SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Establishing a Central Archive for Transit Passenger Data

Gregory Newmark, Ph.D. and Hilary Nixon, Ph.D.

Project 1631
May 2018

This report describes the development of CATPAD, the Central Archive for Transit Passenger Data. CATPAD is a repository that collects, indexes, archives, and makes available online the transit survey instruments, data, and reports collected across the country. CATPAD seeks to overcome the current impediments to accessing disaggregated transit survey data by providing a single, searchable, internet archive to store and disseminate this valuable information. The goal of CATPAD is to make useful data available to inform transit decision making at all levels and to foster ongoing refinement of the nation's transit network.

Study Methods:

This research identified the impediments to making disaggregated transit passenger data available to the public, explored existing national repositories of transportation data, and designed a database structure to facilitate access to the data by anticipated user communities.

Findings:

A perennial problem faced by transportation planners, policymakers, researchers, and community advocates has been accessing the disaggregate passenger survey data collected on the nation's transit systems. While aggregated passenger data such as system-wide passenger boardings and total fare revenues collected are compiled nationally, the detailed individual responses to transit rider surveys typically remain hidden behind the closed doors of an agency.

Key findings include:

1. Impediments to Accessing Transit Survey Data

There are several impediments to accessing transit survey data. There are no nationwide nor systemwide lists of transit survey efforts, which makes finding passenger data challenging to begin with. A second major impediment is that once a survey of interest has been identified, there are typically no organized archives to access the underlying data. Finally, even when the survey has been identified and the data located, transit agency staff can be reluctant to share this information, either out of concern for external criticism or as protocols for protecting personally identifiable information (PII) are not well developed across the industry. Transit agencies are rightfully cautious about sharing information that might

directly or through cross-referencing other sources compromise the identity of survey participants. There is an emerging consensus towards balancing the researcher need for location information and the agency need for protecting PII to recode sensitive locations, such as home and work, from point coordinates to higher level geographies such as census block groups. Questions remain about which locations are sensitive and what is the appropriate level of geography that masks PII.

2. Open Data and Existing Transportation Databases

This research reviewed the National Transit Database (NTD), the National Transit Map (NTM), the Metropolitan Transportation Survey Archive (MTSA), and the Transportation Secure Data Center (TSDC). All the databases share the goal of making transportation data available. The NTD and NTM focus on transit service level information while the MTSA and TSDC focus on disaggregate regional travel surveys. No extant repository expressly considers transit passenger survey data, highlighting the need for CATPAD. The existing repositories employ different means of acquiring data from legal requirement to voluntary sharing. The latter two repositories include individual travel records and have developed protocols to align data across surveys as well as to mask data. The TSDC allows direct access to detailed data, but only via a secure computer network and after research plan approval. These resources demonstrate the trend towards transportation data sharing and different directions to take for CATPAD. Finally, the NTD and NTM offer complementary resources to CATPAD, suggesting that the CATPAD design would benefit from a compatible design.

3. Database Design

Review of existing transportation databases and anticipated CATPAD uses-cases resulted in the design of the CATPAD database. CATPAD has one table that provides basic information on a given survey, such as the name of the effort and its start and end dates. Other tables provide location information, such as state and metropolitan region, transit agencies involved, transit modes surveyed, etc. These tables use agency codes from the NTD to enable the resources to be linked. Additional tables focus on the survey itself, how it was conducted and what type of information was requested of respondents. Finally, CATPAD has a standard file formats for incorporating survey instrument, survey reports, and, most important, the survey data – the disaggregated transit passenger data at the core of this entire effort.

Policy Recommendations:

There are several key policy recommendations regarding this effort: First, CATPAD fills an important unmet need in the transportation planning community. Second, transportation professionals at all levels will benefit from engagement with these detailed transit passenger data. Third, more work needs to be done to determine industry standards for making transit data open through location recoding. Such standards will improve agency comfortability with CATPAD. Fourth, as CATPAD becomes a part of the industry firmament, there may be a need for a broader discussion of standardizing data collection to facilitate comparison. Finally, it is hoped that CATPAD will provide a permanent archival solution to the problem of storing passenger transit data.

About the Authors:

Dr. Gregory Newmark is an assistant professor of Landscape Architecture and Regional & Community Planning at Kansas State University. Dr. Hilary Nixon is a Professor of Urban and Regional Planning at San José State University.

To Learn More

For more details about the study, download the full report at: transweb.sjsu.edu/research/1631

MINETA TRANSPORTATION INSTITUTE

MTI is a University Transportation Center sponsored by the U.S. Department of Transportation's Office of the Assistant Secretary for Research and Technology and by Caltrans. The Institute is located within San José State University's Lucas Graduate School of Business.