

National Accessibility Evaluation Task Report

Task 2.2: Collect Input Datasets

Prepared by the

Accessibility Observatory at the University of Minnesota

September 30, 2016



**ACCESSIBILITY
OBSERVATORY**

UNIVERSITY OF MINNESOTA

Driven to DiscoverSM

1 Overview

The goal of this task is to assemble the data required to deliver national accessibility reports and block-level accessibility datasets to each of the project participants. Subsequent tasks in this project will require data describing the road network and speeds; transit networks and schedules; the pedestrian network; origin and destination locations; and the distribution of employment. All data is required to have full national coverage in a consistent format.

This task is complete. Data has been identified and acquired to meet each of the project's data needs. Some data components will be updated on an ongoing basis to support future annual updates to this project. The following sections provide a description of each data need and the data that was acquired to meet each need.

2 Task Objectives

The goal of this task is to assemble the input data required to deliver national accessibility reports and block-level accessibility datasets to each of the project participants. The following data components are required:

2.1 Road Network Data

Calculation of accessibility by auto will require a detailed dataset representing the road network. This data should be as detailed as possible, including roads of all sizes, capacities, and classifications. It should also be sufficiently up-to-date to provide a reasonable representation of the road network in 2016. To limit the effort required to implement accessibility calculations, it should use a single data format across the entire country.

2.2 Road Speed Data

Calculation of accessibility by auto will also require a dataset describing speeds on the road network. For each road, this dataset should provide a reasonable representation of typical speeds during 2016. It should provide separate speed data for typical weekdays and typical weekend days, and ideally should indicate how road speeds vary over the course of a day due to local congestion. This dataset must be directly compatible with the road network dataset.

2.3 Transit Network and Schedule Data

Calculation of accessibility by transit will require data describing the structure and schedules of transit networks across the country, in a consistent digital format. Because of its widespread adoption in travel planning and analysis tools, the GTFS data format for transit schedules is preferred.

2.4 Pedestrian Network Data

Calculation of accessibility by transit will also require data describing the pedestrian network connecting origin and destination locations to the transit network. This dataset should be as detailed as possible, because the low speeds of pedestrian travel magnify the effects of missing links or unnecessarily circuitous routes. It should also be in a consistent data format with full national coverage, and should be sufficiently up-to-date to provide a reasonable representation of the pedestrian network in 2015.

2.5 Origin and Destination Geography Definitions

To support accessibility calculations at the Census block level, the geography definition dataset should define origin and destination locations based on Census-defined geography. To maximize comparability of data across annual updates, it should be stable year-to-year.

2.6 Employment Distribution Data

Calculation of accessibility to jobs depends on data describing the location of jobs. The employment distribution dataset should provide a reasonable representation of the number of jobs available at individual locations in 2015, and should ideally be available in a consistent data format with national coverage. For accurate calculation of travel times on low-speed networks such as transit and (especially) pedestrian, data should be presented at the highest spatial detail possible. To minimize data processing requirements, this dataset should be directly compatible with the origin and destination geography definitions.

3 Task Status

This task is complete. The datasets collected to meet each data need are identified and discussed below. The project team purchased, installed, and configured a database server that will store and organize all input datasets and streamline many stages of accessibility calculation.

3.1 Road Network Data

After reviewing publicly-available road network datasets, the project team determined that a commercially-sourced road network dataset would be necessary in order to meet the goal of data format consistency across the whole country. The project team reviewed product offerings and pricing from multiple data vendors. Following this review, the University of Minnesota's Department of Civil, Environmental, and Geo- Engineering executed a license agreement with TomTom North America, Inc. for the use of their MultiNet data product. This produce provides a road network dataset with national coverage, including all levels and classifications of roadways. The dataset is updated multiple times a year. After executing this license, the project team received the 2016 road network dataset from TomTom via digital download.

3.2 Road Speed Data

The TomTom data license also includes TomTom's June 2016 Speed Profile data product, which provides speed estimates for every road network link based on data collected by GPS devices over the trailing 24 calendar months. These speed estimates are derived from direct traffic measurements and reflect changing congestion conditions over the course of the day. Separate estimates are provided for each day of the week, as well as for each 5-minute interval during a typical day. The speed data are directly associated with the road network data. The project team received the speed dataset from TomTom via digital download.

3.3 Transit Network and Schedule Data

The project team collected transit schedules in GTFS format directly from transit operators across the country. Schedules were collected in one of three ways. Some transit operators publish GTFS files to a single, publicly-accessible web address; these are automatically downloaded and archived by software created by the project team. Other operators require an interactive, manual download process. Other operators only distribute schedule files on request; the project team contacted these operators by email and/or phone to request the schedule files. Ultimately, schedule availability is based on a decision by each operator to provide digital schedules, and not all operators have made this decision. The project team will update this dataset on an ongoing basis to incorporate schedule revisions from known transit operators, and also to expand the dataset's coverage as new operators who provided digital schedule data are identified.

3.4 Pedestrian Network Data

The project team downloaded an extract of all North America data from OpenStreetMap, a crowd-sourced database of transportation networks with full global coverage. OpenStreetMap provides data in a consistent format in all locations, and is used as a data source for customer-facing travel planning tools implemented by several transit agencies. The project team acquired data as of August, 2015.

3.5 Origin and Destination Geography Definitions

The U.S. Census Bureau's TIGER data products include shapefile datasets of all blocks comprising the entire U.S., as well as aggregations of these blocks into counties, metropolitan areas, and states. The geography defined for the 2010 Census is designed to be stable until the next decennial census. The project team downloaded all 2010 census geography data.

3.6 Employment Distribution Data

The U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) data program provides block-level job estimates covering the entire U.S., excluding of Kansas and Massachusetts. For each block, these data give an estimate of the total number of jobs located there, as well as of jobs segmented by wage level, industry classification, and other factors. This dataset is associated directly with

geographic identifiers from the 2010 census data. The project team acquired LEHD data for 2014, the most recent available.