



Data Management Procedure

Cole Kopca

May 31, 2019

Presented to PacTrans Board of Directors



Origins and Language

Plan to Increase Public Access to the Results of Federally-Funded Scientific Research Results

Version 1.1



December 16, 2015

U.S. Department of Transportation

This plan to the extent feasible and consistent with applicable law and policy; agency mission; resource constraints; U.S. national, homeland and economic security; and the objectives listed below, require digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding to be stored and publically accessible for search, retrieval, and analysis. This plan requires that awardee(s) and/or the respective Operating Administration ensure Public Access to final research data, subject to the above restrictions and those imposed by data quality and the need to protect national/homeland security, individual privacy, and confidentiality.



Items to Consider

- ▶ Final dataset used for analysis and conclusions of PacTrans funded research.
- ▶ If there are issues with sharing your data, we still need to have a record of the meta data but you can specify in the meta data worksheet why the data cannot be made public.
- ▶ If you have questions email pactrans@uw.edu



Before Projects Begin

Review Center Wide Plan

Create Project Level Plan



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



Pacific Northwest Transportation Consortium

Data Management Plan – Project Level

Lead PI Name: _____

Project Title: _____

The federal government is now requiring that all research projects funded with federal dollars, make data available to the public. USDOT states, "...process and archive Digital Data Sets resulting from DOT-funded scientific research in a repository that enables and allows for public access and sharing." However, explicit direction on what portions or iterations of data to be made available remains vague, so our stance is that each PI make an honest effort to provide the data someone would need to repeat your methods, without making data available that is proprietary in nature, that non disclosures have been signed for, or that has privacy issues.

It is important that you refer to the PacTrans Data Management Plan guideline when developing the project-level data management plan, to understand how we will archive and make available, your data. Explicit instruction regarding required supplemental materials (such as metadata) and the process with which you will submit your data will be distributed at a later date.

Please respond to the following questions the best you can at this point (one short paragraph per questions will do).

What data will your project utilize or produce (please include file types and estimated size)?

Based on the description above, which of that data should be made publically available? Which of the data sets cannot be shared to the public? Explain why.

How long will this data remain relevant for and how so how long should be make it available for?

Is there any risk that providing public access to this data could expose you, the PI, or PacTrans to liability issues? If so, please explain.



After Projects Ends

Submit Dataset to PacTrans



Via email, google drive, dropbox, mailed hard drive, etc.

Submit Metadata to PacTrans

PacTrans-Funded Research Project Final Dataset Metadata Worksheet

Please include the following information with the project data submission and any additional details that you think might be useful to anyone searching for or using your data. Remember that, by providing rich supporting documentation for your dataset, you make it easier for other researchers to find, use, and cite your work.

Data files: All data and metadata files should be submitted in a single folder or zip file. Note that there is a 2GB file size limit in the Harvard Dataverse, so any individual data files exceeding this limit should be broken down into smaller files if the data is to be shared publically.

Title: The dataset title should be descriptive of the dataset contents, the project title will not be sufficient in general. For example, for a dataset which contains bicycle volumes collected by volunteers at various intersections in Spokane, WA, an appropriate title might be "Intersection Bicycle Counts in Spokane, WA".

Author: The Author should be the project PI(s), include the ORCID for all authors.

Contact: The contact for the dataset should be the project PI(s). Please include name, affiliation, and email address.

Description: Please describe the dataset and its intended use. Include a brief description of each data file that is submitted, including any metadata files. Details should include where and how the data was collected, processing steps that have been applied, and any known quality or consistency issues. For more complex datasets, consider including a schema describing the structure and relationships present in the dataset.

Date: This is the date at which the Dataset was created. If the data was collected over multiple days, simply enter the start and end dates. If the date(s) at which the data was collected do not coincide with the time period described by the data, please include both the collection date(s) and the date(s) covered by the data.

Period of Relevancy: How long will this data remain relevant for and how so how long should we make it available for?

Subjects: Consider the following subject areas and note all that apply, include additional domain-specific keywords as needed to help people search for your data:

- Physics
- Agricultural Sciences
- Arts and Humanities
- Astronomy and Astrophysics
- Business and Management
- Chemistry
- Computer and Information Science
- Earth and Environmental Sciences
- Engineering



Harvard Dataverse

Each project's dataset will have an entry, created by PacTrans staff, in the Harvard Dataverse repository that includes metadata about the set regardless of data size and sharing concerns.

A screenshot of the Harvard Dataverse 'New Dataset' form. The page header includes the Harvard Dataverse logo and navigation links: Search, About, User Guide, Support, and Cole Kopca. The main content area shows the 'PacTrans Dataverse' (University of Washington) logo and a status of 'Unpublished'. A breadcrumb trail reads 'Harvard Dataverse > PacTrans Dataverse > New Dataset'. Below this, there are tabs for 'Host Dataverse' and 'PacTrans Dataverse'. A note states '*Asterisks indicate required fields'. The 'Citation Metadata' section is expanded, showing fields for Title, Author, Contact, and Description. The Title field contains 'Enter title...' and a button to 'Add "Replication Data for" to Title'. The Author field has 'Name' (Kopca, Cole) and 'Affiliation' (University of Washington) sub-fields. The Contact field has 'Name' (Kopca, Cole), 'Affiliation' (University of Washington), and 'E-mail' (ckopca@uw.edu) sub-fields. The Description field has a note 'This field supports only certain HTML tags.' and a 'Text' sub-field. Plus signs are visible next to the Author, Contact, and Description sections, indicating they can be added to a list.



Harvard Dataverse

Datasets Under 2GB



Data will be made available directly on Harvard Dataverse.

Datasets Over 2GB



Data is too large to store on Harvard Dataverse so will only be kept on PacTrans google team drive.



Linked on Project Profile Page

Harvard Dataverse dataset entries will be linked to PacTrans research project profile pages. If the data is only housed on PacTrans Google Team Drive, there will be a note for people interested in the data to email us directly for access.

The screenshot shows the PacTrans website interface. At the top left is the PacTrans Region 10 logo. To its right is a search bar with the text "Search..." and a "GO" button. Below the logo is a navigation menu with links for "ABOUT PACTRANS", "RESEARCH", "EDUCATION & WORKFORCE DEVELOPMENT", "PARTNERSHIPS & TECH TRANSFER", "PUBLICATIONS", and "NEWS & EVENTS". The main content area is titled "Washington State School Walk Score". It includes a list of project years from 2012-2013 to 2018-2020. The "Small Projects" section is expanded to show "Year 6 (2017-2019)". The project details include: PI: Anne Vernez-Moudon (UW), moudon@uw.edu, ORCID: 0000-0001-8485-4799; Co PIs: none; AMOUNT & MATCH: \$40,000 from PacTrans; \$40,000 Salary Match; PERFORMANCE PERIOD: 8/16/2017 - 8/15/2019; STATUS: Active; CATEGORIES: Safe Routes to School; RESEARCH PROJECT HOT SHEET: Click here; UTC PROJECT DOCUMENTATION: Click here; FINAL PROJECT REPORT: will be available once completed; PROJECT DATA: will be available once completed. A red arrow points to the "PROJECT DATA" text. A "DID YOU KNOW" box on the right contains information about commuting to work in Idaho. Below it is a "MEET THE TEAM" section featuring a photo of Melanie Paredes.