Applying the General Transit Feed Specification to the Global South

Experiences in Mexico City, Mexico—and Beyond

Emily Eros, Shomik Mehndiratta, Chris Zegras, Kevin Webb, and Maria C. Ochoa

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 study analyzed this rapidly evolving transit information sector through the Mexico City, Mexico, experience. The case study illustrated that even a megacity with several transit providers could create a fully functional GTFS feed in a few weeks and obtain the benefits of work done elsewhere. Because of the global open data ecosystem, a range of free or low-cost applications—customer-facing applications and planning tools—can immediately capitalize on these data. However, the Mexico experience revealed a limitation of GTFS in its current form: its inability to easily accommodate semi-structured public transit services common in many cities in the developing world. An adaption to GTFS developed in Mexico City to address this limitation is described. The case study revealed significant untapped potential to maximize the value of the open data ecosystem, particularly for planning and regulatory tools.

In 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 in the United States. In 2005, the Portland, Oregon, transit agency, Tri-County Metropolitan Transit District of Oregon (TriMet), partnered with Google to integrate public transit schedule and route information with Google Maps. In the process, TriMet and Google codeveloped a nonproprietary transit data format, later called 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 information).

To accommodate the varied nature of transit services, an online 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 real-time (GTFS-RT) data feed specification was designed and released by a partnership of agencies, software developers, and Google. GTFS-RT was 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.

CONFLUENCE OF THE TRANSPORT OPEN DATA MOVEMENT AND GTFS

The open data movement arose from philosophical principles of open government, transparency, and accountability, along with practical motivations related to increased returns on public investment, downstream wealth creation, more brainpower for examining complex problems, and enhanced public policy and service delivery (*2*). The open data movement has shifted how transportation agencies communicate with users; agencies are moving from tightly controlling data and the products derived from them toward generating and releasing data with minimal control over the end products. Because of the open data movement, government data releases attempt to follow some key principles intended to empower citizens to use the data: completeness, primacy, timeliness, ease of physical and electronic access, machine readability, nondiscrimination, use of commonly owned standards, licensing, permanence, and nonusage costs (*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 allow transit agencies and operators to engage the power of the software developer community and the citizenry more generally to create new forms of information services about public transportation. GTFS also enables new forms of comparative assessment across

E. Eros and C. Zegras, Department of Urban Studies and Planning, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02138. S. Mehndiratta and M. C. Ochoa, Latin America and Caribbean Region, World Bank Group, 1818 H Street, NW, Washington, DC 20433. K. Webb, Conveyal, 1133 15th Street, NW, 12th Floor, Washington, DC 20005. Corresponding author: E. Eros, ejeros@mit.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2442, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 44–52.
DOI: 10.3141/2442-06

public transportation systems (e.g., benchmarking) and new service modeling possibilities (*5*).

GTFS GOES GLOBAL

The simple file structure of the GTFS prompted rapid global adoption: as of November 2013, Google listed 229 public transit agencies around the world that release official GTFS feeds available for developers to use (*6*). When private transit companies are included in the count, estimates range from 703 (*7*) to 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 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, Mexico.

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 is a single jurisdiction (one mayor) with a population of approximately 8.9 million, 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 on services in the DF, where since 1975 the transportation secretariat (SETRAVI) has regulated both technical and nontechnical 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 considered it important that multimodal journey planners can seamlessly include information on Ecobici station sites and bicycle availability. Expanding the GTFS efforts into the rest of the MCMA will further complicate an already complex institution but will be necessary because more than half the metropolis's 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 because of 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, that is, *colectivos*, for which drivers collect fares and no digital means exist to track ridership.

The colectivos pose a particular challenge. Some 30,000 vehicles in 121 route associations, colectivos account for about 50% of the DF's motorized trips. 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; DGT does not possess up-to-date information about ramal characteristics (route, drivers, vehicles) or demand patterns. This deficit limits regulatory ability.

The route associations have little incentive to provide data; this lack of incentive makes verification difficult. In the spring of 2013, DGT began surveying 10 of the more formalized colectivo corridors. The agency, supported by a nongovernmental organization, collected basic route data (route, travel time, counts) with an iPhone running the MotionX-GPS application (app) along with an Android tablet running a custom-made application. Although successful, this technique is effort intensive, requiring six person days per ramal, including four person days of data collection.

Provision of Non-GTFS-Based Third-Party Information

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 for Android and nine for Apple iOS). Android apps for Mexico City have a combined download count of between 1.1 million and 5.5 million. (Counts are 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 on a map. A few apps also include this information for the electric trolley buses, Suburbanos, and Metrobús. The usefulness of the apps is limited; travel time information may be unreliable and maps outdated. 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 and an Android app, 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 and alighting times, and passenger counts. Users upload this information through WiFi or cell phone networks. The data management portal 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 nongovernmental organizations focused on the role and

TABLE 1 GTFS Data Collection Experiences Across Five Cities

Characteristic	Mexico City	Santiago, Chile	São Paulo, Brazil	Manila, Philippines	Dhaka, Bangladesh
City					
City population (millions)	8.6	4.6	11	12	9.2
Metropolitan population (millions)	18	6	18.8	26	14.5
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
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	1,000,000–5,500,000	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 and 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 are used for planning and analysis in another data format	Not used for planning, used only for display signs at stops	Intended for a jeepney rationalization program and to generate sufficient data to avoid future consulting studies	None yet

NOTE: MIT = Massachusetts Institute of Technology; NGO = nongovernmental organization; AFC = automated fare collection; BRT = bus rapid trasit; SMS = short message service. Data in this table are drawn mainly from secondary sources. Additional information, as available, is discussed in the text.

TABLE 2 Public Transportation Institutionality 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, 13-km 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	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	Direccíon General de Transporte	Regulator (DF)	Microbus (*colectivo*)	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

NOTE: Data are based on discussions between the authors and SETRAVI representatives.

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 include information for about 125 lines, 260 route variations, and more than 5,000 stops, covering all the Metro, Metrobús, RTP, STE, and Suburbano. These agencies already possessed basic schedule information and KML files of routes and stops, so 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 with 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 because precise stop locations and times are a fundamental GTFS building block. Earlier attempts in other cities to model transport without fixed stops required the inclusion of simulated stops in the data feed. These approximated likely boarding and 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 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 and 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 because 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, Philippines), 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 with the GTFS existing frequency-based timetables. This allowed specification of 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, semiformal modes like colectivos, for which precise values are unknown or unrealistic, providing 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 to discuss with agencies the results and a live disruption-tracker tool the team created, release the GTFS feed on its website, discuss with Google the inclusion of data on Google Transit and having Google Transit engineers review and clear the data feed (a prerequisite to publishing in Google Transit), promote the launch of an open trip planner, and showcase the data at a meeting oriented toward software and mobile app developers. Nonprofit organizations and SETRAVI jointly hosted a hackathon challenge in Mexico City to search for mobility solutions to specific problems. From 41 projects, seven winners were selected to enter a funded incubation program. Two of the winning entries used the GTFS feed: a digital map for the metropolitan area transit system and a journey planner. The other projects focused on data collection and cycling tools.

Initial Outcomes

As of November 2013, the GTFS data had been released for only four months. Although short, this period has provided indications 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 the 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 Transit App, and open 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 use. Therefore, SETRAVI can obtain use information only for the official open data site. Early indications, 1 month after GTFS release, show an initial spike of about 120 downloads following the June 21 official data release. Downloads tapered off but reached 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—most from Mexico (637 downloads) with additional concentrations in the United States (25 downloads) and the United Kingdom (7 downloads). Top cities include Mexico City (589 downloads) and Monterrey, Mexico (20 downloads).

Use of GTFS Downloads

Once downloaded, use of the GTFS feed can be monitored by the type and number of apps developed. As of November 2013, five apps made use of the DF data: Transit App, Pdxtrian, Moovit, AGUMóvil, and Hop Stop. Transit App offers web-based trip planning; Hop Stop provides a similar service on Android, iOS, and Windows mobile phone platforms. Pdxtrian enables Android users to locate the nearest stop or 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. Moovit (Android and iOS) 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 that are under way to promote the use of these data. The SETRAVI staff has been following the development of these apps and will continue to promote the GTFS and monitor new products.

Press Coverage and Twitter Traffic

After GTFS feed publication, mainstream and online press covered the event substantially. A preliminary assessment of traditional media shows (*a*) general enthusiasm for the topic and a perception of it being a step toward transparency and the digital era and (*b*) overall low understanding of the transport open data value proposition—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 for examining public reactions to, and uses of, Mexico City's GTFS feed. In the 2-week period after the data launch, users tweeted about the GTFS project more than 80 times. This Twitter traffic is too low for conclusions to be drawn, but the site could provide some sense of public reaction to transit apps in the future.

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, OpenTripPlanner—an open source, rider-oriented, point-to-point itinerary planner—also supports planning analysis (and is then known as OTPA) by generating an accessibility coverage map by using 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 or 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 it can be used to compare various transport scenarios as modeled in GTFS. This feature has been used to show impacts on accessibility caused by disruptions, such as Hurricane Sandy in New York (*17*). Such tools can inform policy makers and planners and contribute to exploration of questions about mobility and accessibility at various spots in the city. Second, analysis of the type and frequency of service interruptions logged in the disruptions feed would facilitate pattern recognition, which could help SETRAVI determine where to make operational improvements.

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

Potential for Regulatory Impact

Along with the overt and direct uses in planning applications, the data availability and transparency enabled by Mexico City's GTFS could have an even stronger impact on relationships between transit regulators and transit operators, particularly the colectivos. 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 about compensation for loss of operating rights. As part of an effort to enhance public transport quality, the city is implementing an ambitious bus rapid transit program (Metrobús's five corridors since 2005) and other pilot projects to formalize bus routes. Every such change implies drawn-out (12 to 18 months) and costly negotiations with the affected route association operating the ramales, often requiring extensive primary data collection on affected drivers, vehicles, and ramales and route and on ridership characteristics. Both the government and route associations have incentives to under- or overreport impact (demand, service levels, traffic volumes), so little trust exists among parties. A GTFS-based system could provide much of the data and facilitate less-contentious data collection and negotiations. In addition to lowering the financial and opportunity costs of negotiations on bus system reform, GTFS-based data could inform longer-range route planning and restructuring. In theory, these data could increase the 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 dispatching efficiency; however, the institutional complexities underlying this loosely regulated system and the benefits of system information that remains opaque cannot be discounted.

DISCUSSION OF RESULTS

This case reveals the interrelated benefits and challenges associated with global adoption of GTFS. Drawing primarily on lessons from Mexico, the discussion also draws from other experience (Table 1).

Measuring Benefits and Impact of GTFS

Ex ante knowledge of the benefits and costs should precede GTFS adoption. The costs, although modest relative to, for example, capital investments, are not trivial, and technical capability and data gathering are required. The benefits can be difficult to quantify, particularly in the short term. Though evidence is limited, experiences suggest three categories of benefits: (*a*) benefits to passengers and potential users of higher-quality information on services, (*b*) benefits to operators and regulators from the use of analytic and monitoring tools, and (*c*) benefits to society of operating in an open data ecosystem.

Users

Although it is spreading quickly, GTFS is a recent phenomenon, established in 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 that provide trip planning and predictions of bus arrival times. Nonetheless, evaluations of the impact for users are limited and are concentrated primarily in the Global North.

Theory suggests that improved access to high-quality transit service information will increase the satisfaction levels of current riders, for example, by reducing real or perceived wait times, and will increase future ridership by improving knowledge about services and service quality relative to alternatives. Little research has been conducted to measure these impacts, even for traditional forms of transit information provision such as static paper maps (*18*). Because of the increasing ubiquity of real-time information, enabled by GTFS-RT and the open data movement, more studies have focused on user impact. The developers of the Seattle, Washington, OneBusAway, a suite of open source tools that deliver real-time bus location information and wait times to mobile devices, implemented an online survey among users. Most (92%) respondents reported that their overall satisfaction with public transit had improved because of OneBusAway; a similar share reported lower waiting times, and modest increases in noncommute 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 used (*20*). Using a panel survey in an effort 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 (2 weeks after the technology debuted), but they 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 bus system ridership found modest average increases in bus use after real-time information was introduced; there was some evidence for higher ridership effects on routes affected in later-stage rollouts, suggesting improved technology or technological diffusion and adoption (as third-party providers entered the game and personal mobile devices improved) (*22*).

This initial evidence suggests promise not just for ensuring a higher-quality product for customers but for attracting and keeping passengers who have a choice. Satisfying current and future riders by choice in societies undergoing rapid motorization may help sustain public transit patronage over time and may result in longer-term values of reduced pollution, reduced congestion, and other externalities. As GTFS and open transit data moves into the Global South, user expectations, needs, and responses may differ from those of the Global North. For example, semiformal services, such as colectivos, typically are not represented on traditional transit maps, where route information comes from experience, others' knowledge (face-to-face social networks), and vehicles, drivers, and fare collectors. In such systems, much benefit may be gained from service maps, although map legibility and interpretability by users must be ensured.

To maximize potential benefits to passengers, digital forms of information services must be matched to the needs and mobile devices of the passenger base. For instance, 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 that tend to use colectivos and 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% in Mexico City) (*15*).

On the other hand, there are more than 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, which offers information services on a range of platforms including SMS, provides an open source option (*19*). Authorities in Santiago, Chile, have developed such services that provide bus arrival time and routing services.

Analysis and Regulatory Tools

Interviews and discussions with agency staff in Mexico indicated that increased transit information could inform internal planning and analysis and could change relationships between regulators and operators, particularly with respect to negotiations about route restructuring. Traditionally, in the DF new route authorizations have not been subject to a structured process that checks 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 (*23*). Especially among the semiformal colectivos, the complex industrial structure matters. Agencies noted that having a comprehensive visual sense of all the routes helped officials anchor discussions with operators about route-related issues in a more technical and less political sphere. Planning tools such as OTPA, which can 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 have been understudied everywhere. Limited evidence from the United States (a single study in Seattle) suggests drivers were particularly in favor of services for disadvantaged groups (e.g., the deaf, the blind) and generally supported service alerts (e.g., breakdowns) but were wary of information that would negatively affect route and ridership perceptions (e.g., service reviews) or publicize past performance data, which could impair a driver's standing even for conditions outside of his or her control (*24*). Among both publicly and privately operated services in the DF, with longstanding, often contentious histories of attempted reforms (*23*), the reaction by relevant agents to such information reforms will vary. The assessment by Lopez Dodero et al. of private bus operators' general willingness to participate in transit improvements (not information services per se) in Mexico may provide some hints (*25*). Lopez Dodero et al. found that unclear legal standing, industrial structure (e.g., owner-operator), poor understanding of operating costs, and other factors negatively correlate with support for transit improvements, and private operators in the MCMA were even less receptive to changes. If information transparency in an open data model could 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 the government to visualize route configuration and better understand duplication or to explore adding or eliminating services from under- or overserved 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. More work is needed on tools that can integrate various data sets (e.g., census, employment, land use) with the detailed transport GTFS feed and that can properly extract and visualize results. The combined analysis of multiple sources of data with the GTFS feed could help urban planning agencies integrate land use and transport within their planning efforts without having to use complex models. This ability could be especially relevant for small cities and cities that lack the capacity to develop and maintain traditional transport models but that need to conduct occasional simple analyses 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, several apps made use of these data to provide value to users. The DF's suite of apps is 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 problems and contexts; as one city develops apps around a particular problem, others can benefit with relatively little additional investment.

This process is not limited to public-facing apps—it includes data collection as well as analysis and planning tools. However, more work is needed, particularly to expand the reach of the open data culture to traditional transport planning tools.

Beyond meeting riders' needs, transit data transparency could prompt a cycle of information availability, public feedback, and government response. Rojas raised these arguments but noted that little research exists on the outcomes of data disclosure initiatives, in terms of both citizen mobility and improved performance from transportation agencies and service providers (*26*). 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 United States that operate with fixed routes and stops. The Mexico City case study highlights a key limitation: the incompatibility of GTFS 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 make up more than 90% of transit trips in Algiers, Algeria; Bamako, Mali; Dakar, Senegal; and Dar es Salaam, Tanzania; and flexible transport services make up more than 70% of trips in Accra, Ghana; Bangalore, India; Caracas, Venezuela; Manila; and Tehran, Iran (*27–30*).

In Mexico a work-around was found by creating a variant to the GTFS feed that is 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, Bangladesh, also faced this challenge. Like Mexico City, Manila chose to avoid schedules, instead providing headway estimates for jeepneys. The Dhaka team included stop location according to where the bus stopped during the data collection ride. Manila's stops were interpolated every 500 m along the route.

These types of work-arounds enable the assembly and release of a GTFS feed, although risking inconsistent and potentially inaccurate information. This possibility may not pose an issue for users unaccustomed to schedule and stop location information. However, at best, these work-arounds require significant time, effort, and cost and could generate data on stops and schedules that may not be meaningful. Developers may become confused about which fields are reliable and which are estimates or constructs. Furthermore, inaccurate data could lower user confidence in the new information, making users skeptical of future information 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 semiformal 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. Transit professionals in several countries have formed an online forum to discuss difficulties incurred in attempts to use a GTFS feed for flexible, non-stop-based services and to discuss possible adaptations to GTFS that would address this gap. Most of these individuals met at a global meeting on the topic in November 2013 in Washington, D.C., where work piloted in Mexico City formed the basis for discussions toward a new or modified standard. These propositions represent the early stages of a dialogue among global transit professionals, and a variant of these extensions is 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 innovation 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 that it is kept current. These efforts require regular updates that reflect changes in schedule service. For formal services, management must prioritize updating the feed. In 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 difficult to identify the appropriate agency that has the combination of ability, authority, and interest to keep the GTFS feed updated. In multijurisdictional metropolitan areas, such as 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 (*23*)]. Government thus could be motivated to maintain the feed (between June and November, data were collected for almost all the DF's colectivos routes). The data challenge is more an institutional than a technological one: most likely, the agencies with the strongest interest in keeping GTFS feeds current will be those that are contemporaneously embracing a broader agenda of system improvement and modernization.

Real-Time Data

Apart from issues related to the GTFS format, many of the important potential benefits may require real-time information, and so GTFS-RT is used. 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; Dhaka and Cebu, Philippines, are 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 data set 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 that have flexible transit systems. The technology can be easily adapted to carry out onboard 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 public transit services, operated or strictly regulated by the public sector, can be relatively easily brought into the GTFS framework, which with an open data approach could quickly lead to development of user-facing information services. Still, use of GTFS would be a challenge for the city's ubiquitous semiformal colectivos, requiring a specification that can accommodate flexible routes, stops, and schedules. For the GTFS benefits to accrue to a large share of the world's population, 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's residents live. Moving toward a GTFS-RT feed for the city will be complicated by this institutionality but possibly enabled through lower-cost AVL approaches based on smartphones. Moving toward a feed of the GTFS-RT type that can generate real-time information for users remains an institutional and technical challenge, as is incorporation of transit services from the rest of the MCMA, beyond the DF.

The Mexico City case offers an ambitious research agenda for furthering the benefits of GTFS in the Global South. More knowledge is needed regarding what types of information users desire and what type of media could best disseminate it—as well as how this information affects short- and long-term demand and user satisfaction. For places that have strong semiformal 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 could help planning authorities and others to better understand dynamic environments through accurate information, equipping them to answer complex questions and to deliver solutions.

ACKNOWLEDGMENTS

The World Bank's Energy Sector Management Assistance Program provided funding for this project. The authors thank Dhyana Quintanar Solares of SETRAVI for her contributions.

REFERENCES

1. *General Transit Feed Spec Changes.* https://groups.google.com/forum/#!forum/gtfs-changes. Accessed July 31, 2013.
2. Janssen, M., Y. Charalabidis, and 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 July 31, 2013.
5. Wong, J. Leveraging the General Transit Feed Specification for Efficient Transit Analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2338,* Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 11–19.
6. *List of Publicly-Accessible Transit Data Feeds.* https://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds. Accessed Nov. 14, 2013.
7. GTFS Data Exchange. *Transit Agencies Providing GTFS Data—By Location.* http://www.gtfs-data-exchange.com/agencies/bylocation. Accessed Nov. 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 Interligado System of 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. Soehodho, S., H. Tetsuro, A. Fujiwara, and C. Montalbo. 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., M. Sohel, 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. *World Internet Users Statistics Usage and World Population Stats.* http://www.internetworldstats.com/stats.htm. Accessed July 27, 2013.
15. *Statistics.* International Telecommunications Union. 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 Aug. 1, 2013.
17. Byrd, A., D. Emory, and K. Webb. Open Trip Planner Analyst: Open Source, Open Data Driven Transportation Network Analysis. Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.
18. Guo, Z. Mind the Map! The Impact of Transit Maps on Path Choice in Public Transit. *Transportation Research Part A,* Vol. 45, 2011, pp. 625–639.
19. Ferris, B., K. Watkins, and A. Borning. OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit. *Proc., 28th ACM Conference on Human Factors in Computing Systems,* Atlanta, Ga., 2010, pp. 1807–1816.
20. Watkins, K., B. Ferris, A. Borning, G. S. Rutherford, and 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,* Vol. 45, 2011, pp. 839–848.
21. Zhang, F., Q. Shen, and K. J. 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., 2008, 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,* Vol. 22, 2012, pp. 146–161.
23. Flores, O. *Expanding Transportation Planning Capacity in Cities of the Global South: Public-Private Collaboration and Conflict in Chile and Mexico.* PhD dissertation. Massachusetts Institute of Technology, Cambridge, 2013.
24. Watkins, K., A. Borning, G. S. Rutherford, B. Ferris, and B. Gill. Attitudes of Bus Operators Towards Real-Time Transit Information Tools. *Transportation,* Vol. 40, 2013, pp. 961–980.
25. Lopez Dodero, A., J. M. Casello, A. R. Molinero Molinero, and D. Vazquez Cotera. Evaluating Private Bus Operators' Willingness to Participate in Transit Improvements in Mexico. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2394,* Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 119–127.
26. Rojas, F. *Transit Transparency: Effective Disclosure Through Open Data.* Kennedy School of Government, Harvard University. http://www.transparencypolicy.net/assets/FINAL_UTC_TransitTransparency_8%2028%202012.pdf. Accessed June 30, 2013.
27. Kumar, A., and F. Barrett. *Africa Infrastructure Country Diagnostic: Stuck in Traffic: Urban Transport in Africa.* World Bank, Washington, D.C., 2008.
28. LTA Academy of Singapore. Passenger Transport Mode Shares in World Cities. *JOURNEYS: Sharing Urban Transport Solutions,* Vol. 7, 2001, pp. 1–11.
29. *Observatorio de Movilidad Urbana para América Latina.* Corporación Andina de Fomento, Caracas, Venezuela, 2010.
30. Cervero, R., and A. Golub. Informal Transport: A Global Perspective. *Transport Policy,* Vol. 14, No. 6, 2007, pp. 445–457.

The findings, interpretations, and conclusions expressed in this paper are those of the authors. They do not necessarily represent the views of the institutions to which the authors belong. Any mistakes are 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.

The Urban Transportation Data and Information Systems Committee peer-reviewed this paper.