Catalogue no. 23-26-0003
ISBN 978-0-660-78102-0

Canadian Public Transit Network Database

Canadian Public Transit Network Database: Metadata Report

Urban Data Lab

Release date: July 30, 2025

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Table of contents

Overview

The Canadian Public Transit Network Database is a consolidated, national-level collection of General Transit Feed Specification (GTFS) data. Data from transit agencies across Canada were gathered from transit provider websites. The release aims to provide a consistent dataset for examining transit coverage and accessibility across jurisdictions, supporting researchers, planners, and developers.

Key processing steps included aggregating GTFS feeds from Canadian providers into a unified dataset, performing basic validation checks to assess compliance with GTFS standards and exporting GTFS data to geospatial format.

The release includes the following resources:

- **GTFS files:** GTFS format .txt files for each transit agency, including stops, routes, trips, stop times, calendars, and additional fields where available.
- **Geospatial national dataset:** a unified geospatial dataset compatible with geographic information systems combining data on stops and routes (in geopackage format).
- **Data validation files:** files summarizing warnings and errors for each transit agency feed.
- **Metadata:** Details on data sources and column name description.

The data are released under an Open Government Licence as supported by the Directive on Open Government.

Background

Purpose

A Canada-wide compilation of public transit information was initially developed for use across a variety of other related Statistics Canada products, such as the Spatial Access Measures. It became apparent that there was a need for a consolidated database of Canadian public transit infrastructure. Hence, the Canadian Public Transit Network Database was conceptualized as a stand-alone public transit database to be made available to researchers, policy makers and the public.

The intent of the dataset is to offer a regularly updated overview of Canadian public transit infrastructure, enabling researchers to make comparisons over time and across different municipalities or transit regions.

The GTFS data standard

General Transit Feed Specification (GTFS) is the open standard format used for public transit data. It is composed of a series of .txt files detailing the services provided by each transit agency, including the location of stops, transit schedules (including daily, weekly or seasonal variation), routes and trips. Initially developed by Google researchers, the GTFS data format is now a simple and accessible worldwide open standard that is commonly used in trip-planning software and applications.

GTFS data require a set of six base .txt files, detailed in Table 1. In addition to these essential base files, many transit agencies release additional information in GTFS format, such as fares, accessibility information, shape files, information on bike allowance (if vehicles are equipped with bicycle racks), frequency of service and transfers.

Table 1
Definitions of GTFS Schedule files

File	Presence	Description
agency.txt	Required	Information about the transit agency such as the name and website.
stops.txt	Required	Contains the geographic coordinates of stops (locations where vehicles pick-up and drop-off riders).
routes.txt	Required	Routes that transit services follow. The file details the id, name and type of each route.
Trips.txt	Required	Combines data from routes, calendar and stop time files to calculate specific journeys that can be taken by a rider.
stop_times.txt	Required	Arrival and departure times at a specific stop for a given trip. Typically, this is the largest and most detailed file. Used for calculating when a service is departing from a certain location.
calendar.txt and/or calendar_dates.txt	Conditionally required	Calendar.txt specifies service dates using a weekly schedule. Calendar_date.txt specifies exceptions (e.g., holidays) to the services defined in calendar.txt. Calendar.txt is required unless all dates of service are contained in Calendar_date.txt.
fare_attributes.txt	Optional	Describes fare information for an agency's routes.
shapes.txt	Optional	Describes the path a vehicle travels along a route.
translations.txt	Optional	Translations of customer-facing dataset values.
attributions.txt	Optional	Dataset attributions

Note: Definitions based on General Transit Feed Specification Reference.

Scope

The Canadian Public Transit Network Database includes all base .txt files (such as routes, stops, and trips) for all feeds. The availability of any additional GTFS .txt files, such as fare information, varies depending on the individual transit providers. Where additional .txt files were available, they were aggregated.

In addition to the static GTFS data many transit agencies release a dynamic format of GTFS called GTFS Realtime. GTFS Realtime provides additional real-time trip updates, such as vehicle positions and service alerts (moved stations, service delays, unexpected interruptions, etc.). GTFS Realtime data require constant refreshing and thus lie outside the scope of this first iteration of the Canadian Public Transit Network Database. The Urban Data Lab at Statistics Canada is investigating the possibility of making real-time data available to users in future iterations.

Data collection and compilation

Data sources

Potential transit agency data sources were compiled from existing databases of transit information, as well as searching for possible feeds using application programming interfaces (APIs) from two open data transit platforms. The first of these is the Mobility Database, an open data platform hosted by MobilityData, a not-for-profit organization dedicated to sharing GTFS data; advocating for best practices in data quality; and providing documentation, information and training. The second, Transitland, is a community-edited, open data GTFS platform.

A concordance file was developed to match feeds between sources and generate a unique ID for each feed. Feeds were downloaded directly from transit agencies with openly available online GTFS data. A small number of agencies were also identified as having GTFS-formatted public transit data, but the data were not openly available through the agency website or an API. In these cases, an attempt was made to contact the agencies by email.

Data compilation and validation

To confirm that all regions were properly covered, API-sourced feeds were manually mapped and checked to confirm that they corresponded to the expected coverage region. If there were any errors, feeds were removed or, wherever possible, replaced by another feed.

The API-derived feeds also occasionally included transit providers based in the United States. These cases were investigated individually. If it was determined that the service was not operating any stops in Canada, these feeds were removed and noted for future inventory purposes.

GTFS feeds were validated using the MobilityData Canonical GTFS Schedule Validator. The validator tests for compliance with the General Transit Feed Specification Reference and GTFS Schedule Best Practices and generates a list of compliance violations or items otherwise affecting data quality. These are classified by severity (errors, warnings and information). A summary of errors, warnings and information notices found for each feed is provided in a data validation CSV file, released in addition to the primary dataset. The data validation CSV file is intended to help researchers assess the quality or flag any potential errors that may be present in feeds included in the Canadian Public Transit Network Database.

The gtfstools package in R was used to validate, manipulate and merge GTFS feeds into the final dataset. One significant issue regarding compilation arose when identifying and dealing with duplicate feeds (i.e., feeds covering the same region). Feeds with identical information were identified and flagged with gtfstools. Cases where two feeds overlapped (for example, one feed is covered by another larger feed) were identified using the QGIS software package. All duplicate cases were investigated individually to ensure that the up-to-date, active and comprehensive feed was selected and that no data were omitted by mistake when dropping duplicate feeds.

Data summary

The final compilation of data contains a total of 139 feeds that cover all 10 provinces and two territories. This total also includes interregional bus, train or ferry services that operate between municipalities or regions.

All agencies included the six required GTFS files. However, some had error warnings that were found during the data validation step. Several optional GTFS files were collected where available but were not present from all data providers. Information on the files provided by each agency is included in the validation summary file included in the download folder.

The final Geopackage file contains a stops layer (points) and a shapes layer (lines). To reduce file size, the shapes layer geometry was simplified using a tolerance of 250m.

Data use

Limitations

Data in the Canadian Public Transit Network Database were compiled from open data sources. Other than a validation check for compliance with GTFS industry standards, data quality is taken “as is.” The accompanying data validation CSV file is intended to provide users with some indication of feed quality. In this version of the dataset, no attempts were made to fix errors or warnings in feeds.

Future improvement

For future iterations of the Canadian Public Transit Network Database, the Urban Data Lab aims to pursue partnerships and discuss data sharing opportunities with transit providers that do not currently upload GTFS data. The aim is to improve the accuracy and coverage of the dataset by ensuring that all municipalities or regions with public transit data are represented.

The Urban Data Lab is also investigating the possibility of publishing GTFS Realtime data in future releases.

Acknowledgement

The Canadian Public Transit Network Database was made possible by the ongoing efforts of organizations that produce and maintain open data and by the cooperation of organizations that have either given permission to include their publicly available data or directly provided their data for release as open data. The contribution and assistance of these organizations are gratefully acknowledged. This data was supported by Housing, Infrastructure and Communities Canada (HICC).

How to cite this dataset

Statistics Canada. (2025). Canadian Public Transit Network Database. https://www150.statcan.gc.ca/n1/pub/23-26-0003/232600032025001-eng.htm

Contact

For queries, corrections or omissions, please contact us at statcan.lode-ecdo.statcan@statcan.gc.ca. Please include the title of the open database in the subject line of the email.