

**Center for Teaching Old Models New Tricks (TOMNET)**  
**A US Department of Transportation Tier 1 University Transportation Center (UTC)**

---

**DATA MANAGEMENT PLAN (DMP)**

**Introduction**

This document constitutes a Data Management Plan (DMP) for the Center for Teaching Old Models New Tricks (TOMNET), a US Department of Transportation Tier 1 University Transportation Center (UTC) led by Arizona State University (ASU). Center partners include Georgia Institute of Technology, University of South Florida, and University of Washington. The mission of the center is to enhance models used for forecasting travel demand and predicting traveler behavior and mobility choices through the explicit incorporation of traveler attitudes, values, perceptions, and lifestyle preferences in such models. In order to achieve this goal, the center will undertake very significant and substantial data collection and data analysis activities. The center is committed to the protection of confidential data, following all protocols for the protection of human subjects and data privacy. At the same time, the center is committed to sharing data, products, publications, and presentations with the transportation community in an effort to enhance the state-of-the-art and the state-of-the-practice of transportation systems planning and modeling. Through the sharing of information, the center hopes to build a repository of resources that researchers and scholars will use for a period of time that extends well beyond the lifetime of the center. The following sections describe the protocols and procedures that will be followed at TOMNET in relation to data protection, data sharing, and data dissemination.

**Data and Products**

TOMNET researchers and scholars will be engaged in considerable data collection activities. In order to better understand the relationship between traveler behavior, mobility choices, and attitudes and values, TOMNET researchers will conduct very in-depth surveys including online surveys, intensive focus groups, and mail-out/mail-back paper surveys. These surveys will be conducted only upon receiving all necessary approvals from the Institutional Review Boards (IRB) at the respective institutions. Survey data will include, but not necessarily be limited to:

- Socio-economic and demographic data of household/persons
- Built environment data including land use and transportation network information
- Travel diary data including detailed information about activities and trips undertaken by individuals over the course of one or more days; activity-trip data would include:
  - Activity purpose
  - Mode of transportation
  - Composition of travel party
  - Origin and destination locations
  - Parking location and price
  - Transit access and egress
  - Transit wait time
  - Start and end times of activity or trip
  - Vehicle used
- Traveler attitudes and values data, which includes information about how individuals feel about various mobility options, built environments, travel experiences, and the environment
- Traveler preferences for different types of built environments, lifestyles, modes of transportation, vehicles and fuel types, and residential and work locations

Through focus groups, TOMNET researchers and scholars will gather additional open-ended qualitative data that will provide richer insights on how and why travelers make various choices, and attitudes-values-perceptions that play a role in shaping mobility choices.

In addition to collecting data through surveys, TOMNET researchers will also use various existing survey and census data sets. These include, for example, the National Household Travel Survey (NHTS) data set, other metropolitan household travel survey data sets, the American Time Use Survey, Consumer Expenditure Survey, and the American Community Survey (ACS) of the Census Bureau. Secondary data sets with information about land use, transportation network attributes, and public health may also be used to help unravel relationships among a multitude of dimensions that define urban spaces and traveler behavior and values. It is anticipated that TOMNET researchers will assemble a variety of integrated databases that fuse or merge information from different data sources. A major undertaking of TOMNET involves the fusion and integration of disparate databases with a view to provide a holistic basis for analyzing the role of attitudes, values, perceptions, and preferences in shaping mobility choices.

In the course of testing the performance of enhanced travel forecasting models in real-world agency settings, it is likely that TOMNET researchers will have to generate synthetic populations using census data sets as well as land use and socio-economic/demographic data sets maintained by metropolitan planning organizations and state departments of transportation. It is anticipated that a variety of synthetic population databases will be generated as part of TOMNET research activities.

The data collected and created/assembled as part of TOMNET research activities will have long-lasting value. Researchers and scholars around the world may be interested in analyzing the data and modeling a variety of relationships to better understand and predict traveler behavior and values. TOMNET believes that it would be of value to preserve the data for long-term access well beyond the life of the TOMNET UTC. The following sections describe the systems that would allow for such long term access.

From a broader perspective, TOMNET will generate a number of products that may also be treated as data for purposes of this data management plan. These products include the following:

- Publications including papers, articles, reports, brochures, policy briefs, survey forms, and newsletters
- Analytical, statistical, and computational models and algorithms
- Program codes, software systems, and geo-spatial visualizations of small and big data
- Descriptions of case studies and model tests documented in reports
- Educational materials including lesson plans, course notes, lecture videos, and course projects

All of these entities will be treated as “data” by the TOMNET team and appropriate protocols and procedures will be followed to store and disseminate the products.

### **Standards and Formats**

TOMNET researchers and scholars will use appropriate industry standards and machine readable formats for all products generated through TOMNET activities. All data sets will be stored and archived in accordance best practice. Databases will be made available and archived in widely used formats such as ASCII (text) and CSV formats. Other data formats will be explored to ensure wide and open access across a number of platforms and software environments. At a minimum, data documentation will be provided in PDF format, with researchers exploring the use of XML and YAML for creating metadata. TOMNET researches will create metadata (i.e., data about the data) in a manner that is easily understood by end

users and the wider community. The Data Documentation Initiative (DDI) Alliance ([www.ddialliance.org](http://www.ddialliance.org)) is an industry standard that will be consulted to ensure that metadata is prepared in a manner that facilitates documentation, discovery, and inter-operability.

The TOMNET team believes in the principles of open access and will generate products in formats that can be readily and easily accessed. All reports, documents, articles, and papers will be made available in PDF format. In addition, where appropriate, selected documents will be made available in HTML format so that individuals can access content even without a PDF reader. Models and algorithms will likely be coded in software specific scripts, although the team will strive to code models and algorithms in the open source statistical programming language R. By doing so, TOMNET will facilitate open access to models and algorithms to the extent possible. However, there may be some models and algorithms that are coded in other software systems, and appropriate documentation in PDF format will be provided so that users can learn how to use the models and algorithms in appropriate software platforms. Similarly, program codes and software systems will be coded in Python, C or C++, C#, or Java; these are standard coding languages and all documentation of the codes will be prepared and made available in PDF format. Educational materials may include presentations in both PowerPoint and PDF formats, videos on a YouTube channel, and other reference documents and lesson plans in MS Word, MS Excel, and PDF formats.

#### **Data Access Policies**

TOMNET researchers will exercise the necessary precautions and follow standard safeguards and IRB protocols to ensure that confidential data is protected and data privacy standards are met. Every survey data collection and data analysis effort will be subjected to IRB scrutiny at the respective institutions to ensure protection of human subjects. All survey data sets will be stripped of all personally identifiable information to create public use data sets that will not compromise the safety or privacy of subjects. These public use data sets will be posted and made available in ASCII and CSV formats (with appropriate metadata) through the TOMNET UTC website ([www.tomnet-utc.org](http://www.tomnet-utc.org)). Names, addresses, and trip origin and destination locations will be stripped from public use data sets that are made available to the community through the web portal.

The full data sets with personally identifiable information will be stored on a secure workstation at each institution with appropriate password protection and firewall security. These data sets will be used by TOMNET researchers and scholars for performing in-depth data analysis and modeling, but results will be published and disseminated so that no personally identifiable information is released. Only aggregate statistical summaries and models will be documented in reports and papers. The confidential data sets will be protected in this manner in perpetuity.

All other products of TOMNET will generally be made available to the broader research community through the center website ([www.tomnet-utc.org](http://www.tomnet-utc.org)). All papers, articles, presentations, policy briefs, and reports will be made freely available to the community through the center website. Similarly, educational materials will be made freely available and no restrictions will be placed on access. Codes, models, and algorithms will also be posted at the center website; these resources will be made available to the research community in the spirit of open access. All codes, models, and algorithms will be open source and distributed under the Apache 2.0 (<https://opensource.org/licenses/Apache-2.0>) or GNU General Public License Version 3 of the Open Source Initiative (<https://opensource.org/licenses/GPL-3.0>).

TOMNET will participate in ensuring that the products of the center are made available through widely used transportation libraries and repositories. The two repositories where TOMNET will make products

accessible are the National Transportation Library (NTL) (<http://ntl.bts.gov>) and the Transportation Research Information Services (TRIS) (<http://www.trb.org/InformationServices/InformationServices.aspx> and <https://trid.trb.org/>). TOMNET leadership will work with UTC program managers to ensure that products of TOMNET that can be accommodated in NTL and TRID are copied from the center website and made available through these repositories.

### **Reuse and Redistribution of Research Data**

The TOMNET team will make research data available to the broader community to foster research in traveler behavior and values, and build community of scholars dedicated to advancing the state-of-the-art in travel behavior research. As mentioned earlier, TOMNET will produce public use versions of data sets in which all personally identifiable information has been stripped away. These data sets will be made available to the research community through the TOMNET website ([www.tomnet-utc.org](http://www.tomnet-utc.org)). Scholars will have to register and complete a simple online form to access the public use data sets. In this way, TOMNET will have a tracking system to track level of usage and interest in the products of the center, and to ensure that there is no inappropriate use of the data. In addition, the public use data sets (together with metadata) will be made available through the ASU Digital Repository (<http://libguides.asu.edu/digitalrepository>). All TOMNET research reports, papers, articles, policy briefs, newsletters, and documents will also be archived in the ASU digital repository so that the materials are accessible to the worldwide community well beyond the life of the center. All of these materials will also be made available via the TOMNET center website. However, by having the materials available through the ASU digital repository, the products of TOMNET will be available to the worldwide community even if and when TOMNET leadership retires and the institutions decide to sunset the TOMNET center website at some point in the distant future. Educational materials will also be deposited in the ASU digital repository, and videos (where applicable) will be made available via a YouTube channel. All webinars will be recorded and links to the webinars will be posted to the TOMNET center website; the webinars will be recorded using the Adobe Connect system housed at ASU and links of recordings will be made available through the TOMNET website.

The confidential data sets that include personally identifiable information will not be made available through open access to the worldwide community. These datasets will remain secure on workstations housed at the center's partner institutions under strict IRB protocols, including password protection, secured access, and firewall protection. However, it may be of value to make such data also available to the broader research community so that studies that require detailed location data can be undertaken in the future by scholars around the world. In order to facilitate this reuse of the data, TOMNET will deposit confidential datasets (i.e., complete raw datasets) with the Transportation Secure Data Center at the National Renewable Energy Laboratory (NREL). The Transportation Secure Data Center (TSDC) at NREL is home to a number of confidential travel survey data sets including GPS-based travel data sets that provide exact locations of individuals throughout the survey day. The TSDC (<https://www.nrel.gov/transportation/secure-transportation-data.html>) is maintained at NREL through a partnership with USDOT and DOE. Researchers and scholars will be able to apply for an account through TSDC and will be allowed secure access to confidential data via specialized protocols implemented by the TSDC. Researchers and scholars will not be able to download data sets, but they will be able to analyze the confidential data online using a number of highly sophisticated statistical and spatial analysis tools that NREL has developed for this purpose.

Through the mechanisms described above, TOMNET will ensure that data and products are available for reuse and redistribution, subject to data privacy laws and human subject protection. Codes, algorithms, and models will be made available to the community through the TOMNET center website using principles

of open source licensing. Scholars will be welcome to download and use the codes and models after completing a simple online form so that there is a tracking system to assess level of interest and usage in the products of the center and prevent inappropriate usage. There are other digital repositories that the TOMNET team will consider for reuse and redistribution of data. Only public use data sets will be made available through such platforms. Examples of such repositories include the Social Science Data Archive ([http://data-archive.library.ucla.edu/da\\_catalog/](http://data-archive.library.ucla.edu/da_catalog/)) and the Inter-University Consortium for Political and Social Research ([www.icpsr.umich.edu/icpsrweb/landing.jsp](http://www.icpsr.umich.edu/icpsrweb/landing.jsp)).