Title

General Transit Feed Specification (GTFS)

Your Name Luyu Liu

Affiliation Auburn University

Email Address luyuliu@auburn.edu

Word Count

3580 words

Abstract

This entry explores the history, structure, applications, and challenges of both GTFS and GTFS-RT and illustrates their impact on transit systems and their role in shaping future transportation analytics. The General Transit Feed Specification (GTFS) is the *de facto* open data standard for the representation, communication, and analysis of public transportation systems. It provides a foundation for trip planning, transit mapping, and analytics applications. GTFS includes several text-based files that describe key transit elements, such as routes, trips, stops, and operating schedules. It is highly standardized, accessible, and flexible, which makes it a valuable resource of transit schedule for developers, transit agencies, and researchers. Beyond the schedule, GTFS Realtime (GTFS-RT) is introduced to provide updates on vehicle positions, trip changes, and service disruptions. GTFS-RT supports better-informed travel decisions in real-time trip planning apps and has become a critical tool for performance monitoring and system optimization.

Main Text

The General Transit Feed Specification (GTFS) is an open data standard for representing, communicating, and analyzing public transportation systems. It enables transit agencies to publish schedule data in a structured way that can be consumed by trip planning applications, mapping services, and transportation analytics platforms. Due to its high adoption rate, especially in the United States, GTFS or GTFS static is now the *de facto* standard for representing digitalized public transit schedule data. Meanwhile, as the demand for real-time data continues to grow and becomes increasingly essential for users' travel planning, the need for more dynamic and timely information has driven the development of *GTFS Realtime (GTFS-RT)*. This extension of the GTFS standard enables transit agencies to share real-time updates, such as vehicle locations, delays, and trip cancellations. This passage introduces the history, structure, and many applications of both GTFS static and GTFS *Realtime* data.

1. GTFS Static

GTFS was initially conceptualized and developed in 2005 between a collaboration between Google and TriMet, the public transit agency from Portland, Oregon, USA (McHugh, 2013). With the launch of Google Maps in 2005, which provided evolutionary experience for driving navigation, there is a strong demand for seamless and automated trip planning and navigation experience for public transit services. To achieve this, a prototype version of GTFS – then called Google Transit Feed Specification – was developed as a standardized format to share the administrative data in the agency's database to Google. Till 2006, the first version of GTFS was published under a Creative Commons License and implemented in multiple cities. By 2010, the GTFS format name was changed to the General Transit Feed Specification and almost 85% of transit miles travelled in the US were covered by open data covered by open data published by transit authorities (Barbeau & Antrim, 2013). GTFS as a data format has been growing rapidly; over 10,000 agencies in more than 100 countries have been using GTFS (GTFS.org, n.d.).

GTFS is organized as multiple relational tables, which resembles the structure of a relational database. Each table defines a separate entity (e.g., stop and route), and tables can be joined by different foreign keys (e.g., trips can be joined to routes with trips' route ID). Below are some necessary tables (GTFS.org, n.d.):

- agency.txt – Basic details about the transit agency, such as its name, website, and time zone. There is usually only one record in the file.
- calendar.txt – The regular operating patterns on a weekly basis. Records in this file represent different plans for different weekdays and their corresponding duration. There are usually only three records for weekdays, Saturday, and Sunday.
- calendar_date.txt – The exception to the regular operating patterns. Records in this file define the specific date of different operating patterns, such as holidays and scheduled maintenance.
- fare_rules.txt and fare_attributes.txt. Fare rules and attributes of the system.
- stops.txt –Transit stops. Records in this file define stop's id, latitude, and longitude coordinates.
- trips.txt – Details of trips. A record in this file defines a single journey of a vehicle along a specific route at a specific time. Note that trips are not only defined by their travel time, but also their direction (*direction_id*, e.g., 0 or 1) and service schedule (*service_id*, e.g., weekdays or Saturday).
- routes.txt – Details of transit routes. A route is defined as a collection of trips, including names, ID, types (bus, train, ferry, etc.), and descriptions. Note that trips with same route ID can still have different geometries and timetable patterns. This also means that routes are a rather loose definition. For example, some routes can have extensions or branches; some bus routes can also have different geometries for different directions due to one-way traffic. Therefore, there will not be an explicitly defined timetable for routes. Instead, the timetable of each trip is explicitly defined in stop_time.txt (discussed below); in fact, even two trips have identical geometries and timetable patterns, their schedule will also be separately defined in the stop_time.txt. This redundancy guarantees the correctness of the schedule information.
- stop_times.txt – Timetables of each trip specifying when vehicles arrive at and depart from stops. A record in this file is a time point when a bus passes by a stop. Note that some agencies may not provide arrival or departure times for every stop, as they are not mandatory in the GTFS. Instead, some systems will only provide time for timepoints, i.e.,

the bus stops at which arrival/departure time is explicitly defined and strictly observed by the transit operators.

- shapes.txt (optional) – Geographic shapes representing the actual paths of transit routes. Many systems are also using the file to interpolate the arrival/departure times between timepoints.

There are other optional files, such as transfers (GTFS.org, n.d.).

GTFS has many advantages, which can be generally categorized into three key perspectives: *standardized, accessible, and flexible*.

Standardized. GTFS, as its name suggests, is a technical standard that precisely identifies and defines the format and structure of all its files. This provides a uniform protocol for agencies and applications to produce and consume schedule data. Therefore, the interoperability of GTFS ensures that transit data can be widely shared, integrated, and used across various systems without requiring custom conversions or modifications. GTFS is also designed to handle transit data of varying complexities, from small local agencies with a few routes to large metropolitan systems with extensive networks. Its structured format allows for efficient storage, processing, and expansion as transit services grow or change over time.

Accessible. GTFS is designed to share the transit schedule as open data, which makes transit information more accessible by allowing agencies to share schedules and route data in an open, machine-readable format. For agencies, even small agencies can publish their transit data without needing expensive custom systems. GTFS consists of several comma-separated values (CSV) files, which are essentially well-defined text files with much smaller sizes. For developers, open-source tools can consume the dataset without major barriers, even for new transit applications. This openness creates a more inclusive transit ecosystem and enables third-party developers, researchers, and policymakers to analyze, enhance, and build upon existing transit information. For commuters, by lowering technical and financial barriers, GTFS helps improve public access to transit schedules and supports the development of new mobility solutions that benefit riders.

Flexible. GTFS files are flexible and context-free, which means that they can accommodate a wide range of transit systems, from large urban networks to small rural routes, without requiring specific adjustments or modifications. The format allows agencies to define their own data attributes, such as custom route types, specific fare structures, or unique trip characteristics, while maintaining compatibility with common tools and applications. This flexibility also means that GTFS can be extended to support emerging needs, such as real-time data or demand-responsive transit, making it adaptable to the evolving nature of public transportation.

Due to the three advantages above, GTFS data is the backbone for trip-planning apps like Google Maps, Apple Maps, and Transit App. These apps use GTFS to provide users with up-to-date trip information, route options, and directions for public transportation. As a part of Mobility as a Service (MaaS) platform, which combines various transport options into a single, seamless service, GTFS enables the integration of public transit, ride-hailing, bike-sharing, and dockless scooter sharing into a unified service. GTFS data can also be integrated with fare systems to track fare structures across different routes and transit modes.

Furthermore, GTFS promotes open data by allowing transit agencies to share their schedules, routes, and operational data with developers and other stakeholders. This transparency fosters innovation and supports the development of third-party applications, websites, and services. Public access to GTFS data further allows citizens and advocacy groups to track services, submit feedback, and suggest improvements. As the foundation of the open-source ecosystem, developers can use GTFS data to build open-source applications that extend transit services in innovative ways, such as crowd-sourced tracking, custom route planning, or transit visualizations.

GTFS also have many applications other than information sharing and trip planning. GTFS has been widely used in the planning and optimization of public transportation as the source of system schedule and geographic information. GTFS data allows transit researchers to analyze the efficiency of their routes, schedules, and stop locations. Agencies can also identify areas with low service coverage or high demand to optimize operations and improve service. Transit planners can also use GTFS to simulate service conditions, identify overcrowding, and plan for future demand. Data analysis helps agencies make data-driven decisions on expanding or reducing services.

Among various topics, accessibility analysis is a key focus in GTFS-based transportation analytics. With different accessibility measures, researchers and urban planners use GTFS data to assess how accessible public transit is to different populations, especially those with limited mobility or in underserved areas. Based on the accessibility analysis, planners can improve equity and service coverage in cities. For example, the application of GTFS data within the 15-minute city framework helps planners identify areas where public transit is inadequate. By optimizing routes and frequencies, planners can ensure residents can reach key services—like schools and grocery stores—within a 15-minute walk, bike ride, or transit trip.

For research purposes, creating GTFS data is a fundamental step for studying transit systems, even when official GTFS data is unavailable. Given that GTFS has become the industry standard for public transit data, researchers have been generating their own GTFS datasets using tools like the National RTAP's GTFS Builder (RTAP, n.d.) or custom scripts, such as developing countries like Ulaanbaatar, Mongolia (J. Kim et al., 2023) and historical simulations like 1930s colonial Seoul (Y. Kim et al., 2022).

In conclusion, GTFS static can serve as a uniform, robust, and context-free information source of public transit schedule, which serves as a foundation for the development of innovative transit applications, data-driven decision-making, and inclusive urban planning. However, there are still some criticisms towards GTFS as a format. A recurring comment is its text-based nature, which was originally meant to be as simple as possible so that agencies could easily edit the data with any editor (e.g., Microsoft Excel), especially for smaller, less-funded agencies (McHugh, 2013). Since all GTFS files are text-based and organized in CSV format, they lack the built-in checks and balances that databases provide, such as data types, relationships between tables, and referential integrity. This makes it easier for issues such as missing values, inconsistent identifiers, or incorrect file formatting to occur, which can lead to faulty or incomplete data being distributed to applications and users.

As a result, thorough validation and regular data audits are necessary to ensure the accuracy and quality of GTFS data, which often requires custom scripts or third-party tools for validation and quality assurance (MobilityData, n.d.). Despite the efforts to make it more applicable and accessible for most agencies, a recent study still found that agencies with higher ridership and those providing lower shares of a region's total vehicle revenue kilometers still tended to adopt

GTFS earlier, which suggests that larger agencies with higher ridership may have had more resources or incentives to implement GTFS (Voulgaris & Begwani, 2023).

2. GTFS Realtime

Despite rich information contained in the GTFS static data, like its name implies, the data only contains static schedule information in the system. Due to the limitations of static data, such as its inability to reflect real-time conditions like delays, cancellations, or unexpected changes, it is not sufficient for providing accurate, up-to-the-second information to passengers.

To address the gaps, GTFS Realtime (GTFS-RT) was introduced by Google in 2010 as an extension of the original GTFS specification. The first major adoption of GTFS-RT occurred when several transit agencies, including San Francisco's Muni, began using it to provide real-time data to passengers. Since then, GTFS-RT has grown in popularity and adoption, with many cities and transit operators worldwide integrating real-time data into their systems.

GTFS-RT consists of three components: *vehicle positions, trip updates,* and *service alerts*. *Vehicle Positions* provide real-time information about the location of vehicles (buses, trains, etc.) on a specific route. It includes the vehicle's current position, speed, and which trip or route it is serving. This data helps applications track vehicles in real time, allowing riders to know exactly where their vehicle is and when it will arrive. This data resembles the Automatic Vehicle Location (AVL) data, which is widely used in transportation performance surveillance and planning.

Trip Updates convey changes to the scheduled trips, including delays, cancellations, or any modifications to the planned schedule. For instance, if a bus is running late due to construction, trip updates will update the arrival/departure time (i.e., estimated time of arrival/departure, or ETA/ETD) at each bus stop in real time according to the time delayed at the moment. In other words, unlike vehicle positions which focus on each bus's status, trip updates focus on each trip's status at each bus stop. Meanwhile, note that the GTFS-RT is a data format that is only for information transmission; the rules for calculating the ETAs are not included in the specification. In fact, multiple third-party transit management software, such as Trapeze, can generate ETAs for trip updates.

Service Alerts are notifications about disruptions or changes in transit service. These could include service interruptions, closures, route diversions, or special alerts about weather conditions or emergencies affecting transit operations.

GTFS-RT is broadcasted by the transit agencies with an set interval, usually from 5 seconds to 2 minutes (Liu & Miller, 2020). To minimize the storage and transmission costs for real-time apps and transit agencies' servers, unlike GTFS static, GTFS-RT is stored and transmitted as Protocol Buffer binary file, whose size is significantly smaller than text-based data. On the other hand, this also means that the data cannot be directly opened by most software or as text file. This could further exacerbate the technology gap for less funded agencies.

GTFS-RT serves as the foundation for real-time trip planning and navigation. This enables users to make better-informed decisions about their travel, such as adjusting their departure times or switching routes in response to unexpected delays. However, the availability of real-time will not always result in reduced waiting time; this is due to delayed buses trying to make up time after being late and low frequency of real-time updates (Liu & Miller, 2020). In some extreme cases, unconditionally believing real-time information will result in consistently missing buses.

GTFS-RT can also serve as an important empirical data source for transit system optimization and surveillance. Researchers and planners have been archiving the historical GTFS files and combined both static and real-time data for various applications such as analyzing on-time performance, calculating reliability, and improving operational efficiency. This includes the spatiotemporal patterns of delays (Park et al., 2020), simulation of general traffic (Jiang et al., 2024), accessibility measures with consideration of real-time data (Liu et al., 2022), and equity analysis of on-time performance (Karner et al., 2016). The extensive archive also allows for thorough longitudinal analysis at a highly localized level over long timeframes (Liu et al., 2024).

The introduction of GTFS-RT represents not just a methodological advancement, but a fundamental shift in understanding how transit systems operate and should be measured. It exposes some major drawbacks of schedule-based analyses. First, it is important to note that schedules alone do not reflect the on-time performance of the system, and there can be systematic errors due to delays. Second, the availability of real-time data fundamentally changed the travel behavior of many transit users, as it allows them to make more informed decisions based on current conditions. Therefore, following the previous changes in the travel behavior, some generally correct assumptions widely adopted by prior studies now become invalid. For example, traditional studies assumed half of the headway as the average waiting time, which is now significantly lower after the adoption of real-time data (Watkins et al., 2011).

In conclusion, while GTFS static data has long been a valuable resource for transit scheduling, its inability to reflect real-time conditions highlights the need for more dynamic data systems. GTFS-RT addresses this gap by offering passengers more accurate and timely information. The introduction of GTFS-RT has revolutionized transit systems by not only improving the passenger experience through better-informed travel decisions but also offering a rich data source for performance analysis and system optimization. We expect to see a higher adoption rate of GTFS-RT data.

It has been more than a decade since the first introduction of GTFS-RT, but there are still some major challenges that prevent more agencies and researchers from utilizing it. A primary issue is its technical barrier. Although a binary format can significantly reduce the transmission costs, it is difficult for many transit agencies, especially those with limited resources, to implement and maintain the system. A GTFS Realtime server may need to handle millions of requests per minute, which requires substantial funding and resources for the development and ongoing maintenance of a reliable and efficient server. The lack of standardized tools and infrastructure for integrating GTFS-RT data with existing systems further complicates its adoption. However, as cities continue to integrate real-time data, further advancements in both technology and research will continue to refine and enhance its utility.

SEE ALSO:

wbieg0864

wbieg0617

wbieg2219

wbieg0348

wbieg0211

wbieg2015

wbieg0019

wbieg0403.pub3

wbieg1114

Further Readings

Wessel, N., Allen, J., & Farber, S. (2017). Constructing a routable retrospective transit timetable from a real-time vehicle location feed and GTFS. Journal of Transport Geography, 62, 92-97.

Liu, L., & Miller, H. J. (2022). Measuring the impacts of dockless micro-mobility services on public transit accessibility. Computers, Environment and Urban Systems, 98, 101885.

Wessel, N., & Farber, S. (2019). On the accuracy of schedule-based GTFS for measuring accessibility. Journal of Transport and Land Use, 12(1), 475-500.

Devunuri, S., & Lehe, L. (2024). A Survey of Errors in GTFS Static Feeds from the United States. Findings.

Liu, L., & Miller, H. J. (2021). Measuring risk of missing transfers in public transit systems using high-resolution schedule and real-time bus location data. Urban Studies, 58(15), 3140-3156.

Key Words

Public transit; Geospatial data science; Urban computing; General Transit Feed Specification.

Reference:

Barbeau, S. J., & Antrim, A. (2013). The Many Uses of GTFS Data – Opening the Door to Transit and Multimodal Applications. *ITS America 2013*. http://prezi.com/-69luw8sfabp/the-many-uses-of-gtfs-data-its-america-april-2013/

GTFS.org. (n.d.). *Why use GTFS? - General Transit Feed Specification*. Retrieved February 12, 2025, from https://gtfs.org/getting-started/why-use-GTFS/

Jiang, S., Sun, Y., Wong, W., Xu, Y., & Zhao, X. (2024). Real-Time Urban Traffic Monitoring Using Transit Buses as Probes. *Transportation Research Record: Journal of the Transportation Research Board*, 03611981241260708. https://doi.org/10.1177/03611981241260708

Karner, A., Golub, A., Chavis, C., at Maryland, N. T. C., & and Technology, O. of the A. S. for R. (2016). *Understanding Regional Disparities in Public Transit Performance Using Realtime Transit Data*. *March*, 28p.

Kim, J., Rapuri, S., Chuluunbaatar, E., Sumiyasuren, E., Lkhagvasuren, B., Budhathoki, N. R., & Laituri, M. (2023). Developing and evaluating transit-based healthcare accessibility in a low-and middle-income country: A case study in Ulaanbaatar, Mongolia. *Habitat International*, *131*, 102729.

Kim, Y., Kim, J., Ha, H. J., Nakajima, N., & Lee, J. (2022). Job Accessibility as a lens for understanding the urban structure of colonial cities: A digital humanities study of the colonial Seoul in the 1930s using GIS. *ISPRS International Journal of Geo-Information*, *11*(12), 614.

Liu, L., & Miller, H. J. (2020). Does real-time transit information reduce waiting time? An empirical analysis. *Transportation Research Part A: Policy and Practice*, *141*, 167–179.

Liu, L., Porr, A., & Miller, H. J. (2022). Realizable accessibility: Evaluating the reliability of public transit accessibility using high-resolution real-time data. *Journal of Geographical Systems*, 1–23.

Liu, L., Porr, A., & Miller, H. J. (2024). Measuring the impacts of disruptions on public transit accessibility and reliability. *Journal of Transport Geography*, *114*, 103769.

McHugh, B. (2013). Pioneering Open Data Standards: The GTFS Story. In *Beyond Transparency: Open Data and the Future of Civic Innovation*. Code for America Press. https://beyondtransparency.org/part-2/pioneering-open-data-standards-the-gtfs-story/

MobilityData. (n.d.). *Canonical GTFS Schedule Validator*. Retrieved February 12, 2025, from https://gtfs-validator.mobilitydata.org/

Park, Y., Mount, J., Liu, L., Xiao, N., & Miller, H. J. (2020). Assessing public transit performance using real-time data: Spatiotemporal patterns of bus operation delays in Columbus, Ohio, USA. *International Journal of Geographical Information Science*, *34*(2), 367–392. https://doi.org/10.1080/13658816.2019.1608997

RTAP. (n.d.). *GTFS Builder*. GTFS Builder. Retrieved February 27, 2025, from https://www.nationalrtap.org/Technology-Tools/GTFS-Builder

Voulgaris, C. T., & Begwani, C. (2023). Predictors of early adoption of the general transit feed specification. *Findings*. https://findingspress.org/article/57722-predictors-of-early-adoption-of-the-general-transit-feed-specification

Watkins, K. E., Ferris, B., Borning, A., Rutherford, G. S., & Layton, D. (2011). Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transportation Research Part A: Policy and Practice*, *45*(8), 839–848.