



## **Data Management Plan for Pacific Northwest Transportation Consortium (PacTrans)**

Lead:

University of Washington, Seattle, Washington 98195

Partners:

Boise State University, Boise, Idaho 83725  
Gonzaga University, Spokane, Washington 99258  
Oregon State University, Corvallis, Oregon 97331  
University of Alaska, Fairbanks, Alaska 99775  
University of Idaho, Moscow, Idaho 83844  
Washington State University, Pullman, Washington 99164

*University of Alaska Fairbanks - Interior-Aleutians Campus is a minority serving institution for  
Alaska Native and Native Hawaiian*

Submitted by

Yinhai Wang, Ph.D.  
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

February 19, 2017

## **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 four primary consortium universities in addition to the UW: Oregon State University (OSU), University of Alaska Fairbanks (UAF), University of Idaho (UI), and Washington State University (WSU). Boise State University (BSU) and Gonzaga University (GU) join PacTrans activities through their state representative universities, i.e., UI, and UW, respectively. Region 10 is characterized by a rapidly growing population, heavy freight movements through urban and rural corridors, topographical constraints imposed by mountains and water, and a lack of transportation infrastructure redundancy. Therefore, PacTrans will undertake the mission of *providing data-driven solutions for the diverse mobility challenges of the 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 promoting accessibility for all, increasing system-wide efficiency, and ensuring reliability across all modes. PacTrans will also emphasize data-driven solutions as a cross-cutting theme, making heavy use of advanced sensing, communications, and data analytics tools in support of its efforts on accessibility, efficiency, and reliability. 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. Environment Data
5. Software tools and model-generated data
6. Other Data

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 project team should describe the anticipated formats that the data and related files will be used in sufficient detail. To make the data sharing easy and convenient, PacTrans will ensure each project team to use 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 should be produced to explain in detail the format of each primary source data. When developing the metadata, guideline in [Project Open Data Metadata Schema](#) 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, or classified information. 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.

At the same time, 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 the efforts that will be taken to provide informed consent statements to participants, the steps that will be taken to protect

privacy and confidentiality prior to archiving the data, and any additional concerns (e.g., embargo periods for the data).

In case if 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.

### **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). Re-distribution of the data is generally not allowed. If, however, a user does see the need for re-distributing a dataset, a written request is needed for the PacTrans Board of Directors and the project team that produced the data to consider the request and make a decision 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 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 create a central data repository and associated data uploading, archival, and management system for PIs to upload and manage their project data, as well as for users to access the datasets. The uploaded data will be stored and maintained for its useful life. During the course of a project, the PIs can choose to store project data in their private computing facilities or at the PacTrans central data repository. However, at the end of the project, the final datasets directly used to support the research products and their supporting documents must be uploaded to the PacTrans central repository before the project can be considered complete.

When developing the central data repository, metadata will be used to ensure its discoverability. The repository will also support the capture and provision of the US Federal Government [Project Open Data Metadata Schema](#). In addition, the repository will support the creation and maintenance of persistent identifiers (e.g., DOIs, handles, etc.) and will provide for maintenance of those identifiers throughout the preservation lifecycle of the data. In particular, the central repository will consider 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.

To ensure the reliability of the data management system, all PacTrans archived data will also be stored in the UW Shared Central File System for Research Collaboration. This is a service sponsored by the UW and should be available even beyond the active period of the FAST Act UTC grant.