

## National Transportation Data Preservation Network (NTDPN) Workshop 2

### *Network Overview and Development Efforts*

Data archiving at the Bureau of Transportation Statistics (BTS) initially focused on small datasets from Department of Transportation (DOT) surveys, but has evolved to include a more diverse collection of data types, including university research data, introducing the need to work with, store, and archive huge databases and overburdening the limited BTS and National Transportation Library (NTL) staff. The introduction of the internet facilitated making data collected by public agencies publicly available at no cost, however storing, managing, and providing data is not free for the provider. Since January 2017, the NTL Data Curator has been working with stakeholders to help bring their institutional repositories into conformance with USDOT recommendations, and realized that while not all research organizations have access to, or the ability to construct, repositories, some large research universities may be able to provide contracted repository services at reasonable costs.

In April 2019, the BTS and NTL sponsored a first of its kind workshop to investigate the possibility of building a National Transportation Data Preservation Network (NTDPN) to provide unified access and discovery for transportation research data through which researchers could combine and reuse data to leverage, rather than duplicate, earlier efforts. The vision for the NTDPN is outlined by two key goals:

- To help users find transportation-related data currently housed in numerous organizational and institutional repositories and archives; and
- To help researchers find reliable homes for the digital data if their organization does not have a repository of its own.

This first workshop brought together a core group of stakeholders to take the first steps supporting this effort and discuss priorities and considerations for transportation data management, the feasibility of building a network of trustworthy repositories for transportation research data, and the rewards and challenges to building such a network (meeting minutes are available at: <https://doi.org/10.21949/1506118>). Identified benefits of the Network include: technology transfer; data and technology reuse; discoverability; data availability; recognition to research contributions; and not duplicating prior efforts. Since this first workshop, the group continued to meet regularly to develop the Network's focus and model moving forward. The group identified potential next steps for the Network to discuss at the second workshop.

### *Next Steps and Action Items Summary*

<b>Action</b>	<b>Description</b>
Recruit additional stakeholders	Recruit additional stakeholders to bring together diverse viewpoints and help achieve the Network goals.
Establish membership criteria	Establish criteria or guidance for repositories to join or participate in the Network.
Develop use-cases	Develop use cases from a variety of perspectives to help recruit members by clarifying stakeholder roles and outlining benefits of joining the Network, and help get executive buy-in.
Establish a Network model	Establish a model to define how the Network will operate. Courtney Mumma is interested in being involved.
Support the development of a metadata catalog	Support the Transportation Research Board (TRB) and Transport Research International Documentation (TRID)/Transportation Research Information Services (TRIS) efforts to further the development of a metadata catalog. The Network should reach out to Bill McLeod, TRIS Manager, to pursue this effort.

Action	Description
Investigate uses for data registries	Identify data registries and determine how they can be used by the Network.
Update the USDOT Public Access Plan	USDOT will update the four-year-old Public Access Plan; expected summer 2020.
Conduct a capabilities assessment	Conduct a capabilities assessment of the Network, its nodes, and its members to determine what each group can or cannot provide, define any gaps in the Network infrastructure, and determine who is best suited for each element to inform a plan moving forward.

### *Network Next Steps Discussions*

Recruit additional stakeholders – Up to this point, a core working group of 13 people has been collaborating, initially to determine if developing this Network was worth pursuing, and now to advance this Network and next steps. Recruitment efforts at the 2020 TRB poster session identified 21 new potential stakeholders.

Establish membership criteria – Discussed criteria and guidance for repositories to join or participate in the Network include:

- Repositories having the [CoreTrustSeal](#) (CTS) is encouraged to ensure data reliability, but not required to join the Network.
- There is no financial commitment to join the Network; only time and expertise.
- The Network has no assumptions about data contributors other than that their data is compliant with the USDOT Public Access Plan.
- Data stored on a website does not count as a repository since the goal is for the Network to provide long-term repository solutions.
- This Network aspires to grow to be a national repository for all transportation data, not just USDOT funded data, but is starting with this subset since the USDOT is a large funder of transportation research.

Develop use-cases – Discussions on the requirements for joining the Network highlighted the immediate need for developing use cases.

Establish a Network model – While different network models have been suggested, no specific model has been identified for the Network to follow and the working group has not determined what type of model will be most appropriate. The Network is technology agnostic and is not trying to create a domain specific repository for transportation, but rather bring repositories together. [Chronopolis](#) and [LOCKSS](#) (Lots of Copies Keeps Stuff Safe) are good examples. Considerations include having a governing body, requiring membership fees, or adopting/emulating existing models.

Support the development of a metadata catalog – The Network could provide a metadata catalog for repositories. A metadata catalog is search engine agnostic and is useful for researchers and repository managers, so this would both add value to the field and be appreciated by researchers. TRID functions as an index and does not host any data itself. TRID could be a central search engine for transportation datasets that search across all repositories.

Investigate uses for data registries – Data registries can help promote findability. Examples include [Schema.org](#) and [FAIRsharing.org](#).

Continue meeting – The Network will reconvene virtually in March 2020.