

MIT Open Access Articles

Applying the General Transit Feed Specification to the Global South

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Eros, Emily, Shomik Mehndiratta, Chris Zegras, Kevin Webb, and Maria Ochoa. "Applying the General Transit Feed Specification to the Global South." *Transportation Research Record: Journal of the Transportation Research Board* 2442 (December 2014): 44–52.

As Published: <http://dx.doi.org/10.3141/2442-06>

Publisher: Transportation Research Board of the National Academies

Persistent URL: <http://hdl.handle.net/1721.1/100710>

Version: Original manuscript: author's manuscript prior to formal peer review

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



**APPLYING THE GENERAL TRANSIT FEED SPECIFICATION (GTFS)
TO THE GLOBAL SOUTH: EXPERIENCES IN MEXICO CITY AND BEYOND**

Originally submitted on Aug 1, 2013

Revised document submitted on Feb 17, 2013

Word Count: 6,314 words + 2 tables = 6,814 words excluding references

Emily Eros (Corresponding author)

ejeros@mit.edu

Massachusetts Institute of Technology

Department of Urban Studies and Planning

77 Massachusetts Avenue

Cambridge, MA 02138

T 603-306-2354 / F 617-253-2654

Shomik Mehndiratta

smehndiratta@worldbank.org

The World Bank Group

Latin America and Caribbean Region

1818 H St, NW

Washington, DC 20433

T 202-458-9691 / F 202-676-9594

Chris Zegras

czegras@mit.edu

Massachusetts Institute of Technology

Department of Urban Studies and Planning

77 Massachusetts Avenue

Cambridge, MA 02138

T 617-452-2433 / F 617-258-8081

Kevin Webb

kwebb@conveyal.com

Conveyal

1133 15th St NW, Floor 12

Washington, DC 20005

T 202-480-9322 / F 202-644-9263

Maria Catalina Ochoa

mchochoa@worldbank.org

The World Bank Group

Latin America and Caribbean Region

1818 H St, NW

Washington, DC 20433

T 202-473-9980 / F 202-676-9594

ABSTRACT

A combination of open data tools and methods, facilitated by data format standardization, has started changing business-as-usual in the transit industry. The General Transit Feed Specification (GTFS) has become the *de facto* standard for releasing public transit route and schedule data. This paper analyzes this rapidly evolving transit information sector through the Mexico City experience. The case illustrates that even a mega-city with several different transit providers can create a fully-functional GTFS feed in a matter of weeks and obtain the benefits of work done elsewhere; thanks to the global open data ecosystem, a range of important free or low-cost applications – customer-facing applications and planning tools – can immediately capitalize on these data. However, the Mexico experience also reveals an important limitation of GTFS in its current form: its inability to easily accommodate semi-structured public transit services common in many developing world cities. An adaption to GTFS developed in Mexico City to address this limitation is described. Finally, the case reveals significant untapped potential to maximize the value of this open-data ecosystem, particularly for planning and regulatory tools.

INTRODUCTION

Over the past decade, government agencies around the world have demonstrated increased willingness to collect and disclose public transit data. The advent of a common data format facilitated this process. Today's *de facto* standard emerged from the USA. In 2005, Portland's (Oregon) transit agency, TriMet, partnered with Google to integrate public transit schedule and route information with Google Maps. In the process, TriMet and Google co-developed a non-proprietary transit data format, later titled the General Transit Feed Specification (GTFS), to standardize and facilitate data release for others to follow suit. GTFS consists of a package of comma-delimited text files, each of which contains one aspect of the transit information and a set of rules on how to record it: six mandatory files (agency, stops, routes, trips, stops times, and calendar) and seven optional files (calendar dates, fare attributes, fare rules, shapes, frequencies, transfers and feed info),

To accommodate the varied nature of transit services, an on-line community process regularly modifies GTFS by adding extensions, optional fields, and additional valid responses (1). The GTFS file format initially managed only static transit information (e.g., routes, stops, and schedules), not dynamic information (e.g., real-time bus locations). In 2011, however, the GTFS-realtime (GTFS-RT) data feed specification was designed and released by a partnership of agencies, software developers and Google. GTFS-RT is designed to provide live updates on transit fleets (e.g., drawing from Automated Vehicle Location [AVL] systems and the static GTFS feed) and be interoperable with GTFS. Although Google does not own or explicitly manage the GTFS, its hosting of the relevant community dialogues institutionalizes modifications.

The Confluence of the Transport Open Data Movement and GTFS

The open data movement arises from philosophical principles of open government, transparency, and accountability, and practical motivations related to increased returns on public investment, downstream wealth creation, more potential brainpower brought to examining complex problems, and enhanced public policy and service delivery (2). For transportation, the open data movement has fundamentally shifted how agencies communicate with users as an increasing number move from tightly controlling data and the products derived from them, towards generating and releasing data with minimal control over the end products. The open data movement has moved governments beyond pure data release to attempt that such releases follow some key principles – completeness, primacy, timeliness, ease of physical and electronic access, machine readability, non-discrimination, use of commonly owned standards, licensing, permanence and non-usage costs – intended to empower citizens to use the data (3).

In transportation, the confluence of open data, GTFS, and increasingly ubiquitous mobile computing, sensing and communication technologies (epitomized by the “smartphone”), has spurred numerous technical innovations from a range of actors. Tools include applications that assist with trip planning, ridesharing, timetable creation, data visualization, planning analysis, interactive voice response, and real-time information provision (4). Together, GTFS and GTFS-RT enable transit agencies and operators to engage the power of the software developer community and citizenry more generally to create new forms of information services about public transportation. GTFS also enables

new forms of comparative assessment across public transportation systems (e.g., “benchmarking”) and new service modeling possibilities (5).

GTFS Goes Global

The GTFS’s simple file structure prompted rapid global adoption: as of November 2013, Google lists 229 public transit agencies around the world that release official GTFS feeds available for developers to use (6). If private transit companies are included, estimates range from 703 (7) or 1,048 (8). GTFS feeds range from covering all public transportation services for a particular region to a single provider.

While concentrated in the Global North, GTFS experiences are also emerging in low- and middle-income cities (Table 1). This paper focuses on one such experience, the recent deployment of GTFS feeds in Mexico City.

DATA COLLECTION AND GTFS FEED GENERATION IN MEXICO CITY

Mexico City (the Federal District or DF) and its metropolitan area (MCMA) epitomize today’s megalopolitan challenges. The DF, itself, represents essentially a single jurisdiction (one Mayor) with approximately 8.9 million persons, yet the broader MCMA encompasses some 40 additional local jurisdictions across two states and another 12 million people, posing institutional and operational challenges for transport and other sectors. This case focuses almost exclusively on services in the DF, where since 1975, the transportation secretariat (SETRAVI) has regulated both technical and non-technical aspects of public transportation planning and policy.

SETRAVI oversees six relevant services in the DF (not including taxis); the government serves as operator (e.g., STE) or regulator (e.g., DGT) (Table 2). Except for a few lines of the Metro (STC), DF services do not extend into the broader MCMA. GTFS data collection included one metropolitan-scale transit service, the single line commuter rail (*Tren Suburbano*), operated privately and regulated by the National Government. The privately-operated Ecobici bike share system was brought along in the process, despite not fitting into the GTFS feed, because the Mexico City authorities consider it important that multi-modal journey planners can seamlessly include information on Ecobici station sites and bicycle availability. Ultimately, expanding the GTFS efforts into the rest of the MCMA, while only further complicating an already complex institutionality, will be crucial, as more than half of the metropolis’ travel demand originates outside the DF where some 70,000 minibuses and buses ply the streets on numerous loosely organized routes.

Existing Agency Capacity and Data Availability

Relative to the “birthplace” of GTFS, the DF offers a much more complex institutional setting, particularly due to the heavy presence of loosely regulated private bus and minibus operators. STE, STC, Metrobús, and the *suburbano* all operate vehicles on dedicated infrastructures, making route and stop location information relatively straightforward. Each of these agencies collects and stores this information as ArcGIS shapefiles or KML files compatible with Google Maps. The agencies also have approximate paper schedules. STC, Metrobús, and the *suburbano* use electronic fare cards and the light rail uses tickets and turnstiles, enabling easy access to data on demand characteristics. Passenger counts are more difficult to obtain for RTP and the DGT-

regulated services, i.e., *colectivos*, where drivers collect fares and no digital means exist to track ridership.

The *colectivos* pose a particular challenge. Some 30,000 vehicles in 121 different route associations, *colectivos* account for about 50% of the DF's motorized trips. The DGT issues route licenses to route associations to operate on a set of route variants (*ramales*). The route associations combine fleet owners and smaller owner-operators who pay fees and dues in exchange for operating privileges and other support. This structure transfers most of the organizational burden and financial risk from the DGT onto *colectivo* unions and operators, but creates challenges in return. Opacity characterizes the system; the DGT does not even possess up-to-date information about *ramal* characteristics (route, drivers, vehicles), much less on demand patterns. This limits regulatory ability.

The route associations have little incentive to provide any data, making verification difficult. The DGT recently began surveying 10 of the more formalized *colectivo* corridors in spring 2013. The agency, supported by a non-governmental organization (NGO), collected basic route data (route, travel time, counts) using an iPhone running the MotionX-GPS app together with an Android tablet running a custom-made application. While successful, this technique remains effort-intensive – six person-days per *ramal* including four person-days of data collection.

Non-GTFS-Based Third Party Information Provision

STE, STC, Metrobús, RTP, and the *suburbano* publish route maps and schedule information on their websites, which software developers have been using to create mobile apps with transit maps and route planning capabilities. As of June 2013, 28 transit-related apps were available for Mexico City (19 Android plus 9 Apple iOS). Android apps for Mexico City have a combined download count of between 1.1 and 5.5 million (similar counts not available from Apple). Most of these apps simply show Metro station locations and trip-planning directions from the user's location – not integrated with, but simply overlaid upon, a map. A few apps also include this information for the electric trolleybuses, *suburbanos*, and Metrobús. The apps remain limited, with no reliable travel time information and oft-outdated maps. None of the apps include information for the regular bus or *colectivo* services.

Project Overview and Outputs

Within this context, Mexico City began generating its first GTFS feed by enrolling the relevant agencies in the process. Supported by a grant, the team developed a web-based Data Management Portal (DMP) and an Android application, TransitWand, to carry out the complementary fieldwork. TransitWand uses GPS signals to track a user's location and allows a user to mark stops, boarding/alighting times, and passenger counts. Users upload this information through WiFi or cell phone networks. The DMP was used to create, upload, view, and edit route data and convert them to GTFS. In April 2013, a series of workshop sessions with government officials and NGOs focused on the role and potential uses of open transit data and how to use the Android app and GTFS editor.

After the workshop, the transit agencies used the GTFS editor to convert their data into GTFS format. The collected data includes information for about 125 lines, 260 route variations, and over 5,000 stops, covering all the Metro, Metrobús, RTP, STE, and

suburbano. Since these agencies already possessed basic schedule information and KML files of routes and stops, the conversion process was relatively straightforward. Frequency data were more difficult to obtain, however, in some cases having to be estimated or substituted with default values.

For *colectivos*, route and stop location information was collected using TransitWand for nearly 1,100 *ramales* by October 2013. These remain excluded from the current GTFS feed, because of incomplete data and the challenges to making the data GTFS-compatible. *Colectivos* do not consistently adhere to fixed stop locations, routes, schedules, consistent headways, or trip times; the current GTFS design cannot model such networks.

The lack of formal stops along *colectivo* routes presents a significant challenge as precise stop locations and times are a fundamental GTFS building block. Prior attempts in other cities to model transport without fixed stops required the inclusion of “simulated stops” in the data feed. These approximated likely boarding/alighting locations at regular intervals, near intersections, or other obvious locations. This approach produced a data feed that could be used by any existing GTFS consumer. However, it burdens the feed producer to create and maintain a very large number of simulated stops in order to simulate continuous boarding and alighting along a given route. This is both impractical and semantically inaccurate.

A proposed extension to GTFS was developed as part of this project to allow the feed producer to indicate that continuous boarding and alighting is allowed along a given route (16). This change allows the feed to define a minimum of two stops along a trip and indicate that the GTFS consumer should interpolate intermediate boarding/alighting points between these stops at a specified frequency along the trip shape.

An additional modification was proposed to allow the inclusion of localized vehicle type names, considered necessary since the predefined list of GTFS mode types inadequately represents many transport options found around the globe, including *colectivos*. Accurate description of vehicle type is often critical for public communication. This change allowed feeds to include both language- and location-specific terminology to describe transport infrastructure.

Through subsequent discussion with the GTFS community, including GTFS producers in locations with similar needs (e.g., Manila), a refined version of the stop interpolation and vehicle type localization modifications are being proposed for formal inclusion in the GTFS protocol.

Schedules and journey times were modeled using GTFS’ existing frequency-based timetables. This allowed specifying relative travel times between stop locations and service headways on a given route for a specific period of the day. However, the current GTFS protocol defines timetables, including frequencies, to the second. For dynamic, semi-formal modes like *colectivos*, for which precise values are unknown or unrealistic, the ability to define journey time and headway variance could be a useful GTFS modification.

After creating the initial GTFS feed, SETRAVI held a second series of meetings and workshops from 18-22 June 2013: discussing with agencies the results and a live disruption-tracker tool the team created, releasing the GTFS feed on its website, discussing with Google data inclusion on Google Transit and having Google Transit engineers review and clear the data feed (a prerequisite to publish in Google transit),

promoting the launch of an open trip planner, and showcasing the data at a meeting oriented towards software and mobile-app developers. Additionally, non-profit organizations and SETRAVI jointly hosted a hackathon challenge in Mexico City searching for mobility solutions to specific problems. From 41 projects, seven winners were selected to enter a funded incubation program. Two of the winners use the GTFS feed: a digital map for the metropolitan area transit system, and a journey planner. The other projects focus on data collection and cycling tools.

Initial Outcomes

As of November 2013, only four months have passed since the GTFS data have been released. Though it is still early, this provides some indication of possible uses and outcome of the data.

Tracking GTFS Downloads

Monitoring actual GTFS data use is difficult. The DF's GTFS feed can be downloaded from GTFS Data Exchange and the DF government's official open data website. It will also feed into Google Transit (pending licensing agreements), open trip planners such as The Transit App, and any other apps developed from the data. Usage statistics cannot be obtained for third-party redistribution sites and Google will not share metrics on Google Transit usage. Therefore, SETRAVI can only obtain use-information for the official open data site. Early indications, reviewed one month after GTFS release, show an initial spike of about 120 downloads following the June 21 official data release. Downloads taper off but reach a total of about 300 unique IP addresses within the first week and 683 for the first month. The IP addresses come from 11 countries, the vast majority from Mexico (637 downloads), with additional concentrations in the USA (25) and the UK (7). Top cities included Mexico City (589 downloads) and Monterrey (20).

Use of GTFS Downloads

Once downloaded, the GTFS feed use can be monitored by the type and number of apps developed. As of November 2013, five apps made use of the DF data: The Transit App, Pdxtrian, Moovit, AGUMóvil, and Hop Stop. The Transit App offers web-based trip planning, while Hop Stop provides a similar service on Android, iOS, and Windows mobile phone platforms. Pdxtrian enables Android users to locate the nearest stop/station for their route. AGUMóvil (Android, iOS, and Blackberry) incorporates trip planning capabilities with traffic updates and road congestion information, and allows users to report potholes or lighting failures. Finally, Moovit (Android and iOS) currently provides real-time information for the *Métrobus* service. Moovit generates its data by sensing the location and movement of other individuals using the app. Additional apps and tools may be developed as a result of various events underway to promote the use of these data. SETRAVI staff have been following the development of these apps and will continue to promote the GTFS and monitor new products.

Press Coverage and Twitter Traffic

During the period after GTFS feed publication, mainstream and online press covered the event substantially. A preliminary assessment from traditional media reflects (i) a general enthusiasm with the topic and a perception of it being a step towards transparency and the digital era, and (ii) an overall limited understanding of the transport open data value

proposition – given that the GTFS release was repeatedly portrayed as the launch of a mobile app instead of the launch of an open data feed. Monitoring social media offers another means to examine public reactions to, and uses of, Mexico City’s GTFS feed. Within the two-week period after the data launch, users tweeted about the GTFS project over 80 times. This Twitter traffic is too low to draw conclusions, but the site could provide some sense of public reaction to transit apps moving forward.

Forward-Looking Applications

The GTFS data have yet to play a significant role in internal data management and analysis. However, SETRAVI has been exploring two potential avenues. First, Open Trip Planner – an open source, rider-oriented, point-to point itinerary planner – also supports planning analysis (OTPA) by generating an accessibility coverage map utilizing the same underlying transport network used for the journey planner. This analysis can be used to measure raw travel time accessibility or, combined with demographic/employment information to generate aggregate transit opportunity accessibility indicators (e.g. “100,000 jobs can be reached in 20 minutes by transit from this location”), or used to compare various transport scenarios as modeled in GTFS. This feature has been used to show accessibility impacts of disruptions, such as Hurricane Sandy in New York (17). Such tools can inform policymakers and planners and help explore questions about mobility and accessibility at different spots in the city. Second, analyzing the type and frequency of service interruptions logged in the disruptions feed would facilitate pattern recognition, potentially helping SETRAVI determine where operational improvements could be made.

In addition, collecting spatial data for each of the 1,100 *ramales* is enabling creation of the first comprehensive database of the microbus routes and stops. Previously non-existent, this basic information should be valuable as SETRAVI tries to formalize and improve bus and rapid transit services.

Potential for Regulatory Impacts

Despite the overt and direct uses in planning applications, the data availability and transparency enabled by Mexico City’s GTFS may have an even stronger impact upon relationships between transit regulators and transit operators – particularly the *colectivos*. As discussed, the *colectivo* system is characterized by information opacity, disorganization, poor integration with other transit modes, and a lack of regularization. Preliminary interviews with government and route association personnel suggest that the enhanced transparency and data availability facilitated by GTFS-based platforms may help the city with at least one key regulatory task: negotiations with route associations on compensation for loss of operating rights. As part of an effort to enhance public transport quality, the city is implementing an ambitious BRT program (*Metrobús*’ five corridors since 2005) and other pilot projects to formalize bus routes. Every such change implies drawn-out and costly negotiations (12-18 months) with the affected route association operating the *ramales*, often requiring extensive primary data collection on drivers, vehicles and *ramales* affected, and route and ridership characteristics. Both government and route associations have incentives to under-/over-report impacts (demand, service levels and traffic volumes), so little trust exists amongst parties. A GTFS-based system could provide much of the data and facilitate easier, less contentious data collection and

negotiations. Beyond lowering the financial and opportunity costs of negotiations over bus system reform, GTFS-based data could inform longer-range route planning and restructuring. In theory, these data can increase government's regulatory abilities and institutional capacity. *Colectivo* operators appear willing to participate in GTFS-related data collection efforts, expecting to gain in terms of costs and efficiency of dispatching; however, the institutional complexities underlying this loosely regulated system and the benefits to some actors of system information remaining opaque cannot be discounted.

DISCUSSION

This case reveals the interrelated benefits and challenges associated with global adoption of GTFS. While drawing primarily on the lessons from Mexico in this respect, the discussion also draws from other experiences (see Table 1).

Measuring Benefits and Impact of GTFS

Ex-ante knowledge of the benefits and costs should, theoretically, precede GTFS adoption. The costs, though modest relative to capital investments and the like, are not trivial, requiring technical capability and data gathering. The benefits can be difficult to quantify, particularly in the short term. Though the evidence remains limited, experiences thus far suggest three categories of benefits: (i) to passengers and potential users of higher-quality information on services; (ii) to operators and regulators from the use of analytic and monitoring tools; and (iii) to society more generally of operating in an open-data ecosystem.

Users

Though spreading fast, GTFS remains a recent phenomenon, only having been established in mid-2006. GTFS adoption, and the open data movement that has accompanied it in many contexts, has resulted in numerous new forms of transit information provision, including third-party outward-facing apps providing trip planning and bus arrival time predictions. Nonetheless, impact evaluations for users remain limited, and concentrated primarily in the Global North.

Theory suggests that improved access to high-quality transit service information will increase current riders' satisfaction levels – by, for example, reducing real and/or perceived wait times – and raise future ridership by improving knowledge about services and/or service quality relative to alternatives. Limited research has been conducted to measure these impacts, even for “traditional” forms of transit information provision such as static paper maps (18). With the increasing ubiquity of real-time information, enabled by GTFS-RT and the open data movement, a growing number of studies have focused on user impacts. The developers of Seattle's (Washington) OneBusAway, a suite of open source tools to deliver real-time bus location information and wait times to a range of mobile devices, implemented an on-line survey among OneBusAway users. The great majority (92%) of respondents reported that their overall satisfaction with public transit had improved as a result of using OneBusAway; a similar share reported lower waiting times; and, modest increases in non-commute trip-making were also reported (19). A follow-up survey, carried out at bus stops, found OneBusAway users to have lower real and perceived wait times, although no difference in self-reported “aggravation” levels; potentially confounding variables, such as income or employment were not included (20).

Using a panel survey, attempting to capture before and after effects of real-time passenger information on a university shuttle service, Zhang *et al.* found no quantifiable impact on rider frequency or mode choice (two weeks after the technology debuted), but did find increases in overall satisfaction and feelings of security in using the shuttles after dark (21). A longitudinal, route-level analysis of the Chicago Transit Authority's (CTA) bus system ridership, found modest average increases in bus use after introducing real-time information, with some evidence that routes affected in later stage rollouts had higher ridership effects, suggesting improved technology and/or technological diffusion and adoption (as third-party providers entered the game and personal mobile devices improved) (22).

This initial evidence suggests important promise, not just in ensuring a higher quality product for customers, but in attracting and keeping passengers who have a choice. Satisfying current and future "choice" riders in societies under rapid motorization might help sustain public transit patronage over time, implying longer term value in reduced pollution, congestion and other externalities. That said, as GTFS and open transit data moves into the Global South, user expectations, needs, and responses may differ from the Global North. For example, semi-formal services, such as *colectivos*, typically are not even represented on "traditional" transit maps, with route information coming from experience, others' knowledge (face-to-face social networks) and/or the vehicles/drivers/fare collectors. In such systems, much benefit may be gained from simply providing service maps, although map legibility and interpretability by users must be understood.

Relatedly, to maximize potential benefits to passengers, digital forms of information services must be matched to the needs and endowments of the passenger base. For instance, currently information from the DF's GTFS data is available primarily on the internet and via smartphone apps. But only 37% of Mexico's population has regular access to the internet (14), and web-based information may not reach the majority of transit users – especially low-income populations who tend to ride *colectivos* and currently have little access to route and schedule information. Moreover, many apps typically developed from GTFS feeds require a smartphone. This excludes much of the city's population, as smartphone penetration reaches just 8% nationwide (and is likely less than 20% even in Mexico City) (15).

On the other hand, there are over 21 million mobile phone connections in the DF alone, creating potential for SMS-based services to reach a much greater share of Mexico City's population. An SMS-based data service for the *colectivo* system would bridge a major information gap, reaching a large pool of potential users. Seattle's OneBusAway, offering information services on a range of platforms including SMS, provides an open source option (19). Authorities in Santiago, Chile have developed exactly such services providing bus arrival time and routing services.

Analysis/Regulatory Tools

Interviews and discussions with agency staff in Mexico indicate that increased transit information may inform internal planning and analysis and has the potential to change relationships between regulators and operators, particularly with respect to negotiations over route restructuring. Traditionally, in the DF new route authorizations were not subject to a structured process that checked for service duplication or for verification

between the proposed fleet and demand. In reality, the process was more political (both among operators and between operators and authorities) than technical (see, e.g., (22)). Especially among the semi-formal *colectivos*, the complex industrial structure matters. Agency officials noted that just having a more comprehensive and visual sense of all the routes helped anchor discussions with operators on route-related issues in a more technical and less political sphere. Planning tools such as OpenTripPlanner Analyst, that can help quantify the incremental accessibility benefits of particular routes to the system can further the role of technical analysis for route-related issues.

The implications of improved customer information on operators are understudied everywhere. Limited evidence from the USA (a single study in Seattle) suggests drivers were particularly in favor of services for disadvantaged groups (e.g., deaf, blind), generally supported service alerts (e.g., breakdowns), but were wary of information which would negatively impact route and ridership perceptions (e.g., “service reviews”) or publicize past performance data, which might impair the driver’s standing even for conditions outside of his control (23). Among both publicly and privately operated services in the DF, with longstanding often contentious histories of attempted reforms (e.g., (22)), the reaction by relevant agents to such information “reforms” will vary. López *et al.*’s (24) assessment of private bus operators’ general willingness to participate in transit improvements (not information services, per se) in Mexico may provide some hints; they find unclear legal standing, industrial structure (e.g., owner-operator), poor understanding of operating costs, among other factors to negatively correlate with support for transit improvements, with private operators in the MCMA even less receptive to changes. That said, if information transparency in an “open data model” can improve the level of trust in government, it could increase the political feasibility of system reform.

For broader planning purposes, the GTFS feed in the web portal allows government to visualize route configuration and better understand duplication, or to explore adding or eliminating services from under/over served areas. Moreover, in combination with the appropriate OTPA tool, GTFS can allow planning agencies to carry out advanced accessibility analysis and to visualize results via indicators or heat-map like images. Clearly more needs to be done on the tools side, to integrate different data sets (e.g. census, employment, land uses) with the detailed transport GTFS feed, and to properly extract and visualize results. The combined analysis of multiple sources of data with the GTFS feed yields the potential to help urban planning agencies in integrating land-use and transport within their planning efforts without having to use complex and onerous models. This may be especially relevant for small cities, and/or cities without the capacity to develop and maintain traditional transport models and who need to occasionally conduct simple analysis of accessibility and public transport demand.

Operating in an Open-Data Ecosystem

An open-data ecosystem can, in theory, lower the barrier to innovation and enhance cross-fertilization of tools, approaches and ideas. In Mexico City, for example, almost immediately after the release of the GTFS feed, a number of apps made use of these data to provide value to users. The DF’s suite of apps is already growing; all have been created by American, Canadian, and Israeli developers as transfers of previously-existing apps into the Mexico City environment. The nature of the GTFS format facilitates easy innovation transfers between different problems and contexts; as one city develops apps

around a particular problem, others can benefit with relatively little additional investment.

This is not limited to public-facing apps – it includes data collection as well as analysis and planning tools. However more must be done, particularly in expanding the reach of this open-data culture to “traditional” transport planning tools.

Beyond riders, transit data transparency may enable a cycle of information availability, public feedback, and government response. Rojas (25) raises these arguments but notes that little research exists on the outcomes of data disclosure initiatives, in terms of both of citizen mobility and improved performance from transportation agencies and service providers. The early evidence from the DF case suggests benefits: the GTFS feed process created an opportunity for SETRAVI to integrate all transport agencies into one mobility-related project. An outstanding question, however, relates to how the private sector will react to such information, how it will influence subsequent service reforms and, ultimately, whether and how it might influence metropolitan-scale service coordination and integration.

Key Challenges and Strategies

GTFS, now the *de-facto* standard for digital transit data release, was designed to accommodate scheduled transit systems in the USA that operate with fixed routes and stops. The Mexico City case study highlights a key current limitation: GTFS’ incompatibility with flexible services that operate without fixed stops or schedules. Many cities across Latin America, Africa and Asia share this predicament; research indicates that flexible transport services constitute over 90% of transit trips in cities like Algiers, Bamako, Dakar, and Dar es Salaam, and over 70% in cities like Accra, Bangalore, Caracas, Manila, and Tehran (26, 27, 28, 29).

In Mexico a work-around was found by creating a variant to the GTFS feed based on defining fixed stops at regular intervals combined with the possibility for users to assess travel times and connections from any point between stations. Headway estimates, based on existing knowledge (including vehicle counts and speed data) substituted for schedules. Teams working in two cities described in Table 1, Manila and Dhaka, also had to deal with this challenge. Like Mexico City, Manila chose to avoid schedules, instead providing headway estimates for their *jeepneys*. In terms of stop locations, the Dhaka team included stop location based on where the bus stopped during the data collection ride. Manila’s stops were interpolated every 500 meters along the route.

These types of workarounds enable the assembly and release of a GTFS feed, but possibly risking inconsistent and potentially inaccurate information. This may not pose any issue for users unaccustomed to schedule and stop location information. However, at best, these workarounds require significant time/effort and cost – generating data on stops/schedules that may not be meaningful. Developers may become confused about which fields may be reliable and which are estimates/constructs. Furthermore, inaccurate data could lower users’ confidence in the new information, making them skeptical of information in the future and further reducing trust in operators and transit agencies, potentially turning the data collection process into a net loss.

Modifying GTFS to adapt to the range of semi-formal transit services must be a priority if this information specification is to bring benefits to much of the world. Adaptations could take the form of a more flexible version of the GTFS, with optional

fields for stop locations, schedule data, and frequencies – or at least a means to encode assumptions and estimates included in a GTFS. Currently, transit professionals from several countries have formed an online forum to discuss difficulties incurred from attempting to use a GTFS feed for flexible, non-stop-based services and to discuss possible adaptations to GTFS that would address this gap. A majority of these individuals met at a global meeting on the topic in November 2013 in Washington, DC. At this meeting the work piloted in Mexico City formed the basis for discussions towards a new, or modified standard. While these propositions represent the early stages of a dialogue amongst global transit professionals, a variant of these extensions are now being proposed for formal incorporation into GTFS. Through these exchanges it has become apparent that not only might a “flexible” GTFS help cities of the Global South, but it also might spur transit industry innovations and enable formalization of flexible transit services in the Global North.

Another important challenge has to do with maintaining the GTFS data-feed and ensuring it is kept current. This requires regular updates to reflect any changes in schedule service. With formal services, this means ensuring that management prioritizes updating the feed. In the case of Mexico City, SETRAVI, as the DF’s transport authority, provides such an authorizing environment, although not for the entire MCMA. For less formal services, such as the DF’s *colectivos*, it may be more challenging to identify the appropriate agency that has the combination of ability, authority and the interest in keeping the GTFS feed updated. In multi-jurisdictional metropolitan areas, again like the MCMA, such challenges may be compounded. The DF is considering a broader formalization effort, with a more active regulatory role envisioned for government (although the DF’s transit regulatory history tends to be marked by such ebbs and flows; 22). This may incentivize government to maintain the feed (again, between June and November data have been collected for more or less all of the DF’s *colectivos* routes). Ultimately, the data challenge is more an institutional than technological one – most likely, the agencies with the strongest interest in keeping GTFS feeds current will be those who are contemporaneously embracing a broader agenda of system improvement and modernization.

Real-Time Data

Quite apart from the issues related to the GTFS format, many of the important potential benefits may well need real-time information, thus the use of GTFS-RT. Real-time information systems need not be as costly as the AVL equipment installed on buses in New York, São Paulo, and Santiago. Instead, transit agencies could use mobile technology to generate live data on transit systems; Cebu and Dhaka are already experimenting with this approach. For as little as \$63 per unit, transit agencies in Dhaka can place a smartphone on a bus and begin receiving real-time information about the vehicle’s location. These live data can create a robust dataset capable of characterizing a city’s transit system over time and of delivering up-to-the-minute information about current system dynamics. This information would be especially useful in low-and middle-income cities with flexible transit systems. The technology can be easily adapted to carry out on-board surveys and passenger counts as well.

SUMMARY AND CONCLUSIONS

GTFS, GTFS-RT and the contemporaneous “open data” movement are transforming public transportation in cities around the world. The Mexico City case reveals the possibilities and challenges of this transit data specification in one of the largest megacities’ of the Global South. The case shows that the city’s multiple different public transit services, operated or strictly regulated by the public sector, can be relatively easily brought into the GTFS framework and that, accompanied by an “open data” approach can quickly lead to development of user-facing information services. Nonetheless, GTFS proves challenging for the city’s ubiquitous semi-formal *colectivos*, requiring a specification that can accommodate flexible routes, stops and schedules. For GTFS’ benefits to accrue to a large share of the world’s citizens, such modifications will be critical. For the Mexico City case, additional challenges arise from the broader metropolitan area’s complex institutionality, as the current GTFS effort covers service over an area where less than half the megalopolis’ residents live. Moving towards a GTFS-RT feed for the city will also be complicated by this institutionality, but possibly enabled through lower-cost AVL approaches based on smartphones. Moving towards a GTFS-RT-type feed, capable of generating real-time information for users, will remain an institutional and technical challenge; as will incorporating transit services from the rest of the MCMA (beyond the DF).

This case indicates an ambitious research agenda for furthering the benefits promised by GTFS in the Global South. From the user perspective, more knowledge is necessary on the types of and media for information desired and impacts on short- and longer-term demand and user satisfaction. For places with strong semi-formal and informal systems, an important question is whether GTFS and the open data movement might accelerate reforms, changing the industrial structure, regulatory power, and broader system accountability and even ushering in new service innovations. Finally, this information may aid planning authorities and others to better understand dynamic environments through accurate information, equipping them to better answer complex questions and attempt to deliver solutions.

ACKNOWLEDGEMENTS

The authors wish to thank the World Bank’s Energy Sector Management Assistance Program (www.esmap.org) for providing the funds to finance this project and Dhyana Quintanar Solares (Director of Planning, SETRAVI, Mexico City) and her team for the leadership and commitment in carrying out this project. The findings, interpretations and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the institutions the authors belong to. All remaining mistakes remain those of the authors – who administered the project on behalf of the funding agency, were part of the implementation team, and supported a preliminary evaluation.

REFERENCES

- (1) Google Groups. *General Transit Feed Spec Changes*. <https://groups.google.com/forum/#!forum/gtfs-changes>. Accessed July 31, 2013.

- (2) Janssen, M., Y. Charalabidis, A. Zuiderwijk. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, Vol. 29, No. 4, 2012, pp. 258-268.
- (3) Sunlight Foundation. *Ten Open Data Principles*. <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/> Accessed July 30, 2013.
- (4) Antrim, A., and S. Barbeau. *The Many Uses of GTFS Data: Opening the Door to Transit and Multimodal Applications*. <http://www.locationaware.usf.edu/wp-content/uploads/2010/02/The-Many-Uses-of-GTFS-Data-%E2%80%93-ITS-America-submission-abbreviated.pdf>. Accessed 31 July, 2013.
- (5) Wong, J. Leveraging the General Transit Feed Specification (GTFS) for Efficient Transit Analysis. *Transportation Research Record*, 2013 (forthcoming).
- (6) GoogleTransitDataFeed. *List of Publically-Accessible Transit Data Feeds*. <https://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds>. Accessed November 14, 2013.
- (7) GTFS Data Exchange. *Transit Agencies Providing GTFS Data – By Location*. <http://www.gtfs-data-exchange.com/agencies/bylocation>. Accessed November 14, 2013.
- (8) Google Maps. *Transit – Cities Covered*. <https://www.google.com/landing/transit/cities/index.html>. Accessed June 30, 2013.
- (9) Webb, M., ed., *Jane’s Urban Transport Systems 2009-2010*, 28th ed. Janes Information Group, London, 2009.
- (10) Muñoz, J. C. and A. Gschwender. Transantiago: A Tale of Two Cities. *Research in Transportation Economics Vol. 22, No. 1*, 2008, pp. 45–53.
- (11) Hidalgo, D. Citywide Transit Integration In A Large City: The Case Of The Interligado System, São Paulo, Brazil. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2114*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 19-27
- (12) Based on population estimates and Soehodho, S., Tetsuro, H., Fujiwara, A., and Montalbo, C. Transportation Issues and Future Condition in Tokyo, Jakarta, Manila, and Hiroshima. *Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 5*, 2005, pp. 2391-2404.
- (13) Mahmud, S., Sohel, M., and M. I. Anwar. *A Preliminary Feasibility Study of Bus Rapid Transit System in the Context of Present Road Network in Dhaka*. <http://ssrn.com/abstract=2009988>. Accessed July 31, 2013.
- (14) Internet World Stats. *World Internet Users Statistics Usage and World Population Stats*. <http://www.internetworldstats.com/stats.htm>. Accessed July 27, 2013.
- (15) International Telecommunications Union. *Statistics*. <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>. Accessed July 27, 2013,
- (16) Webb, K. *GTFS – Internationalization*. <https://github.com/conveyal/gtfs-internationalization/blob/master/README.md>. Accessed August 1, 2013.
- (17) Byrd, A., Emory, D., and K. Webb. Open Trip Planner Analyst: Open Source, open data driven transportation network analysis. Poster P13 - 6039 , Session 616 (Open Data in Transportation). *Transportation Research Board Conference 2012*, Washington, DC.

- (18) Guo, Z. Mind the Map! The impact of transit maps on path choice in public transit. *Transportation Research Part A*, 45, 2011, pp. 625-639.
- (19) Ferris, B., K. Watkins, A. Borning. OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit. in Proceedings of CHI 2010, Atlanta, GA, USA, April 10-15, 2010, pp. 1807-1816.
- (20) Watkins, K., B. Ferris, A. Borning, G.S. Rutherford, D. Layton. Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transportation Research Part A*, 45, 2011, 839-848.
- (21) Zhang, F., Shen, Q., & K. Clifton. Examination of Traveler Responses to Real-Time Information About Bus Arrivals Using Panel Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2082, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 107–115.
- (22) Tang, L, and P. Thakuriah. Ridership effects of real-time bus information system: A case study in the City of Chicago. *Transportation Research Part C*, 22, 2012, 146-161.
- (22) Flores, O. Expanding transportation planning capacity in cities of the Global South: Public-private collaboration and conflict in Chile and Mexico. PhD dissertation, MIT, Cambridge, 2013.
- (23) Watkins, K., A. Borning, G.S. Rutherford, B. Ferris, B. Gill. Attitudes of bus operators towards real-time transit information tools. *Transportation*, 40, 2013, 961-980.
- (24) Lopez, A., J. Casello, A. Molinero, D. Vazquez. Evaluating Private Bus Operators' Willingness to Participate in Transit Improvements in Mexico. *Transportation Research Record* (forthcoming), 2013.
- (25) Rojas, F. *Transit Transparency: Effective Disclosure through Open Data*. Transparency Policy Project, Harvard Kennedy School, Cambridge, MA, USA. http://www.transparencypolicy.net/assets/FINAL_UTC_TransitTransparency_8%2028%202012.pdf. Accessed June 30, 2013.
- (26) Kumar, A. and F. Barrett. *Africa Infrastructure Country Diagnostic: Stuck in Traffic: Urban Transport in Africa*. The World Bank, Washington, DC, 2008.
- (27) LTA Academy of Singapore. "Passenger Transport Mode Shares in World Cities," *JOURNEYS: Sharing Urban Transport Solutions* No. 7, 2001, p.p. 1–11.
- (28) Corporación Andina de Fomento. *Observatorio de Movilidad Urbana para América Latina*. Corporación Andina de Fomento, Caracas, 2010.
- (29) Cervero, R. and A. Golub. Informal Transport: A Global Perspective. *Transport Policy* Vol. 14, No. 6, 2007, p.p. 445–457.

LIST OF TABLES AND FIGURES:

TABLE 1 GTFS Data Collection Experiences Across Five Cities

TABLE 2 Public Transportation Institutionalization in Mexico City

TABLE 1 GTFS Data Collection Experiences Across Five Cities

		Mexico City	Santiago	São Paulo	Manila	Dhaka
City characteristics	City population (thousands)	8,600	4,600	11,000	12,000	9,200
	Metro population (thousands)	18,000	6,000	18,800	26,000	14,500
	Primary transit modes	Microbus, Metro, Bus	Bus, Metro	Bus, Metro	Jeepney, Light rail	Rickshaw, Bus
	Regulatory scale	National	National	National, Local	National, Local	Local
	Transit trips per day (millions)	15 (9)	5.25 (10)	8.7 (11)	3.6 (12)	11.2 (13)
	Internet access (national level) (14)	37%	59%	46%	32%	5%
	Mobile phone ownership (national level) (15)	83%	118%	124%	99%	56%
	Smartphone ownership (national level) (15)	13%	18%	28%	14%	<1%
Transit data outputs	Project initiator	City Government, World Bank	City Government	City Government	City Government, World Bank	MIT, Urban Launchpad, Kewkradong
	NGO assistance	Yes	No	No	Yes	Yes
	Method	Android app (TransitWand), iPhone app (MotionX-GPS), data management portal	AVL, AFC	AVL, AFC	GPS	Android app (Flocktracker)
	Data collection started	2012	2007	2008	2006	2012
	GTFS released	2013	2013	2012	2012	2013
	Routes included	475	376	1,329	906	78
	Difficulties encountered	Fixed stop locations, fixed schedules and headways, vehicle type	Group taxis not included due to their flexible operations	None – no flexible transit services	Fixed stop locations, fixed schedules and headways, vehicle type	Fixed stop locations, fixed schedules and headways, lack of agency websites

		Mexico City	Santiago	São Paulo	Manila	Dhaka
GTFS outcomes	Open data access	Yes	Yes	No	Yes	Forthcoming
	Number of apps using GTFS or other transit data for trip-planning	28	18	22	6	0
	Number of Android app downloads	1.1 - 5.5 million	480,000 – 1,850,000	62,000 - 280,000	11,000 - 50,000	Not applicable
	Types of external apps	Trip planners, BRT arrival time predictor	Trip planning, bus arrival times, electronic fare card services, and SMS-based route/arrival time service	Trip planning, bus arrival times	Forthcoming	Paper-based bus map
	Internal applications	Open Trip Planner Analyst extension, regulatory impacts	GTFS is used for display signs at stops and may be used for automated sign and map printing in the future. The transit data is used for planning and analysis in another data format.	Not used for planning, only for display signs at stops	Intended for a jeepney rationalization program and to generate sufficient data to avoid future consulting studies	None yet

The data in this table are drawn mainly from secondary sources. Additional information, as available, is discussed in the text.

TABLE 2 Public Transportation Institutionalility in Mexico City

Service	Formal name	Government Role (jurisdiction)	Service type	Formed	Structure	Daily load
STE	Servicio de Transportes Eléctricos del Distrito Federal	Operator (DF)	Electric trolleybus and light rail	1946	8 lines, 290 buses, 13km light rail	241,000
RTP	Red de Transporte de Pasajeros del Distrito Federal	Operator (DF)	Diesel bus	2000	100 routes, 1,400 buses	750,000
STC	Sistema de Transporte Colectivo	Operator (DF)	Metro (heavy rail)	1967	12 lines, 195 stations, 300+ trains	4,200,000
Metrobús	Sistema de Corredores de Transporte Público de Pasajeros del Distrito Federal	Regulator (DF)	Bus rapid transit (BRT)	2005	4 lines, 138 stops, 365 buses	700,000
Tren Suburbano	Ferrocarril Suburbano de la Zona Metropolitana del Valle de México	Regulator (Federal)	Suburban railway	2008	1 line, 7 stations	134,000
DGT	Dirección General de Transporte	Regulator (DF)	Microbus (<i>colectivo</i>)	1970s	121 routes, 1,227 variations, 9 concessions, 28,000 buses	8,700,000
Ecobici	Ecobici	Regulator (DF)	Bikesharing	2010	275 stations, 4,000 bicycles	25,000

Data based on discussions between the authors and SETRAVI representatives