NREL
NATIONAL RENEWABLE ENERGY LABORATORY

# Archiving Data from New Survey Technologies: Lessons Learned on Enabling Research with High-Precision Data While Preserving Participant Privacy

## Preprint

J. Gonder and E. Burton
*National Renewable Energy Laboratory*

E. Murakami
*U.S. Department of Transportation*

*To be presented at the 10th International Conference on Transport Survey Methods*
*Leura, Australia*
*November 16-21, 2014*

**NREL is a national laboratory of the U.S. Department of Energy**
**Office of Energy Efficiency & Renewable Energy**
**Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

**Conference Paper**
NREL/CP-5400-62901
November 2014

Contract No. DE-AC36-08GO28308

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at http://www.osti.gov/scitech

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

*Cover Photos: (left to right) photo by Pat Corkery, NREL 16416, photo from SunEdison, NREL 17423, photo by Pat Corkery, NREL 16560, photo by Dennis Schroeder, NREL 17613, photo by Dean Armstrong, NREL 17436, photo by Pat Corkery, NREL 17721.*

NREL prints on paper that contains recycled content.

## ABSTRACT

During the past 15 years, increasing numbers of organizations and planning agencies have begun collecting high-resolution Global Positioning System (GPS) travel data. Despite the significant effort and expense to collect it, privacy concerns often lead to underutilization of the data. To address this dilemma of providing data access while preserving privacy, the National Renewable Energy Laboratory, with support from the U.S. Department of Transportation and U.S. Department of Energy, established the Transportation Secure Data Center (TSDC). Lessons drawn from best-practice examples from other data centers have helped shape the structure and operating procedures for the TSDC, which functions under the philosophy of first and foremost preserving privacy, but doing so in a way that balances security with accessibility and usability of the data for legitimate research. This paper provides details about the TSDC approach toward achieving these goals, which has included creating a secure enclave with no external access for backing up and processing raw data, a publicly accessible website for downloading cleansed data, and a secure portal environment through which approved users can work with detailed spatial data using a variety of tools and reference information. This paper also describes lessons learned from operating the TSDC with respect to improvements in GPS data handling, processing, and user support, along with plans for continual enhancements to better support the needs of both data providers and users and to thus advance the research value derived from such valuable data.

## 1. INTRODUCTION

Departments of transportation and metropolitan planning organizations (MPOs) regularly collect travel survey data for purposes such as developing/updating transportation demand forecasting models and identifying transportation needs within a survey region. These surveys have been conducted for many decades and historically include mail-out/mail-back travel diaries supplemented by computer-assisted telephone interviews. In the late 1990s, several MPOs began investigating the use of Global Positioning System (GPS) technology to improve the accuracy and completeness of personal travel data collection [1]. Continually improving GPS accuracy and declining equipment costs have made it practical to include a GPS data-collection component in many modern surveys.

The expanded application of vehicle- and person-based GPS instruments along with other enabled new technologies (such as smart phone data collection) have created different opportunities and challenges relative to traditional diary-based and computer-assisted telephone interview survey techniques. The increased spatial and temporal precision of the data can help improve the overall quality and confidence in survey results, and it can open up new application opportunities for the data in both traditional and nontraditional fields [2]–[8]. However, the increased data precision (sufficient to identify specific house or business locations at trip ends) also creates an inherent privacy concern, particularly when the spatial data are linked to demographic information collected during the survey effort. This confidentiality issue must be addressed if the archived data are to provide ongoing research and analysis value.

With data-collection costs ranging from hundreds of thousands to millions of U.S. dollars per study and limited available funding for conducting surveys, it is important to maximize each

study's benefit through preserving the data and responsibly making it available for ongoing research. Indeed, recent years have seen growing recognition of the value of and growing support for "open data" policies wherever possible. A May 2013 White House Executive Order on Open Data Policy referenced the tremendous public value that has been derived from the decision to make GPS technology itself open to anyone and cited that experience as part of the motivation for increasing open data in government [9]. Researchers supported with federal funds must now create a data-management plan in advance of conducting the research. In addition, data-curation and institutional repositories are now considered "essential infrastructure" [10], [11].

Recognition of the value to archive and make accessible household travel behavior survey data led (more than a decade ago) to the creation of the Metropolitan Travel Survey Archive at the University of Minnesota [12]. The Metropolitan Travel Survey Archive has received intermittent funding from the U.S. Department of Transportation (DOT) through the Federal Highway Administration (FHWA) and start-up funding from the Bureau of Transportation Statistics. However, the aforementioned privacy concerns from transportation agencies that had incorporated GPS into their travel behavior studies made these agencies reluctant to include a copy of the GPS data in the Metropolitan Travel Survey Archive, even if only for archival purposes and not for data access.

In response to this open data impediment and consistent with the recommendations of a 2007 National Research Council report about resolving the conflict between data utilization and confidentiality protection [13], the U.S. Department of Energy's (DOE's) National Renewable Energy Laboratory (NREL) began partnering in late 2009 with DOT/FHWA to develop a data center that allows access to highly detailed records of travel in time and space in a way that maintains respondent anonymity. Ongoing DOT and DOE support since then has established the resulting Transportation Secure Data Center (TSDC) [14] as one of NREL's significant archiving efforts for sensitive transportation data. Others include the National Fuel Cell Technology Evaluation Center (NFCTEC; formerly the Hydrogen Secure Data Center, HSDC), the Fleet DNA medium- and heavy-duty vehicle drive cycle repository, and the Fleet Sustainability Dashboard (FleetDASH) [15]-[17].

The remainder of this paper provides further details about the approach, structure, and contents of the TSDC and lessons learned through the process of developing and operating it.

## 2. DESIGNING THE TSDC

When developing the basic design for the TSDC, NREL and FHWA considered lessons from best-practice examples at other data centers. These included examples from other NREL secure transportation data centers (such as those mentioned above) as well as examples from analogous data centers, such as the Census Bureau's Research Data Center program, with its long established system of providing researchers access to highly confidential data [18]. One of the security features noted when benchmarking the Research Data Center program was the requirement that users travel to specific locations to access the data. Unfortunately, even with the steady addition of new sites, the travel requirement presents a cost and time inconvenience for many researchers. Further benchmarking revealed one example repository at the NORC Data Enclave that permitted restricted remote access for researchers. This enclave stores social science

micro-data records, such as those from the Annie E. Casey Foundation's Making Connections Survey on topics including economic hardship in families and those from a National Science Foundation Survey of Earned Doctorates [19]. Another recent virtual research data center example provides access to sensitive Medicare and Medicaid program data [20].

For the TSDC approach, NREL and FHWA decided that the availability of secure remote access to a single data center would give the greatest benefit to both data providers and users. On the provider side, this relieves the burden from each MPO or transportation agency of having to store and protect data, respond to data-sharing requests, and/or set up their own secure data center. Users also benefit from not needing to travel, from finding data in a single location, and from working with data they would not otherwise be able to access. As described later, accomplishing this vision required implementing technical controls to prevent removal of data through the remote connection and providing appropriate tools on the remote site for use in conducting analyses.

An important step in the TSDC development process was establishing an advisory group to help provide oversight, to give technical input, and to represent the interests of various stakeholders. This group includes data providers and users who work in industry, academia, and government. Input from the advisory group helped drive the TSDC development philosophy to first and foremost implement security measures for protecting data and preserving participant privacy. At the same time, advisory committee members representing the user community help ensure that protections are implemented in a way that still permits researchers access to critical data elements in the most user-friendly manner possible (while working within the privacy-protection constraints).

Consultation with the advisory group on how to best strike the balance between privacy protection and usability led ultimately to organizing the TSDC into three distinct sections: (1) a secure enclave for raw data (with no external access), (2) a public download area for cleansed data, and (3) a secure portal environment for controlled access to detailed data. For the first level of external access through the public download area, the data is "cleansed" by removing latitude/longitude spatial details to prevent the identification of individual participants. This approach offers strong data protection while satisfying the needs of many users who simply require aggregated driving information or individual vehicle time versus speed profiles. The public download area allows these users quick and easy access to the data they need without requiring them to go through the more involved approval and connection procedure to access the secure portal environment. Security is also enhanced by providing these users access to only the data details they require. The next section of this paper provides additional information about these three TSDC sections.

## 3. TSDC DATA HANDLING AND ORGANIZATION

### 3.1. TSDC Section 1: Secure Enclave for Storing and Processing Raw Data

NREL drew upon best-practice lessons from similar data-archiving activities to establish key TSDC features, including security such as physical building-badge access, an on-site security force, and individual account passwords; redundancy, including maintaining backup copies; and data handling, such as database management along with processing the data for quality control and information addition. With respect to data handling, significant parallel learning and cooperative development have occurred with the aforementioned commercial vehicle Fleet DNA project, which also works with extensive amounts of GPS data. Fig. 1 illustrates the processing procedure that has evolved to support both data projects.

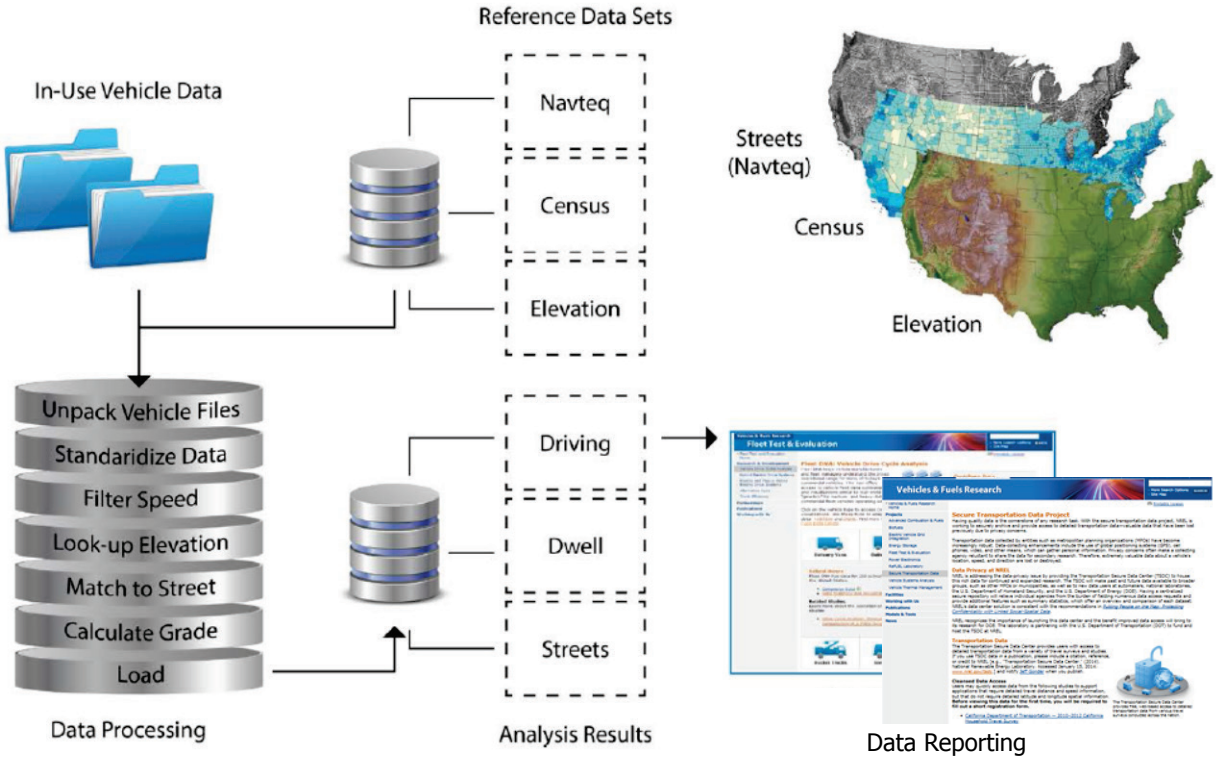


**Fig. 1.** Shared data-processing procedures for Fleet DNA and the TSDC

Processing routine subcomponents that have evolved during the course of the TSDC and Fleet DNA projects include the initial steps to structure and standardize data formats. The latest update has been to transition from a traditional fixed-table database structure to a NoSQL data-storage approach. This change was motivated by the growing number of data sets and particularly data types being stored. Each archived study typically contains time-series data (e.g., recorded at one-second intervals) that include time stamps, latitude/longitude coordinates, travel speeds, and measures of GPS quality, such as the number of satellites contributing to the coordinate determination. Study results are also summarized at trip and tour levels, as well as at the level of overall metadata (e.g., demographic information for participating households and vehicles, and specific details on the original study). A traditional database structure worked well initially for

storing these data, with satisfactory performance for the time it takes to read entire tables from the disk when sub-setting the data for various analyses; however, a growing number of columns within the data tables and a growing amount of variation or customization from study to study leads to processing inefficiencies for traditional databases that require a fixed structure across all data tables.

One driver of data-channel variability among studies is the varied inclusion of parameters from participating vehicles' on-board data bus. This is frequently the case for Fleet DNA data sets, and on-board data are also included in the TSDC, such as from a subsample of participants' vehicles in the large 2010–2012 California Household Travel Survey [21]. The NoSQL approach organizes the data such that only the information needed for a particular subset analysis gets read off the disk, and it permits flexibility in the number of parameters stored for each study and each participant.

The next step in the processing routine following data organization and format standardization is GPS speed filtration. Anyone who has spent much time analyzing raw GPS driving data can readily appreciate the need for this quality-control step. Over the years, NREL has developed and refined techniques for identifying/correcting GPS measurement errors, which can result from equipment cold-starts (in which recording does not begin until after a vehicle/person has already started moving), signal dropouts, signal noise/jumps, and measurement drifts. If uncorrected, quality issues with the raw data can lead to unreasonable calculations of driving distances, speeds, and accelerations in second-by-second GPS data. In addition to enabling the processed GPS data to be used for things such as vehicle powertrain simulations, the quality-control steps have sometimes identified data-recording errors that had not been flagged from the original study. For example, in one study approximately 2% of trips were inadvertently assigned duplicate identification numbers, resulting in huge jumps in the GPS point recordings (actually made by two different vehicles driving across town from one another). In another study, 5% to 10% of the GPS data-collection devices were inadvertently set to record speed data in kilometers per hour rather than miles per hour. When the data was interpreted in the expected unit of miles per hour, some of the vehicle speeds appeared high. The error was identified and corrected by comparing point-to-point distance calculations to the speed readings. Learning from these experiences, NREL developed several visualizations and quantitative metrics to monitor the automated quality-control routines themselves, which are used to flag data for manual inspection and to continually improve data-processing accuracy and speed.

In addition to quality-control processing, NREL performs several steps that add further value to the GPS travel profiles. These include providing accompanying reference information (e.g., geospatial economic, demographic, and land-use data) and linking the GPS points to underlying road network and elevation data (using reference layers from NAVTEQ/Nokia/HERE and from the U.S. Geological Survey, respectively). The data-linking processes require additional quality-control steps but can produce useful information, such as the distribution of travel on different road types and relative utilization of specific roads in the network. NREL's elevation linking and filtering procedure adds a road grade estimate to each GPS data point. It is described in more detail in separate publications [22], [23].

### 3.2 TSDC Section 2: Public Download Area for Cleansed Data

Fig. 2 shows a screenshot of the TSDC website, from which cleansed versions of the data are publicly available for download. As described previously, the intent with the public download area is to make available data versions that do not have privacy concerns but that are sufficient for many applications. These data include aggregated statistics, trip-driving distances, and second-by-second vehicle-speed profiles (absent latitude/longitude coordinate details). Following guidance from the TSDC advisory committee, the public download area also excludes data such as vehicle model (though make and year are included, and car/truck class is appended when possible), because information on rare vehicle models could theoretically be combined with other data sources to violate the anonymity of an individual participant. The committee also recommended a cautious initial approach to exclude NREL's point-by-point road-grade estimate from the cleansed data because of concerns about grade potentially being used to backfill latitude/longitude locations into some travel profiles. In addition to these data cautions, users are required to register and accept a point-and-click legal disclaimer and use agreement (also pictured in Fig. 2) stating that he/she will not attempt to identify individual or personal information from the data.

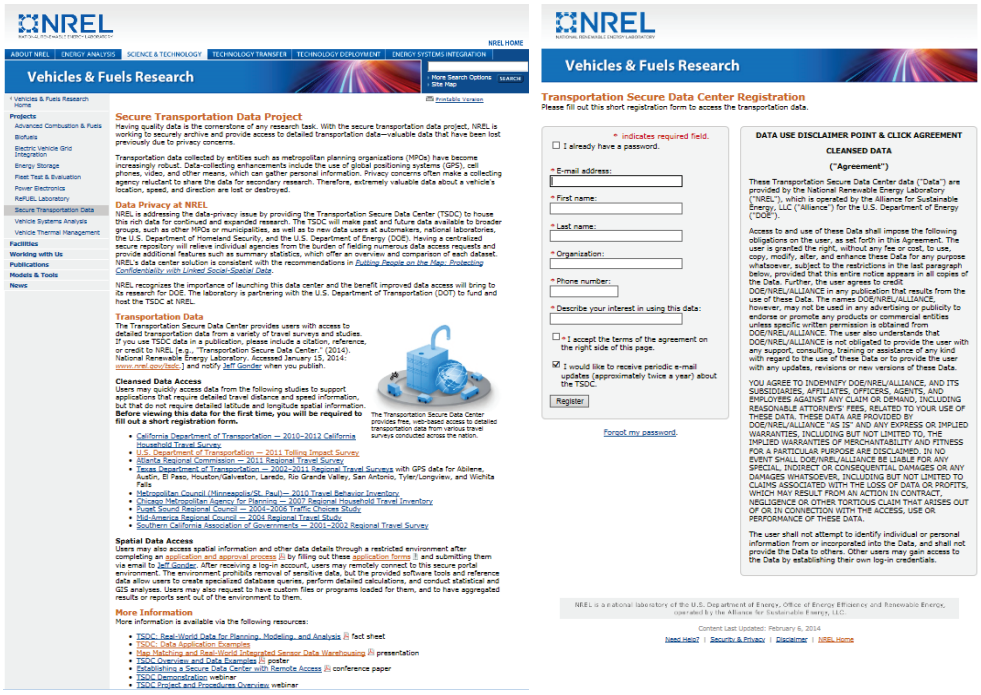


**Fig. 2.** TSDC website and downloadable cleansed data registration form [14]

After completing the simple registration form and logging in, users can view a more detailed summary description of the studies from which data has been made available. Several files are also readily available to download for each study. These typically consist of:

- The final report—Providing extensive details about the original study and the context of the collected data
- Multiple data files—Typically in comma-separated variable (.csv) format and generally including:

  - Detailed point-by-point travel data for each trip by each vehicle/participant throughout the entire study period, both in raw/original form and in a processed form with NREL's quality-control and filtering routines applied.
  - Demographic data on participants (bins for age, income, number of vehicles in household, etc.) and their vehicles (year, make, and vehicle class/fuel economy when possible).
- A data dictionary—Providing a general overview of the study and data, any issues of which to be aware, definitions of each variable in the downloadable data files, and summary statistics (about driving distances, speeds, etc.).

NREL, FHWA, and the TSDC advisory group also considered processing options to generate other types of cleansed data. One considered approach was to code trip ends to the centroid of a geographic region, such as a traffic analysis zone or TAZ. This would make some spatial data available for download, simply with obfuscated trip ends. However, this option was ultimately turned aside for failing to strike the desired balance between privacy protection and usability. Even partially obfuscated spatial data could present an added privacy concern when placed alongside other details such as driving speed/distance, household/vehicle demographics, and trip purpose [24]. Usability of the obfuscated spatial data would also be limited by the inability to link actual trip ends back to land-use information from sources such as parcel data and Google Maps. Instead, it was decided to provide only spatial data through the secure portal environment described below—relying on rigorous user screening, technical controls, and a legal agreement for privacy protection, and providing numerous geographic information system (GIS) tools and reference data to support user analyses within the environment.

### 3.3. TSDC Section 3: Secure Portal Environment for Detailed Data

While developing the operating details for the secure portal environment, NREL, FHWA, and the TSDC advisory group again considered what lessons could be drawn from related best-practice examples. It was found that the National Household Travel Survey (NHTS) had established a process for researchers to request the "NHTS DOT file." This example presented a situation similar to the TSDC goals—permitting restricted access to greater geographic and demographic details from the original data (that *could* be abused to identify a participant) in order to permit legitimate research/analyses that could not otherwise be accomplished without the detailed data. The NHTS procedure requires users to describe the analysis they wish to conduct and sign a confidentiality agreement before receiving the detailed data. The TSDC adopted these as well as additional privacy-protection steps as summarized below:

- Applicants must complete an analysis description form describing:
  - The analysis to be conducted
  - Other data sources considered to support the analysis and why they are insufficient
  - The output results the user anticipates wanting to remove from the secure environment
- Applicants must sign a data use and disclaimer agreement including:
  - Confidential data-protection legal language
  - An explicit pledge to not attempt to identify individual participants
  - An additional signature required from his/her university advisor or line manager

- Users must also complete a Condition of Use for Cyber Resources form before they may establish a connection account to the NREL virtual machines hosting secure portal environment
- After users complete the paperwork, the advisory group reviews the application and provides an access recommendation; DOT and NREL then make the final approval or denial decision
- Once approved, users may interface with the data only through the secure portal environment:
    - The environment prohibits data transfer (Clipboard sharing, local drive access, and internet connection are all disabled.)
    - NREL reviews any externally developed user code or base files before loading them into the environment for use and similarly audits aggregated results that a user wishes to remove from the environment before providing them to the user.

With the secure portal environment, the TSDC created a more secure mechanism for controlled access to sensitive data than the approach of sending the data directly to approved users. Maintaining the data behind computer firewalls avoids the potential for intentional or unintentional re-sharing of the data to unauthorized users after giving an authorized user full control of a copy of the database.

Although operating within these security measures inevitably places some constraints on data users, NREL has also sought to include features and functions within the secure portal environment to enhance its usability for researchers. These efforts have included providing a variety of software tools to support a range of user preferences and abilities. The available software includes free and open-source tools for database query, programming, and analysis, such as PostgreSQL/PostGIS/QGIS, GRASS, Python, and R. The database-integrated free and open-source tools help provide computational efficiency for large GIS analyses (important when a data set contains millions of GPS data points), but they require some programming expertise. Users are therefore also given access to the commercial ArcGIS software from ESRI, which provides a user-friendly integrated system for working with GIS data through a graphical user interface.

To further support an assortment of spatial analyses, NREL has included a variety of additional GIS reference information in the secure portal environment. These data include Census Topologically Integrated Geographic Encoding and Referencing (TIGER) system files showing water bodies, roads, landmarks, and political boundaries (such as county, census tract, and block group). When available, additional reference details accompany specific data sets. For example, accompanying the Traffic Choices Study data, the TSDC also includes archived road speed and UrbanSIM data for the Puget Sound region (grid-based information about demographic, economic, and land-use characteristics) [25]. As mentioned earlier, users can ask NREL to load additional tools or reference files into the secure portal environment to support a particular analysis. Fig. 3 and Fig. 4 illustrate a few examples of the available tools and reference information.

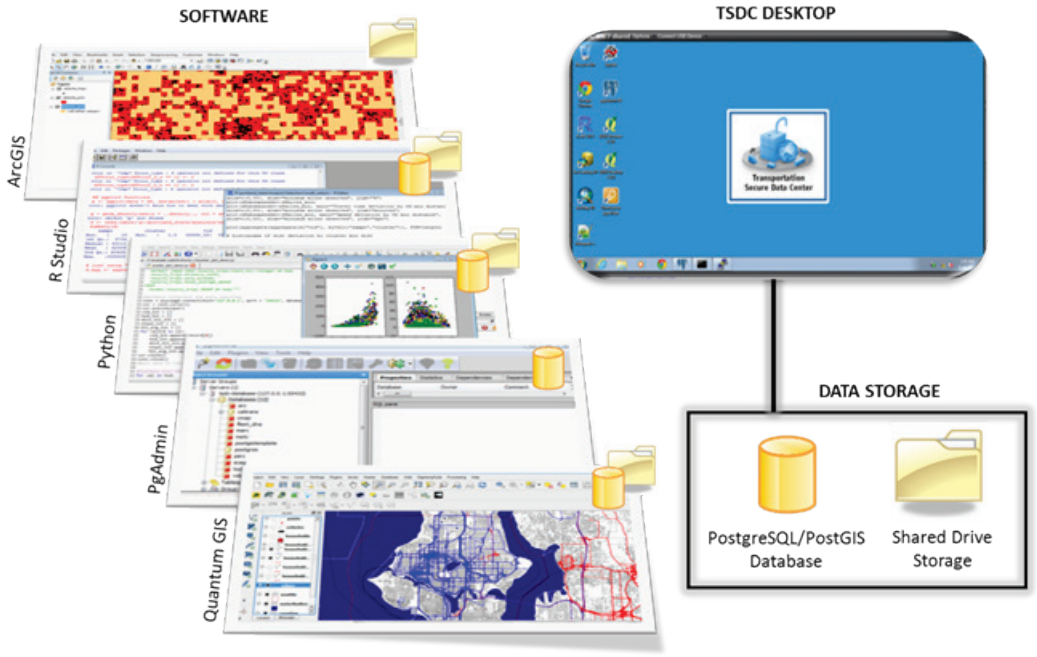


Fig. 3. TSDC software and supporting data resources

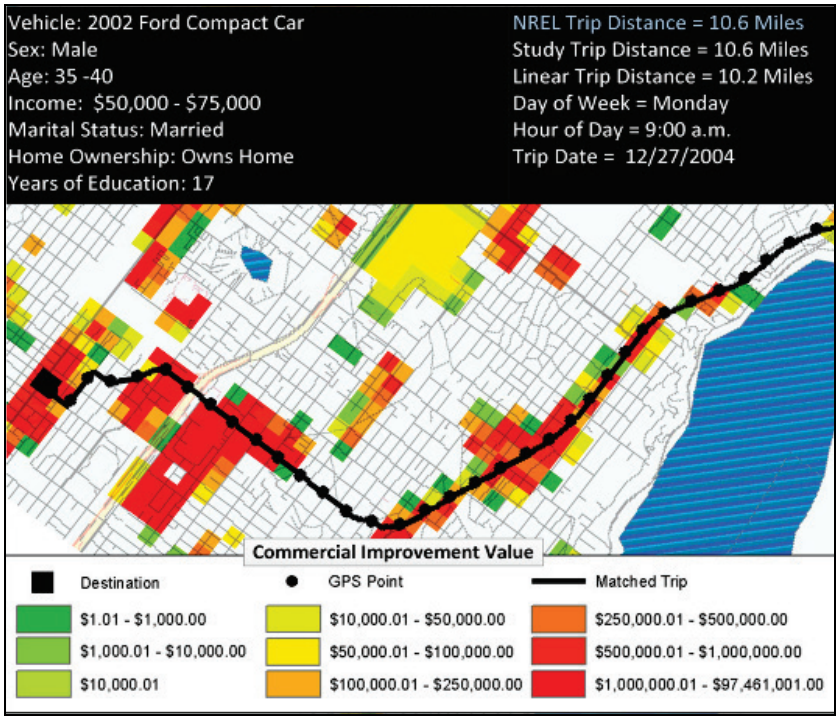


**Fig. 4.** Example of demographic, processed trip, and regional data available in the TSDC

## 4. OPERATING THE TSDC—WORKING WITH DATA PROVIDERS AND USERS

In each year of its initial operation, the TSDC has gradually expanded the number of data sets that it houses and the number of users it supports. Some data providers were more willing than others to be among the earliest participants; but as the data set count and successful track record have grown, more and more providers have become interested in participating. Providers state that they appreciate being able to point interested users to the TSDC rather than individually hosting something similar and interfacing with users themselves. Some providers have also been motivated by the opportunity to leverage the processing, filtering, and data-addition functionalities that NREL has developed for TSDC data sets to take advantage of these enhancements to their own data.

Several of the data sets included in the TSDC were collected as add-on samples to traditional diary-based regional travel surveys for assessing issues such as trip underreporting. These GPS samples typically include second-by-second data for several hundred to several thousand vehicles and/or persons recorded from one day to one week. A few of the TSDC participants providing these types of data include the Atlanta Regional Commission, California Department of Transportation, Chicago Metropolitan Agency for Planning, Mid-America Regional Council, and Texas Department of Transportation. Altogether, the TSDC second-by-second vehicle-based data from these providers' studies total more than one million miles of driving.

The aforementioned Traffic Choices Study conducted by the Puget Sound Regional Council represents another large data set hosted in the TSDC. The Traffic Choices Study collected data from 2004 to 2006 to evaluate travel behavior changes in response to time- and location-variable road tolling. The particular value of this data set is that it contains GPS data from more than 400 vehicles spanning a period of 18 months, including baseline data collected prior to the 7-month experimental tolling period.

As the number of TSDC data sets has grown, so too has the user base of researchers accessing the data. NREL has correspondingly shifted some of the TSDC focus from the formational activities described in the previous sections to ongoing data center operations. These efforts include not only user support, but also leveraging user experiences and feedback to implement ongoing enhancements. The varied research topics that TSDC users have explored include travel time variability, driving behavior, mass transit opportunities, electrified vehicle analysis, emissions and air quality modeling, and vehicle design and durability testing. The TSDC website includes a subpage on "Data Application Examples" that further describes user research topics and provides links to completed publications that have leveraged GPS travel data [14].

In addition to this wide variety of research topics, NREL has found wide variability in user experience and comfort levels working with the data and with tools provided in the TSDC secure portal environment. Although many users can work fluently with database tools and open-source software, a number of others have little experience working with large databases and are more comfortable with commercial software packages. In an effort to better support more novice users, as well as provide increased convenience for all users, NREL has been developing feature enhancements in an interactive user manual and data-visualization tool for the secure portal environment.

Fig. 5 displays images from this interactive tool, including a set of drop-down menus for quickly exploring data tables and their contents for each study. The tool also helps users to create basic graphs and to query database variables for visualization in the graphs without requiring any SQL programming knowledge. Users may additionally download preformatted reports using the tool.

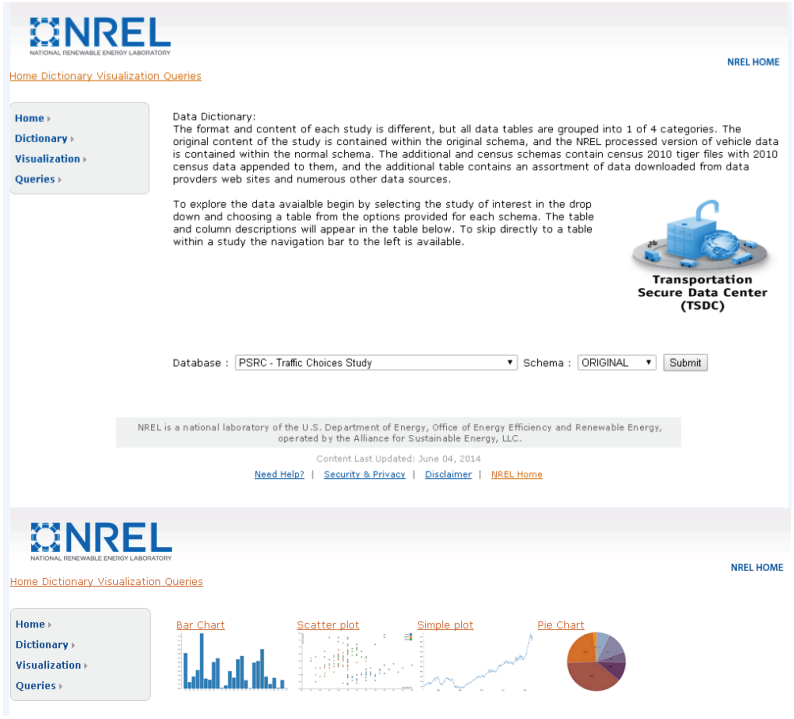


**Fig. 5.** Images from the interactive user manual and data visualization tool for the TSDC secure portal environment

In addition to providing access to analysis tools and reference information, NREL staff field questions and help users get started with their analyses; however, NREL staff do not have an unlimited amount of time to support user projects, and ultimate responsibility falls to the users themselves to make sure they have the required skill sets and can perform the work necessary to accurately complete their analyses. To help users get started, NREL supplies example code for software tools within the secure portal environment that demonstrates basic functions such as querying data, performing calculations, and generating plots.

NREL currently does not charge any fees from TSDC users (or data providers) and hopes to continue operating without imposing such fees—simply relying upon the modest operating budget provided by DOT and DOE, along with leveraging operating efficiencies from NREL's other sensitive data-archiving activities. One area in which budget limitations create a challenge for maximally supporting all users is that of making costly commercial software tools available in the secure portal environment. The expense for many tools can quickly become prohibitive to support network licenses accessible to all environment users. NREL has prioritized availability of an ArcGIS license given that software tool's prevalence for GIS analysis, although only a single license seat is available so far. This limits the number of simultaneous secure portal users who can access the tool, which has become an occasional issue as the user base for this environment has grown (though there are no such limits on the open-source GIS tools available

in the environment). NREL does plan to add another ArcGIS license to help with this, and will continue exploring licensing options for other frequently requested commercial tools.

When it comes to the secure portal application process itself, the vast majority of researchers who have applied have ultimately been granted accounts to use the environment. The typical process after receiving a prospective user's application includes an initial screening to check for missing items or insufficient answers. NREL and/or FHWA then iterate with the applicant to fill in any omissions and provide any needed elaboration, such as about the purpose, methods, or desired outputs of the proposed research or to clarify what can or cannot be removed from the secure portal environment. After the applicant submits any needed revisions, the remaining steps of advisory committee review/approval and user account creation typically take less than two weeks. The only occasions thus far when prospective users' applications have not been approved involved situations in which the proposed analysis did not actually require access to the data in the secure portal environment, or when users were unable or unwilling to remedy incomplete items identified during the initial application screening. The full application packet and instructions about the process are posted on the TSDC website [14].

Among the topics for future enhancements, NREL plans to establish a collaborative environment in which users can share results, communicate, and answer questions from each other. Initial indications suggest that current users would be receptive to such a concept, because several already expressed a willingness to share code and intermediate results that they have generated through the course of completing their analyses. New users will no doubt also benefit from the opportunity to leverage previous users' experience to come up to speed more quickly and to build on others' results.

## 5. SUMMARY

During the past 15 years, increasing numbers of organizations and planning agencies have begun collecting high-resolution GPS travel data. Despite the significant effort and expense to collect it, privacy concerns often lead to underutilization of the data. Through the TSDC effort described here, NREL, DOE, and DOT/FHWA have partnered to address these concerns, support the needs of numerous data-starved applications, and increase research returns from the original data-collection investment. The TSDC has been developed and organized with the intent to first and foremost preserve privacy, but to do so in a way that balances security with accessibility and usability of the data for legitimate research needs. The TSDC structure helps support this goal by providing a secure enclave with no external access for backing up and processing raw data, a publicly-accessible website for downloading cleansed data, and a secure portal environment through which approved users can work with the detailed spatial data using a variety of tools and reference information.

The TSDC team applied lessons from best-practice examples at other data centers when developing the structure and operating procedures for the TSDC, and the team relies upon an advisory committee representing diverse stakeholder interests to endorse any procedure changes and to review applications from prospective secure portal environment users. Several years of operating the TSDC have also generated useful lessons in areas such as GPS data handling, processing, and user support. As the TSDC continues to grow, NREL will apply these and

ongoing lessons toward continual enhancements to better support the needs of data providers and TSDC users and to thus advance the research benefit derived from such valuable data.

## 6. ACKNOWLEDGEMENTS

The authors are grateful for initial start-up support for the TSDC provided by internal NREL funding and by the DOT FHWA Office of Operations, and also for continued support throughout the project from the DOT FHWA Office of Planning and the DOE Vehicle Technologies Office. Further thanks go to Brennan Borlaug, Adam Duran, Arnaud Konan, Avinash Pallapu, and Eric Wood for their contributions toward GPS data processing, analysis, and user support, and to TSDC users for their helpful inputs and collaboration on continual project improvements.

NREL is managed and operated by the Alliance for Sustainable Energy under DOE Contract No. DE-AC36-08-GO28308.

## 7. REFERENCES7

[1] Murakami, E. and Wagner, D. (1999) Can using Global Positioning System (GPS) improve Trip reporting? *Transportation Research* C, 7, pp. 149-165.

[2] Gonder, J., Markel, T., Thornton, M. and Simpson, A. (2007) Using Global Positioning System travel data to assess real-world energy use of plug-in hybrid electric vehicles. *Transportation Research Record (TRR), Journal of the Transportation Research Board (TRB)*, No. 2017, Sustainability, Energy and Alternative Fuels, p. 26.

[3] Barth, M. and Boriboonsomsin, K. (2008) Real-world $CO_2$ impacts of traffic congestion. Paper #08-2860. *Proceedings of the TRB 87th Annual Meeting; January, Washington, DC.*

[4] Tate, E., Harpster, M. and Savagian, P. (2008) The electrification of the automobile: From conventional hybrid, to plug-in hybrids, to extended-range electric vehicles. SAE Publication 2008-01-1315. *Proceedings of SAE Congress 2008; April, Detroit, MI.*

[5] Earleywine, M, Gonder, J., Markel, T. and Thornton, M. (2010) Simulated fuel economy and performance of advanced hybrid electric and plug-in hybrid electric vehicles using in-use travel profiles. *Proceedings of the 6th IEEE Vehicle Power and Propulsion Conference (VPPC); Sept.1-3, Lille, France.*

[6] Kahn, M. and Kockelman, K. (2012) Predicting the market potential of plug-in electric vehicles using multiday GPS data. *Journal of Energy Policy*, February.

[7] Attibele, P., Makam, S. and Lee, Y. (2013) A comparison of real-world and accelerated powertrain endurance cycles for light-duty vehicles. *Proceedings of Innovative Automotive Transmissions, Hybrid and Electric Drives; May.*

[8] Neubauer, J. and Wood, E. (2014) Impact of range anxiety and home, workplace, and public charging infrastructure on simulated battery electric vehicle lifetime utility. *Journal of Power Sources*, 257, July 1; pp. 12-20.

[9] Burwell; S., VanRoekel, S., Park, T. and Mancini, D. (2013) Open data policy—Managing information as an asset. Memorandum for the Heads of Executive Departments and Agencies, M-13-13. Executive Office of the President of the United States, May 9. http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

[10] Lynch, C. (2003) Institutional repositories: Essential infrastructure for scholarship in the digital age. *Libraries and the Academy*, 3:2, April; pp. 327-336.
[11] Cragin, M., Palmer, C., Carlson, J. and Witt, M. (2010) Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368:1926, September 13, pp. 4023-2038. doi: 10.1098/rsta.2010.0165 Phil. Trans. R. Soc. A.
[12] "Metropolitan Travel Survey Archive." (2014) University of Minnesota http://www.surveyarchive.org/. Accessed September 6, 2014.
[13] National Research Council. (2007) Putting people on the map: Protecting confidentiality with linked social-spatial data, in: M.P. Gutmann and P.C. Stern (Eds), Panel on Confidentiality Issues Arising from the Integration of Remotely Sensed and Self-Identifying Data. Committee on the Human Dimensions of Global Change. Division of Behavioral and Social Sciences and Education. (Washington, DC: The National Academies Press). http://books.nap.edu/openbook.php?record_id=11865.
[14] "Transportation Secure Data Center." (2014) National Renewable Energy Laboratory. Accessed September 6, 2014: www.nrel.gov/tsdc.
[15] "National Fuel Cell Technology Evaluation Center." (2014) National Renewable Energy Laboratory. Accessed September 6, 2014: www.nrel.gov/hydrogen/facilities_nfctec.html.
[16] "Fleet DNA: Vehicle Drive Cycle Analysis." (2014) National Renewable Energy Laboratory. Accessed September 6, 2014: www.nrel.gov/fleetdna.
[17] "Fleet Sustainability Dashboard." (2014) Accessed September 6, 2014: federalfleets.energy.gov/FleetDASH.
[18] "Research Data Center Program." U.S. Census Bureau, Center for Economic Studies. Accessed December 5, 2011: http://www.ces.census.gov/index.php/ces/researchprogram.
[19] "NORC Data Enclave." Accessed December 5, 2011: http://www.dataenclave.org/index.php/home/welcome.
[20] "CMS Virtual Research Data Center." Accessed September 20, 2014: http://www.resdac.org/cms-data/request/cms-virtual-research-data-center.
[21] California Department of Transportation. (2013) 2010-2012 California Household Travel Survey Final Report. Prepared by NuStats, LLC, in association with GeoStats, Franklin Hill Group, and Mark Bradley Research & Consulting, June 14.
[22] Wood, E., Burton, E., Duran, A. and Gonder, J. (2014) Appending High-Resolution Elevation Data to GPS Speed Traces for Vehicle Energy Modeling and Simulation. Technical Report. NREL/TP-5400-61109. (Golden, CO: National Renewable Energy Laboratory, June). http://www.nrel.gov/docs/fy14osti/61109.pdf.
[23] Wood, E., Burton, E., Duran, A. and Gonder, J. (2014) Contribution of road grade to the energy use of modern automobiles across large datasets of real-world drive cycles. *Proceedings of the 2014 SAE World Congress; April, Detroit, MI.* Preprint available at http://www.nrel.gov/docs/fy14osti/61108.pdf.
[24] Elango, V. V., Khoeini, S., Xu, Y. and Guensler, R. (2013) Longitudinal Global Positioning System travel data and breach of privacy via enhanced spatial and demographic dnalysis. *Transportation Research Record (TRR), Journal of the Transportation Research Board (TRB)*, 2354, p. 86.
[25] Puget Sound Regional Council. (2008) Traffic Choices Study—Summary Report. A Global Positioning System Based Pricing Pilot Project: Evaluating Traveler Response to Variable Road Tolling Through a Sample of Volunteer Participants. Seattle, WA: April.