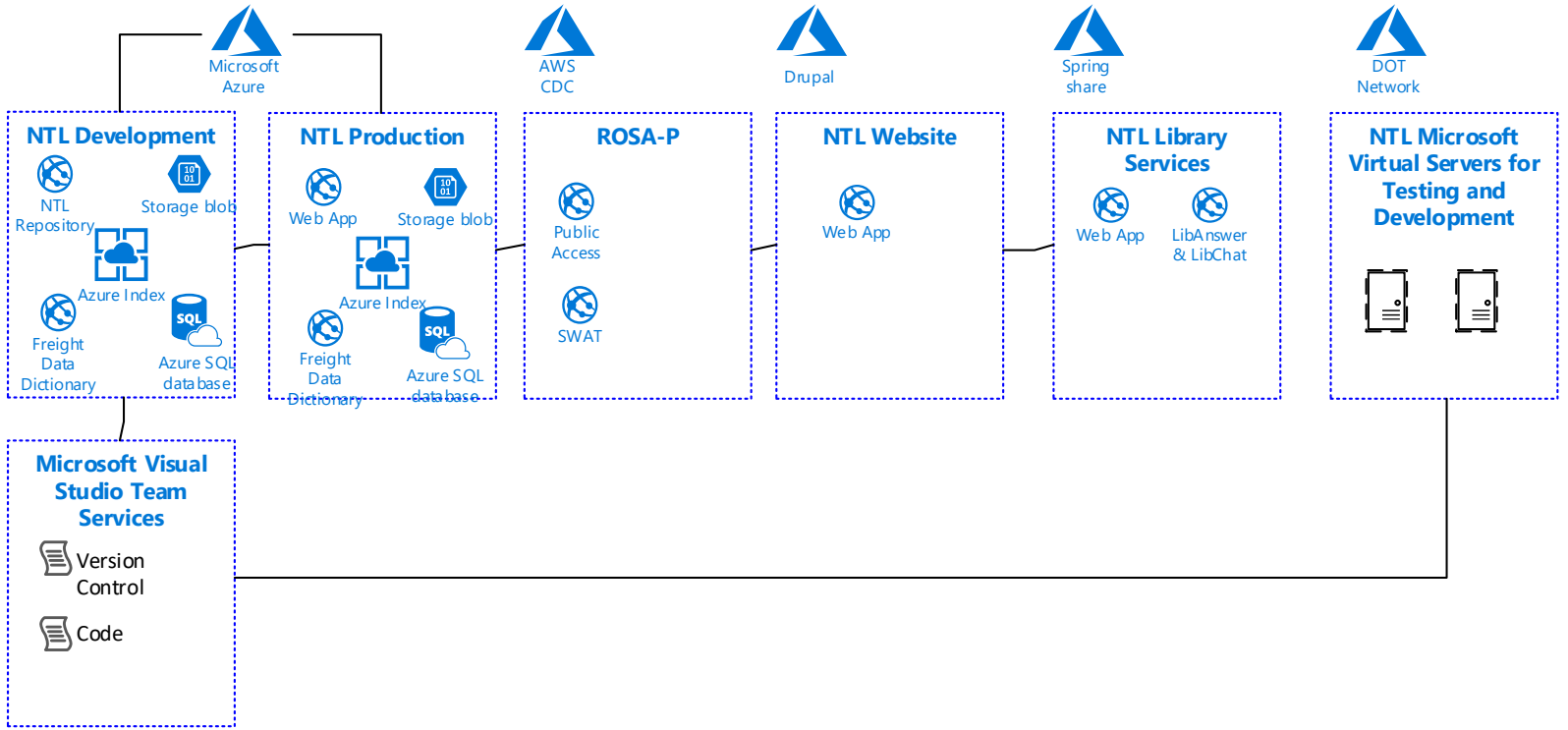
Xin Wang, Systems Librarian
Vinod Koduri, Senior Web Engineer

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

NTL IT System Improvements

- NTL Repository 2.0 in Azure
- ROSA_P (NTL DC for public access)
- NTL Repository 3.0 in AWS
- Mediated Submission APP

NTL Infrastructure Environment



U.S. DEPARTMENT OF TRANSPORTATION

NATIONAL TRANSPORTATION LIBRARY REPOSITORY

Who?

As part of the Bureau of Transportation Statistics, the National Transportation Library maintains a broad collection of statistical and transportation information. The goal is to provide information to DOT staff, transportation researchers, and policymakers at all levels.

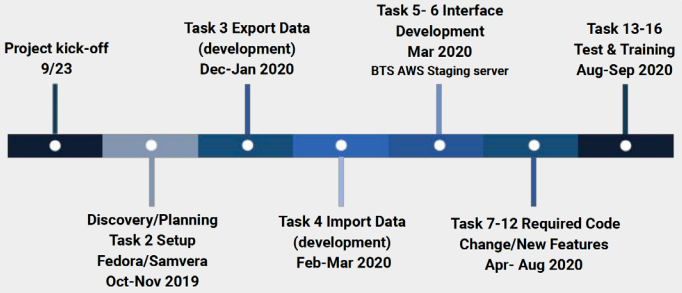
What?

- Build a new repository with Fedora in AWS
- Utilize ROSA-P interface and Workroom to handle current functionality during development
- Integrate with required internal and external interfaces
- Simple and intuitive user experience

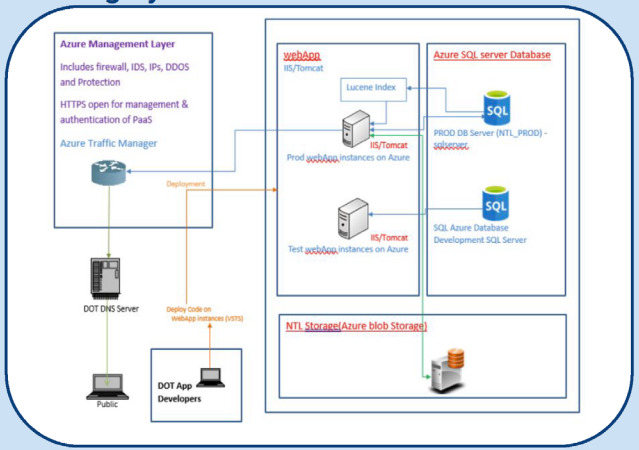
Why?

Modernize the NTL repository system to improve business user experience and realize potential cost savings.

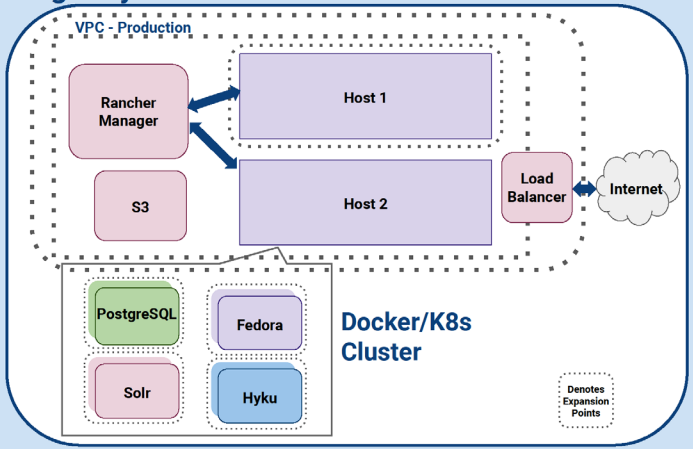
When? 12-month development timeline, to launch September 2020



Existing System Architecture



Target System Architecture



NTL Submissions

Xin Wang, Systems Librarian
Vinod Koduri, Senior Web Engineer

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

Background

Submissions page is designed to replace the email submission system where NTL catalogers used to receive data and documents thru emails.

- Improved metadata accuracy
- More efficiency and less work for catalogers
- More flexibility for catalogers
- Improved analysis and decision making

Features

- Let users input all the required metadata fields to submit
- Catalogers can view/edit/update/delete the user summitted records
- All the work is archived for the future needs
- Records can be directly transferred to workroom

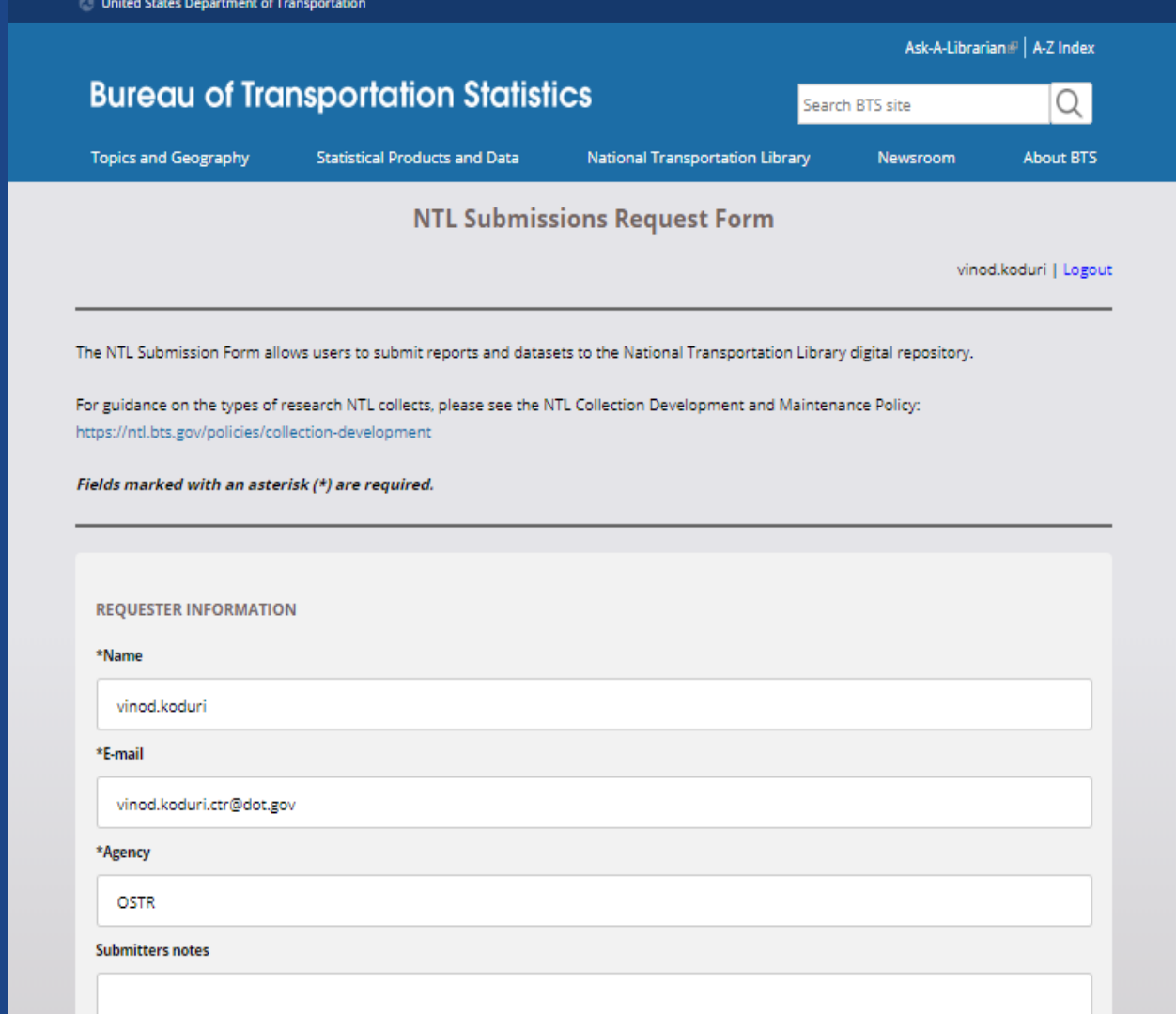
Where is it now

- Development completed
- Completed user testing and changes/fixes
- Added New standard BTS headers and footers
- Production ready in couple of weeks

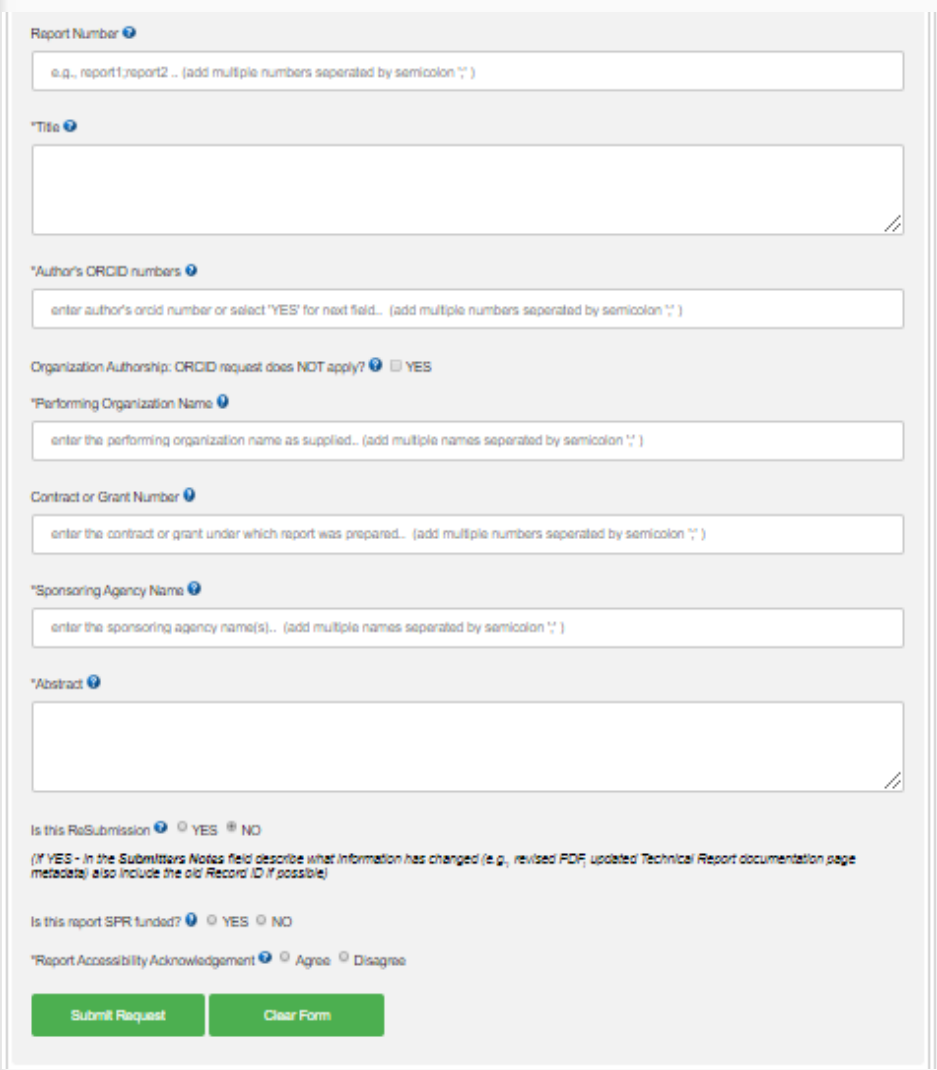
More Updates..

- UI changes on how the data is displayed
- Integrate spell check functionality

Submission home page



Submission home page



Reviewer Page

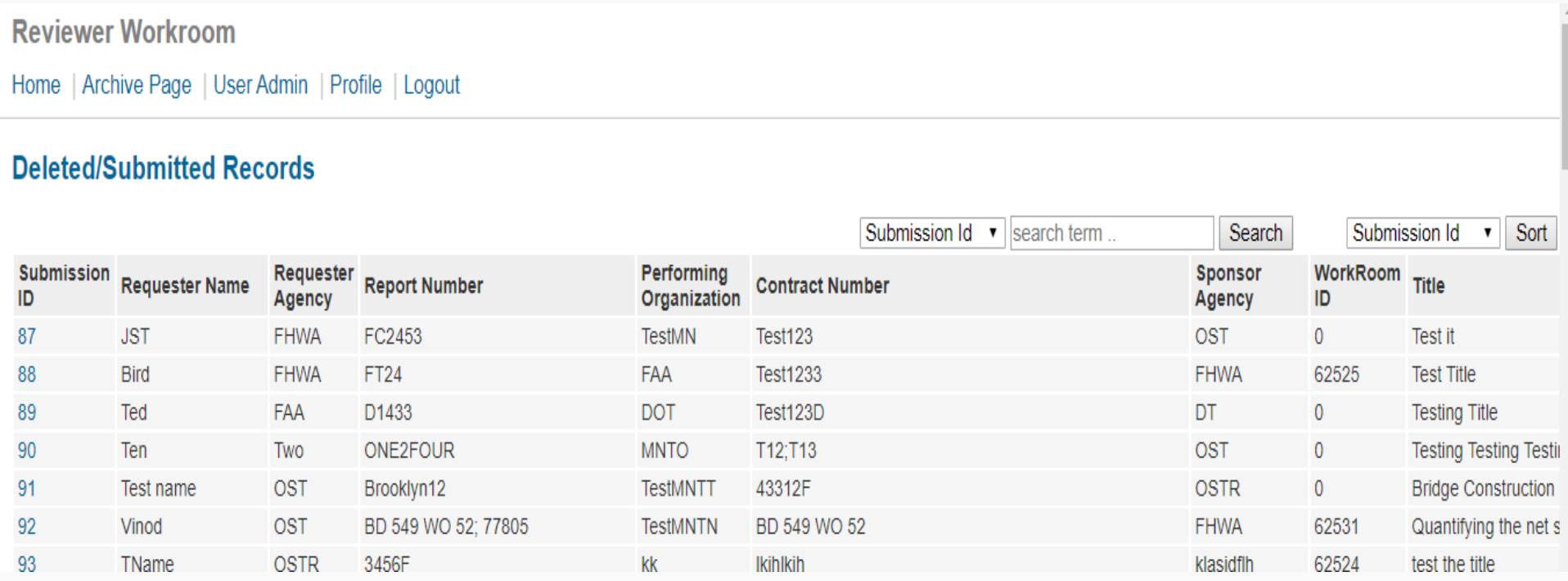
Reviewer Workroom

Home | Archive Page | User Admin | Profile | Logout

Pending Records

Submission ID	Requester Name	Requester Agency	Report Number	Title	ReSubmit	View/Edit | Delete	Transfer to Workroo
109	test name	OST	F3334	Testing		view/edit | delete	transfer
110	Rales	FHA	test123	Test title		view/edit | delete	transfer
124	xin.wang	OSTR		ddddddddddddddddddddddddd		view/edit | delete	transfer
125	judith.salter	OSTR	Submissions_Test_Firefox_Judith_1	Submissions_Test_Firefox_Judith: Test 1		view/edit | delete	transfer
150	vinod.koduri	OSTR		testing		view/edit | delete	transfer
155	vinod.koduri	OSTR	test	ttestupdate		view/edit | delete	transfer

Archive Page



Submission ID	Requester Name	Requester Agency	Report Number	Performing Organization	Contract Number	Sponsor Agency	WorkRoom ID	Title
87	JST	FHWA	FC2453	TestMN	Test123	OST	0	Test it
88	Bird	FHWA	FT24	FAA	Test1233	FHWA	62525	Test Title
89	Ted	FAA	D1433	DOT	Test123D	DT	0	Testing Title
90	Ten	Two	ONE2FOUR	MNTO	T12;T13	OST	0	Testing Testing Testi
91	Test name	OST	Brooklyn12	TestMNTT	43312F	OSTR	0	Bridge Construction
92	Vinod	OST	BD 549 WO 52; 77805	TestMNTN	BD 549 WO 52	FHWA	62531	Quantifying the net s
93	TName	OSTR	3456F	kk	lkihlkih	klasidflh	62524	test the title

2019 Intern Projects @ NTL

Legacy Data Rescue Research
Persistent Identifier LibGuide

Lisa A. Curtin
MSIS ‘20, University of Tennessee-Knoxville
USDOT STIPDG Intern, Summer 2019
https://orcid.org/0000-0003-1137-7789

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

Legacy Data Rescue

"Rescuing data usually expands the temporal and/or spatial coverage of a database's holdings which can be accessed for study." - Griffin, 2015

Legacy data are data collected or compiled in the past, stored in an obsolete format.

Obsolete formats include data on paper, and other physical storage media such as video and audio cassettes, magnetic tape, floppy disk, etc.

Data rescue is the process of making legacy data accessible.

This may be as simple as scanning documents or copying data from an old storage device to a newer one.

It may be as complex as using deep-learning or crowd-sourcing to translate legacy data into machine-readable data.

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

NTL's Current Capability for Data Rescue

Using Adobe Acrobat DC Pro on scanned PDFs

- Enhance scans
- Recognize text (OCR)
- Highlight text
- Export selection as .docx
- Format text in Word
  - "Normal" text formatting
  - Remove artifacts
  - Standardize font & text size
  - Remove spacing errors and OCR noise
- Paste .docx content into Excel

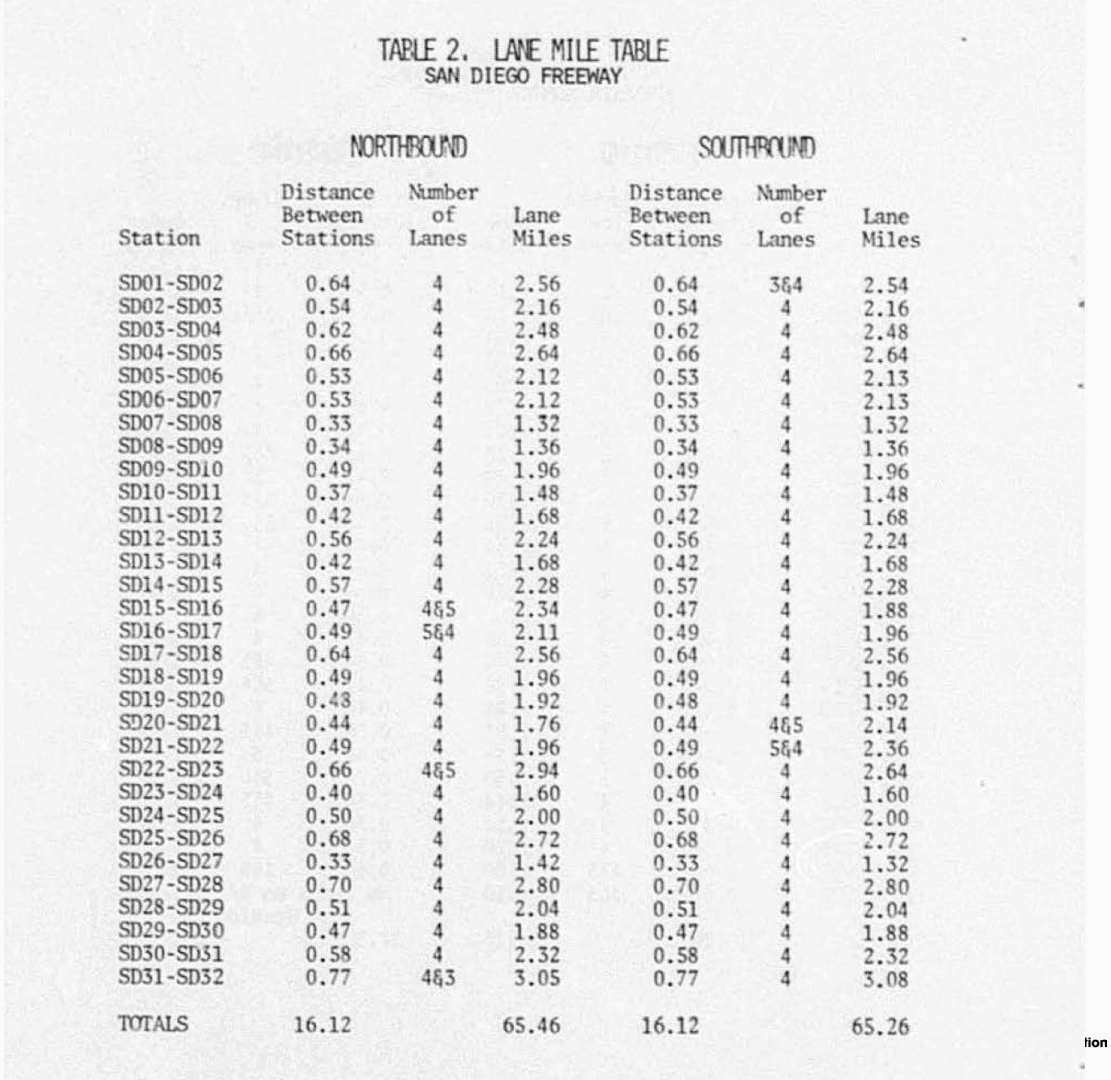


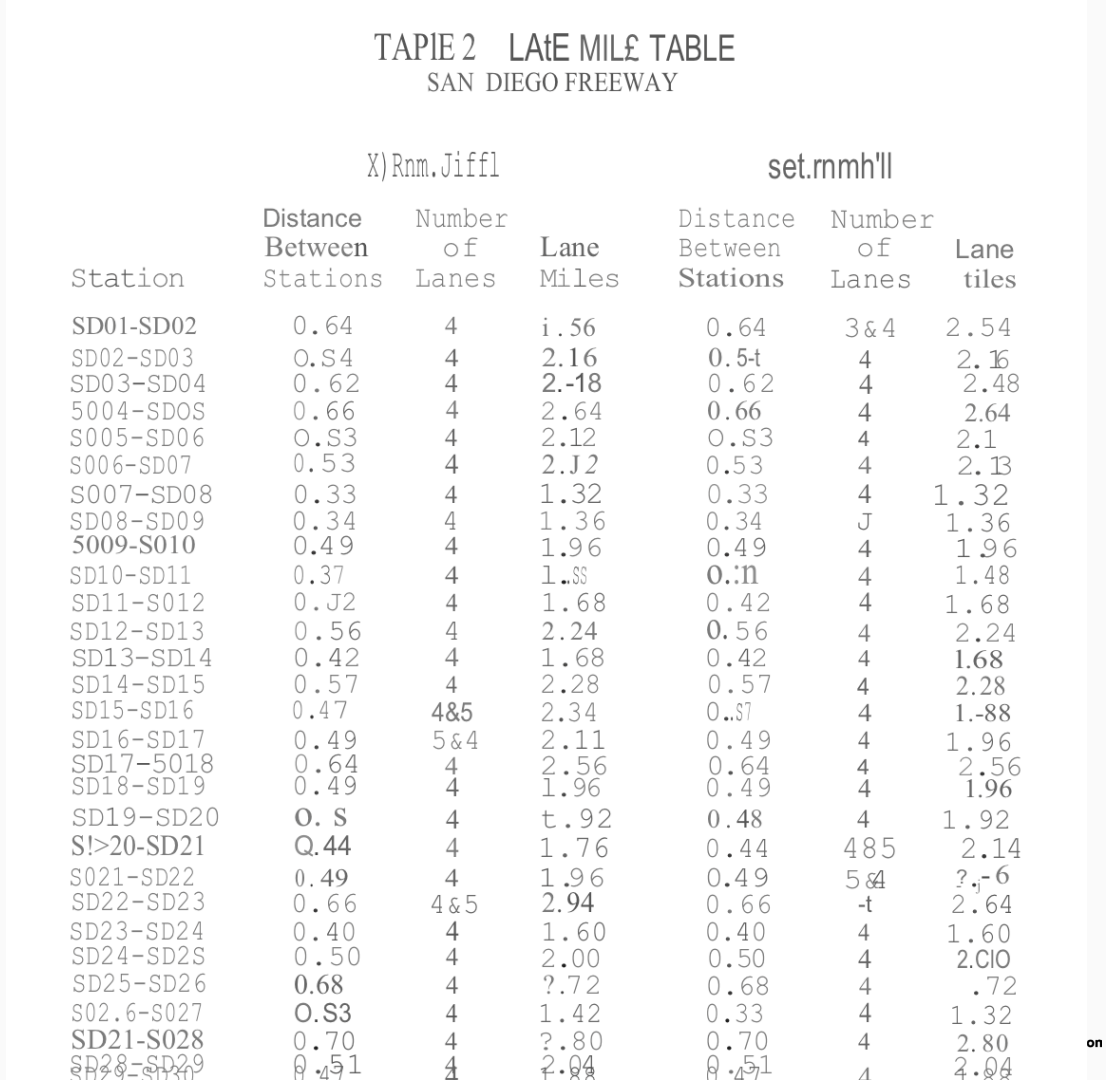
TABLE 2. LANE MILE TABLE
SAN DIEGO FREEWAY

	NORTHBOUND			SOUTHBOUND		
Station	Distance Between Stations	Number of Lanes	Lane Miles	Distance Between Stations	Number of Lanes	Lane Miles
SD01-SD02	0.64	4	2.56	0.64	3&4	2.54
SD02-SD03	0.54	4	2.16	0.54	4	2.16
SD03-SD04	0.62	4	2.48	0.62	4	2.48
SD04-SD05	0.66	4	2.64	0.66	4	2.64
SD05-SD06	0.53	4	2.12	0.53	4	2.13
SD06-SD07	0.53	4	2.12	0.53	4	2.13
SD07-SD08	0.33	4	1.32	0.33	4	1.32
SD08-SD09	0.34	4	1.36	0.34	4	1.36
SD09-SD10	0.49	4	1.96	0.49	4	1.96
SD10-SD11	0.37	4	1.48	0.37	4	1.48
SD11-SD12	0.42	4	1.68	0.42	4	1.68
SD12-SD13	0.56	4	2.24	0.56	4	2.24
SD13-SD14	0.42	4	1.68	0.42	4	1.68
SD14-SD15	0.57	4	2.28	0.57	4	2.28
SD15-SD16	0.47	4&5	2.34	0.47	4	1.88
SD16-SD17	0.49	5&4	2.11	0.49	4	1.96
SD17-SD18	0.64	4	2.56	0.64	4	2.56
SD18-SD19	0.49	4	1.96	0.49	4	1.96
SD19-SD20	0.48	4	1.92	0.48	4	1.92
SD20-SD21	0.44	4	1.76	0.44	4&5	2.14
SD21-SD22	0.49	4	1.96	0.49	5&4	2.36
SD22-SD23	0.66	4&5	2.94	0.66	4	2.64
SD23-SD24	0.40	4	1.60	0.40	4	1.60
SD24-SD25	0.50	4	2.00	0.50	4	2.00
SD25-SD26	0.68	4	2.72	0.68	4	2.72
SD26-SD27	0.33	4	1.42	0.33	4	1.32
SD27-SD28	0.70	4	2.80	0.70	4	2.80
SD28-SD29	0.51	4	2.04	0.51	4	2.04
SD29-SD30	0.47	4	1.88	0.47	4	1.88
SD30-SD31	0.58	4	2.32	0.58	4	2.32
SD31-SD32	0.77	4&3	3.05	0.77	4	3.08
TOTALS	16.12		65.46	16.12		65.26

NTL's Current Capability for Data Rescue

Using Adobe Acrobat DC Pro on scanned PDFs

- Enhance scans
- Recognize text (OCR)
- Highlight text
- Export selection as .docx
- Format text in Word
  - "Normal" text formatting
  - Remove artifacts
  - Standardize font & text size
  - Remove spacing errors and OCR noise
- Paste .docx content into Excel



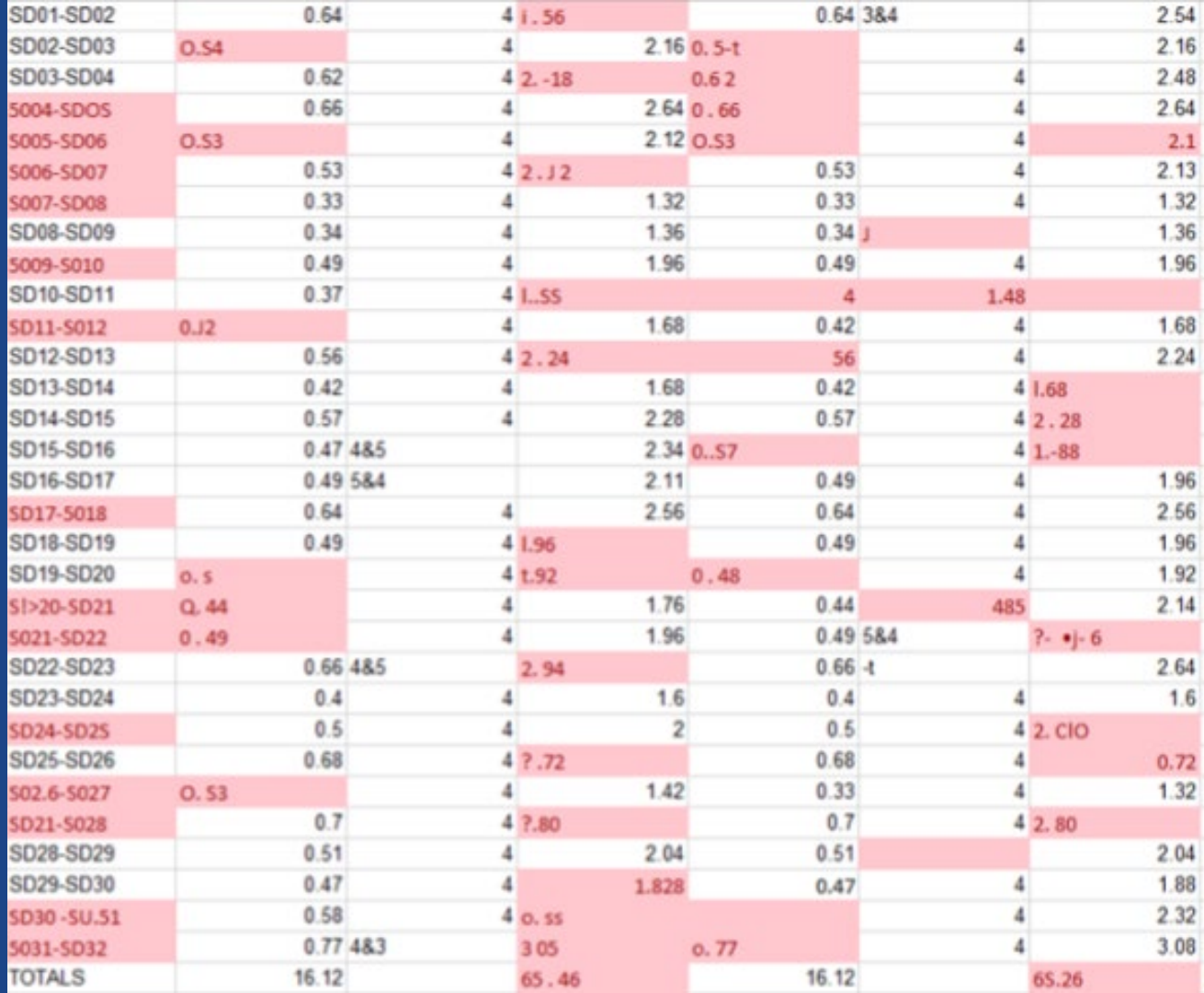
TAPIE 2 LAtE MIL£ TABLE
SAN DIEGO FREEWAY

Station	X) Rnm. Jiffl Distance Between Stations	Number of Lanes	Lane Miles	set.rnmh'll Distance Between Stations	Number of Lanes	Lane tiles
SD01-SD02	0.64	4	i.56	0.64	3&4	2.54
SD02-SD03	O.S4	4	2.16	0.5-t	4	2.16
SD03-SD04	0.62	4	2.-18	0.62	4	2.48
5004-SDOS	0.66	4	2.64	0.66	4	2.64
S005-SD06	O.S3	4	2.12	O.S3	4	2.1
S006-SD07	0.53	4	2.J2	0.53	4	2.13
S007-SD08	0.33	4	1.32	0.33	4	1.32
SD08-SD09	0.34	4	1.36	0.34	J	1.36
5009-S010	0.49	4	1.96	0.49	4	1.96
SD10-SD11	0.37	4	1.$$	0.:n	4	1.48
SD11-S012	0.J2	4	1.68	0.42	4	1.68
SD12-SD13	0.56	4	2.24	0.56	4	2.24
SD13-SD14	0.42	4	1.68	0.42	4	1.68
SD14-SD15	0.57	4	2.28	0.57	4	2.28
SD15-SD16	0.47	4&5	2.34	0..$7	4	1.-88
SD16-SD17	0.49	5&4	2.11	0.49	4	1.96
SD17-5018	0.64	4	2.56	0.64	4	2.56
SD18-SD19	0.49	4	1.96	0.49	4	1.96
SD19-SD20	O. S	4	t.92	0.48	4	1.92
S!>20-SD21	Q.44	4	1.76	0.44	485	2.14
S021-SD22	0.49	4	1.96	0.49	5&4	?.-6
SD22-SD23	0.66	4&5	2.94	0.66	-t	2.64
SD23-SD24	0.40	4	1.60	0.40	4	1.60
SD24-SD2S	0.50	4	2.00	0.50	4	2.CIO
SD25-SD26	0.68	4	?.72	0.68	4	.72
S02.6-S027	O.S3	4	1.42	0.33	4	1.32
SD21-S028	0.70	4	?.80	0.70	4	2.80
SD28-SD29	0.51	4	2.04	0.51	4	2.04

NTL's Current Capability for Data Rescue

Using Adobe Acrobat DC Pro on scanned PDFs

- Enhance scans
- Recognize text (OCR)
- Highlight text
- Export selection as .docx
- Format text in Word
  - "Normal" text formatting
  - Remove artifacts
  - Standardize font & text size
  - Remove spacing errors and OCR noise
- Paste .docx content into Excel



SD01-SD02		0.64		4	i . 56		0.64	3&4			2.54
SD02-SD03	O.S4			4		2.16	0. 5-t		4		2.16
SD03-SD04		0.62		4	2. -18		0.6 2		4		2.48
5004-SDOS		0.66		4		2.64	0 . 66		4		2.64
S005-SD06	O.S3			4		2.12	O.S3		4		2.1
S006-SD07		0.53		4	2 . J 2			0.53	4		2.13
S007-SD08		0.33		4		1.32		0.33	4		1.32
SD08-SD09		0.34		4		1.36		0.34	J		1.36
5009-S010		0.49		4		1.96		0.49	4		1.96
SD10-SD11		0.37		4	I..SS			4		1.48	
SD11-S012	0.J2			4		1.68		0.42	4		1.68
SD12-SD13		0.56		4	2 . 24			56	4		2.24
SD13-SD14		0.42		4		1.68		0.42	4	I.68	
SD14-SD15		0.57		4		2.28		0.57	4	2 . 28	
SD15-SD16		0.47	4&5			2.34	0..S7		4	1.-88	
SD16-SD17		0.49	5&4			2.11		0.49	4		1.96
SD17-5018		0.64		4		2.56		0.64	4		2.56
SD18-SD19		0.49		4	I.96			0.49	4		1.96
SD19-SD20	o. s			4	t.92		0 . 48		4		1.92
S!>20-SD21	Q. 44			4		1.76		0.44		485	2.14
S021-SD22	0 . 49			4		1.96		0.49	5&4		?- •j- 6
SD22-SD23		0.66	4&5		2. 94			0.66	-t		2.64
SD23-SD24		0.4		4		1.6		0.4	4		1.6
SD24-SD25		0.5		4		2		0.5	4	2. CIO	
SD25-SD26		0.68		4	?.72			0.68	4		0.72
S02.6-S027	O. S3			4		1.42		0.33	4		1.32
SD21-S028		0.7		4	?.80			0.7	4	2. 80	
SD28-SD29		0.51		4		2.04		0.51			2.04
SD29-SD30		0.47		4		1.828		0.47	4		1.88
SD30 -SU.51		0.58		4	o. ss				4		2.32
5031-SD32		0.77	4&3		3 05		o. 77		4		3.08
TOTALS		16.12			65 . 46			16.12			6S.26

Other Options for Legacy Data Rescue

Commercial solutions in the form of paid services or dedicated OCR software such as ABBYY or OmniPage

DIY/open-source solutions including Python scripts for “fixing” OCR’d text

“Scholarly” solutions focus on document analysis and table detection using OCR/analytics, deep learning AI, and/or crowdsourcing

Solutions from other Gov. Agencies

NSF awards grants for legacy data rescue projects

USGS uses OmniPage to OCR their documents, has dabbled in crowdsourcing

NLE contracts w/Internet Archive for digitization

NLM’s PMC converts documents to HTML

NIST contracts with PMC to convert documents to HTML

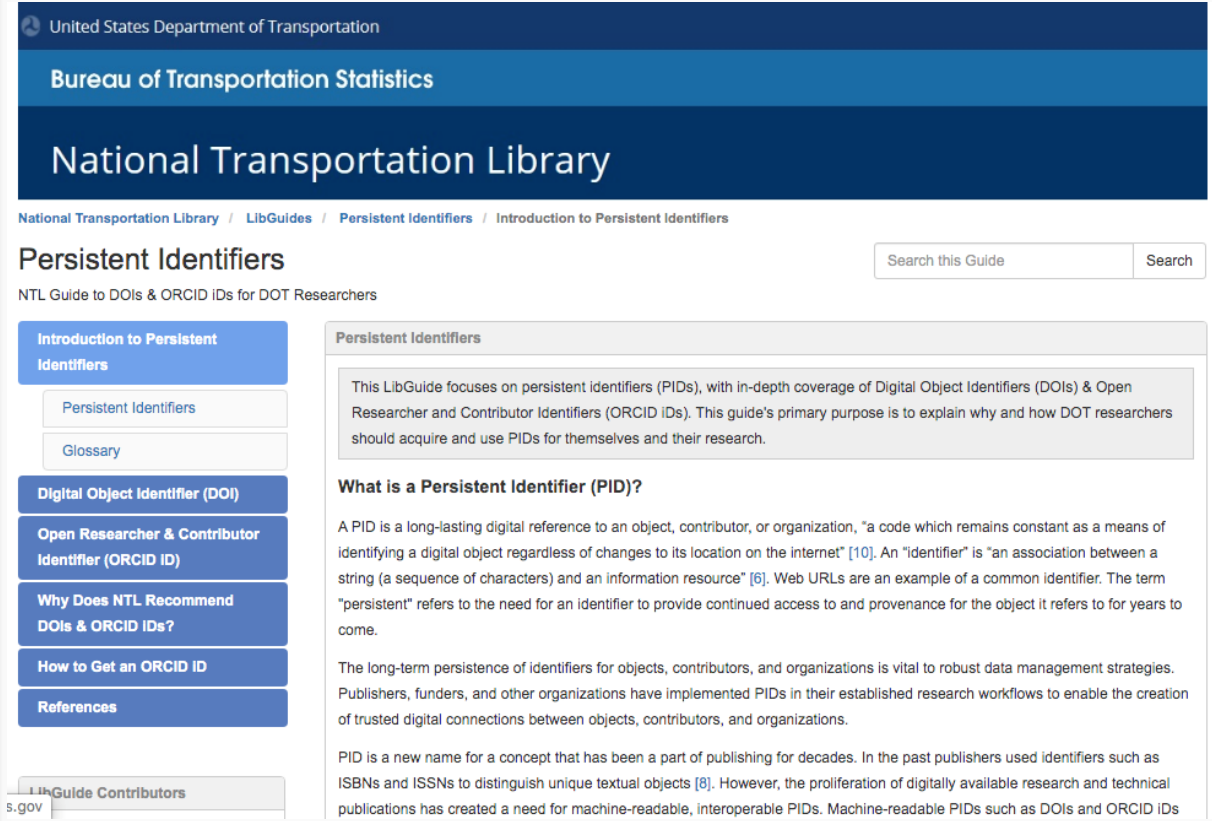
LC and Smithsonian offer crowdsourcing platforms

Persistent Identifier LibGuide

Created a resource for use in socializing DOIs & ORCiDs with researchers across USDOT

DOIs & ORCiD IDs are required by the USDOT 2015 Public Access Plan but remain unknown to and underused by USDOT researchers

transportation.libguides.com/persistent_identifiers



Thank you!

Working at NTL for the summer was a great experience.

- Learned from the NTL staff and met other information professionals in DC
- Learned to use LibGuides and developed professional research skills
- Chose related courses this year including GIS, information architecture, metadata, web design
- Inspired to pursue an additional government information center experience—spring practicum at the Dept. of Energy's Office of Scientific and Technical Information (OSTI) here in Oak Ridge, Tennessee

Feel free to contact me with any questions or future opportunities!

lisa.curtin2@gmail.com
lcurtin1@vols.utk.edu

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

Questions?

Ed Strocko (Ed.Strocko@dot.gov)
Mary Moulton (Mary.Moulton@dot.gov)
Jesse Long (Jesse.Long.ctr@dot.gov)
Leighton Christiansen (Leighton.Christiansen@dot.gov)
Xin Wang (Xin.Wang@dot.gov)
Vinod Koduri (Vinod.Koduri.ctr@dot.gov)

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

Come see us at TRB 2020!

- Exhibit Hall Booth 835
- Section 508 Compliance and Document Accessibility: How to Make Your Reports Accessible for Everybody
- Research Data Management for State DOTs
- National Transportation Data Preservation Network (NTDPN) Workshop 2
- Moving Forward and Advancing Society with Transportation Libraries, Information, and Data Poster Session 1486

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