Building a National Transportation Data Preservation Network Workshop

Detailed Proceedings

May 2019

U.S. Department of Transportation
John A. Volpe National Transportation Systems Center

Volpe

Building a National Transportation Data Preservation Network
Held at RDA's 13th Plenary, Philadelphia, PA

Loews Hotel, Congress Room C
April 5, 2019
9:00 AM – 3:30 PM

Detailed Proceedings

Participants

- Jacob Carlson http://orcid.org/0000-0003-2733-0969, *University of Michigan*
- Emily Chan https://orcid.org/0000-0002-3736-4858, *San Jose State University*
- Charles Ducker https://orcid.org/0000-0002-7535-5258, *USDOT (OST)*
- Kendra Levine https://orcid.org/0000-0002-2479-3952, *UC Berkeley ITS Library*
- Shashi Nambisan https://orcid.org/0000-0001-9232-945X, *University of Alabama*
- Hilary Nixon https://orcid.org/0000-0001-5378-3473, *San Jose State University*
- Steven Polunsky http://orcid.org/0000-0003-1219-0764, *University of Alabama*
- Rolf Schmitt, *USDOT (OST)*
- Lisa Kay Schweyer, *Carnegie Mellon University*
- Michael Witt http://orcid.org/0000-0003-4221-7956, *Purdue University (PURR)*
- Leighton Christiansen https://orcid.org/0000-0002-0543-4268, *USDOT (OST)*
- Julianne Schwarzer, *USDOT (Volpe)*
- Angela Berthaume, *USDOT (Volpe)*

Meeting Notes

Background on the Bureau of Transportation Statistics (BTS) and National Transportation Library (NTL)

- Data archiving at BTS initially focused on small datasets from Department of Transportation (DOT) surveys and has evolved to include university research data. The archiving initiative was inspired by a travel demand characteristics survey in Baltimore in the 1970s that had valuable data, but the data disappeared. While BTS was focused on compiling statistics and surveys, a small effort to obtain and share non-statistical information evolved into the NTL, which would house small datasets from surveys and other data products to preserve research documents created by and for the DOT. Two years later, the NTL was in the legislation.
- Survey data is generally in a consistent, neat format with responses collected from distributed questionnaires and is relatively easy to archive and store. BTS and NTL are no longer solely focused on survey data. Now BTS and transportation researchers collect and store other data types, such as GPS records from trucks or ships and administrative records such as credit card use for travel patterns. This has introduced the need to work with, store, and archive huge databases. For example, one effort collects the record or every tenth airline ticket purchased in the country and extracts key information about travel patterns and prices. This dataset will be useful to have archived over time. The FAA data project, System Wide Information Management (SWIM), captures and makes available all information transmitted through the air traffic control system. Preserving these records will enable experiments in creative use of the information over time.
- This shift in focus is larger than the limited BTS and NTL staff, and these organizations cannot tackle issues themselves. So, they looked to develop networks among other institutions that can

house large datasets, and determine how to divide challenges in a consistent fashion. The additional data generated by university research is an added complexity but is also what helps fund repositories.

- BTS was created at the onset of the internet when it became easy to make data available to the public at no cost. In those days, government agencies were supposed to recover the cost and users had to buy data, but the Internet changed all this. Now data collected by public agencies are available for free to the user, but not free to the provider. Storing and providing data costs money. The free data model is also affected by increased reliance on private data vendors who sell restricted data to public agencies. BTS and NTL need to figure out how to afford what everyone wants to accomplish and leverage programs to provide data efficiently and effectively, and find funding to make data Findable, Accessible, Interoperable, and Reusable (FAIR).
- One participant noted how grateful they are for the BTS and NTL.

Workshop Background

- 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 [evaluation guidelines](#) researchers and repository owners can use to judge conformance with the USDOT Public Access Plan. These guidelines are based on the [Data Seal of Approval](#) "Core Trustworthy Data Repository Requirements" and the [CoreTrustSeal](#) (CTS).
- 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 for their research. However, access to trustworthy repositories varies widely.
- Since January 2017, the NTL Data Curator has been working with several academic libraries and university transportation centers (UTCs) to help them bring their institutional repositories into conformance with the USDOT recommendations. During discussions with academic transportation researchers and repositories, the NTL realized that standing up a repository was not possible for some research organizations, while large research universities might be able to provide contracted repository services at reasonable costs.
- The major goal of this first workshop is to bring together repository managers with large transportation research data collections to discuss creating a repository network or consortium. During this first workshop, interested parties would discuss feasibility, rewards, and challenges to building such a network of trustworthy repositories for transportation research data.
- This workshop is intended to:
  - Allow participants to discuss preservation, metadata, and repository workflows and needs;
  - Allow participants to discuss how a network of transportation research repositories could work together;
  - Discuss the successes and failures of similar networks in other disciplines;
  - Discuss roles and service models for smaller research units who become members of the network; and,
  - Prepare for next steps in improving long-term preservation of federal funded transportation research data.
- Every federally-funded research report needs to be preserved through the NTL, but researchers are only required to provide a link to the data, not the actual dataset. Data is stored in repositories. But

holding the report in NTL and data in a repository can create difficulties for accessing research. A network of repositories can help alleviate some of these difficulties. Creating a central database would not be a viable solution. Institutions have their own systems and support other entities such as State organizations and MPOs that do not have the resources to create their own repositories.

- The workshop is proposing to focus on research data as part of the USDOT Public Access Plan, focusing on small data sets from research projects that do work with or preserve data in repositories as an easy and achievable first step for the group. This effort will push the existing mandate to the next logical step. Solutions discussed can be leveraged and scaled to tackle larger datasets. Today will focus on discussing the pros and cons of models that can potentially be scaled up for future efforts.

Considerations, Challenges/Barriers, and Opportunities

- <u>Defining nodes and stakeholders</u> – Defining the scope and purpose is important when building a network, but first the nodes must be defined to properly discuss challenges and barriers. The textbook network modeling approach defines the nodes (who they are, their needs, and their contributions), defines the node requirements (compared to current capabilities and gaps), and defines the communication between nodes. There are many different stakeholders involved, each with ideas about their own nodes. The group should discuss how these nodes interact to provide context for the various actors in the network.
  - o *NTL* – This agency could serve as the head node, providing guidance and direction.
  - o *"Haves;" Universities receiving money from USDOT to do transportation research* – These institutions oversee repositories and submit data to them. They are dealing with protocols and standards. They collect anything and everything across the university.
  - o *"Have-nots;" Researchers/groups not receiving DOT funding*
  - o *Repositories* – The network should agree on a set of principles for repositories. Data should be as FAIR and trustworthy as possible. Repositories can be a voice for researchers' concerns about embargos and scooping, and advocate for FAIR, trustworthy, reusable, and sharable data. The DOT should include a list of trustworthy repositories. Per the Public Access Plan, the NTL data curator needs to approve the repository documentation for trustworthiness. This will help build a database of available repositories for researchers to submit data and search for data.
  - o *State DOTs*
  - o *Funding organizations*
  - o *Librarians*
  - o *Principle Investigators (PIs)*
  - o *Media/the public*
  - o *Governance team* – A governance team including all the players and their perspectives.
- <u>Competition</u> – All the nodes or stakeholders involved are sharing and competing in the same field. These parties are all working to advance the same goal but they are also all competing for the same research dollars.
- <u>Usability and accessibility for users</u> – How can repositories organize the variety of data inputs from a mix of users to parse out transportation relevant data and make it easily accessible for transportation researchers? Something that is easily findable is more useful and brings value to researchers. Findable information can be reused and cited.

- o Developed solutions should not be limited to DOT-funded research datasets. Some repositories archive data from multiple sources, some of which are from research funded from other federal agencies or other sources. End users are interested in access to and using/reusing the data and don't necessarily care who funded the study, so it is important to not lose track of non-DOT research.
  - o Metadata is important. Users shouldn't have to know something is in a repository to use it. A metadata catalog tool would be useful. Metadata driven search engines search various repositories to find related data.
  - o Consider reusability. Archived datasets are not necessarily the highest quality from a reusability perspective. Researchers are simply following requirements to archive data and are not considering how the data can be reused in the future. This should receive more consideration. The importance and value of reusable data needs to be communicated.
  - o A repository finder could be a useful tool. Many researchers don't know what repositories are available to submit data to or understand what options are available to them. Identifying, collecting key information on, and cataloging available data repositories can help researchers identify and learn about options for repositories. Ways for researchers to find repositories can be leveraged to develop ways for users to find data. Transport Research International Documentation (TRID) may have already done this for datasets, so partnering with the Transportation Research Board (TRB) may be useful.
- Data management plans (DMPs) – University Transportation Center (UTC) agreements should include language specifying who is responsible for the data, whether it is the lead university or a partner. The current DMPs for UTCs can be found in the NTL Repository & Open Science Access Portal (ROSA P). DMPs should include a specific plan for data sharing or justification as to why data cannot be shared, and a specific preservation plan including what repository will be used or a justification for the determination to not preserve the data. Model language and a list of prompt questions should be developed as a template on data management and guidance for UTCs.
- Measuring/tracking research impact – It is not clearly demonstrated how data is reused for future research and projects. Technology (tech) transfer requires sharing research products and data resulting in greater impacts. Requiring tech transfer might help change the culture and expand research benefits to other organizations.
  - o There is a need to document the benefits and impacts of research efforts. Research impact could be an automated process to reduce the burden by automatically tracking citation counts or viewers, reuses, etc. for datasets and papers. This would only provide one perspective of impact, but it is tangible and easily automated.
  - o Demonstrating the number of times a report or dataset has been cited shows the value of the research, but this can be difficult because the information can be stored in different places, so tracking efforts will not be showing a complete record.
  - o TRID does not engage in tracking citations.
  - o There are concerns about researchers being scooped by sharing data. Researchers should be assured data is limited to the public until publication. Researchers should be encouraged to cite data as a first class citable object, as if it were a paper.
  - o With the volume of research submitted to TRB, there is an opportunity to work with TRB to develop a template for citations. TRB could offer guidance on appropriate dataset citations.
- Identifiers – Different entities have different identifiers. Individual researchers are identified through Open Researcher and Contributor ID (ORC-ID) unique identifiers, reports are identified

through the Digital Object Identifier (DOI) system. Technical report documentation pages include some identifiers. The Crossref Funder Registry contains numerous funders, including State DOTs and USDOT uniquely identified by mode. Everyone should be aware of these funder identifiers, and they should be used in the data hub.

- <u>Legality</u> – Data is kept by the NTL for five years. The DOT needs feedback if it's worth archiving for longer. Data will be housed at different repositories by the various nodes, not the DOT, but should the NTL be the public access point for archived data?
- <u>Duplicity of datasets</u> – Not all data may be appropriate for or worth archiving or storing in a repository. If a project manipulated an archived dataset or tabulated other available datasets to be in a format conducive to answer that research problem, is that worth archiving or is that just creating a duplicative dataset? Many researchers would be interested in the source data to use in their own way. However, one valuable part of sharing data is researchers getting different results from different manipulations. Duplicative datasets can be useful for data validation. Subsets of data can take months to manipulate and clean, and they can be useful to others. Long term preservation of data subsets is different than storage. The original datasets are preserved but subsets can saved as needed if they are determined to be of value and worth paying to preserve after five to ten years.
- <u>Confidentiality</u> – Some data is proprietary or was collected under a promise of confidentiality and cannot be sent for storage. Confidentiality has become harder to protect and is a growing issue in the public sector. Personally Identifiable Information (PII) levels should be included in data collection efforts and repositories should identify what levels can be handled. With the public availability of a variety of datasets and tools, data scientists are getting very good at reverse engineering data leading to confidentiality breaches. This is a downside of democratizing data. The Census Bureau has expressed concerns about producing public datasets in efforts to protect respondents from being identified through reverse engineering.
- <u>Costs</u> – Data management and storage costs come out of the research budget and impact/reduce the amount of research researchers can conduct with funding. The European Union (EU) includes data storage and management costs that are two to five percent of the research budget. One participant estimated storage costs were about a third of their cost to preserve data to be $1 per GB per year, inclusive of services and hardware. Another participant noted that sixty percent of their repository costs are spent towards servicing the data. Different types of costs were discussed, including:
    - <u>Technology refresh</u> – This is a huge challenge for electronic records. Formats change with time (the group discussed data stored on 3.5 inch floppy disks) and it can be costly to keep archive data usable.
    - <u>Data storage costs</u> – Repositories can more easily budget for data storage costs based on assumed intellectual records of research and usage demands.
    - <u>Data management and services costs</u> – Maintaining the link between datasets and reports is important and should be part of the data management process. Links get broken when sites or systems are updated or changed, or during technology refreshes. DOI metadata for data and publications with citations need to be updated so they can get picked up by Scholix. It is important to acknowledge the staff efforts needed to maintain repositories. These costs are difficult to predict and budget.
    - <u>Opportunity costs</u> – Research impact is reduced by spending research dollars on data management. This reduces the amount of student funding etc. which reduces the overall research produced.

- Funding mechanisms – Several ideas for funding mechanisms were discussed, including:
  - Secure other funding sources, such as legislation for funding from Congress for a dedicated digital archiving budget so funding does not come out of the research budget.
  - Change expectations and require research proposals to designate funding for data management and include a DMP. UTCs should require a detailed DMP to be reviewed by the NTL data curator and include requirements that a certain percent of funding are designated to data management.
  - The Public Access Plan report is three years old and needs to be reviewed. The policy should include language that proposals should or must include line items for DPMs and long term storage costs.
  - Establish a cost sharing mechanism.
- Repository scale – Repositories have frameworks specific to their institution and the specific datasets they typically collect. Opening repositories to datasets outside the institution may be problematic and/or costly. Storage and staffing costs should be considered.
  - There should be classes of certified repositories. The "haves" are those organizations or institutions that either have repositories in their university or have a means to send data to a repository (such as R1 UTCs and partners). The "have-nots" do not have these resources available for sending in data for long term storage. How can those who are not getting DOT funding get connected with the larger network of research and datasets? Ways to build in funding mechanisms for those that do not have access to repositories should be determined.
- Repositories going out of business – Do management plans have a backup contingency? The NTL is working on agreements across transportation libraries and repositories to be a repository of last resort. Sharing copies of research and data can provide some sort of backup. Backup contingencies may be outside the scope of a DMP and should instead be addressed by using certified trustworthy digital repositories.
- Misrepresentation of data – Data can be misrepresented, and with data being more available, it is more open to abuse. One solution could be including a publication or longer footnotes about what individual data points are, what they represent, and how they can and can't be used.
  - This problem is not unique to transportation data.
  - A major challenge is how to measure and communicate data quality. Some fields have extensive literature on these topics, such as survey data with documentation on theory and practice in communicating survey error, but other data sources have less. All data sources have error, and it is important to communicate what researchers can do with the data without overextending it and outline how it can be useful to end users.
- Transparency for data processing procedures – Some datasets include multiple rounds of calculations and data tweaks which can lead to discrepancies among datasets. For example, there are instances where bridge locations do not line up with road and rail intersections when overlaying datasets from Federal Highway Administration (FHWA), Federal Rail Administration (FRA), and the National Bridge Inventory (NBI). Not all datasets clearly document data processing procedures. This documentation is particularly important for data being used to inform or develop policies and for data from private sources which are not as transparent as public sources.
- Types of data – There are several types of data to consider in the network. Data can be actively collected (such as through surveys) or passively collected (such as Global Positioning System (GPS) data). Data can be from public sources at the Federal, State, or local level, or from the

private sector being organized, random, or crowdsourced. Data can be collected on a continuous or episodic frequency.

National Transportation Data Preservation Network Functions
- As a note: the bullets below summarize discussion topics related to the network functions. This first meeting of the group served as a brainstorming session and the group was constrained by limited time. Some areas could benefit from additional discussion and research.
- Network or Community of Practice (CoP) scope –
  - *Data preservation* – Preserve final reports/products and raw datasets or data that ensures the project reproducibility and reuse.
  - *Datasets* – Focus on managing datasets that supported publications. Start with USDOT funded Public Access data and final products, and then figure out how to expand, first to other transportation datasets, and then to others. The group should start small first. To keep scope narrow, it might be wise to first focus on one particular area of transportation to make this first round more achievable. Expansions need to be defined and will help determine who should be involved in future meetings. If the scope is too broad, the network format will get much more complicated and larger. Larger datasets also have additional issues such as privacy, PII, disclosure agreements, third party data, etc.
  - *Tracking research impact and output* – Maximize and measure data sharing.
  - *Data accessibility and discoverability* – Data needs to be usable and citable. The citation of transportation researcher's and others' data should be encouraged and promoted. The value of reusing data needs to be demonstrated. Data needs to be consistent as well.
  - *Compliance plus usefulness* – Data should be FAIR and trustworthy, and repositories should be certified. Go beyond checking the legal requirements box, and ensure data is input into the network in the most useful way to then be available to users in the most useful way. This also helps ensure data is not misused.
  - *Actionable, achievable items*
- Questions to narrow focus – To better understand what is needed from the network, the group needs to consider the flow of information from the data sources to the data uses.
  - *Who has the data?* These are the nodes (USDOT, States, academia, private sector, etc.).
  - *What happens to the data?* This is what occurs between data holders/providers and data users. This can be indexing, creating a new repository, etc.
  - *Who uses the data?* This denotes access level (temporal, spatial, unit resolution, etc.). It is important to know who the end users are to better prepare data and understand what they want/need from the network. This is also important to make sure data is not misused.
- Levels of consideration for data management – There are two levels of issues to address:
  - Data to the repositories – How to prepare and package data so it is discoverable and most useful to end users?
  - Data in the repositories – How to manage and store the data for future use? This is what the group should focus on.
- Potential models – The goal is to create a single search engine for users to search metadata to find the available data of interest. Use a spoke-and-hub network to connect users to data/information. Ensure the model is scalable to be expanded for future use.
  - The National Science Foundation (NSF) uses a spoke-and-hub format.
  - TRID also uses a spoke-and-hub method. TRID focuses on research reports but also indexes some datasets occasionally and could be a good example for research metadata.

TRID has defined processes for resource management and partnering. TRID links out to journals and NTL datasets and does not hold or duplicate any data. This network will be different from TRID since TRID does not include datasets, only research products, and is only an index, not housing any information.

- o The TRB tool only includes active or currently closed out projects.
- o Science.gov is a good example to emulate. It is a federated search engine that has been set up and is maintained.
- o ScholarWorks and Research Hub were also discussed.
- o Additional research and discussion on potential models could be valuable.

- <u>Leverage existing infrastructure</u> – Capitalize on the fact that many users rely on search engines such as Google to find information, and expose datasets to these established search engines so data is findable. This would be simple for repositories and could be an easy first step to build into data preparation. Interfaces exist at low costs. Everything submitted to DOT is in Google dataset searches already. This is not a replacement strategy for a federated repository but a supplemental effort to make information more accessible.
  - o ROSA P has the highest number of users in the transportation community, then Google and TRID on a similar but smaller scale of users. This means ROSA P is a known, good resource that avoids noise from Google searches. Marketing and outreach efforts are important to ensure this repository is discoverable.

- <u>Examples from other disciplines</u> – The network infrastructure can be built based on other areas with demonstrated successes. This network will still need its own champions, perhaps repository managers or librarians, to advocate for the developed network and processes making data sharable and findable. Implementation of the network will be the difficult part. While other areas have spent considerable time and money developing their systems, potential good examples include:
  - o Medical field – Medicine has good examples with large datasets and a vast variety of topic areas and researchers. The scope and impact are on different levels than the transportation field, however, and medicine has a much higher management budget.
  - o Agriculture – Agriculture repositories could be good examples to build from since this field has good examples at a mix of different levels.
  - o NSF Council of Data Facilities, social sciences, and astronomy also have mature models that may be worth mining for ideas.

- <u>Develop use cases</u> – Have motivating use cases that clearly identify the value of the network for different users to encourage stakeholder participation and help develop the network. Use cases would help demonstrate the value of the network rather than it simply being a requirement, debunk myths, and demystify the processes. They would also help others better understand data management and develop common terminology and dialog. Use cases could include how the network would be utilized from different roles/perspectives including:
  - o Users utilizing a portal looking for data and how they would use different searches to find the data of interest.
  - o Users preparing and packaging data themselves to enter into the system/portal in a way that meets the sponsor's expectations while being accessible for other users.
  - o State DOTs implementing the Public Access Plan, demonstrating the value in making data more usable and accessible to allow new research and new data to be generated.
  - o Internal DOT staff using the network to provide data and proof of value of research efforts to justify a need to fund or provide resources for proposals.

- Media and legislators using the network to follow-up on stories or do additional research on a topic by providing a centralized data source.
- Explanations of the importance of the network from the perspective of funders, librarians, PIs, students looking for research opportunities, the media, the public, etc.

- Clarify the purpose/role of repositories – Researchers view their research as living datasets but repositories view research as snapshots supporting discrete research projects. Educate researchers on what repositories are, why they are important, and clarify they are not a platform for further live analyses.

Next Steps and Meeting Outcomes

- Engagement – This meeting brought together researchers and repository managers. Making these connections and providing different perspectives and contexts was a key meeting takeaway.
- Network focus – Develop a National Transportation Data Preservation network or CoP based on suggestions and discussions throughout the day. The network will have a tight scope on data that are produced by research funded by the DOT in accordance with its Public Access Plan. The main goals are to ensure data preservation, and measure and maximize the impact of data and data sharing. In tandem, initiate a culture change in the field and name champions for outreach and education. With the Administration interested in measuring the impact of research funding, now is a good time to establish metrics to demonstrate the value of funding efforts.
- Conduct a capabilities assessment – Engage repository managers who are responsible for entering data into the systems. It is important to understand the capabilities and interests of repositories to determine the target audience and gaps in network infrastructure to inform a plan moving forward.
  - UTCs get together at the Council of University Transportation Centers (CUTC) TRB meeting. Repository managers and UTC directors could attend to discuss their current capabilities and identify gaps in capabilities that could be facilitated by the network.
  - Survey universities (perhaps the UTC managers or DOT funded universities to start) to determine which are using repositories and what kind of repository they are using. Distributing a survey prior to CUTC meeting would provide a broader reach and inform discussions in the in-person meeting by providing a background understanding.
    - The survey would benefit from engaging both the repository managers and UTC directors.
    - The group needs to be cognizant of the Paperwork Reduction Act if distributing surveys. Channeling the request through CUTC could alleviate these concerns.
    - Some participants did not like the idea of a survey prior to the meeting, preferring in-person discussions that provide more detail and context.
  - **Leighton will identify universities conducting DOT funded research, locate the repository managers, and lead this effort.**
  - **Leighton will look into securing funding for another invitational meeting like this one to bring UTC managers to TRB to interact with other stakeholders.**
  - **Michael Witt will assist in these efforts.**
- Public Access Plan report updates – The Public Access Plan report is three years old and needs to be reviewed and updated. **Leighton and Charles will lead this review**.
- Identify partners – This network is more of a CoP comprised of several partners. Potential partners for phase 2 of these efforts include:
  - AASHTO COD Committee on Data Management and Analytics

- o AASHTO RAC Research Advisory Committee (Shashi can help coordinate)
- o State DOTs
- o Inform Leighton of any other potential contacts
- <u>Send feedback</u> – If anyone has any feedback on statistical produces or use of State data, let BTS know but sending an email or going through Leighton. This sort of feedback lets BTS know what they're doing well and what they need to work on.
- <u>Next meetings</u> – The group will reconvene virtually late-May or early-June 2019. **Leighton will distribute a Doodle poll for availabilities.**

Appendix – Email forwarded by Shashi about ASCE and data archiving

[Data Availability Statement Policy](#)

ASCE is taking steps to improve the availability and reproducibility of work published in its journals. ASCE is introducing a new policy requiring authors to specify the availability of data, computational models, code, and other electronic materials used in work submitted for publication. The new policy is intended to (1) make it easier for readers and others to use, reproduce, and extend published work; (2) increase the quality of submissions; and (3) further encourage people to use work published in the ASCE Journals and cite that work. This new policy is a step forward to allow readers, authors, and our field to practice more open and reproducible science (Rosenberg, D.E.., D. W. Watkins 2018) (Govindaraju, R.S., M Hantush, X. Chu 2019).

When submitting a new manuscript, authors will include a new section titled "Data Availability Statement" before the "Acknowledgments." Within this section, authors will select one or more of the following statements verbatim, with the inclusion of all citations to data, code, or models.

- Some or all data, models, or code generated or used during the study are available in a repository or online in accordance with funder data retention policies (Provide full citations that include URLs or DOIs.)
- Some or all data, models, or code used during the study were provided by a third party. (List items). Direct requests for these materials may be made to the provider as indicated in the Acknowledgments.
- Some or all data, models, or code generated or used during the study are available from the corresponding author by request. (List items).
- Some or all data, models, or code generated or used during the study are proprietary or confidential in nature and may only be provided with restrictions (e.g. anonymized data). (List items and restrictions).
- All data, models, and code generated or used during the study appear in the submitted article.
- No data, models, or code were generated or used during the study (e.g., opinion or dataless paper).

Available data and code must be cited in the reference section. Authors may update or change their statement if information changes during the course of peer review. It may not be changed after the paper is accepted.

References

Rosenberg, David E., and David W. Watkins 2018, "New Policy to Specify Availability of Data, Models, and Code", J. Water Resour. Plann. Manage., 2018, 144(9): 01618001, https://doi.org/10.1061/(ASCE)WR.1943-5452.0000998.

Govindaraju, Rao S., Mohamed Hantush; and Xuefeng Chu, "New Policy for Transparency of Data, Models, and Code", J. Hydrol. Eng., 2019, 24(3): 01618001, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001785