Data Management Plan for
Pacific Northwest Transportation Consortium (PacTrans)

Lead:
University of Washington, Seattle, Washington 98195

Partners:
Northwest Indian College, Bellingham, Washington 98226
Portland State University, Portland, Oregon 97201
University of Alaska, Anchorage, Alaska 99508
University of Idaho, Moscow, Idaho 83844
Washington State University, Pullman, Washington 99164

Northwest Indian College is a minority serving Tribal college

Submitted by

Yinhai Wang, Ph.D., P.E., Dist.M.ASCE, F.IEEE, F.ITE
Department of Civil and Environmental Engineering
University of Washington
Seattle, WA 98195-2700

Tel: 206-616-2696
Fax: 206-543-1543
Email: yinhai@uw.edu

August 9, 2023

PacTrans Data Management Plan

Pacific Northwest Transportation Consortium (PacTrans) is the U.S. Department of Transportation (USDOT) University Transportation Center (UTC) for Federal Region 10. Led by the University of Washington (UW), PacTrans includes five consortium universities in addition to the UW: Northwest Indian College (NWIC), Portland State University (PSU), University of Alaska Anchorage (UAA), University of Idaho (UI), and Washington State University (WSU). Region 10 is characterized by a rapidly growing population with heavy freight movements through rural and urban corridors. The area is prone to earthquakes, tsunamis, wildfires, landslides, and sea level rising, all of which greatly impact the region's transportation network. Therefore, PacTrans will undertake the mission of ***developing human-centered and transformative multimodal mobility solutions for an equitable Pacific Northwest*** through research, education, technology transfer, and workforce development efforts in collaboration with university, transportation agency, and industry partners.

Specifically, PacTrans research will center on: (1) human-system integration; (2) equity and accessibility; (3) safety, reliability, and resiliency, and (4) multimodal connectivity. PacTrans will also emphasize data-driven solutions as a cross-cutting theme, making heavy use of advanced sensing, communications, and data analytics tools. Clearly, data play a critical role in PacTrans activities and will be managed carefully from the beginning of this center grant.

Types of Data Anticipated

Given the mobility focus of PacTrans, PacTrans funded research and other activities will gather data broadly from relevant areas, including infrastructure, user, vehicle, environment, etc. Additionally, there are also software tool or model generated data to be collected and applied in research and educational activities. Below are examples of the anticipated types of data:

1. Infrastructure and control data
2. Road user data
3. Vehicle data
4. Survey data
5. Environment data
6. Software and model-generated data
7. Survey data
8. Guidelines, examples, handouts, brochures, and posters
9. Lesson plans, syllabi and/or course materials
10. Other data

PacTrans also funds an education project. For this project, data collected or generated may include, but are not limited to:

1. Educational materials to support student research, syllabi and course modules in PacTrans-related topics, K-12 outreach, educational software, recordings of seminars and other professional events, etc.
2. Continuing education, lecture notes, and other curriculum or training materials

3. Performance metrics of PacTrans education, such as: number of attendees in K-12 programs; transportation related courses offered by faculty affiliated with PacTrans, number of undergraduate and graduate students participating in PacTrans activities; number of transportation related degree and non-degree programs
4. Performance metrics of PacTrans diversity, such as number of female, minority, or disabled students in PacTrans research and demographics of PacTrans event attendees
5. Performance metrics of PacTrans workforce development, such as number of attendees at training sessions, webinars, career-building activities, or online courses and hours of technical assistance offered to agencies through project meetings and training workshops

All datasets directly used to support important conclusions of PacTrans research products will be preserved for long-term access. These datasets are required deliverables to PacTrans project PIs. Data-specific restrictions for release, if any, should be clearly documented and submitted as part of the proposal to PacTrans.

Proposed Standards and Machine-Readable Formats

The collected data may be in or converted into electronic form. This can be in various formats, ranging from records/tables in relational database, to text files, to Excel spreadsheets or pdf files, among others. For each project, the research team will describe the anticipated formats that the data and related files will be used in sufficient detail in their pre-project, project level data management plan. For guidance on project level data management plans, refer to the US DOT Public Access Guidance webpages, especially "Creating Data Management Plans for Extramural Researchers" at https://doi.org/10.21949/1520571. To make the data sharing easy and convenient, PacTrans will ensure each project team uses platform-independent and non-proprietary formats to ensure maximum utility of the data. If this is impossible, the specific data standards and formats employed should be clearly documented, together with the rationale for using those standards and formats. A metadata file will also be produced to explain in detail the format of each primary source data. When developing the metadata, guideline in Project Open Data Metadata Schema (DCAT-US Scheme v 1.1) will be followed as much as possible. When preparing the data format document, the following will be considered:

1. List the format(s) of data to be collected and indicate if they are open or proprietary.
2. If proprietary data formats are used, discuss the rationale for using those standards and formats.
3. Describe how versions of data will be signified and/or controlled.
4. If the file format used is not standard to the area of research, describe the specific format that will be used for the project.
5. List the documentation that will be created to help other users understand the datasets.
6. Indicate what metadata schema will be used to describe the data. If the metadata schema is not one standard, discuss the rationale for using that schema.
7. Describe how will the metadata be managed and stored.
8. Indicate what tools or software is required to read or view the data.
9. Describe the quality control measures.

Policies for Sharing

In general, data from research projects funded wholly or in part by PacTrans must be made publicly accessible. Exceptions to this policy are data that contain personally identifiable information, confidential business information, classified information, or data that was obtain through an agreement that it would not be made public. In these cases, notes are needed to explain why the entire or part of the datasets cannot be public accessible. Typically, the level of sharing depends on the nature of the data. For example, software tools that implement the model/algorithms of research should be shared after the intellectual property issue is properly addressed. Simulation data and other types of model-generated data can be shared without any restrictions. Infrastructure and control data can be shared upon obtaining approval from transportation management agencies who manage the infrastructure (such as city or state Departments of Transportation). Road user and vehicle data can be shared after removing personal identifiable information. Environment data can be shared upon the approval of the data owner. Other types of data may also be shared at appropriate levels depending on the way of data collection, content of the data, and their actual formats.

Protecting research participants and guarding against the disclosure of identities and/or confidential business information is an essential norm in scientific research. If needed, proper documents will be prepared to address these issues and outline: (1) the efforts that will be taken to provide informed consent statements to participants, (2) the steps that will be taken to protect privacy and confidentiality prior to archiving the data, and (3) any additional concerns (e.g., embargo periods for the data).

In cases where it is impossible to deidentify the data in a manner that protects privacy and confidentiality while maintaining the utility of the dataset, the necessary restrictions on access and use should be clearly stated. In matters of human subject research, the informed consent forms should describe how the collected data will be shared with the research community and whether additional steps, such as an Institutional Review Board (IRB), may be used to protect privacy and confidentiality. Any research involving First Nations or tribal populations or members should adhere to the [CARE Principles of Indigenous Data Governance](#).

Policies for Re-use, Re-distribution, and Production of Derivatives

For a dataset generated by PacTrans-funded projects and/or activities, re-use of the dataset is allowed once the dataset is officially published with a digital object identifier (DOI) or other type of persistent identifier issued by a repository. Re-distribution of the data is generally not allowed. If, however, a user sees a need to distribute a dataset, a written request must be submitted to the research team that produced the data, as well as the PacTrans Board of Directors. These requests will be considered and decisions on a case-by-case basis. Production of derivatives based on the published datasets are generally allowed, but production of the derivatives based on the software packages (i.e., new development based on the source codes) is generally not allowed in order to protect the intellectual property (IP) of the project team. Again, PacTrans is willing to consider special requests on a case-by-case basis for production of derivatives based on software packages in order to maximize use of PacTrans research products while protecting PacTrans researchers' IPs.

Data Storage and Sharing

PacTrans will continue to leverage the data storage and sharing system that was established during the FAST Act grant. The uploaded data will be stored and maintained for its useful life. At the end of each project, the final datasets that directly supported the research products and their supporting documents will be uploaded to the following two places and links will be provided via the project pages on the PacTrans website, and submitted to National Transportation Library (NTL), Transportation Research International Documentation (TRiD), etc.

1. PacTrans collection on the Harvard Dataverse; and
2. UW Shared Central File System.

<u>PacTrans collection on the Harvard Dataverse</u>

PacTrans established a collection in the Harvard Dataverse at the beginning of the FAST Act grant. At the conclusion of each project, the metadata and datasets for each respective project are uploaded to a project level entry in this collection. When the data size exceeds the limits of a single entry in the Dataverse, as much data as possible is uploaded and note is added indicating that the rest of the data is available by contacting PacTrans staff. When the data cannot be made publicly available for any of the reasons previously discussed, an entry is still created that contains all of the metadata but a note is provided explaining why the data cannot be made publicly available.

<u>UW Shared Central File System</u>

At the beginning of the FAST Act grant, PacTrans also created a UW Google Shared Drive. This folder contains a summary sheet of all the projects and related data that are stored in the Drive, and each project has a subfolder where project level data management plans, metadata worksheets, and the final datasets are stored. This system serves as a backup to the Dataverse entries and also allows staff to store data that is larger than what is allowed on the Dataverse collection.

This repository ensures discoverability, and supports the capture and provision of the US Federal Government [Project Open Data Metadata Schema](#) (DCAT-US Scheme v 1.1). In addition, the repository supports the creation and maintenance of persistent identifiers (e.g., DOIs, handles, etc.) and provides for maintenance of those identifiers throughout the preservation lifecycle of the data. In particular, the repository considers the following:

1. Provide a URL for each PacTrans-funded project as the access point for data sharing.
2. Indicate the approximate time period between data collection and submission to the archive.
3. Identify where data will be stored prior to being sent to an archive.
4. Indicate how back-up, disaster recovery, off-site data storage, and other redundant storage strategies will be used to ensure the data's security and integrity.
5. Implement procedures to protect data from accidental loss or malicious modification or deletion prior to receipt by the archive.
6. Identify and implement back-up, disaster recovery, off-site data storage, and other redundant storage strategies and solutions to ensure the data's security and integrity for the long-term.
7. Specify the duration for the chosen archive to retain the data.
8. Indicate if the chosen archive employs, or allows for the recording of, persistent identifiers linked to the data.

9. Ensure the chosen data repository meets the criteria outlined on the [Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan](#) page.

Policies for IP, Copyright, and License protection

The project PIs and their institutions own the research data they generate. Intellectual Property (IP) developed includes, but is not limited to inventions, patent applications, patents, software, or other legally protectable information, will be owned by the participating party or parties whose employees or agents make or generate the IP. All rights to IP will be maintained by the parties who are responsible for its creation and will be subject to each organization's policies of rights assignation. Therefore, ownership may reside with a single entity or multiple parties. In the event that program IP is jointly held, each contributing organization will have equal rights to the technology, unless otherwise agreed upon and such an agreement is memorialized in writing. Background IP, program IP and other proprietary or confidential information disclosed by any participant to another will be treated as confidential under provisions of the agreements to which each is a party. The IP office of each PacTrans member university will assess patentability and commercial potential and initiate pursuit of patent or other legal protection of their PacTrans-sponsored research when appropriate.

Further researchers are reminded:

1. Data, as a collection of facts, cannot be copyrighted under US copyright law.
2. PacTrans research carried out under a US DOT University Transportation Centers (UTC) program grant is federally funded. As stated in grant language and referenced in the University Transportation Centers (UTC) Grant Deliverables and Reporting Requirements: For Grants Awarded in 2023 Funded by the Bipartisan Infrastructure Law (BIL):
    a. Researchers must comply with the US DOT Public Access Plan, meaning, among other requirements, research data must be shared with the public, either by the researchers or by US DOT;
    b. That by accepting US DOT funding through this grant, researchers have granted to US DOT a comprehensive non-exclusive, paid-up, royalty-free copyright license for all research outputs (publications, datasets, software, code, etc.). This includes all rights under copyright, including, but not limited to the rights to copy, distribute, prepare derivative works, and the right to display and/or perform a work in public; and,
    c. In accordance with Chapter 18 of Title 35 of the United States Code, also known as the Bayh-Dole Act, where PacTrans elects to retain title to any invention developed under this UTC grant, US DOT retains a statutory nonexclusive, nontransferrable, irrevocable, paid-up license to practice or have practiced for or on behalf of the United States any such invention throughout the world.

None of the general IP provisions above negate US DOT's non-exclusive rights nor PacTrans' obligations.