

# Tier-1 University Transportation Center on Telemobility

## Master Data Management Plan

Version 1.2

October 1, 2020

Approved October 1, 2020

### **Lead Institution**

Northwestern University

### **Partner Institutions**

University of California, Berkeley

University of Texas, Austin, McCombs School of Business

## Introduction

This document serves as the Tier-1 University Transportation Center on Telemobility's Data management Plan (DMP), describing our policies for handling data both during and after our research program. Tier-1 University Transportation Center on Telemobility [Telemobility UTC] is a USDOT funded research center composed of a consortium of universities: Northwestern University (lead), University of California, Berkeley, and University of Texas, Austin McCombs School of Business. The main research priority of Telemobility UTC is aimed at assessing how ICT technologies and e-commerce processes will impact the demand for travel and mobility in the future, taking into consideration lessons learned from this dramatic natural experiment that is the COVID-19 pandemic.

We view the DMP as a living document that will be updated as our project evolves. The DMP includes discussions of handling, storage, dissemination, and sharing of research results. OMB Circular A-110 defines research data as: "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples)." All of our efforts will comply with the U.S. Department of Transportation (DOT) Public Access Plan, and this DMP will be available via the NICR website ([www.cutr.usf.edu/nicr](http://www.cutr.usf.edu/nicr)). Telemobility UTC benefits from helpful Data Management Planning Resources organized by the Northwestern University Transportation Library. The Telemobility UTC Director and Executive Committee will be responsible for implementation of the DMP with input from all project principal investigators (PIs).

In coordinating research, education, and workforce training, the Telemobility UTC will ensure that data of all types will be managed and organized for security, consistency, and public dissemination when appropriate. This document details the general requirements for all activities funded by the Telemobility UTC, whether it's conducting research, educating students, training professionals, or providing outreach to connect industry, government and academia.

Telemobility UTC researchers will follow the guidelines and policies in this Master Data Management Plan (DMP). Researchers wanting more guidance on the U.S. DOT Public Access Plan can find guidance at <https://ntl.bts.gov/public-access>.

## 1. Data Description

The Telemobility UTC expects a variety of data to be gathered or generated depending on the types of research and outreach activities. We envision that the research projects will rely upon a wide range of quantitative and qualitative transportation data, including use of existing data sets, collection of new data, and simulationS. These efforts are divided into different activities:

Research projects – these are projects led by one of the PIs within the Telemobility UTC.

- Publications: working papers, white papers, research reports, research presentations, and open access articles
- Travel surveys, interviews, expert panels, and focus groups.
- Tolling data including rates, speeds, volumes and automatic vehicle identification (AVI) data from sensors.

- Databases and spreadsheets including metadata.
- Vehicle/traveler location data at various temporal resolutions, including Global Positioning Systems (GPS) and mobile phone data.
- Data describing transportation network/supply characteristics.
- Images including maps.
- Simulated data produced via simulation tools.
- Traveler activity and transport demand data including origins and destinations.
- Traffic count, occupancy and speed data at various temporal resolutions.
- Geographic information systems (GIS) spatial data.
- Demographic, socioeconomic, and census data.
- Transit vehicle, taxi, and transportation network company (TNC) fleet position, passenger ons/off, and transit schedule data.
- Source code for developing algorithms, simulations and other tools/products.
- Private data shared by agencies or private companies

Educational activities – these are long and short courses and research projects designed to disseminate knowledge from the Telemobility UTC’s research activities to students attending the partner institutions.

- Course notes and syllabi with consideration of involvement from different consortium institutions
- Summary of attendees and outcomes
- Project deliverables
- Video/webcasts of presentations, lectures

Outreach activities – these are workshops, webinars, invited speaker seminars, and advisor board meeting minutes.

- Videos or webinars of presentations, seminars, and or invited talks
- Conference proceedings
- Meeting minutes
- Contact lists for all involved active faculty, students, staff

All data collection efforts involving human subjects at NICR will follow each university’s Institutional Review Board (IRB) and Human Research Protection Program (HRPP) policies. Northwestern University policies are listed here: <https://www.irb.northwestern.edu/human-research-policies-guidance-oversight/>.

In limited cases, it is possible that data providers or project partners may place legal restrictions on openly sharing data. In these cases, principal investigators (PIs) will provide the details and rationale for lack of public access.

## 2. Standards Used

Telemobility’s policy is to use data in open, standard formats and to provide appropriate metadata and data dictionaries so that others may understand our data. A variety of file formats will be generated from Telemobility UTC research, which we will seek to migrate to a single stable format in accordance with Library of Congress standards<sup>2</sup>. Except for private data shared by agencies and private companies,

and certain videos or webcasts, all data produced from the Telemobility UTC's research will be made publicly accessible.

Code can be of any language. Test instances should be able to run specific input data to get expected output data. Test instances should allow researchers at a different consortium institution or partner to apply at their scale.

We expect that data will be collected and generated in a range of formats, including but not limited to:

- .docx, .xlsx, .pptx
- .xml
- .pdf/a
- .ascii
- .txt
- .csv
- .ascii
- .tiff
- .jpeg
- .png
- .avl
- .gps
- .mxd, .yr, .gdp, .shp, .dbf
- .html
- .php
- .mov
- .mpeg
- .mp4

As described above, any data used in Telemobility UTC projects will be converted to open access formats so that the public may easily access the data. Data tables and databases will include necessary metadata and/or data dictionaries so that it is understandable to future users.

PIs will use the following guidelines related to data acquisition or collection, as applicable:

1. Project open data formats. In limited cases, PIs will explain the rationale for use of proprietary data format and/or convert data to open format.
2. Data version control or indication.
3. Metadata schema or data dictionary to make the data understandable by other researchers, and management or storage of the same.
4. Are any tools or software required to read or view the data?
5. Data quality control measures.

In process research-project data will be hosted on a central server at Northwestern University Transportation Center for the duration of the study, while a permanent version will be added to the Arch data repository at Northwestern University following completion of the study (see "5. Archiving of Data"). PIs should refer to the U.S. Federal Government Project Open Data Metadata Schema, chosen by the U.S. DOT as the preferred schema in their Public Access Plan, to develop the metadata.

### **3. Access Policies**

Telemobility UTC's policy is to openly share data generated/used under our projects and programs funded by the U.S. DOT and our partners, subject to any confidentiality, privacy, security or human subjects restrictions. Telemobility UTC welcomes other researchers to pursue the principles of Reproducible Research whereby they may verify findings and build upon them. University campus IRB and HRPP policies will be followed and no private, confidential, or personally identifiable information (PII) will be revealed.

Telemobility UTC PIs will be responsible for protecting privacy and confidentiality according to their

campus IRB and HRPP policies. Each research-project PI will address the following information, as applicable:

- Identify data that will be publicly shared and how it will be accessed.
- Identify restrictions on data access.
- Follow campus IRB and HRPP policies, and methods for protecting PII, confidentiality and security.
- Identify additional steps beyond campus IRB and HRPP policies to protect privacy and confidentiality.

#### **4. Re-Use, Redistribution, and Derivative Products Policies**

Telemobility UTC's policy is to make project data available for open sharing under the Creative Commons Zero (CC0) universal public domain dedication. Under CC0, data and derivative products will be available for reuse and redistribution without restriction. In rare cases, as described above, PIs for individual research projects may face restrictions due to use of proprietary data, and this will be disclosed in individual project work plans.

#### **5. Archiving and Preservation Plans**

Telemobility UTC's policy is to archive and preserve project data, with appropriate disaster recovery procedures to ensure that data will remain openly available and recoverable. We intend to use "Arch," an open access repository for the research and scholarly output of Northwestern University, as the repository for Telemobility UTC project data: <https://arch.library.northwestern.edu/>. Arch provides a flexible set of metadata values for each data set deposit. Telemobility UTC will work with Northwestern librarian staff to develop internal documentation and workflows to ensure that our data deposits include all required elements from the U.S. Federal Government Project Open Data Metadata Schema.

Individual PIs at partner universities may propose to use institutional data archives such as the Texas Data Repository (TDR) at the University of Texas, Austin or the Merritt Repository at the University of California, Berkeley.

Data sets deposited into Arch are subject to level-one digital preservation, as outlined in Northwestern's digital preservation policy: <https://www.library.northwestern.edu/about/administration/policies/digital-preservation-policy.html>.

Arch will retain data files for as long as dependably possible (indefinitely). Telemobility UTC will follow Northwestern guidelines on recommended file formats for long-term preservation (e.g. TXT, CSV, ODS for data): <https://www.library.northwestern.edu/about/administration/policies/file-format-recommendations.html>.

Each data set in Arch is automatically assigned a Digital Object Identifier registered with DataCite: <https://datacite.org/>. And Arch is also indexed by the Registry of Research Data Repositories: <https://www.re3data.org/repository/r3d100012925>.

Prior to data archival, as previously stated in Section 2, Telemobility UTC project data will be stored on a central server at the Northwestern University Transportation Center. If the data storage requirements

exceed NUTC's server capacity, Telemobility UTC will use Northwestern's Research Data Storage Service, the University's central storage platform, as a secondary storage service. The Research Data Storage Service provides two storage zones based on data regulation guidelines: 1) Non-Audited: Appropriate for non-regulated research data.; and 2) Audited: Appropriate for the storage of most legally and contractually restricted data, such as Personal Health Information (PHI) and Personally Identifiable Information (PII), or must be in compliance with regulations such as HIPPA and FISMA.

Telemobility research collaborators inside or outside of Northwestern University will share data files and project documents under 15G using Box, as recommended by Northwestern University.

## REFERENCES

- [Guidelines for Evaluating Repositories for Conformance with the DOT Public Access Plan](#)
- [Plan to Increase Public Access to the Results of Federally Funded Scientific Research Results](#)
- [Creating Data Management Plans for Extramural Research](#)
- National Science Foundation [Dissemination and Sharing of Research Results](#)
- Office of Management and Budget. Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals and Other Non-Profit Organizations ([OMB Circular A-110](#))