Slide Title: U.S. Open Science Policy Perspectives & Transportation

Speaker text:
Thank you to Caroline for the nice introduction, and to the workshop organizers for inviting me to speak to you today. It is an honor and pleasure to be part of this great panel.

I will give a brief overview of Open Science policies and practices within the United States government. While most of my examples will come from the Department of Transportation, I will include some discussion of government-wide activities and resources, including data.gov.

[Next slide]

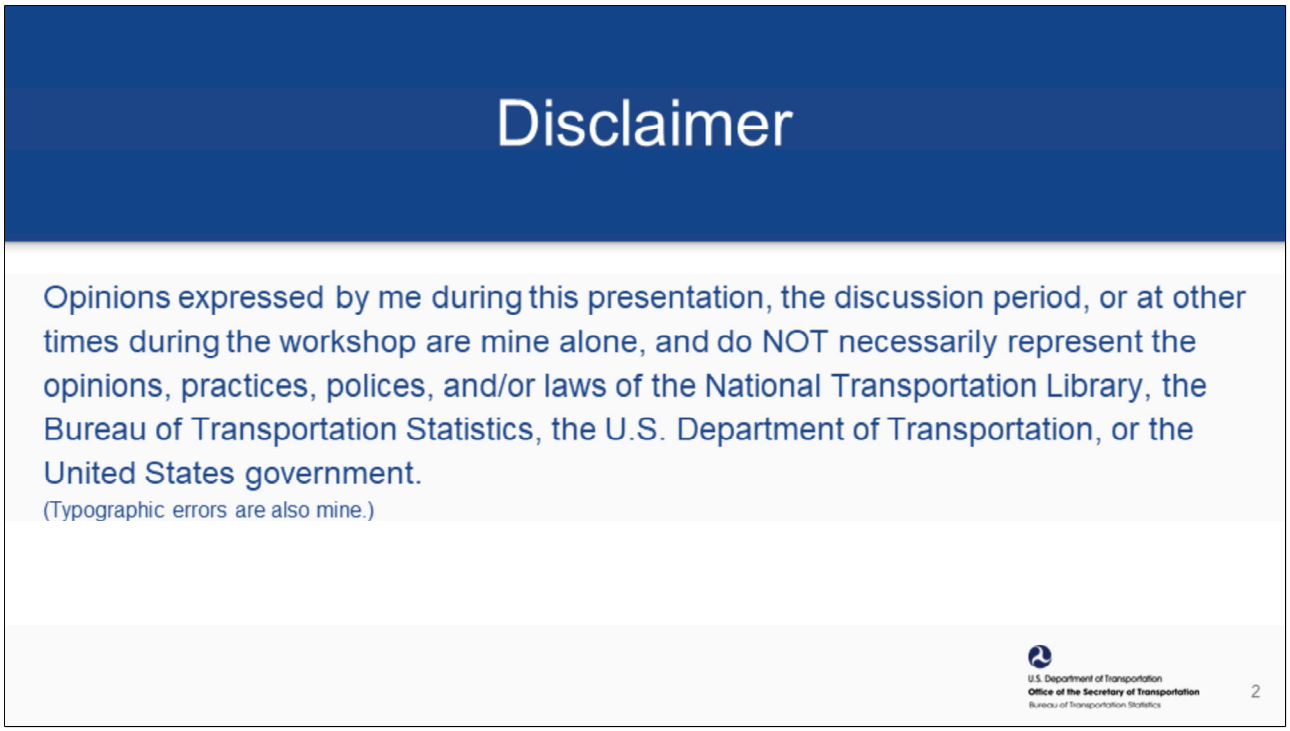
[Slide text not presented orally:
Leighton Christiansen https://orcid.org/0000-0002-0543-4268
Data Curator, National Transportation Library,
Bureau of Transportation Statistics, U.S. Department of Transportation
leighton.christiansen@dot.gov

Presented to: Transportation Research Board Annual Meeting 2021
1467 – Open Science in Transportation: Challenges and Opportunities in a COVID-19

Era
2021-01-21]

The digital object identifier for this presentation is: https://doi.org/10.21949/1520725

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



Slide Title: Disclaimer

Speaker text:
Before I go farther: Opinions and typos are my own.

[Next slide]

[Slide text not presented orally:
Opinions expressed by me during this presentation, the discussion period, or at other times during the workshop are mine alone, and do NOT necessarily represent the opinions, practices, polices, and/or laws of the National Transportation Library, the Bureau of Transportation Statistics, the U.S. Department of Transportation, or the United States government.
(Typographic errors are also mine.)]

Slide Title: Contents

Speaker Notes:
Here are the broad themes I will touch on.

Let's get right to it.

[Next slide]

[Slide text not presented orally:
Here are the broad themes I will touch on.

- Sharing U.S. Research before "Open Science"
- Opening U.S. Government-Funded Science
  - Policies 2005 to 2020
  - Practices
  - Technology:
    - Data.gov
    - U.S. DOT Systems

    - Resources
- Challenges
- Conclusions
- Supplemental Slides
- Links to resources]

Slide Title: Sharing U.S. Research before "Open Science"

Speaker notes:
The U.S. Government has been sharing publications and research outputs with the public, in an organized way, since 1861. Two bodies responsible for sharing are the U.S. Government Publishing Office (GPO) and the National Technical Information Service (NTIS).

I want to point out that the 1950 law that established NTIS specifically calls for the sharing of scientific, technical, and engineering information with the public.

[Next Slide]

[Text of slide not presented orally:
U.S. Government Publishing Office (GPO)
https://www.gpo.gov/

- Opens March 4, 1961 as Government Printing Office
- Printing and binding for the Senate and House of Representatives, the Executive Branch, and the federal Judiciary.
- Embraces digital future, and rebranded Government Publishing Office in 2014
- GPO Style Manual: https://www.govinfo.gov/content/pkg/GPO-STYLEMANUAL-2016/pdf/GPO-STYLEMANUAL-2016.pdf

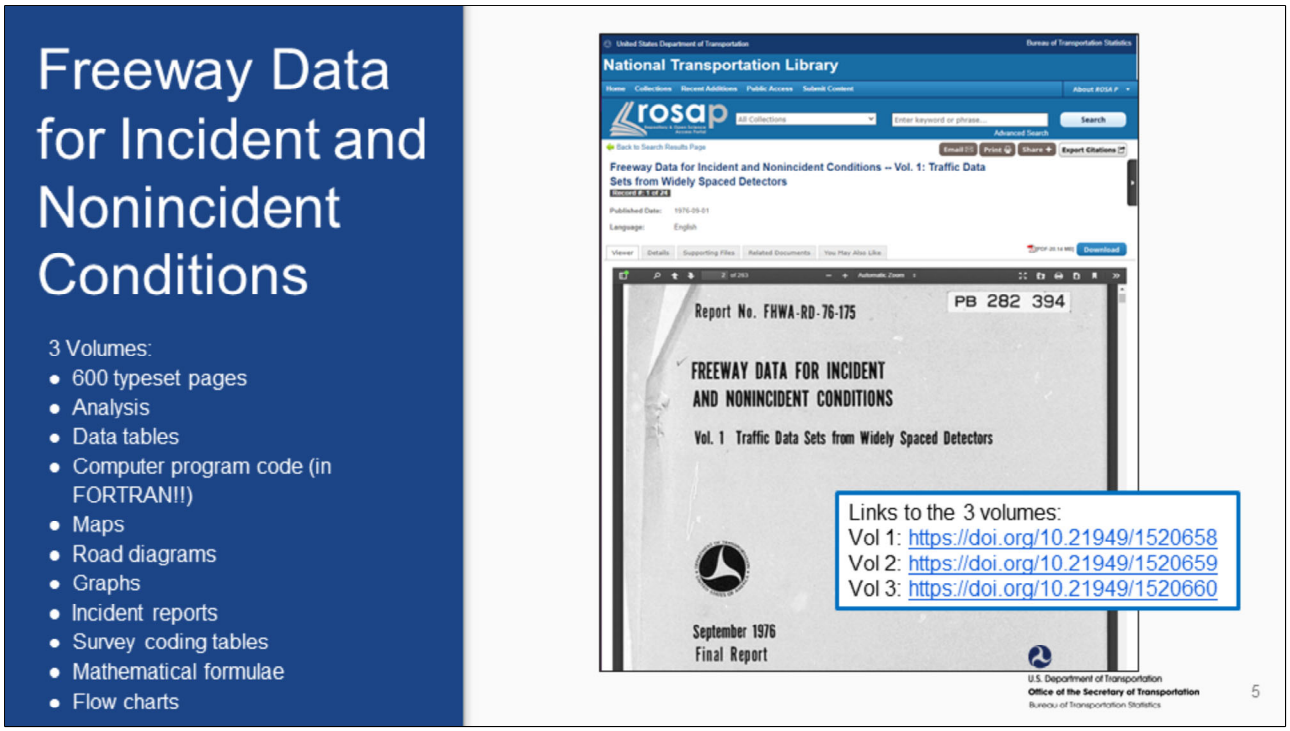
National Technical Information Service (NTIS)
https://www.ntis.gov/

- Established by law on September 9, 1950, as “Publication Board”
- Clearinghouse for the collection and dissemination of scientific, technical, and engineering information (STEI)
- Federal agencies are required to send a copy of their STEI products to NTIS
- NTIS catalogs, organizes, preserves and disseminates to public online through National Technical Reports Library (NTRL) https://ntrl.ntis.gov/NTRL/]

[Expanded background text:
The U.S. Government has been sharing publications and research outputs with the public, in an organized way, since 1861. Two of the many bodies responsible for sharing information and research are the U.S. Government Publishing Office (GPO) and the National Technical Information Service (NTIS).

I won’t spend much time on these, except to point out that the 1950 law that established NTIS specifically calls for the sharing of scientific, technical, and engineering information (STEI) with the public.

NTIS holds millions of records and also distributes scanned copies of many of its holdings. These can be accessed through a number of NTIS interfaces, including the National Technical Reports Library (NTRL) https://ntrl.ntis.gov/NTRL/]



Slide Title: Freeway Data for Incident and Nonincident Conditions

Speaker notes: One report I accessed from NTIS is the 3-volume “Freeway Data for Incident and Nonincident Conditions”. This 1976 research project was funded by DOT’s Federal Highway Administration, and carried out by the California state DOT.

Why talk about such an old report in this space? Because these three volumes have every type of research output we find challenging in relation to Open Science. These outputs include:

- Data tables
- Computer program code (FORTRAN!!)
- Maps
- Road diagrams

And while it is nice to be able to brag a little that U.S. DOT has been sharing reports, data, and computer code for 40 or 50 years, we need to temper that pride, because much of that content is “Imprisoned in PDF.”

[Next slide]

[Text of slide not presented orally:
Full list of research file types in teh 3 reports includes:

- 600 typeset pages with
- Analysis
- Data tables
- Computer program code (FORTRAN!!)
- Maps
- Road diagrams
- Graphs
- Incident reports
- Survey Coding tables
- Mathematical Formulae
- Flow charts

Links to the 3 volumes:
Vol 1: https://doi.org/10.21949/1520658
Vol 2: https://doi.org/10.21949/1520659
Vol 3: https://doi.org/10.21949/1520660]

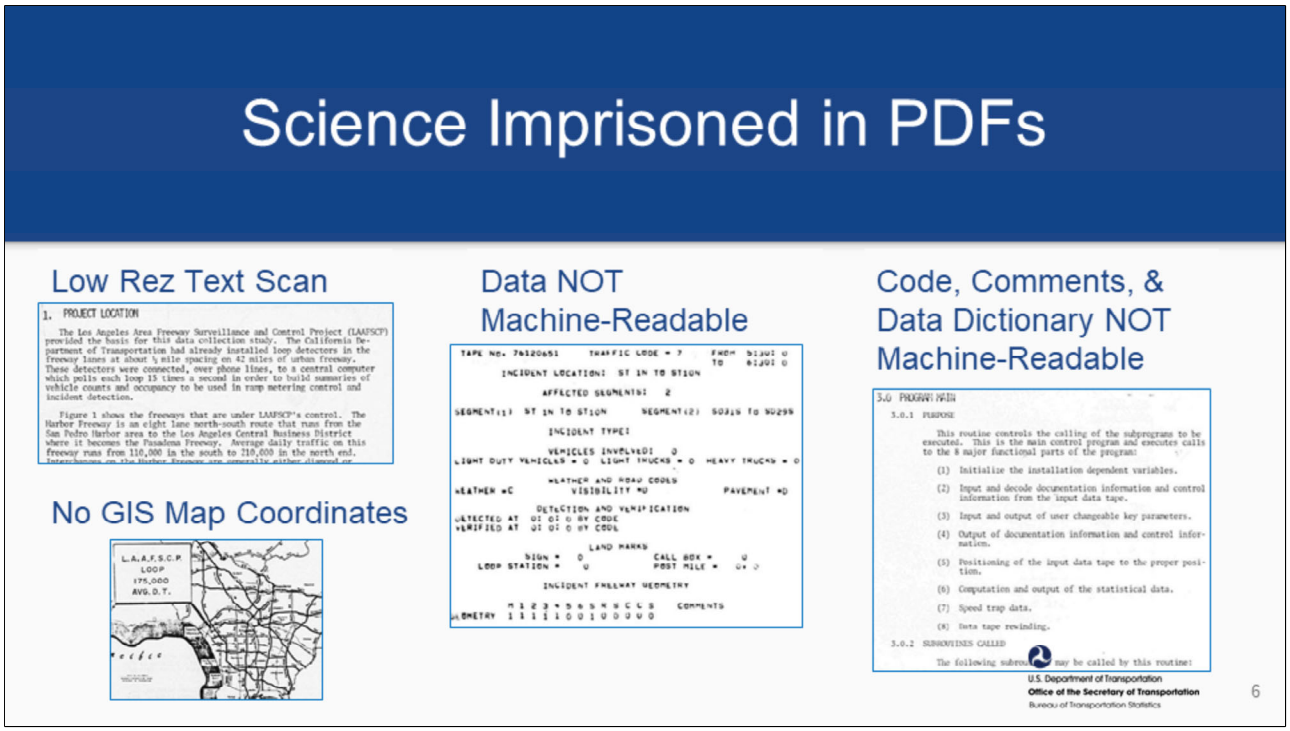
[Expanded background text:
Among the scientific and technical reports you might access from NTIS is the 3-volume report entitled: “Freeway Data for Incident and Nonincident Conditions” from 1976 and 1977. This research project was funded by the Federal Highway Administration of U.S. DOT, and carried out by the California state Department of Transportation.
In fact, when the National Transportation Library received a request for this publication, we had to contact NTIS to get a scanned copy so that we could add it to our Repository & Open Science Access Portal (ROSA P) (pictured here).

Why talk about such an old report in this space? Because these three volumes have every type of research output we are talking about, or finding challenging, in relation to Open Science. These outputs include:

- 600 typeset pages with
- Analysis
- Data tables
- Computer program code (FORTRAN!!)
- Maps
- Road diagrams
- Graphs
- Incident reports
- Survey Coding tables

- Mathematical Formulae
- Flow charts

And while it is nice to be able to brag a little that U.S. DOT has been sharing scientific reports, data, and computer code for 40 or 50 years, as you might imagine, we need to temper that pride, because much of that content is “Imprisoned in PDF.”]



Slide Title: Science Imprisoned in PDF

Speaker notes: If you cannot see much detail on this slide, that is expected. While the scans of the “Freeway Data’ reports exists, they are rather low resolution.

I won’t spend much time here talking about the challenges of opening legacy science, other than to say that the problems of 45 years ago are still with us: How do we make our science discoverable and useful to humans, while also machine-readable and useful to computers?

[Next slide]

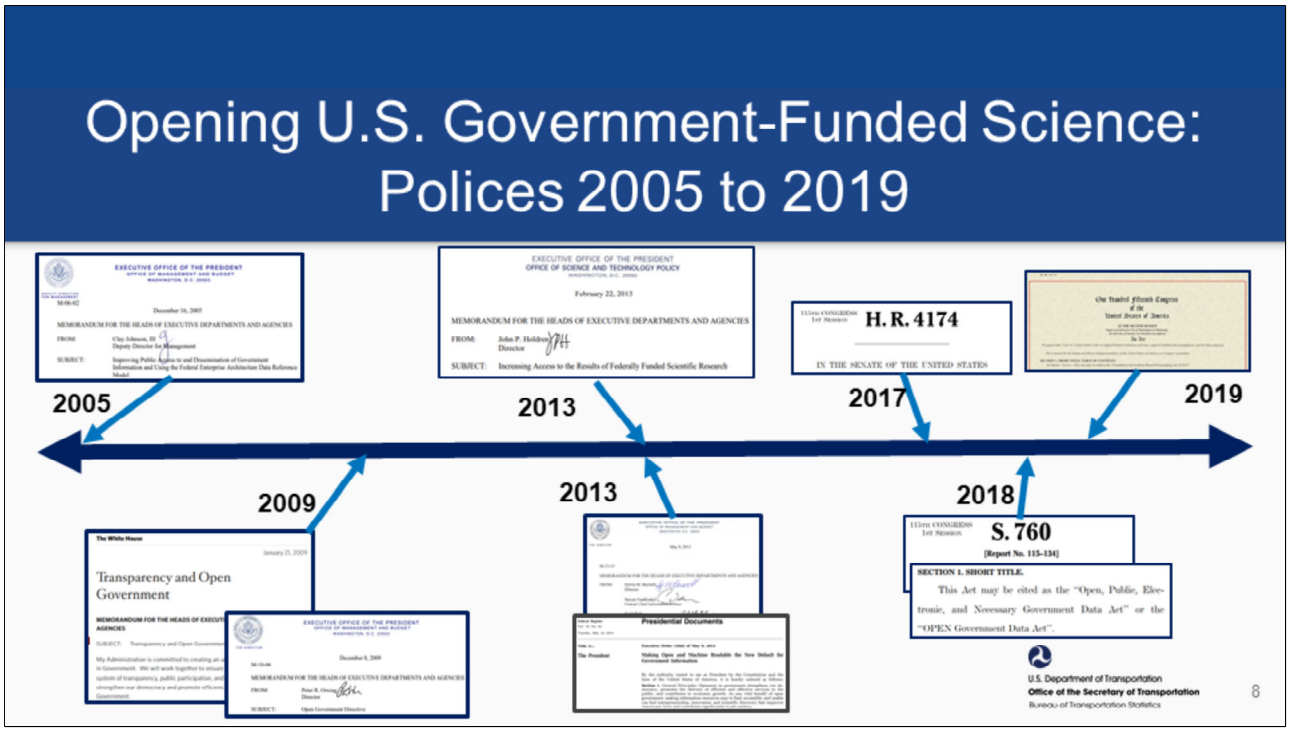
Text of slide not presented orally:

- Low Rez Text Scan\
- No GIS Map Coordinates
- Data NOT Machine-Readable
- Code, Comments, & Data Dictionary NOT Machine-Readable

Slide Title: Open Science: U.S. Federal Policy & Implementation

Speaker notes: Now we can turn our attention to the recent past, the present, and the hopeful future by taking a quick look as U.S. Federal Open Science policies and implementation efforts. This will include overviews of Policies, Practices, Technology, and Resources.

[Next slide]



Slide Title: Opening U.S. Government-Funded Science: Policies 2005 to 2019

Speaker notes: Over the past 15 years we have seen many executive orders, policies, and laws, that seek to make our government more transparent. These included directives for increased public access to federally funded publications and datasets. It is hoped that opening research outputs to broader public use will have social, economic, and research benefits, especially as data is re-used in novel ways.

A small sample of these include:

- 2009 “Transparency and Open Government”
- 2013 “Increasing Access to the Results of Federally Funded Scientific Research,”

Let us focus for a second on the most recent , the Foundations for Evidence-Based Policymaking Act of 2018. Title II is the Open, Public, Electronic, and Necessary (OPEN) Government Data Act, which requires that non-restricted U.S. government data be available in machine-readable formats.
But new policies are not the only reason for the U.S. shift towards Open Science. Federal policy had to change because scientific research practice is changing in the digital era.

[Next slide]

[Text of slide not presented orally:
List of policies and memos picture on the screen:

- Office of Management and Budget (OMB) Memo M-06-02, "Improving Public Access to and Dissemination of Government Information and Using the Federal Enterprise Architecture Data Reference Model," December 16, 2005, https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2006/m06-02.pdf
- White House "Transparency and Open Government" memo, January 21, 2009, https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government
- Office OMB memo M-10-06, "Open Government Directive", December 8, 2009, https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2010/m10-06.pdf
- White House Office of Science and Technology Policy (OSTP) memo "Increasing Access to the Results of Federally Funded Scientific Research," also knows as "The Holdren Memo," February 22, 2013, https://web.archive.org/web/20130308142014/https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- OMB memo M-13-13, "Open Data Policy – Managing Information as an Asset," May 9, 2013, https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2013/m-13-13.pdf
- White House Executive Order 13642 "Making Open and Machine Readable the New Default for Government Information," May 9, 2013, https://www.govinfo.gov/content/pkg/FR-2013-05-14/pdf/2013-11533.pdf
- House of Representatives bill H.R. 4174, "Foundations for Evidence-Based Policymaking Act of 2018," introduced October 31, 2017, https://www.congress.gov/bill/115th-congress/house-bill/4174
- U.S. Senate bill "S.760 - Open, Public, Electronic, and Necessary Government Data Act," introduced March 29, 2017, https://www.congress.gov/bill/115th-congress/senate-bill/760/text , and,
- Public Law No. 115-435 "Foundations for Evidence-Based Policymaking Act of 2018," signed into law on January 14, 2019, https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf

[Expanded background text:

- Over at least the past 15 years we have seen a number of White House executive orders, policies, and laws, that seek to make the operations of the government more transparent. This included executive orders calling for increased public access to federally funded publications, research, and data, so that citizens have as much access as possible to the products they fund through taxes. It is hoped that opening research outputs to broader public use may also have social, economic, and research benefits,

especially as data is re-used in novel ways, perhaps not considered by the original researcher.

- A non-exhaustive list of these order, memos, and laws include: (see list above)

Let us focus for a second on Public Law 115-435, the **Foundations for Evidence-Based Policymaking Act of 2018 (HR 4174).** Title II of the Act, includes the Open, Public, Electronic, and Necessary (OPEN) Government Data Act. The OPEN Government Data Act requires that non-restricted U.S. government data be available in machine-readable formats, and that each federal agency have a Chief Data Officer. This act is consistent with the spirit of the memos and policies that preceded it, including the U.S. DOT Public Access Plan.]

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

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

The authors note that science will now be conducted in a “fully digital world” and that data is “an endless fuel for creativity.”
The authors list seven guiding principles for the new research reality. Among these are:

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

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

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

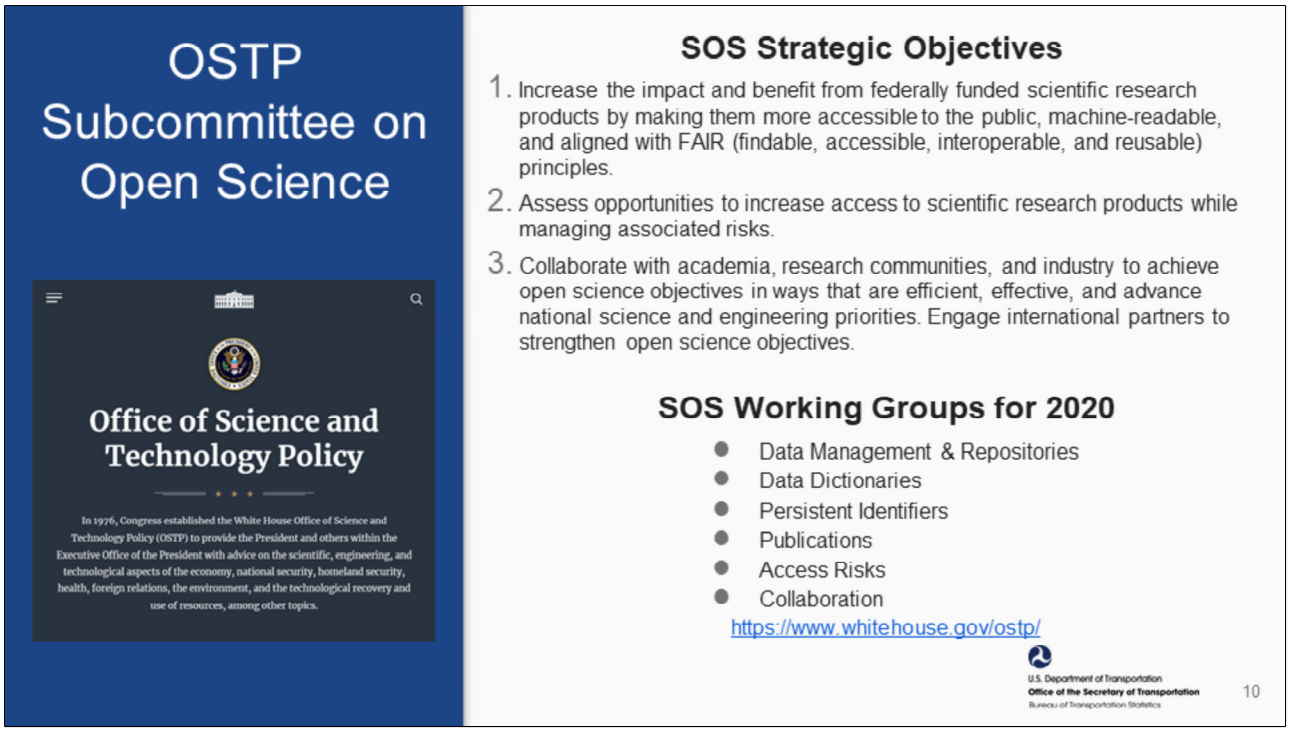
[Next slide]

[Text of slide not presented orally:

Full list of Guiding Principles:

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

URL to “Harnessing the Power of Digital Data” report:
https://www.nitrd.gov/Publications/PublicationDetail.aspx?pubid=25]



Slide Title: OSTP Subcommittee on Open Science Workplan 2020

Speaker notes:

The Subcommittee on Open Science (SOS) is working on federal implementation.

The Strategic Objectives boil down to:

- Increase public access,
- Data is Machine-readable,
- Align with FAIR principles,
- Manage risks,
- And Collaborate with stakeholders.

In SOS we come create tools to aid open science. One is a list of basic data management plan sections, so that researchers funded by more than one agency will satisfy all with a single DMP.

DOT representatives serve in SOS, and DOT practice and policy are informed by this work.

For example:

[Next slide]

[Text of slide not presented orally:
The 2020 SOS Strategic Objectives were:

- Increase the impact and benefit from federally funded scientific research products by making them more accessible to the public, machine-readable, and aligned with FAIR (findable, accessible, interoperable, and reusable) principles.
- Assess opportunities to increase access to scientific research products while managing associated risks.
- Collaborate with academia, research communities, and industry to achieve open science objectives in ways that are efficient, effective, and advance national science and engineering priorities. Engage international partners to strengthen open science objectives.

To achieve these objectives, the SOS is divided in to 6 Working Groups. For 2020 these were:

- Data Management & Repositories
- Data Dictionaries
- Persistent Identifiers
- Publications
- Access Risks
- Collaboration

The SOS is made of many folks who either had a handing in writing their agency public access plans, or are now responsible for implementing those plans. We often refer to Public Access and Open Science in our discussions. We try to come up with tools and recommendations that make it easier to implement public access and open science.
One recent outputs have included

- A publication describing the desirable characteristics of repositories that may be used to preserve and share government data;
- A harmonized list of basic data management plan sections, so that researchers funded by more than one agency will satisfy all with a single DMP; and,

- A survey of agency Persistent Identifier usage and future implementation plans.]

[Expanded background text:
One federal-wide group working on open science implementation is the White House Office of Science and Technology Policy (OSTP) Subcommittee on Open Science (SOS), which has been in operation for a few years now.

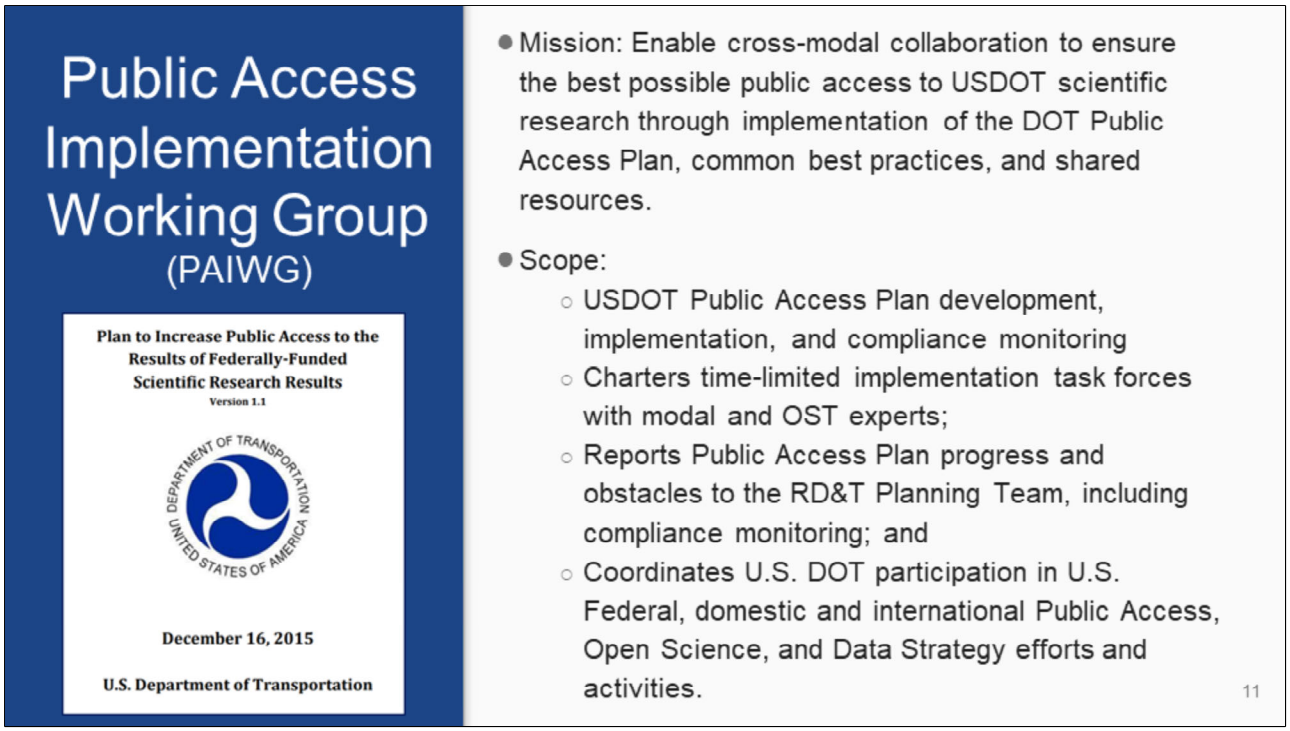
The 2020 SOS Strategic Objectives were:
- Increase the impact and benefit from federally funded scientific research products by making them more accessible to the public, machine-readable, and aligned with FAIR (findable, accessible, interoperable, and reusable) principles.
- Assess opportunities to increase access to scientific research products while managing associated risks.
- Collaborate with academia, research communities, and industry to achieve open science objectives in ways that are efficient, effective, and advance national science and engineering priorities. Engage international partners to strengthen open science objectives.

To achieve these objectives, the SOS Working Groups for 2020 were:
- Data Management & Repositories
- Data Dictionary
- Persistent Identifiers
- Publications
- Access Risks
- Collaboration

The SOS is made of many folks who either had a handing in writing their agency public access plans, or are now responsible for implementing those plans. We often refer to Public Access in our discussions. We try to come up with tools and recommendations that make it easier to implement public access.
Some recent outputs have included
- A publication describing the desirable characteristics of repositories that may be used to preserve and share government data;
- A harmonized list of basic data management plan sections, so that researchers funded by more than one agency will satisfy all with a single DMP; and,
- A survey of agency Persistent Identifier usage and future

implementation plans.

DOT representatives serve on most of these working groups. DOT actions and policy reviews are informed by this effort to standardize federally-funded research.]



Slide Title: Public Access Implementation Working Group (PAIWG)

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

Let us leave the policy and practice realm, and take a look at technologies.

[Next slide]

[Slide text not presented orally:

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

- Scope The Public Access Implementation Working Group (PAIWG):
    - Owns USDOT Public Access Plan development, implementation, and

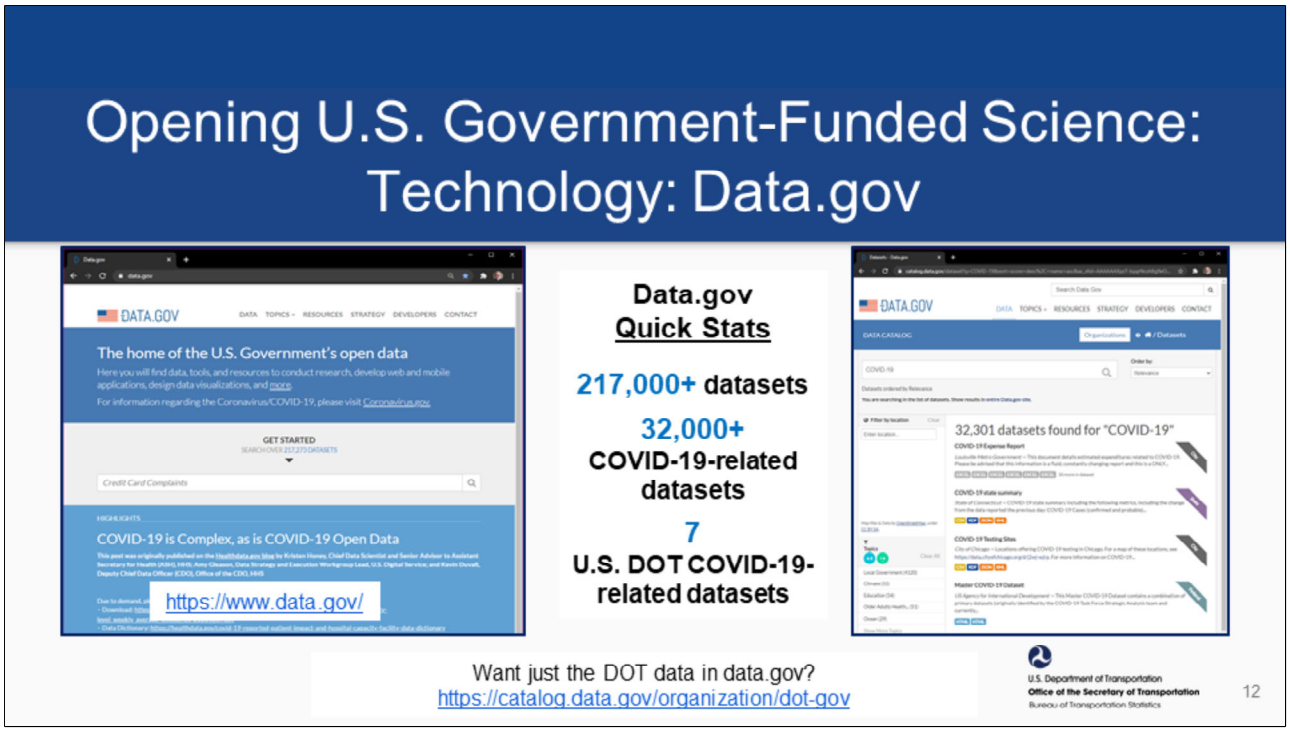
compliance monitoring across all categories of public access outputs, including consistent-facing communications and inputs to implementation support resources;
    - Charters time-limited implementation task forces with modal and OST experts;
    - Reports Public Access Plan progress and obstacles to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

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

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

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

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



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

Speakers notes;

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

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

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

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

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

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

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

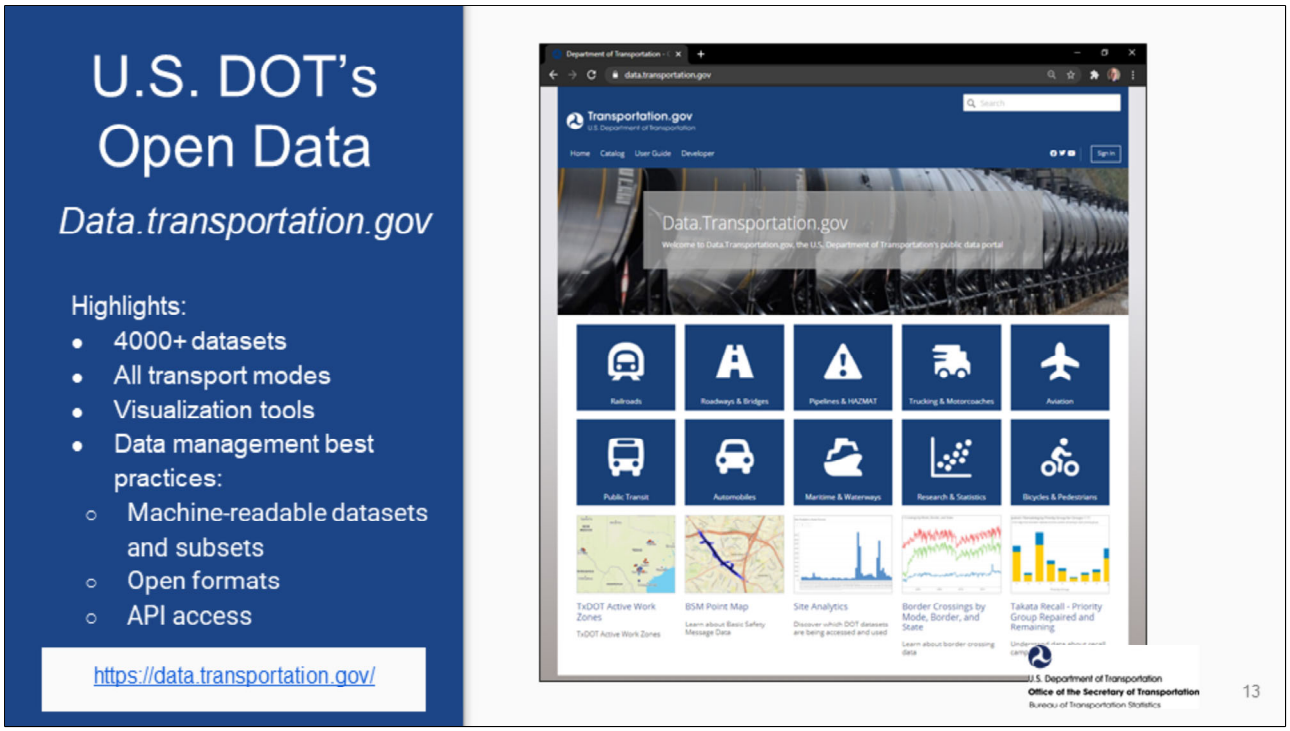
Speaking of DOT data, let’s now take a quick look at DOT’s data inventory

[Next slide]

[Expanded background text:

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

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



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

Speaker notes:
Data.gov harvests metadata from DOT's DATA.TRANSPORTATION.gov, a data catalog, warehouse, and visualization suite.
Most of the 4000-plus datasets are available for public download.

Some of the highlights include:
- All transportation modes are represented
- Data visualization tools are built in

We should now visit DOT's open scientific research report portal.
[Next slide]

[Slide text presented orally:
Be sure to visit https://data.transportation.gov

List of data.transportation.gov highlights
Highlights:
- 4000+ datasets
- All transport modes

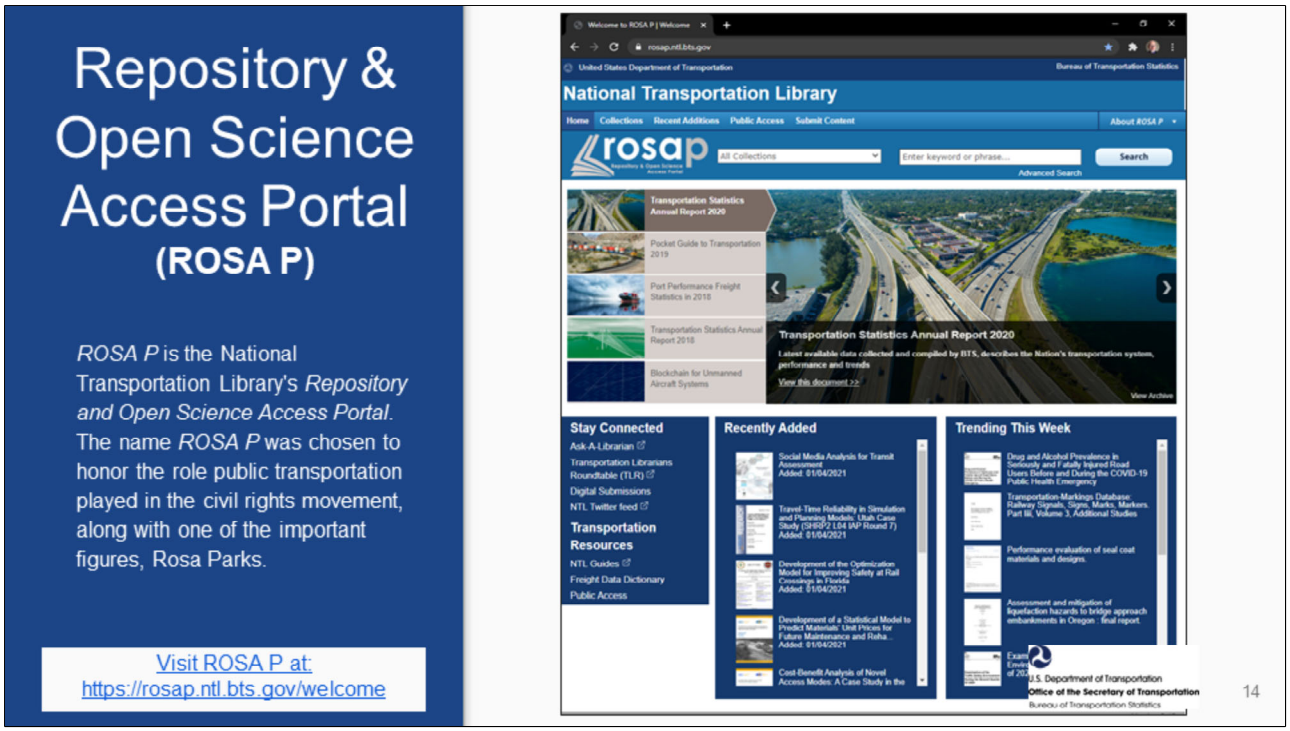
- Visualization tools
- Data management best practices:
    - Machine-readable datasets and subsets
    - Open formats
    - API access]

[Expanded background text:
The U.S. DOT data catalog from which data.gov pulls is called DATA.TRANSPORTATION.gov, a data catalog, data warehouse, and data visualization suite. U.S. DOT launched its first data inventory in September 2010. Data.transportation.gov now utilizes the data visualization platform Socrata. As of January 2021, Data.transportation.gov contains records of more than 4000 datasets. Most of these are available to the public for download.

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

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

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



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

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

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

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

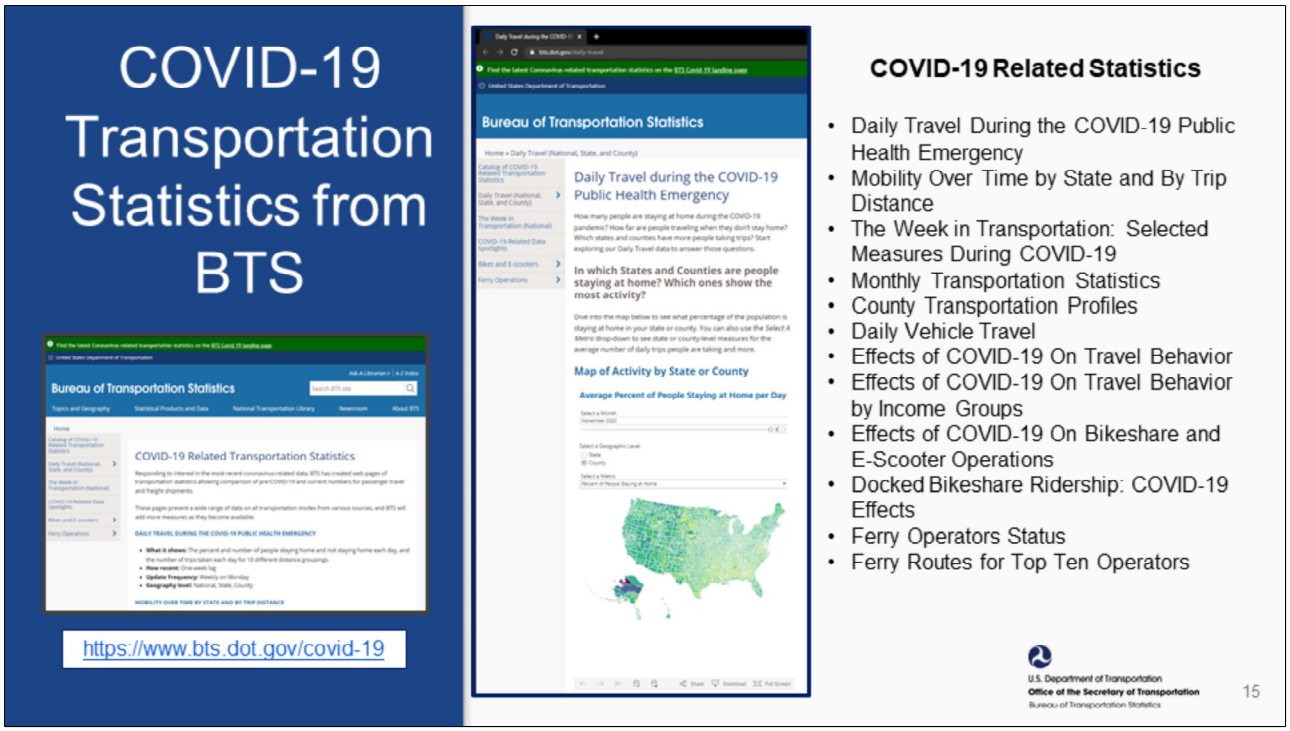
[Next slide]

[Slide text not presented orally:
ROSA P is the National Transportation Library's Repository and Open Science Access Portal. The name ROSA P was chosen to honor the role public transportation played in the civil rights movement, along with one of the important figures, Rosa Parks.
Visit ROSA P at: https://rosap.ntl.bts.gov/welcome]

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

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

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



Slide Title: COVID-19 Transportation Statistics from BTS

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

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

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

[Next slide]

[Slide text not presented orally:
COVID-19 Related Statistics:

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

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

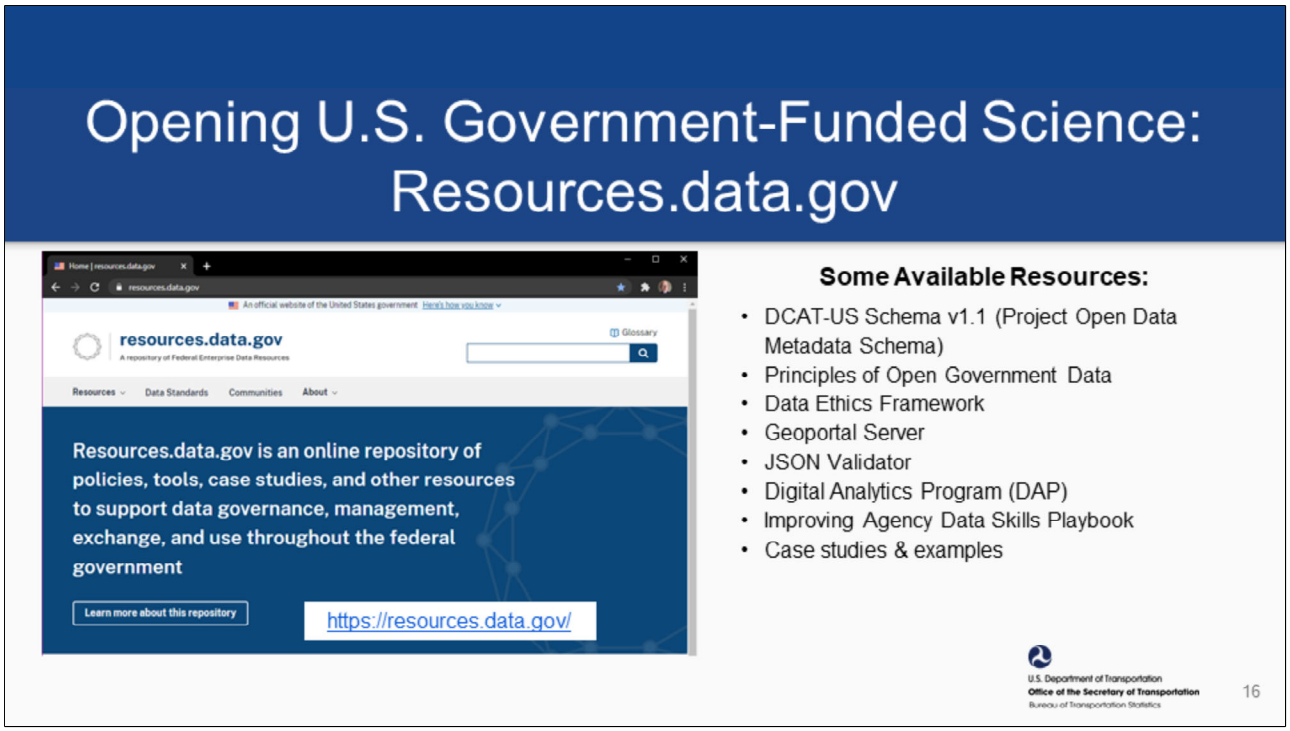
https://www.bts.dot.gov/covid-19]

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

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

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

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



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

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

Some of the available resources include:
- The DCAT-US Schema v1.1
- The Data Ethics Framework
- And Case studies & examples

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

[Next slide]

[Slide text not presented orally:
Some of the available resources include:
- The DCAT-US Schema v1.1 (Project Open Data Metadata Schema)
- The Principles of Open Government Data
- The Data Ethics Framework
- A Geoportal Server
- A JSON Validator

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

https://resources.data.gov/]

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

Speaker notes:
Most of the presentation paints a rosy picture of U.S. Government progress towards Open Science.

But we should take a moment to outline the challenges we face. On the slide, I have sketched some of the challenges among the four broad themes I discussed, and within the context of COVID-19.

But for sake of time, I will discuss only a handful.
Under Policy Challenges we should note that socialization and implementation of new open science policies can be uneven within departments and across government, as we have seen as DOT, and we are working hard to bring each research office into compliance.

Under Practice Challenges I want to acknowledge that Culture change is hard. New practices challenge existing workflows, and integrating Open Science practices can seem like a lot of work for little pay-off.

Under Resource Challenges I want to remind folks that open science and the data skills needed for open science is new work. However, there can be bureaucratic resistance to creating the new positions needed to support open science.

So what have these challenges meant during COVID-19?

- There are great examples of U.S. agencies moving quickly to open access to COVID-19 research. For instance, in March 2020, the National Library of Medicine (NLM) worked with publishers and scholarly societies to open up access to coronavirus research through PubMed Central.
- On the other hand, COVID-19 has also showed many departments our shortcomings in relationship to our vision of opening U.S. science. I think we will need to take a hard look at the COVID-19 experience, and shift resources so that we can prepare for the next time. And there will be a next time. We live on an active planet, in a dynamic universe, building fragile infrastructures.
    - I hope that one way we use all of the COVID-19 data we collect , in every discipline, will be to make the changes needed to avoid or mitigate the levels of suffering we have seen this time.

On that note, let me wrap up.

[Next slide]

[Slide text not presented orally:
Full list of challenges:

Policy Challenges::

- Policy writing can take time and is not always responsive to events. There are exceptions of course.
- Policy directions can change with leadership changes. This can especially impact policies that are aimed into the future, or take time to reach fruition.
- The return on investment for Open Science policies can be hard to measure. And we work in a world where ROI is a key metric for future funding.
- The socialization and implementation of new open science policies can be uneven within departments and across government. The implementation of the DOT public access has been uneven, and we are working hard to bring each research office into compliance.

Practice Challenges

- Culture change is hard. New practices challenge existing workflows, and integrating Open Science practices can seem like a lot of work for little pay-off.
- Researcher resistance to openness is real. For some researchers and some disciplines it is hard to embrace openness, and its potential benefits. Some folks tend to focus on the potential misfortunes.
- Retraining and reskilling of existing employees may be needed, but can be difficult to achieve. Resistance can come from both leadership and employees.

Technology Challenges

- Purpose built existing infrastructures may not adapt well to Open Science needs.
- System integrations can be complex., especially when combining systems built in-house with third-party systems.

Resource Challenges

- For many U.S. research offices, funding has been flat for years. Which of course means a real decline in overall funding. When folks are measuring every minute of research effort in relationship to cost, it can be hard to convince them to add the time to make their research open, by documenting data and code, planning for long-term data preservation, etc.
- Open science and the data skills needed for open science is new work. However, there can be bureaucratic resistance to creating the new positions needed to support open science.
- Creating Open Science resources, such as metadata standards, training materials, open interfaces and repositories, all take real time. And that time is often hard to find within existing work days.

So what have these challenges meant during COVID-19?

- There are great examples of U.S. agencies moving quickly to open access to COVID-19 research. For instance, In March 2020, the National Library of Medicine (NLM) worked with publishers and scholarly societies to open up access to coronavirus research through PubMed Central. https://www.nih.gov/news-events/news-releases/national-library-medicine-expands-access-coronavirus-literature-through-pubmed-central

- COVID-19 has also showed many departments our shortcomings in relationship to our vision of opening U.S. science. I think we will need to take a hard look at the COVID-19 experience, and shift resources so that we can prepare for the next time. And there will be a next time. We live on an active planet, in a dynamic universe, building fragile infrastructures. I hope that one way we use all of the COVID-19 data we collect , in every discipline, will be to make the changes needed to avoid or mitigate the levels of suffering we have seen this time.]

[Expanded background text:

Most of the presentation paints a rosy picture of progress towards Open Science in the United States Government.

But we should take a moment to outline the challenges we face as well. And it is in this slide where you may hear many more opinions from me. Let us look at some of the challenges among the four broad themes I discussed, and within the context of COVID-19.

I will limit myself to one or 2 bullets per heading.

Policy Challenges::

- The socialization and implementation of new open science policies can be uneven within departments and across government. The implementation of the DOT public access has been uneven, and we are working hard to bring each research office into compliance.

Practice Challenges

- Culture change is hard. New practices challenge existing workflows, and integrating Open Science practices can seem like a lot of work for little pay-off.

Technology Challenges

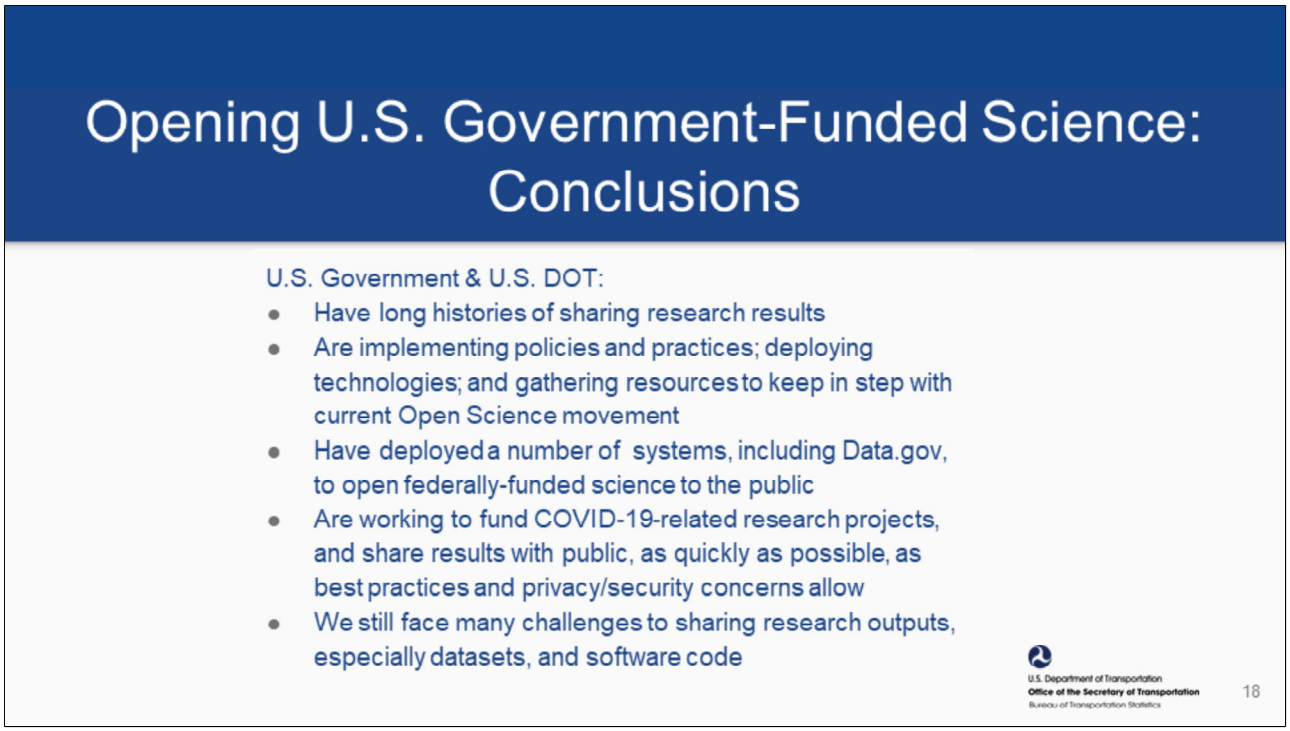
- System integrations can be complex., especially when combining systems built in-house with third-party systems.

Resource Challenges

- For many U.S. research offices, funding has been flat for years. Which of course means a real decline in overall funding. When folks are measuring every minute of research effort in relationship to cost, it can be hard to convince them to add the time to make their research open, by documenting data and code, planning for long-term data preservation, etc.
- Open science and the data skills needed for open science is new work. However, there can be bureaucratic resistance to creating the new positions needed to support open science.

So what have these challenges meant during COVID-19?

- There are great examples of U.S. agencies moving quickly to open access to COVID-19 research. For instance, In March 2020, the National Library of Medicine (NLM) worked with publishers and scholarly societies to open up access to coronavirus research through PubMed Central. https://www.nih.gov/news-events/news-releases/national-library-medicine-expands-access-coronavirus-literature-through-pubmed-central
- COVID-19 has also showed many departments our shortcomings in relationship to our vision of opening U.S. science. I think we will need to take a hard look at the COVID-19 experience, and shift resources so that we can prepare for the next time. And there will be a next time. We live on an active planet, in a dynamic universe, building fragile infrastructures. I hope that one way we use all of the COVID-19 data we collect , in every discipline, will be to make the changes needed to avoid or mitigate the levels of suffering we have seen this time.]



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

Speaker notes:
I hope this presentation has provided you some understanding of how the U.S. Government & U.S. DOT:
● Have long histories of sharing research results
● Are keeping in step with current Open Science movement
● Are funding and sharing COVID-19-related research
● And that We that still face many challenges

Thank you so much for your time an attention.

[Next slide]

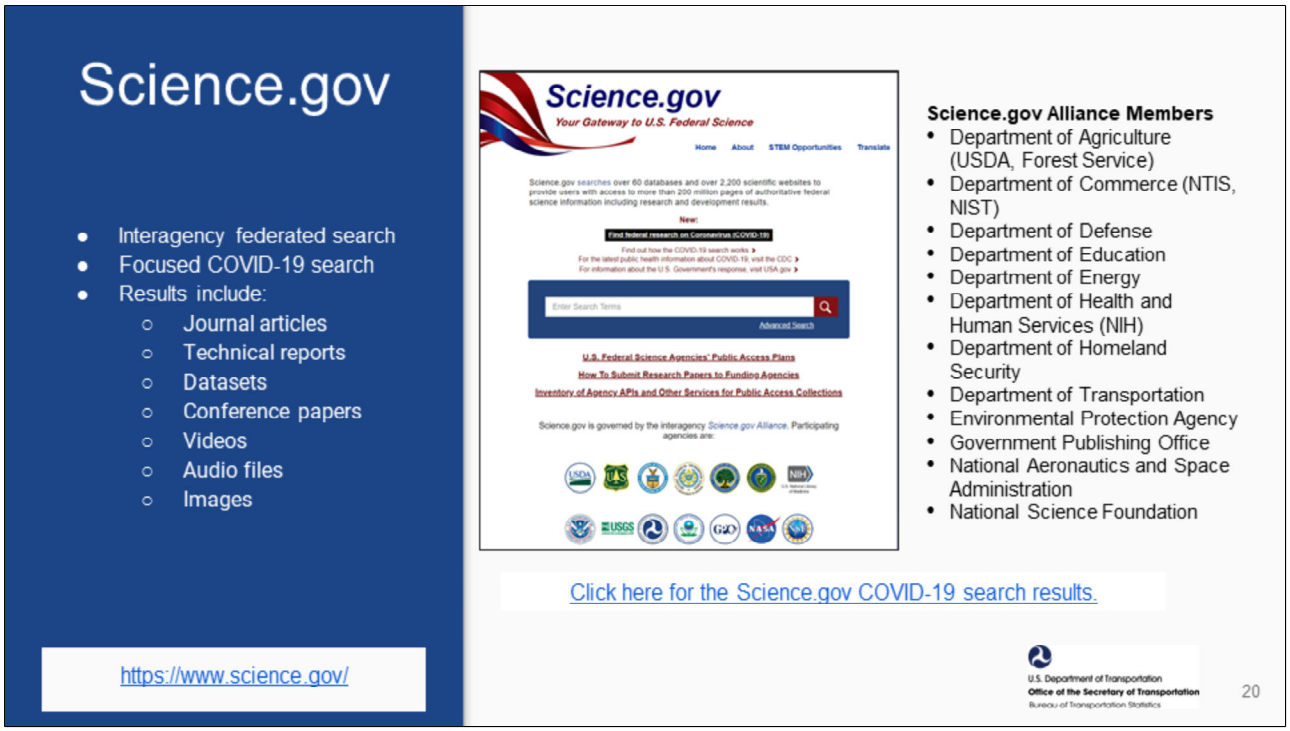
[Slide text not presented orally:
I hope that my presentation has provided you some understanding of how the U.S. Government & U.S. DOT:

- Have long histories of sharing research results
- Are implementing policies and practices; deploying technologies; and gathering resources to keep in step with current Open Science movement
- Have deployed a number of systems, including Data.gov, to open federally-funded science to the public

- Are working to fund COVID-19-related research projects, and share results with public, as quickly as possible, as best practices and privacy/security concerns allow
- We still face many challenges to sharing research outputs, especially datasets, and software code]

Slide Title: Supplemental Slides

Slide text: The following Supplemental Slides were intended for the presentation. However, they were trimmed from the presentation in order to remain in the 10 minute time limit.



Slide Title: Science.gov

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

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

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

[Next slide]

[Slide text not presented orally:
Scince.gov URL: https://www.science.gov/

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

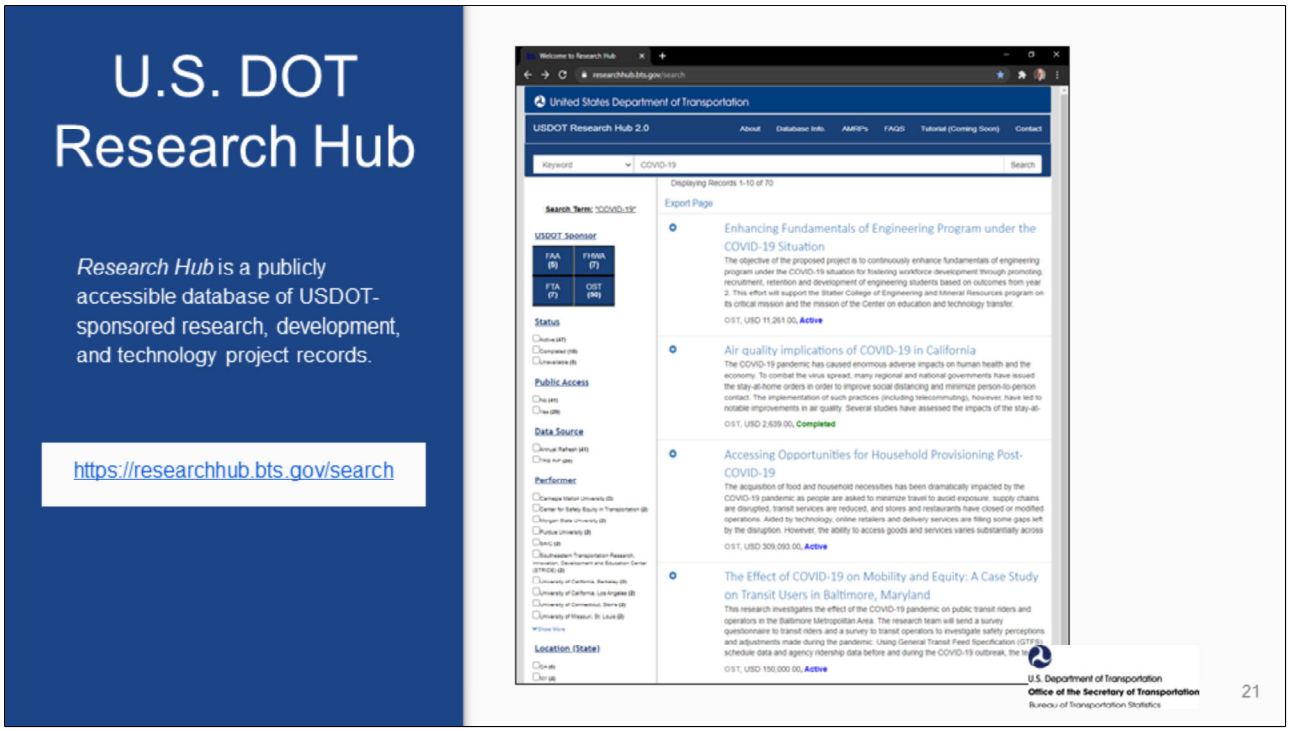
- Department of Agriculture (USDA, Forest Service)
- Department of Commerce (NTIS, NIST)
- Department of Defense
- Department of Education
- Department of Energy
- Department of Health and Human Services (NIH)
- Department of Homeland Security
- Department of Transportation
- Environmental Protection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

Click here for the Science.gov COVID-19 search results:
https://www.science.gov/scigov/desktop/en/service/link/runSearch/fullRecord:%22Coronavirus%22%20OR%20%222019-nCoV%22%20OR%20%222019nCoV%22%20OR%20%22COVID-19%22%20OR%20%22SARS-CoV-2%22

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

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

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



Slide Title: Research Hub

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

The USDOT Research Hub is a publicly accessible database of USDOT-sponsored research, development, and technology project records. The database acts as a central repository for information on active and recently completed projects from USDOT's Operating Administrations, providing a comprehensive account of the Department’s research portfolio at the project level.

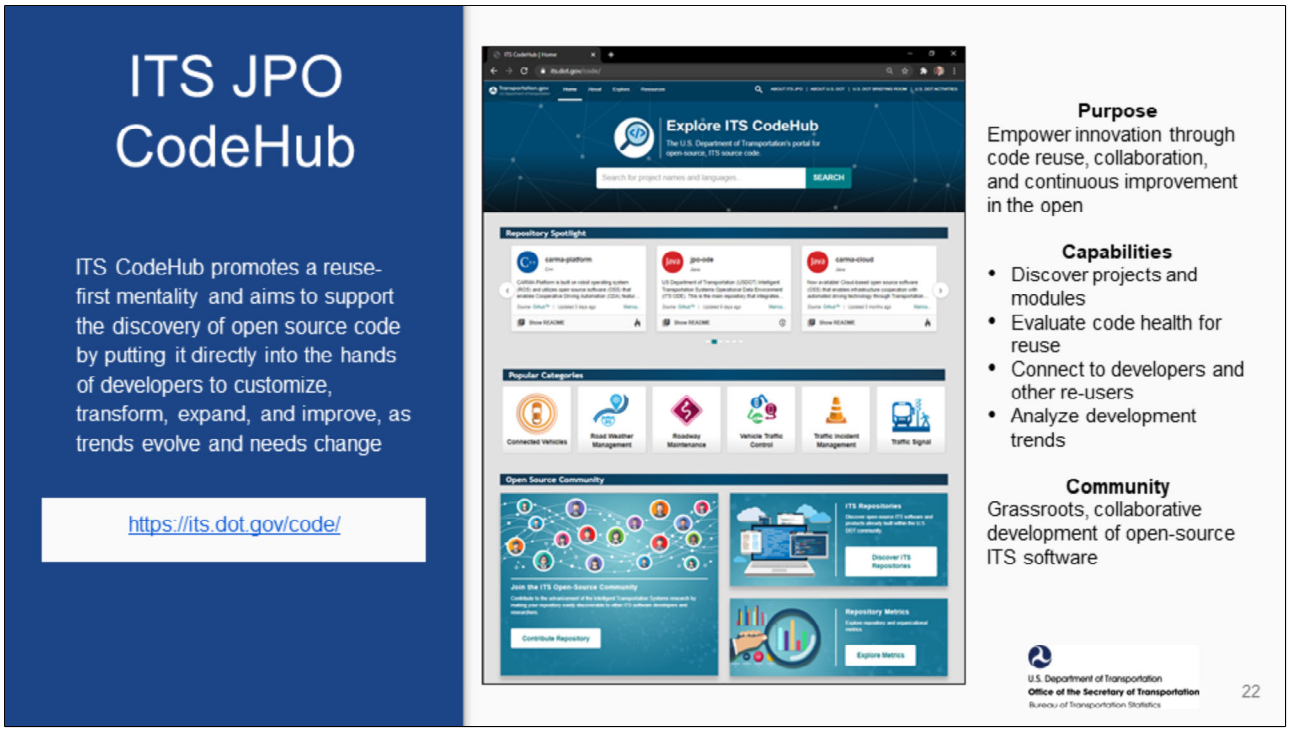
A quick search of Research Hub shows a number of active COVID-19 related projects, and at least 1 that is already completed.

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

[Next slide]

[Slide Text of slide not presented orally:
Research Hub is a publicly accessible database of USDOT-sponsored research, development, and technology project records.

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



Slide Title: ITS JPO CodeHub

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

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

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

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

The Purpose
Empower innovation through code reuse, collaboration, and continuous improvement in the open...
ITS CodeHub's primary objectives are to:

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

The Capabilities
More than just a catalogue of software development projects...
ITS CodeHub goes beyond cataloging U.S. DOT-funded software development projects. ITS CodeHub offers the transportation community the following capabilities:

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

The Community
The community plays a pivotal role in the development of open-source products—from discovery, to quality assurance, to coding and development, to adoption and integration.
ITS CodeHub fosters a community for the grassroots, collaborative development of open-source ITS software among the U.S. DOT, state and local agencies, researchers, and companies.

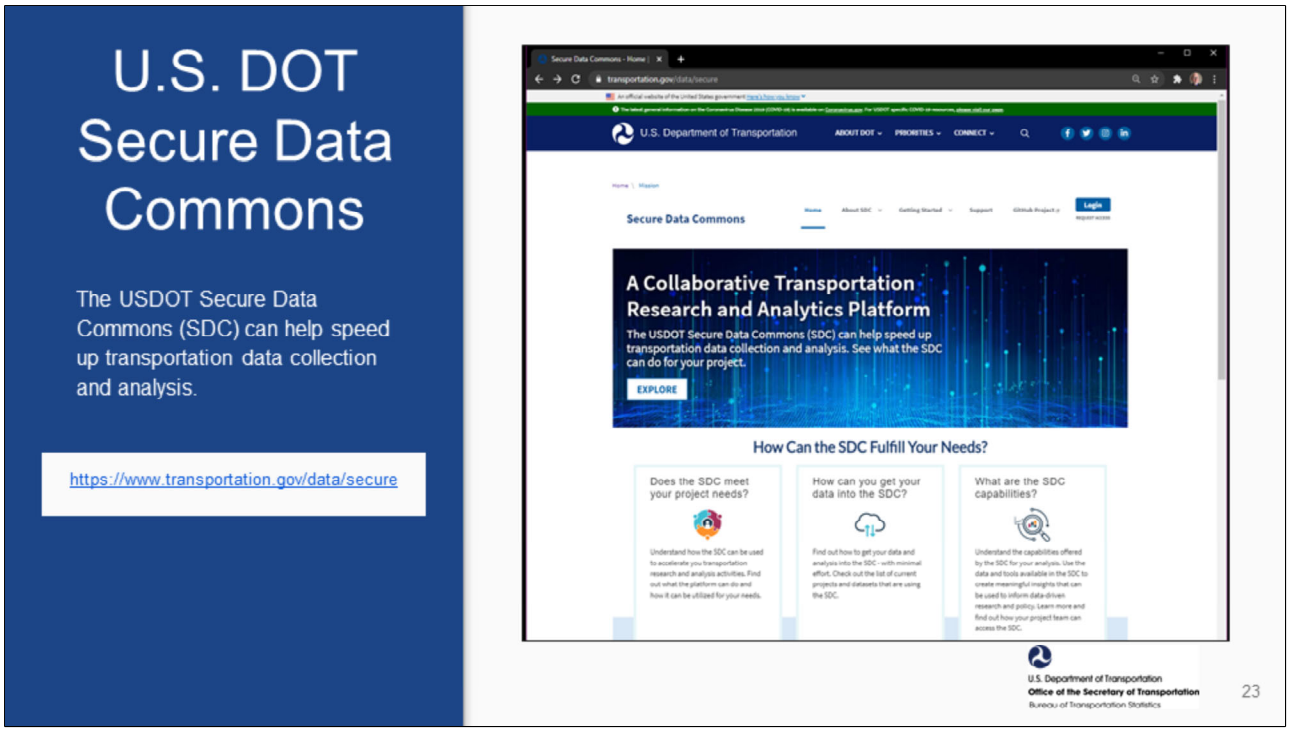
Contact Information
For more information about the ITS CodeHub, please contact the ITS JPO Data Program support team at: data.itsjpo@dot.gov.

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

[Slide Text of slide not presented orally:
ITS CodeHub promotes a reuse-first mentality and aims to support the discovery of open source code by putting it directly into the hands of developers to customize, transform, expand, and improve, as trends evolve

and needs change

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



Slide Title: U.S. DOT Secure Data Commons

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

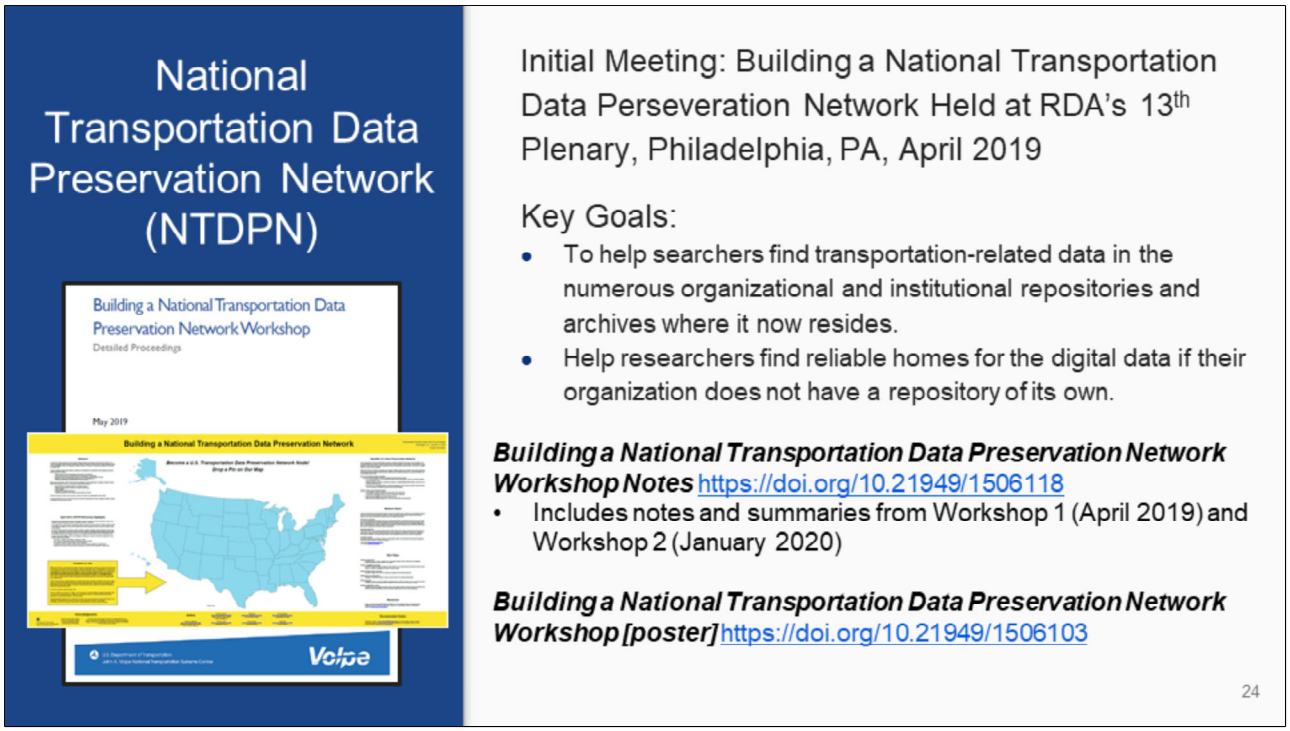
The Secure Data Commons (SDC) is a cloud-based analytics platform that enables traffic engineers, researchers, and data scientists to access transportation-related datasets. The U.S. Department of Transportation (USDOT) created the SDC to provide a secure platform for sharing and collaborating on research,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

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

[Next slide]

[Slide text not presented orally:
The USDOT Secure Data Commons (SDC) can help speed up transportation data collection and analysis.

https://www.transportation.gov/data/secure
]



Slide Title: National Transportation Data Preservation Network (NTDPN)

Speaker notes:
The transportation research community faces a number of challenges. These include:

- There is no domain-specific data repository in the U.S.
    - And we are NOT trying to build one!!
- U.S. DOT-funded transportation research data already exists in scores of academic repositories
- These repositories operate under a variety of policies and practices
- Researcher may not know which repository to use

Our goals include:

- Improve discovery, sharing, and preservation of U.S. DOT-funded data
- Without building one repository to rule them all

One possible solution is a National Transportation Data Preservation Network (NTDPN)
In April 2019, BTS sponsored, and NTL, with help from the team at the Volpe center, organized a first of its kind workshop, to investigate the possibility of building a National Transportation Data Preservation Network.

A National Transportation Data Preservation Network would provide unified access and discovery for transportation research data, where data are clearly described and defined. As a result, researchers would be able to combine and reuse data, opening up new paths of inquiries that leverage (rather than duplicate) earlier efforts.

Goals:

- help searchers find transportation-related data in the numerous organizational and institutional repositories and archives where it now resides
- help researchers find reliable homes for the digital data if their organization does not have a repository of its own.

:

The Network Vision includes the following focus areas:

Data Access & Preservation

- The network will connect transportation data from across the U.S. to be easily discovered across various repositories and institutions.
- Network participants will encourage adherence to the FAIR Principles << https://www.force11.org/group/fairgroup/fairprinciples >> ensuring data will be uniformly described and identified to allow researchers to easily find useful and related data, with an understanding of the quality and utility of the data.
- Long-term preservation and access can be guaranteed through CORETRUSTSEAL, or other, repository certification.

Networking Infrastructure

- A network action group is currently exploring which network model to adopt.
- One potential model is a hub-and-spoke architecture (like the NSF model), where member nodes feed into the discovery and access points.
- This model extends existing repositories and efforts, allowing for participation by researchers regardless of their institutional affiliation. The network will also feed into other established discovery tools for data and transportation.

Community of Practice (CoP)

- The network will support a CoP for transportation research, with collaboration through data management, shared governance, education, compliance, and reuse.

We believe the creation of such a network as providing a large number of benefits to the transportation research community.

Today, transportation data in the U.S. is stored and managed in disparate ways across numerous platforms. This can make finding data a challenge. Further, as best practices for data preparation, documentation, and preservation are well established, these practices are unevenly implemented across the transportation research domain. This can have a negative impact on long-term data accessibility. Making data findable, accessible, interoperable, and reusable (FAIR), produces

many benefits: researchers creating data receive credit through data citation; greater ease in identifying data gaps; and improved findability of previously collected data for reuse.

Why a Data Preservation Network is Important and Key Benefits of a NTDPN:

1. Improve tracking and coordination of data as there currently is no central repository for U.S. transportation research data;
2. Identify partners willing to help archive and preserve collected data even without access to an institutional repository; and,
3. Encourage adoption of repository standards which makes data easier to find, for people and computers, through the use of robust metadata among other benefits.

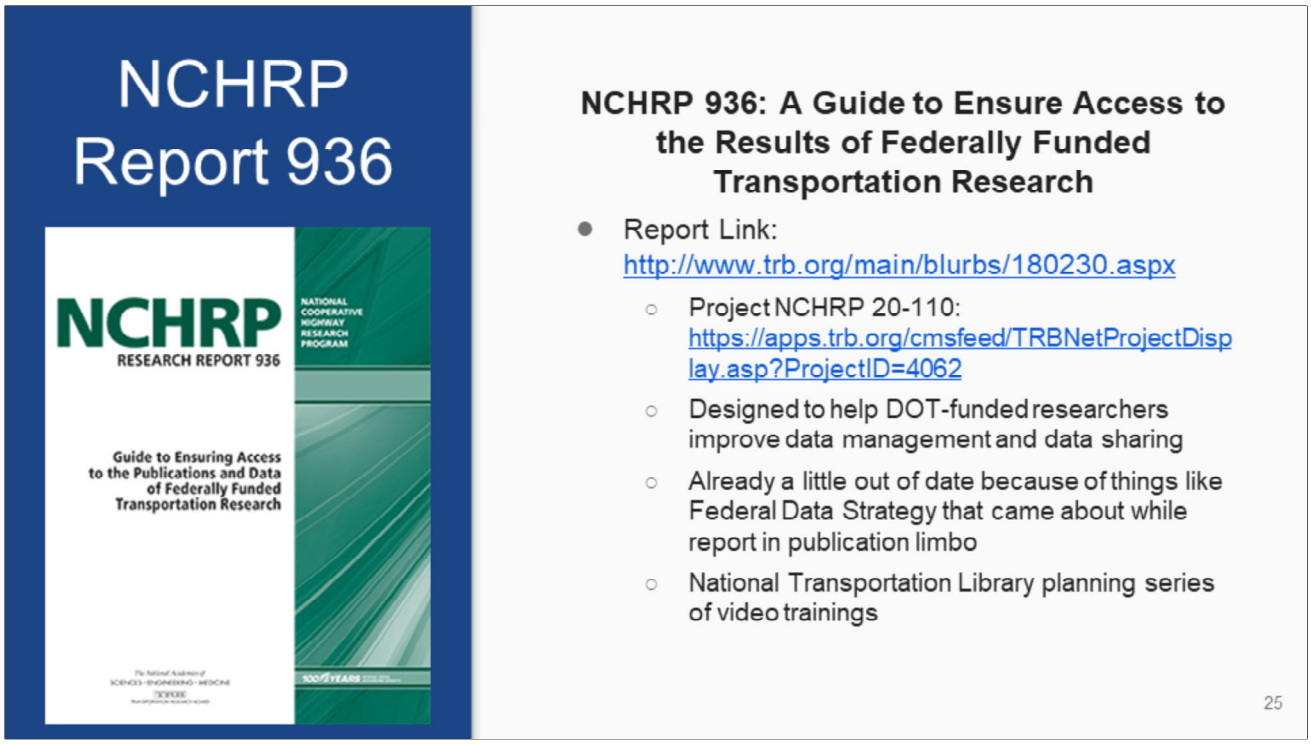
Benefits of Joining a Data Preservation Network, include:

--Help establish repository standards for transportation data preservation;
--Contribute to developing research norms for accessing and citing data;
--Improve data availability for the widest possible use; and,
--Meet like-minded individuals who care about transportation data preservation.

IF you would like to read more about out activities so far, please see the meeting summaries and our poster listed on this slide.

Building a National Transportation Data Preservation Network Workshop https://doi.org/10.21949/1506118

Building a National Transportation Data Preservation Network Workshop [poster] https://doi.org/10.21949/1506103



Slide Title: NCHRP Report 936

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

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

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

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

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

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

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

Slide Title: Links to resources

Slide text:

United States. Government Publishing Office (GPO). (2021). Washington, D.C. https://www.gpo.gov/
United States. Government Publishing Office (GPO). (2021). GPO Style Manual. Washington, D.C. https://www.govinfo.gov/content/pkg/GPO-STYLEMANUAL-2016/pdf/GPO-STYLEMANUAL-2016.pdf
United States. National Technical Information Service (NTIS). (2021). Washington, D.C. https://www.ntis.gov/
United States. National Technical Information Service (NTIS). (2021). National Technical Reports Library (NTRL). Washington, D.C. https://ntrl.ntis.gov/NTRL/
United State. Department of Transportation. Federal Highway Administration. (1976). Freeway Data for Incident and Nonincident Conditions – Vol. 1: Traffic Data Sets from Widely Spaced Detectors. Washington, D.C. https://doi.org/10.21949/1520658
United State. Department of Transportation. Federal Highway Administration. (1977). Freeway Data for Incident and Nonincident Conditions – Vol. 2: Traffic Data Sets from Closely Spaced Detectors. Washington, D.C. https://doi.org/10.21949/1520659
United State. Department of Transportation. Federal Highway Administration. (1977). Freeway Data for Incident and Nonincident Conditions – Vol. 3: FORTRAN Program Documentation for Analyzing Individual Data Sets. Washington, D.C. https://doi.org/10.21949/1520660

Slide Title: Links to resources

Slide text:
United States. Office of Management and Budget (OMB). (2005). Memo M-06-02, "Improving Public Access to and Dissemination of Government Information and Using the Federal Enterprise Architecture Data Reference Model." https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2006/m06-02.pdf
United States. White House. (2009). "Transparency and Open Government." [Memorandum]. https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government
United States. Office of Management and Budget (OMB). (2009) Memo M-10-06, "Open Government Directive." https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2010/m10-06.pdf
United States. White House. Office of Science and Technology Policy (OSTP). (2013). "Increasing Access to the Results of Federally Funded Scientific Research." [Memorandum.] https://web.archive.org/web/20130308142014/https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
United States. Office of Management and Budget (OMB). (2013). Memo M-13-13, "Open Data Policy – Managing Information as an Asset." https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2013/m-13-

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

Slide Title: Links to resources

Slide text:
United States. House of Representatives. (2017). H.R. 4174, "Foundations for Evidence-Based Policymaking Act of 2018." https://www.congress.gov/bill/115th-congress/house-bill/4174
United States. Senate. (2017). S.760 - "Open, Public, Electronic, and Necessary Government Data Act." https://www.congress.gov/bill/115th-congress/senate-bill/760/text
United States. (2019). Public Law No. 115-435 "Foundations for Evidence-Based Policymaking Act of 2018." https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf
United States. National Science and Technology Council. Interagency Working Group on Digital Data. (2009). Harnessing the Power of Digital Data for Science and Society. Washington, D.C.
https://www.nitrd.gov/Publications/PublicationDetail.aspx?pubid=25
White House. Office of Science and Technology Policy (OSTP). (2020).
https://www.whitehouse.gov/ostp/
United States. Department of Transportation. (2015). "Plan to Increase Public Access to the Results of Federally-Funded Scientific Research."
https://doi.org/10.21949/1520559
United States. General Services Administration. Technology Transformation Services. (2021). "Data.gov." https://www.data.gov/

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

Slide Title: Links to resources

Slide text:
United States. Department of Transportation. Bureau of Transportation Statistics. National Transportation Library. (2021). “Repository & Open Science Access Portal (ROSA P).” https://doi.org/10.21949/1398953
United States. Department of Energy. Office of Scientific and Technical Information. (2021). “Science.gov.” https://www.science.gov/
United States. Department of Transportation. Office of the Assistant Secretary for Research and Technology. (2021). “Research Hub.” https://researchhub.bts.gov/search
United States. Department of Transportation. Bureau of Transportation Statistics. (2021). “COVID-19 Related Transportation Statistics.” https://www.bts.dot.gov/covid-19
United States. General Services Administration. Technology Transformation Services. (2021). “resources.data.gov.” https://resources.data.gov/
United States. Department of Health and Human Services. National Institutes of Health. (2020). “The National Library of Medicine expands access to coronavirus literature through PubMed Central.“ https://www.nih.gov/news-events/news-releases/national-library-medicine-expands-access-coronavirus-literature-through-pubmed-central

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