National Transportation Data Preservation Network Workshop: Summary May 2019

Prepared for: U.S. DOT, Bureau of Transportation Statistics, National Transportation Library
Prepared by: U.S. DOT, John A Volpe National Transportation Systems Center

Background

Since January 2017, the National Transportation Library (NTL) Data Curator has been working with academic libraries and university transportation centers (UTCs) to help bring their institutional repositories into conformance with the USDOT recommendations, and realized not all research organizations have access to, or the ability to construct, repositories. Additionally, while every federally-funded research report must be preserved through the NTL, researchers are only required to provide a link to the data stored in repositories, not the actual dataset, which can create difficulties for accessing research. A network of repositories could help alleviate some of these difficulties.

Workshop

On April 5, 2019, USDOT held a workshop titled: *Building a National Transportation Data Preservation Network.* The workshop was collocated with the Research Data Alliance (RDA) 13th Plenary. 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 transportation research data.

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, and their capabilities to help properly frame discussions.
- Prioritizing creating Findable, Accessible, Interoperable, and Reusable (FAIR) data, with a large discussion focus on data usability.
- Funding mechanisms and costs associated with data storage and management, with a large discussion focus on changing researcher expectations and proposal requirements for data management plans (DMPs) and funding.
- Measuring and documenting research impacts to demonstrate the value of research funding.
- Concerning topic areas including research scooping, data misrepresentation/misuse, and confidentiality.
- Partnering opportunities and potential network model examples.

Next Steps

Based upon the 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 the following, while defining actionable, achievable targets:

- Data preservation of materials useful for the research project, with an initial narrow focus on USDOT funded Public Access data and final products;

- Research impact tracking to maximize and measure data sharing, starting with implementing an automated citation or viewership count to provide one tangible perspective of impact; and
- Legal compliance plus usefulness, namely data accessibility and discoverability, by providing FAIR, trustworthy data and certifying repositories to ensure data is the most useful for end users.

The group will reconvene virtually in late-May or early-June 2019 to address the key action items established in the first meeting and follow-up on additional action items discussed to continue developing a plan moving forward.

Action Items

Through this effort, the group established several action items and potential ideas to pursue, summarized in the following table:

Action	**Description**
Continue the discussion	Hold a series of webinars as well as an in-person meeting at the next Transportation Research Board (TRB) conference in January 2020.
Conduct a capabilities assessment	Engage repository managers responsible for entering data into systems to understand capabilities and interests to determine the target audience and gaps in network infrastructure to inform a plan moving forward. Survey the target group to provide a background understanding and inform in-person discussions. USDOT will identify parties of interest and lead this effort. USDOT will try to secure funding for an invitational travel meeting bringing repository managers and research data librarians to the Council of University Transportation Centers (CUTC) TRB meeting to interact with UTC managers and other stakeholders.
Update the USDOT Public Access Plan	USDOT will review and update the three-year-old Public Access Plan.
Identify partners and stakeholders	This network includes a variety of partners. Additional stakeholders to involve include: the American Association of State Highway and Transportation Officials (AASHTO) Committee on Data Management and Analytics, AASHTO Research Advisory Committee (RAC), and State DOTs. As the effort matures, other partners may be added.
Establish a network model	Develop a spoke-and-hub network to connect users to data/information. Network design should consider the potential for future expansion. Other disciplines may have established networks to emulate or build off of.
Develop use-cases	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.
Leverage existing search engine infrastructure	Expose data to established search engines by including metadata to make data findable and accessible through established systems.
Develop a sustainable funding mechanism for data storage and management	Shift expectations and require project proposals to include funding for long-term storage costs as well as data sharing and preservation. Proposals should detail who is responsible for the data. Develop guidance for universities including a template to include in DMPs. Establish a cost

	sharing mechanism. Investigate alternative funding sources dedicated to digital archiving.
Increase data usability	Increase data usability by developing a metadata catalog tool to search various repositories using metadata driven search engines, developing a repository finder to increase researchers' unawareness of potential repositories available to submit data to, and effectively communicating the importance and value of reusable data to improve data quality.
Alleviate confidentiality concerns	Document Personally Identifiable Information (PII) 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.

Conclusion

This workshop was a crucial first step in developing a plan for transportation data storage and preservation and forming a National Transportation Data Preservation network. The meeting facilitated connections and networking between researchers and repository managers, providing invaluable perspectives for the group and a better, holistic understanding of the issues at hand. Moving forward, the group plans to continue to meet and communicate, connecting additional stakeholders to expand the initiative and achieve a more robust transportation data storage and preservation network.