Building a National Transportation Data Preservation Network

Abstract

In early 2019, USDOT sponsored a workshop titled: Building a National Transportation Data Preservation Network. The event brought together thirteen librarians, researchers, repository managers, and others involved in data management efforts to discuss feasibility, rewards, and challenges to building a network or consortium of trustworthy repositories for transporta-

The group discussed several topics related to priorities and considerations for transportation data management during the day-long workshop, including:

- Defining relevant nodes and stakeholders, their needs and contributions;
- Funding mechanisms and costs associated with data storage and management;
- Measuring and documenting research impacts to demonstrate the value of research funding; Research scooping, data misrepresentation/misuse, and confidentiality; and,
- Partnering opportunities and potential network model examples.

Based upon this discussion, USDOT will work with the participants and other stakeholders to establish a National Transportation Data Preservation network or Community of Practice (CoP) focusing on:

- Data preservation of materials useful for the research project; Research impact tracking to maximize and measure data sharing;
- Providing FAIR, trustworthy data; and,
- Certifying repositories to ensure data is the most useful for end users.
- The group continues to meet to enact next steps. The group is looking for new stakeholder to join the effort.
- This poster describes the work of the group to date, and invites other stakeholders to become engaged in building a national transportation data preservation network.

April 2019, NTDPN Workshop Highlights

- The major goal of this first workshop was to bring together repository managers with large transportation research data collections to begin discussing the creation of a repository network or consortium.
- It was noted that in the United States, transportation research is spread out among more than 51 states or regions, scores of universities and colleges, and many private entities. Most of these organizations receive some type of federal funding
- The USDOT Public Access Plan includes a guidance website to safeguard federally-funded transportation research data for long-term preservation and reuse. Among the various best practices and guidelines offered in these pages are evalua-
- tion guidelines researchers and repository owners can use to judge conformance with the USDOT Public Access Plan. • The group discussed feasibility, rewards, and challenges to building such a network of trustworthy repositories for trans-
- portation research data, including: preservation, metadata, and repository workflows and needs;

for their research.

- how a network of transportation research repositories could work together;
- the success and failures of similar networks in other disciplines; • contracting and service models for smaller research units who become members of the network; and,
- preparing for next steps in improving long-term preservation of federal funded transportation research data

Invitation to Join

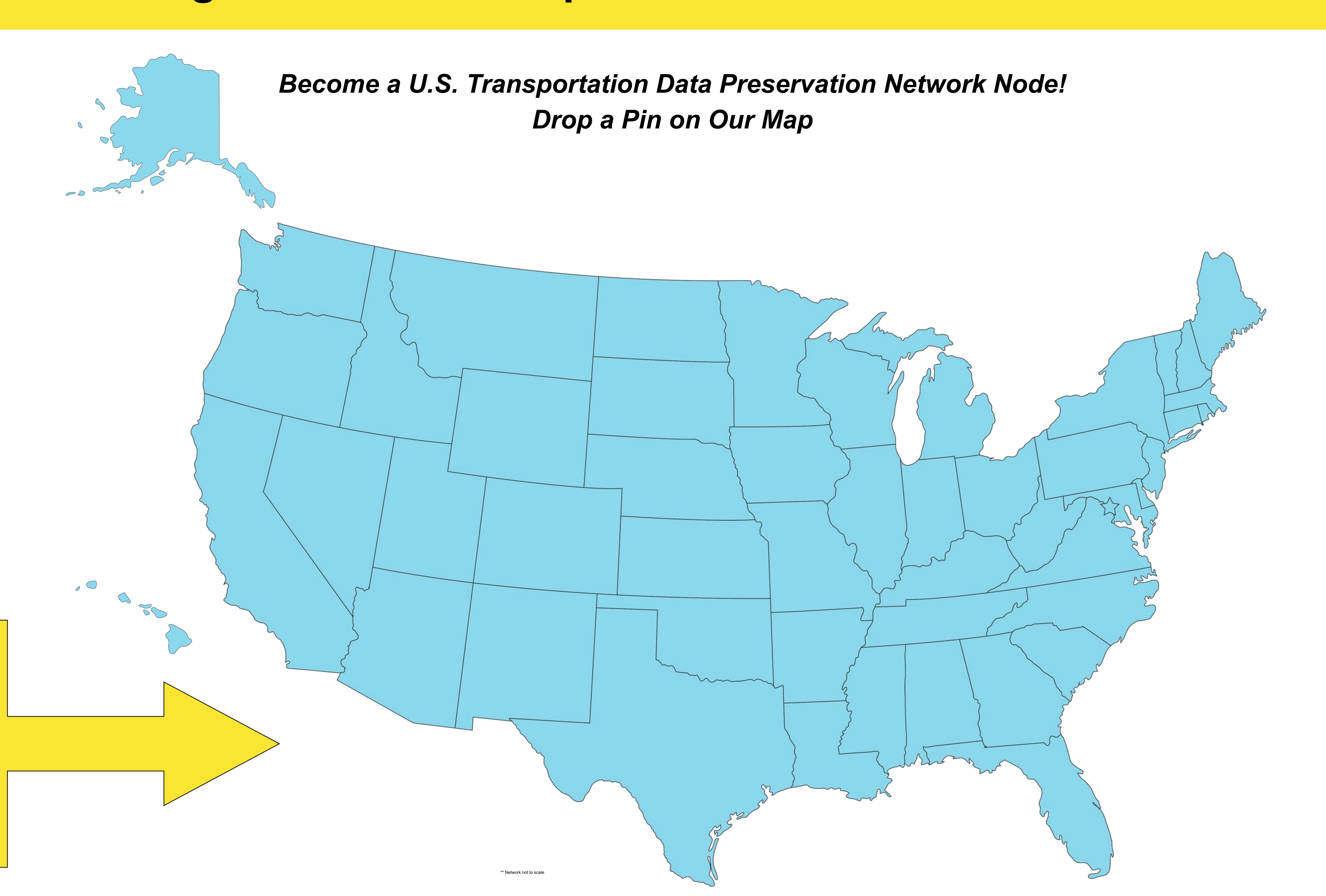
While there has been a lot of talk about the benefits of "big data," the transportation research community and industry are still in their infancy of using data to drive informed decision making and setting better policy. Simultaneously, we are on the cusp of a transportation revolution with new technologies and concepts for alternative modes of transportation from Hyperloop and High-Speed Rail to commercial space and the integration of drones into the National Airspace System, not to mention Autonomous Vehicles of all varieties now joining the surface and marine transportation networks. From geospatial maps and transit routes, to safety data collected about pedestrians and bicyclists along a particular route, transportation-related data can be used for multiple purposes. Therefore, transportation data should be preserved for future uses, and widely accessible, in usable formats.

Until now, there has been no dedicated location in the United States, either individually or federated, for the curation, preservation, and sharing of transportation-related data generated during federal, state, industry, and academic research. With the guidance of the National Transportation Library of the United States Department of Transportation, several academic and public institutions are working to design and establish a data preservation network capable of providing all the benefits and services the transportation field needs.

So, what do we need to make this happen? You.

This is our invitation to you to join us in creating -- from the ground up -- the data preservation network transportation needs as we move into a more data-centric era. Please join us in this effort and help us to make transportation data curation, preservation, and access best practices our common practice!

Take your first step: Please choose a numbered pin and place it on the map at the right in the approximate location of your organization. Then write that same number on the back of your business card and drop it in the envelope. No business card? Enter your map pin number and your contact information on the sign up sheet on the table.



Benefits of a Data Preservation Network

Today, transportation data in the United States 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 (or *FAIR*), produces many benefits. These include researchers who create data receiving credit through data citation; greater ease in identifying data gaps; and, improved findability of previously collected data for reuse.

Why is a Data Preservation Network Important?

- There are three key benefits to a National Transportation Data Preservation Network:
- 1. As no central repository for U.S. transportation research data currently exists, a network can *improve tracking*
- 2. Researchers without access to an institutional repository can easily identify partners willing to help archive and
- 3. The network can encourage adoption of repository standards, which among other benefits, make data easier to find, for people and computers, through the use of robust metadata.
- Benefits of Joining a Data Preservation Network?
- 1. Help establish repository standards for transportation data preservation; 2. Contribute to developing research norms for accessing and citing data;
- 3. Improve data availability for the widest possible use; and
- 4. Meet other like-minded individuals who care about transportation data preservation.

Network Vision

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.

Data Access & Preservation

The network will aggregate transportation data from across the United States to be easily discovered across various repositories and institutions. Network participants will adhere to FAIR Principles, so that data will be uniformly described and identified. This standardization of formats and descriptors allows researchers will easily be able to find useful and related data, with an understanding of the quality and utility of the data. Because networked repositories will be CORETRUSTSEAL² certified, long-term preservation and access will also be guaranteed.

The network will use 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

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

1: FAIR Principles: https://www.go-fair.org/fair-principles 2: CoreTrustSeal: https://www.coretrustseal.org/

Next Steps

Continue the Discussion:

Description: Hold an in-person meeting at the Transportation Research Board (TRB) 2020 Annual Meeting. You are invited! Ask a poster author for room details.

Conduct a Capabilities Assessment:

Description: Engage repository managers at public, academic, and university based transportation research organizations to understand capabilities and interest in network building.

Update the USDOT Public Access Plan:

Description: Engage with USDOT as it reviews and updates the 2015 Public Access Plan.

Description: This network includes a variety of partners. Reach out to additional stakeholders.

Identify partners and stakeholders:

Description: Develop use-cases for different roles/perspectives to clarify the purpose, role, and value of the network for different users, encourage stakeholder participation, and develop common terminology and dialog.

Description: Document sensitive information levels in data collection efforts and identify what levels repositories can handle to alleviate confidentiality concerns with publicly available data at risk for reverse engineering

Resources

Carlson, Jacob http://orcid.org/0000-0003-2733-0969, et. al. 2019. Building a National Transportation Data Preservation Network Workshop. United States. Department of Transportation. Bureau of Transportation Statistics. https://doi.org/10.21949/1506118

Acknowledgements

Leighton L Christiansen

National Transportation Library, USDO

leighton.christiansen@dot.gov

UC Berkeley ITS Library

Steven Polunsky

http://orcid.org/0000-0003-1219-0764

University of Alabama

Michael Witt http://orcid.org/0000-0003-4221-7956 Purdue University

Recommended Citation