
Pacific Southwest Region 9 University Transportation Center Data Management Plan

April 2017

Creator
Genevieve Giuliano



Commented [LLC1]: Since this is a UTC-level DMP, it might be useful to include an Introduction before the “Data Description” section that:

- 1.Describes the purpose or mission of PSR9 UTC;
- 2.State that this DMP has been written to comply with US DOT funding requirements in the “Plan to Increase Public Access to the Results of Federally-Funded Scientific Research”;
3. Informs PIs they are expected to follow the guidance and rules laid out in the DMP;
4. Informs PIs of the need to create Project level DMPs to fill in project data details that are not included in this general UTC DMP, and they can follow the guidance at https://ntl.bts.gov/publicaccess/creatingaDMP_extramural.html for help;
 - a.Language might be: “each research project completed under this DMP will include a short DMP recording project specific information, such as data type, file formats, PIs and other data managers, etc., as needed.”
5. States the contact person or office at PSR9 to be contacted for questions about this DMP or about writing project-level DMPs

Commented [SH2R1]: Resolved, see revised Disclaimer section and PSR section

Disclaimer

This Data Management Plan has been written to comply with U.S. Department of Transportation funding requirements in the “Plan to Increase Public Access to the Results of Federally-Funded Scientific Research”.

All Pacific Southwest Region University Transportation Center principle investigators funded by USDOT are expected to follow the guidance and rules laid out in this DMP. Detailed instructions will be provided upon the receipt of a research award, and the PI will be required to acknowledge compliance with the DMP requirements.

For any questions regarding this document or how to comply, please contact Victoria Deguzman at (213) 821-1025, or via email at victoriv@price.usc.edu.

Pacific Southwest University Transportation Center

The Pacific Southwest Region University Transportation Center (PSR UTC) conducts an integrated, multidisciplinary program of research, education and technology transfer aimed at improving the mobility of people and goods throughout the region. Our program is organized around four themes: 1) technology to address transportation problems and improve mobility; 2) improving mobility for vulnerable populations; 3) Improving resilience and protecting the environment; and 4) managing mobility in high growth areas.

Region 9 includes four diverse states as well as the Pacific Island territories. The region is home to eight metropolitan areas in excess of one million in population (including the nation's second largest), four of the seven most visited U.S. cities on the globe, four of the nation's 10 busiest airports, the nation's largest port complex, and the largest high tech region on the planet. At the same time the region is home to vast, sparsely settled desert regions, and some of the most remote pacific islands. The region has the nation's highest proportion of non-native born populations, large concentrations of native American and native island populations, and concentrations of extreme poverty and disadvantage.

The PSR UTC supports research that addresses the unique combination of problems in the region, serves as a clearinghouse for curriculum to improve transportation education programs, addresses workforce development through targeted training, community college partnerships and professional development, and conducts a comprehensive, multimedia dissemination program.

The PSR UTC is structured to have a truly regional focus and is intended to build a strong university-government-industry partnership. Our consortium of universities and community colleges, together with partnerships with state Departments of Transportation (DOTs) and Metropolitan Planning Organizations (MPOs), forms a region-wide network to guide the development and implementation of the center's research, education, and technology transfer programs.

Data Management Plan

Data description

Transportation data often becomes inaccessible over time for the following reasons: 1) data has a format no longer supported by current software; 2) there is no information on the source or structure of the data; 3) data resides with the researcher and is not available to others. Historical data can be very valuable. Data may be used in new ways, or to answer new questions. Different datasets may be merged or used for comparative analysis. With this Data Management Plan, our research data will be able to accumulate over time and be accessible to future researchers.

As part of compliance with the PSR UTC Data Management Plan (DMP), PIs will write data descriptions for funded projects. This should distinguish between newly collected data and data being re-used from other projects as well as the actual observations and generated data to be submitted to a data repository. The PI is responsible for compliance with the DMP.

Some but not all types of data that PSR UTC gather are listed in below:

Travel surveys: On-line or on-paper surveys administered to random samples of individuals or households as to daily travel patterns, attitudes and perceptions, and demographic characteristics. Datasets include individual and/or household records with names, addresses, and all other personal identifiers redacted.

Professional surveys: On-line or on-paper surveys administered to random samples or purposive samples of individuals acting in their professional capacity and/or as representatives of their organizations. Datasets include individual or organization records with individual names, addresses, and all other personal identifiers redacted.

Interviews: Structure, semi-structured, or open-ended interviews conducted by phone or in-person with human subjects and/or key informants on topics relating to travel behavior, transportation policy, transportation practice, or similar topics. Datasets include transcriptions of interviews with name, addresses, and all other personal identifiers redacted. Any video and audio recordings will be granted an exception to data sharing requirements owing to privacy concerns and requirements of Institutional Review Boards.

Video observations: Video recordings of the operation of transportation facilities, such as streets, roundabouts, bike trails, and transit stations, including the movement of people and vehicles through these facilities. Datasets include video files as well as manual or automated coding of the video.

Vehicle/person location data: Time-stamped vehicle or person location collected by mobile device or GPS loggers. High-resolution data will be granted an exception to data sharing requirements owing to privacy concerns and requirements of Institutional Review Boards. Datasets summarizing the high-resolution data will be made available.

Traffic volume data: Proprietary datasets collected using cell phone tracking and Bluetooth signals, licensed from third parties. Datasets summarizing the licensed data will be made available.

Vehicle activity data: Time-resolved data on vehicle trajectory (speed, attitude, and position of vehicles) and other parameters such as energy use and emissions. Includes data collected through instrumentation of test vehicles and data produced through simulation models.

Infrastructure data: Data points that map out the location of infrastructures such as warehouses, gas stations, etc. Highway, railway and public transportation routes are also included.

Demographic data: Data on individual and household characteristics as well as vehicle ownership and

Commented [LLC3]: This section provides a good description of the types of data that will be generated and the methods. It does also raise potential privacy concerns and discusses some remediation.

What I feel is missing includes:

- 1.The potential value of the data over the long-term.
- 2.Indication of responsible parties.
 - a.For this you could say: "Project PIs or Project teams will be required to indicate who on the team is responsible for managing project-generated data." The PI would then include this in their project level DMP.
- 3.Description of how you will check adherence to the data management plan.
 - a. This could include a statement that someone in the UTC will review each project-level DMP, on some regular basis; or that the UTC will provide training on DMP requirements, etc.

Commented [SH4R3]: Section revised to address these comments

other travel-related characteristics acquired from state agencies and/or commercial vendors. Exceptions to data sharing requirements will be granted when access to such datasets is restricted and special permission is required for access.

Industry data: Data on business operations acquired from state agencies and/or commercial vendors. Exceptions to data sharing requirements will be granted when access to such datasets is restricted and special permission is required for access.

Modeling and simulation data: Input data and outputs for models such as CUBE, TRANSCad, or VISSIM; input data and output for simulation models created by PSR researchers.

Data format and metadata standards

Data from PSR projects will be stored in non-proprietary formats, such as txt, csv, mp3, dat, JPEG, etc. We will generally follow the format of the NREL TSDC server http://www.nrel.gov/transportation/secure_transportation_data.html and/or the FHWA Research Data Exchange <https://www.its-rde.net/>

In some cases, using proprietary data formats is unavoidable (e.g. shp or msd files); in such cases the rationale for using those standards and formats will be explained.

The Metadata schema PSR will be using is described on this website:

http://wiki.datadryad.org/Metadata_Profile. Dryad is a conformant data repository of the USDOT and the metadata schema they use is assumed to meet the requirements.

Policies for access and sharing

All projects involving human subjects will abide by the requirements of the Institutional Review Board (IRB) of the institutions of the Principal Investigators of the projects. These projects must have an IRB-approved protocol for ensuring informed consent of participants and protecting privacy and confidentiality. Data will be shared only after redaction of all individual identifiers, including names, residential addresses, geo-coordinates of residences, and email addresses.

Projects using proprietary data from commercial or public sources will abide by all conditions and requirements imposed on the use of the data. If the source organization prohibits the public sharing of the data, the project will be granted an exception from data sharing requirements.

Programming code developed by a project will be archived with the data if it is required to access the data.

The Principal Investigators of the project will be responsible for acquiring IRB approval and adhering to IRB and other data sharing requirements. The PIs must report IRB approvals and other data sharing requirements in their project proposals and progress reports.

Policies for re-use, redistribution, derivatives

Intellectual property rights will generally be held by the Principal Investigators of the projects and/or their home institutions. However, data transferred to the archive becomes part of the public domain. Copyrights may apply to data from some projects, such as those using copyrighted instruments or proprietary data sources. Any copyrights applying to the data will be identified. Projects using proprietary data from commercial or public sources will abide by all conditions and requirements imposed on the use of the data.

Materials generated under the project will be disseminated in accordance with University/Participating institutional and USDOT policies.

Commented [LLC5]: The statements on file types and proprietary formats are good.

Some things I feel are missing include:

1. The creation of contextual documentation, such as data dictionaries defining the variables; README.txt files giving rationale for the project and explaining methodologies; code books defining how data was analyzed; etc;
2. Mention of quality control measures, or expectation that PIs will include QC measure in project level DMP.

Commented [SH6R5]: Resolved, see "Processes to ensure project-level adherence" section

Commented [LLC7]: We suggest that the 2 places that say that Dryad is a "US DOT approved repository" should be reworded to say "Dryad policies are conformant to the requirements enumerated by the US DOT Public Access Policy."

- The DOT does not actually "approve" one data repository over another. We simply evaluate them to see if they are conformant to our guidelines and requirements.
- Further, while a repository may be conformant, if cannot be assumed that their metadata schema meets the needs of Project Open Data metadata schema: <https://project-open-data.cio.gov/v1.1/schema/>
- It will be important for you to make a comparison.

Commented [SH8R7]: We went to re-examine the POD of Dryad and found out the only issue we may have is USDOT's POD require bureauCode and programCode which we don't deem as necessary to us since we are a university institution. So we think Dryad do meet those requirements.

Please let us know if anything else needs to be addressed.

Research data which documents, supports and validates research findings will be made available after the main findings from the final research data set have been accepted for publication.

Plans for archiving and preservation

PSR will archive all data on Dryad, whose policies are conformant to the requirements enumerated by the US DOT Public Access Policy. Their policies are informed by the Open Archival Information System (OAIS) reference model (ISO 14721:2012) which defines "long term" to be a period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community. They have sufficient capability to ensure compliance with legal regulations, and maintains all applicable licenses covering data access and use, including, if applicable, mechanisms to protect privacy rights and maintain the confidentiality of respondents.

We have reviewed Dryad's policies for archiving and preservation < <http://datadryad.org/pages/policies> >. We feel they offer adequate services, such as detailed data curation instructions and security plans to ensure they conform with the US DOT Public Access Policy.

Processes to ensure project-level adherence

There are four stages that PSR UTC will perform to ensure that each project funded by US DOT complies with the DMP requirements proposed by US DOT:

Proposal stage: PIs commit to conforming to the DMP in the proposal

Award stage: PIs receive DMP compliance instructions as part of the award letter, and PIs must acknowledge the award letter and commit to all requirements in the letter by signature. Contents included in the DMP compliance instructions are: a) Data descriptions as indicated in the first section; b) Data formats and reasons for necessary proprietary formats, if applicable; c) Contextual documentation, such as data dictionaries defining the variables; README.txt files giving rationale for the project and explaining methodologies; code books defining how data was processed; d) Quality control measures; e) If applicable, explanations on why certain data sets cannot be shared.

Project completion stage: PIs submit data to PSR UTC director and PSR UTC will conduct an internal review to ensure compliance and then submit the data to Dryad within 60 days.

Commented [LLC9]: This section is perhaps a little thin.

It would be nice if this area included:

- Discuss your chosen data archive's policies and practices for back-up, disaster recovery, off-site data storage, and other redundant storage strategies to ensure the data's security and integrity for the long-term.
- Indicate how long the chosen archive will retain the data.
- Indicate if the chosen archive employs, or allows for the recording of, persistent identifiers linked to the data.
- Discuss how your chosen data repository meets the criteria outlined on the Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan page.

To answer the first point you might state: "We have reviewed Dryad's policies for back-up and disaster recovery << provide link to policy >> on this DATE. We feel they offer adequate protections, such as...(list a few)."

It is important to the DOT that we can be confident from your DMP that you have created these policies thoughtfully, and will follow through on managing the data as good stewards. The evidence for this is your DMP.

A UTC-level DMP can be a bit longer than 3 pages if you need more space give us that confidence.

Language to include might be: In project-level DMPs, PIs should describe how:

- Describe how data will be protected from accidental or malicious modification or deletion prior to receipt by the archive.
- Describe 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.

Commented [SH10R9]: Resolved; various sections revised