

10th International Conference on Transport Survey Methods

## Archiving data from new survey technologies: enabling research with high-precision data while preserving participant privacy

Jeffrey Gonder<sup>a</sup>, Evan Burton<sup>a</sup>, Elaine Murakami<sup>b,\*</sup>

<sup>a</sup> National Renewable Energy Laboratory (NREL) Transportation Center, 15013 Denver West Parkway, Golden, CO 80401 USA

<sup>b</sup> U.S. Department of Transportation (DOT) Federal Highway Administration, 1200 New Jersey Ave, SE, Washington, DC 20590, USA

---

### Abstract

Despite the significant effort and expense to collect high-resolution Global Positioning System (GPS) data in travel surveys, privacy concerns often lead to its underutilization. This paper describes development of the Transportation Secure Data Center (TSDC) to address this dilemma of providing data access while preserving privacy. The TSDC operating structure was developed in consultation with an advisory committee and includes: 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.

© 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of International Steering Committee for Transport Survey Conferences ISCTSC

*Keywords:* Travel data; GPS; Transportation Secure Data Center; TSDC; Data privacy

---

### 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

---

\* Corresponding author. Tel.: +1-303-275-4462; fax: +1-303-275-3765.  
E-mail address: [jeff.gonder@nrel.gov](mailto:jeff.gonder@nrel.gov)

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 non-traditional fields—from assessing day-to-day variations in travel time, mode and routes to analysing the impact of driving behaviour on vehicle component durability and real-world energy consumption [2]–[14]. 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 [15]. 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” [16], [17].

Recognition of the value to archive and make accessible household travel behaviour survey data led (more than a decade ago) to the creation of the Metropolitan Travel Survey Archive at the University of Minnesota [18]. 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 behaviour 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 [19], 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 centre 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) [20] 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 commercial vehicle drive cycle repository, and the Fleet Sustainability Dashboard (FleetDASH) [21]–[23].

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 centres. These included examples from other NREL secure transportation data centres (such as those mentioned above) as well as examples from analogous data centres, such as the Census Bureau's Research Data Center program, with its long established system of providing researchers access to highly confidential data [24]. 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

Download English Version:

<https://daneshyari.com/en/article/1106769>

Download Persian Version:

<https://daneshyari.com/article/1106769>

[Daneshyari.com](https://daneshyari.com)