Transportation Data as Disruptive Innovation in Mexico City

by Emily J. Eros

B.A. in Geography
Dartmouth College
Hanover, NH (2009)

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of

Master in City Planning
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2014

©2014 MIT. All rights reserved.

The author hereby grants to MIT permission to reproduce and to
distribute publicly paper or electronic copies of the thesis document
in whole or in part in any medium now known or hereafter crafted.

Signature redacted

Signature redacted

Certified by

Certified by

Accepted by

Accepted by

Signature redacted

Signature redacted
Transportation Data As Disruptive Innovation in Mexico City

by Emily J. Eros

Submitted to the Department of Urban Studies and Planning on May 21, 2014 in partial fulfillment of the requirements for the degree of Master of City Planning

ABSTRACT
Despite the popularity of big data and smart city initiatives in rich countries, relatively few city governments in the Global South possess even basic information about public transportation routes and operations within their jurisdictions. The growing ubiquity of affordable mobile phones and internet-capable devices has enabled some developing cities to begin collecting and compiling these data. This thesis uses a 2013 data collection project undertaken within the Federal District of Mexico City as a case study to examine the role of information as a disruptive innovation in the transportation sector; it explores the potential impacts of transportation information on microbus regulators, owners/operators, and users. To do so, it draws from literature reviews, interactions with transportation agency staff, and interviews with microbus operators. Findings suggest that increased static information may increase government power with respect to microbus operators, particularly during franchising negotiations, but that it may offer limited benefits to users. Dynamic (i.e., real-time) sensors could benefit regulators, owners, and users alike, but would require genuine support or tolerance from microbus drivers. If the government continues to expand its current franchise attempts, then real-time data collection would be achievable and could offer benefits to all involved parties. Otherwise, it would be unfeasible to incentivize drivers to allow live-tracking to take place on their vehicles. Regardless, the case study suggests that transportation information can play a significant role in changing the regulatory dynamics within the Global South and encourages further efforts in the field.

Thesis Supervisor: Chris Zegras, Associate Professor of Urban Studies and Planning
Keywords: GTFS, microbus, open data, transportation, Mexico City
ACKNOWLEDGEMENTS

This thesis was made possible with the help of many people and organizations:

Shomik Mehndiratta and Catalina Ochoa, who included me in the World Bank's data collection project in Mexico City.

The Mexico City project team: Dhyana Quinatar, Mariana Campos, Natalia Gutiérrez, Kevin Webb, Neil Taylor, and Ian Stott – who did the hard work of data collection and let me pester them with questions about project details.

Abel Lopez Dodero, who provided enormous help with organizing interviews, connecting me with microbus drivers, and following up with advice and suggestions.

Chris Zegras, who provided guidance, patiently read drafts, and supplied extensive feedback throughout the entire thesis process.

Sarah Williams, for contributing advice, acting as thesis reader, and including me in her digitalmatatus project in Nairobi.

Onésimo Flores Dewey, who shared his extensive knowledge about Mexico City's political and transportation dynamics.

The Singapore-MIT Alliance for Research and Technology, whose Future Urban Mobility program provided funding to support this thesis.

In addition to all the academic help, special thanks to Katherine for her unquestionable support and unyielding optimism even in the face of Icelandic weather, and to the Boston climbing community for keeping me on belay (literally and emotionally) throughout grad school.

Any errors in this document are mine; the findings, interpretations and conclusions expressed in this paper are my own and do not necessarily represent the views of the funding agencies or institutions involved in the data collection project.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ............................................................... 7

ONE | INTRODUCTION ............................................................... 11

TWO | TECHNOLOGICAL AND THEORETICAL CONTEXT ............... 13

DATA: BIG, SMART, AND OPEN ............................................... 13
PUBLIC TRANSPORTATION INFORMATION ............................... 16
THERE'S AN APP FOR THAT .................................................... 21
MIND THE GLOBAL GAP ........................................................ 24
CREATIVE DISRUPTION ......................................................... 33
DISRUPTIVE INNOVATION IN TRANSPORTATION INNOVATION ................. 38
METHODS .................................................................................. 40

THREE | EMPIRICAL CONTEXT .................................................... 42

PUBLIC TRANSPORT IN THE DF: A CYCLICAL CENTURY .......... 42
CURRENT STATE OF PUBLIC TRANSPORTATION OPERATIONS .............. 46
2013 DATA COLLECTION AND DISSEMINATION PROJECT ................. 54

FOUR | POTENTIAL IMPACTS OF INFORMATION ......................... 57

DATA CHANGES AND POTENTIAL DISRUPTION ....................... 57
TRANSPORTATION USERS ..................................................... 57
TRANSPORTATION REGULATORS ............................................ 61
TRANSPORTATION OPERATORS ............................................. 64
PATHS FORWARD ................................................................. 68
SCENARIO 1: NO ADDITIONAL MAJOR INTERVENTION ................. 68
SCENARIO 2: GOVERNMENT INTERVENTION FOR DATA COLLECTION .............. 69
SCENARIO 3: DATA COLLECTION AS PART OF INCREMENTAL FORMALIZATION .... 70
LIMITATIONS .......................................................................... 71

FIVE | SUMMARY AND CONCLUSION ....................................... 73

BIBLIOGRAPHY | WORKS CITED ............................................. 76
ONE | INTRODUCTION

In today’s information age, technology grows more powerful, compact, affordable, and widespread at unprecedented rates. We hold, at our fingertips, the ability to generate and access almost limitless information through our internet-capable devices. Increased information often carries with it the promise of solutions to everyday problems in a variety of sectors. The thesis focuses on transportation.

Technological improvements and the open data movement have worked in tandem to create a wide range of information services for travellers in the USA and other rich countries. In numerous USA cities, online public transportation information provides directions, live traffic updates, and even next-arrival times for city buses. Meanwhile, a gap separates many developing countries, whose transit agencies often lack even basic information for the services and routes they oversee. Maps are nonexistent. Data are unavailable. This exacerbates the already-difficult processes of planning, regulating, improving, and even simply using public transportation (sometimes simply referred to as “transit” in this thesis) in these areas.

Until recently, the equipment required for transportation data collection posed prohibitively high costs for many areas in the Global South. Nowadays, however, the growing ubiquity of low-cost mobile phones and internet-capable devices has enabled an increasing number of places to collect, compile, and release transit data.

As countries step into the open data movement, the next question is, “So what?” Does introducing these new forms of information affect the way a city operates? If so, how? Which dynamics and relationships are affected? These questions form the basis of this thesis. Specifically, I aim to test the hypothesis that new forms of public transit information can act as a type of disruptive innovation, breaking impasses both within and between groups of regulators, operators, and users.

Using a recent transportation data collection project in Mexico City as a case study, I explore the potential impacts of transit information on regulators, operators, and users of the city’s privately owned and operated microbus system. The initial results offer support for the hypothesis, suggesting that information can act as a type of disruptive innovation capable of altering the organizational structure of transportation services within a city. In doing so, it may increase the relative power of city government, which currently struggles to exercise authority over the microbuses.

The thesis consists of five chapters, including this introduction. Chapter Two provides theoretical framing. It discusses the origins of the “smart city” and open
data movements which led to the widespread creation of open transit data feeds around the world, then explores the role of information as a disruptive innovation within the transportation sector. Chapter Three discusses the historic context and structure of public transportation within Mexico City from the 1900s to present. I provide an account of the transportation data collection project undertaken in Mexico City’s Federal District and the research methods I employed. Chapter Four analyzes the potential impacts of information upon, within, and between the relevant categories of stakeholders: regulators, operators, and users. It evaluates potential scenarios moving forward. A final chapter concludes, placing this study within a broader context and exploring future directions for research.
Data: big, smart, and open

The past decade witnessed significant innovations in information and communication technology (ICT). This led to more compact, powerful, and ubiquitous devices, used for both personal and commercial applications. Microprocessors and memory devices grew in capacity and shrunk in size and price. Global personal computer ownership rose from 7% in 2004 to 20% at the close of 2013 (Heggestuen, 2013). Smartphone penetration leapt from zero to 22% (Heggestuen, 2013). Internet access grew from 14% to 36% (World Bank Group, 2014). All around the world, individuals and businesses became more connected.

As technology spread, urban areas themselves became increasingly “smart” – rich with smartphones, smartcards, and other equipment capable of sensing, conveying, and transmitting large volumes of data. Private technology companies like IBM, Cisco, and Siemens capitalized on this moment; they promoted the idea of the “smart city” and “big data” initiatives, persuading cities that large volumes of data would enable them to do more with less (Townsend, 2013) in nearly any application, whether it involved transportation, environmental sustainability, energy grid systems, or even parking meters. It became increasingly fashionable for cities to partner with tech companies and brand themselves as smart cities. Regions from Dubai to Southampton to Santander, as well as countless others, evoked this terminology (Evers, 2013; iTinnovation, 2014; Sambidge, 2014). Smart Cities author Anthony Townsend reports that a realistic estimate would value the smart city industry and its infrastructure at $100 billion from 2010 to 2020 (Mathis, 2014).

Despite this hype, related literature reports vague and varied understandings of what constitutes a smart city (Hollands, 2008). Atlantic Cities found that cities and firms used the term to describe “anything from complex networks of government-controlled sensors and cameras to a parking meter that sends you a text message when you run out of time on the meter” – possibly explaining the industry’s astronomic estimated value (Mathis, 2014). In the academic community, general consensus agrees that, at minimum, a smart city must use ICTs for well-managed self-monitoring in order to improve the city’s functioning. Each author takes a different angle, however. Definitions vary from highly technocentric (Hall, 2000) to highly human-centric (Dirks & Keeling, 2009; Papa, et al., 2013; Rios, 2012). Some highlight elements of efficiency (Komninos, 2013), e-governance (Paskaleva, 2009; Van der Meer & Van Winden, 2003), or creativity (Komninos, 2013). Others focus on relationships between technology, management improvements, and
policy initiatives (Chourabi et al., 2012). Meanwhile, another group of academics expresses skepticism. Hollands describes the term as “self-congratulatory rhetoric of the smart-label bandwagon” and asks, “What city does not want to be smart or intelligent?” (Hollands, 2008). Yet, regardless of whether their efforts generate progress or simply act as a branding mechanism, cities continue to launch projects and initiatives around the world. These actions evidence technology’s powerful allure; it promises to help disentangle the complex problems caused by rapid urbanization, using information as a key input.

In addition to collecting vast amounts of data as part of smart city initiatives, public agencies and other organizations progressively began releasing this information to the public. Some analysts (Hollands, 2008) suggest this information availability is key to the “true” smart city movement. In the mid-2000s, the information technology community called for increasing availability of such data in an effort to increase transparency, accountability, returns on public investment, downstream wealth creation, and service delivery. At the same time, academics like Elinor Ostrom likewise offered philosophical support for data sharing; Ostrom posited that the knowledge commons is different from other public goods because the use of information enhances the existing resource pool rather than depleting it (Chignard, 2013). The term “open data” gained traction and, in December 2007, 30 open information advocates met in Sebastopol, California to define the concept of open public data and convince USA presidential candidates to adopt the cause (Chignard, 2013). The group conceptualized public agencies’ data as a form of common property, from which the public can generate value and creative solutions through participation and collaboration. They pushed for more than just a simple data release, but a set of principles that data should strive to meet: completeness, primacy, timeliness, accessibility, machine-readability, non-discrimination, permanence, and conformity to non-proprietary and common data standards (Tauberer, 2014). These efforts were largely successful. The following year, newly-elected President Obama began his first day in office by signing the Memorandum on Transparency and Open Government (Obama, 2009). The memorandum echoed the Sebastopol group’s principles and required the development of an Open Government Directive, which required federal agencies “to take immediate, specific steps to achieve key milestones in transparency, participation, and collaboration” (Orszag, 2009). The USA government also launched a portal to access open government information, data.gov, in 2009.

Beyond the USA, over forty other countries have pursued open data initiatives (US Government, 2014); a database of open data catalogs from around the world lists 377 registered catalogs from local, regional, and national sources (Datacatalogs.org, 2014). Internationally, the open data movement found early support from the scientific community. In 2004, Ministers of Science from all OECD countries
signed a declaration stating that all publically funded archive data should be made free and available for public use (OECD, 2004). Five years later, many countries began to launch open data initiatives.

Several of these initiatives stemmed from political action. New Zealand’s federal government was an early mover, launching an open data portal in 2009 (Guy, 2009). In Canada, the three major national political parties, as well as one minor party, each introduced party platforms or government resolutions for open data in 2010 (Green Party of Canada, 2010; Liberal Party of Canada, 2010; Parliament of Canada, 2010). Government agencies at multiple scales created open data sites soon afterwards.

In other countries, like the UK, the push for open data largely came from below. Frustrated by the lack of publically-available mapping resources (and the cost of purchasing them), Steve Coast launched a collaborative initiative, OpenStreetMap, in 2004 in order to crowdsource this data for public consumption. This type of project spurred questions and criticisms of the government for not making publically-funded data available to the public. The media helped to spread these concerns. The Guardian newspaper’s technology section began an influential “Free our Data” campaign in 2010 that prompted the British government to begin publishing increasing amounts of Crown data (The Guardian, 2010).

Little research documents open data initiatives in developing nations. However, a review of individual government open data websites reveals that many countries - (including Argentina, Brazil, Chile, Costa Rica, India, Kenya, Morocco, and Uruguay) launched open data portals as early as 2011 – very shortly after first-movers from the developed countries. It is unclear which factors prompted these efforts, and to what extent different groups and organizations were influential.

Irrespective of geography, the open data movement and smart city initiatives meant that government agencies held increasingly rich stores of data concerning their operations, evidencing an increasing willingness to release these data to the public. This shift towards open data has begun to change how many public agencies engage, communicate, and share information with users. Previously, agencies had a tendency to tightly control information and products created from it; many of them greeted early open data initiatives with suspicion and resistance. The USA Department of Defence, for example, expressed security and privacy concerns in the face of Obama’s open data policies (Sternstein, 2010). The Singaporean government took a similar stance (Hicks, 2010).

Several concerns could motivate this desire for control: data are expensive, information is power, inaccurate data can cause more harm than good, data can pose security risks, individuals could use data improperly, etc. The open data
movement, along with crowdsourcing and lowered costs for data collection, began to change this dynamic. As agencies released their data into the information commons, users stepped into the role of active participants and empowered collaborators in a variety of sectors – public transportation included.

Public transportation information

Public transportation (also referred to simply as “transit” in this thesis) includes any mode of transportation available to the public, provided that passengers do not hold an ownership stake in the vehicle or service (Zegras, et al., Forthcoming). This can include a wide variety of vehicles: subways, commuter rail, BRT, minibuses, taxis, jitneys, pedicabs, bikeshares, etc. These systems vary in their degree of “formality”¹ (here defined as the level of regulatory oversight) and in their degree of flexibility (referring to whether services are provided along fixed routes or vary based on demand) (Cervero, 1997; Zegras, et al., Forthcoming). As a result, public transit can be highly regulated, with fixed-routes (e.g. the London Underground); informal and flexible (e.g. Uber, a smartphone-based ridesharing service in which drivers do not hold taxi licenses); regulated and flexible (e.g. licensed taxis and MBTA’s “The Ride”, an on-demand transport service for individuals with restricted mobility); or unregulated, with fixed routes (e.g. private vehicles operating along fixed routes).

The type and degree of information available to describe these transit systems also varies widely. Table 1 shows several examples of public transit data types and their characteristics. Several of these aspects could be placed along a continuum based on their precision (e.g. exact vehicle locations vs. general route maps), level of formalization (e.g. personal experience vs. official agency databases), use of technology (e.g. physical records vs. crowdsourced data), or scale of stakeholders (e.g. individuals vs. large government organizations). Moving towards digital and more formal information collection, storage, and dissemination can enable some stakeholders to more easily access and use transportation data. This does not always hold true, however. For example, app-based location information offers little to no benefit to users who do not own smartphones or do not have internet access; these individuals would likely be better served by paper-based maps and schedules. Thus, the characteristics of information collection and dissemination have important ramifications for how much information is available/accessible to whom, and what uses it may play.

¹ Informal” is a difficult term. The lines between “formal” and “informal” are blurry and transit services often fall along a spectrum of formality (rather than a binary categorization) depending on levels of enforcement, etc. Thus, many services characterized as “informal” might be more accurately called “semi-formal”. For the purpose of this thesis, I use “informal” to refer to transit that is largely or entirely unregulated by government agencies, recognizing the limitations of the term.
<table>
<thead>
<tr>
<th>Data type</th>
<th>Producer</th>
<th>Means of collection</th>
<th>Means of storage</th>
<th>Consumer</th>
<th>Means of delivery</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle type, number, and capacity</td>
<td>Owners, concession holders, operating agency</td>
<td>Purchasing records, physical observation, crowdsourcing</td>
<td>Physical records, digital databases</td>
<td>Government regulator</td>
<td>Digital or physical exchange</td>
<td>Service analysis, evaluation, &amp; regulation</td>
</tr>
<tr>
<td>Route</td>
<td>User, vehicle operators, transit operating agency</td>
<td>Individual experience, direct observation, AVL equipment or GPS devices, crowdsourcing</td>
<td>Personal memory, physical maps, spatial records (e.g. GIS), digital databases or feeds, hard-copy records</td>
<td>Users, operating agencies, regulators, developers</td>
<td>Personal communication, digital exchange, websites, physical maps and signage, web-based apps, mobile phones</td>
<td>Trip planning &amp; navigation, system monitoring &amp; analysis, operations planning, regulation</td>
</tr>
<tr>
<td>Schedule</td>
<td>Operating agency, regulator</td>
<td>N/A: a schedules are set, not collected</td>
<td>Digital or physical records, data feeds</td>
<td>Users, operating agencies, vehicle operators, regulators</td>
<td>Personal communication, physical schedules and signage, websites, web apps, digital exchange</td>
<td>Trip planning &amp; navigation, operations management &amp; planning</td>
</tr>
<tr>
<td>Travel time</td>
<td>Users, vehicle operators, operating agency</td>
<td>Personal experience, AVL/GPS equipment, crowdsourcing</td>
<td>Personal memory, physical records, digital databases</td>
<td>Users, operating agencies, regulators</td>
<td>Personal communication, printed schedules, web-based apps, digital exchange</td>
<td>Trip planning, system analysis, operations planning</td>
</tr>
<tr>
<td>Frequency</td>
<td>Operating agencies, uses, vehicle operators</td>
<td>Direct observation (vehicle counts), AVL/GPS devices, crowdsourcing</td>
<td>Physical records, digital records, personal memory</td>
<td>Operating agencies, regulators, users (self or others)</td>
<td>Physical or digital exchange, internet devices, personal communications</td>
<td>System analysis, operations planning, regulation, trip planning</td>
</tr>
<tr>
<td>Data type</td>
<td>Producer</td>
<td>Means of collection</td>
<td>Means of storage</td>
<td>Consumer</td>
<td>Means of delivery</td>
<td>Use</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------</td>
<td>---------------------------------------------------</td>
<td>---------------------------------------</td>
<td>------------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>Speed</td>
<td>Users, vehicle operators, operating agencies</td>
<td>Direct observation, AVL/GPS equipment, crowdsourcing</td>
<td>Personal memory, digital databases</td>
<td>Users, operating agencies, regulators</td>
<td>Personal communication, digital exchange</td>
<td>Trip planning, operations planning, law enforcement</td>
</tr>
<tr>
<td>Reliability</td>
<td>User, operating agency, regulators</td>
<td>Direct observation, AVL/GPS equipment, crowdsourcing</td>
<td>Personal memory, digital databases</td>
<td>Users, operating agencies, regulators</td>
<td>Personal communication, digital exchange</td>
<td>Trip planning, operations planning, service evaluation</td>
</tr>
<tr>
<td>Current conditions and service updates</td>
<td>Users, vehicle operators, regulators and other actors</td>
<td>Direct observation, AVL/GPS equipment, planned system disruptions</td>
<td>Digital records, data feeds, social media, personal memory</td>
<td>Users, vehicle operators, operating agencies</td>
<td>Personal communication, physical or digital signage, digital exchange</td>
<td>Trip planning &amp; navigation, operations management &amp; planning</td>
</tr>
</tbody>
</table>

Derived from Zegras, et al., Forthcoming
These types of transportation data play a key role in service operation, regulation, and planning. They also facilitate and encourage transit usage. Providing accurate and usable information to the public will, in theory, increase system understanding and accessibility for potential users (Hall, 1983) by raising awareness about different route and mode options and assisting passengers to navigate the system (Lyons, 2006). System ridership may, in turn, increase over time as passengers expand their destination choices and/or substitute public transportation for previously-used transport modes (Abdel-Aty, et al., 1996). At an even broader level, increased public transportation information can improve overall perceptions about a city in general (Freska, 1999).

Despite these benefits, publically-accessible/customer-facing data (e.g. routes, schedules, frequencies) remained limited and difficult to use until the past decade. Until recently, the public could only access information from printed maps and schedules, some transit agency websites, and physical signs posted along physical routes. As the open data and smartphone “app” eras emerged, many transit agencies expressed hesitations about releasing their schedule and route information in other digital forms to the public. New York City’s Metropolitan Transit Agency (MTA), for example, cited security and terrorism risks, as well as concerns about data integrity – regular service disruptions can make parts of its static transportation feed unreliable. Some agencies expressed fear that developers might use information to create apps that would spread erroneous transit information or otherwise blemish their agencies’ reputations (Rojas, 2012). Legal barriers and monetary motivations acted as another source of resistance; some agencies outsourced their data management to firms who claimed contractual rights to the information, and others simply didn’t want to lose advertising revenue on their own websites and trip planning services (Roush, 2012). For these reasons, most transit agencies took a default position of data protection rather than release.

As the open data movement became more popular in the USA and Europe, the public transportation industry experienced a shift in attitudes about transit data. However, even as agencies grew amenable to opening their data, they lacked a common digital format in which to release it. This prevented agencies from building upon other agencies’ and third-party developers’ advances with respect to data collection, use, and dissemination. Furthermore, developers often needed to spend valuable time developing methods to “screen scrape” (i.e. copy and convert to a useable format) schedule information off of transit websites, and to create individual and specific applications for different contexts (Roush, 2012).

---

2 Some exceptions did occur. For example, Roush notes that University of California students collected and released San Francisco’s transit information online in 1994 (Roush, 2012).
Portland, Oregon, eliminated many of these difficulties in 2005. The city’s transit agency, TriMet, partnered with Google to integrate public transit schedule and route information with Google Maps, enabling residents to search for transit directions online in the same way they could search for driving directions. In the process, TriMet and Google co-developed a non-proprietary transit data format, the General Transit Feed Specification (GTFS), to standardize and facilitate data release for other cities that might wish to follow suit.

The GTFS format offers simplicity and ease of adoption. It consists of a ZIP file (a compressed package of individual files) that contains a series of comma-delimited text files, including six mandatory files and seven optional files. Each file details one aspect of the transit information: stops, routes, trips, etc. Together, the files can describe most formal transit systems in their entirety. This simple format means that many transit agencies can use existing route maps and schedules to produce a GTFS feed in-house, using basic spreadsheet software, in the span of one or two days. Free tools also exist online for converting schedules to GTFS format. Small transit agencies that lack this technical capacity can generally commission private assistance at relatively low costs, typically around $3,000-5,000 (Wong, et al., 2013).

GTFS initially managed only static transit information (e.g. routes, stops, and schedules) as opposed to dynamic information (e.g. real-time bus locations and service-alert disruptions). In 2011, Google and a partnership of USA transit agencies and developers launched the GTFS-realtime (GTFS-RT) data format to once again facilitate live updates on transit fleets, drawing from Automatic Vehicle Location (AVL) technology that typically uses an on-board GPS to determine and transmit the vehicle’s geographic coordinates to a radio receiver, enabling remote-tracking. Together, GTFS and GTFS-RT act as a means for cities to integrate their public transit information into Google Maps, to incorporate real-time data on vehicles’ locations, and to enable independent software developers to create customer-facing websites and mobile applications using the data.

The Portland case reduced barriers to transit data that were both technical and political in nature. The GTFS format provided an easy path for agencies to store their data. More importantly, however, the growing push for open data raised pressures on transit agencies to release their information, and Google provided influence and pervasiveness that may have been necessary to “break the logjam” of agency hesitation (Roush, 2012). After about 2007, agencies in the USA began to follow Portland’s example and release their own transit data in the GTFS format. As of June 2013, an estimated 283 public transit agencies around the world

---

2 Originally called the “Google Transit Feed Specification”. Its name was changed in 2009.
released official GTFS feeds (Google Maps, 2013). When considering unofficial or private sources, the global number of GTFS feeds surpasses 700 (GTFS Data Exchange, 2014) or 1,000 (Google Maps, 2013).

**There’s an app for that**

Enabled by the growing prevalence of open and consistent transit data, developers created a wide variety of tools to benefit transit users, planners, and regulators. These apps serve a variety of purposes. Antrim and Barbeau (Antrim & Barbeau, 2013) list several categories and examples, shown in Table 2. As a result of these and other innovations, individuals have a range of tools at their disposal to make it easier to locate, understand, use, and improve transportation services. For example, OpenTripPlanner’s analyst extension (OTPA) can use a city’s GTFS feed to generate an accessibility coverage map. This analysis can measure raw travel time accessibility and can be combined with demographic and/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”). It can also be used to compare various transport scenarios as modeled in GTFS (e.g. accessibility impacts from adding a Metro extension). OTPA users have used this feature to show accessibility impacts of disruptions, such as Hurricane Sandy in New York (Byrd, et al., 2012). These tools can inform policymakers and planners and help explore questions about mobility and accessibility at different locations in the city.

---

4 Data taken from the Google Transit list of city providers, with layer-only feeds removed.
Table 2: Transportation tools and examples that use GTFS data

<table>
<thead>
<tr>
<th>Category</th>
<th>Purpose</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trip planning and maps</td>
<td>Plans a trip from one location to another using public transportation</td>
<td>Google Maps, OpenTripPlanner, Bing Maps, Hopstop</td>
</tr>
<tr>
<td>Ridesharing</td>
<td>Connects user with potential ridesharing matches</td>
<td>Parkio, Avego</td>
</tr>
<tr>
<td>Real-time info and/or mobile apps</td>
<td>Provides transit information on mobile devices (including real-time info)</td>
<td>Google Maps, Transit App for iOS, RouteShout, Tiramisu OneBusAway, NextBus, TransLoc</td>
</tr>
<tr>
<td>Data visualization</td>
<td>Provides graphic visualizations of transit routes, stops, and schedule data</td>
<td>Walk Score, Apartment Search, Mapnificent</td>
</tr>
<tr>
<td>Accessibility</td>
<td>Assists transit riders with disabilities in using public transportation</td>
<td>Sendero Group BrailleNote GPS</td>
</tr>
<tr>
<td>Interactive voice response (IVR)</td>
<td>Provides transit information over the phone via an automated speech recog.</td>
<td>BusLine, TransitSpeak, TravelSpeak</td>
</tr>
<tr>
<td>Timetable creation</td>
<td>Creates a printed list of the agency’s schedule in a timetable format</td>
<td>TimeTablePublisher</td>
</tr>
<tr>
<td>Planning analysis</td>
<td>Assists transit professionals in assessing the current or planned transit network</td>
<td>OpenTripPlanner Analyst Extension, Graphserver, Transit Boardings Estimation and Simulator Tool, TransCAD 6.0</td>
</tr>
</tbody>
</table>

Derived from Antrim and Barbeau, 2013

Policymakers and transit professionals typically expect improved access to high-quality transportation information to increase users’ satisfaction levels and to raise transit ridership (Ferris, et al., 2011). Initial research confirms this to some extent. Amongst Seattle bus-riders using OneBusAway, Ferris et al. found a substantial increase in overall satisfaction levels and a decrease in perceived wait times (Ferris et al., 2011). Brakewood et al. found a similar decrease in Boston, where perceived “usual” wait times declined amongst passengers using real-time information systems for commuter rail services (Brakewood, et al., 2014). On university campuses, real-time information for private shuttles decreased perceived wait times and passenger frustrations in Cape Town (Vanderschuren & de Vries, 2013) and increased passenger satisfaction at the University of Maryland (Zhang, et al., 2008), although the latter study found no quantifiable impact on rider frequency or mode choice. Transit data may also increase actual service performance; Ji et al. found that bus drivers at Ohio State University used real-time information in...
order to adjust their driving and better adhere to their schedules, improving transit reliability around campus (Ji, et al., 2014). Finally, ridership did increase (either from mode shift or additional trips) in at least one case; Tang and Thakuriah (Tang & Thakuriah, 2012) analyzed route-level monthly average weekday ridership on the Chicago Transit Authority (CTA)’s bus system for 2002 through 2010. Results suggest that a modest increase in ridership could be attributed to the CTA’s bus tracker service.

In addition to assisting riders, transportation information can also impact operator actions. Lee et al. surveyed 40 drivers in Baltimore about their reactions to real-time data; respondents stated that they would attempt to get back on schedule if they were running behind (31 of 40 respondents) or ahead (30 of 40 respondents) (Lee, Chon, Hill, & Desai, 2001). Watkins et al. conducted semi-structured interviews with 253 bus drivers in greater Seattle and found similar results; many drivers reported positive attitudes towards real-time information availability and stated that they would alter how they drive in order to adhere to their schedule (Watkins, et al., 2013). Ji et al. attempted to measure these effects empirically (Ji et al., 2014). They analyzed AVL data in order to investigate drivers’ responses to real-time location information and found that drivers did use real-time information in order to better adhere to schedules; they adjusted dwell times at time-point stops and changed their speed along roadways in order to provide more on-time service. Exogenous factors (e.g. traffic volumes, space at stops, heavy passenger demand at busy stops, and roadway geometry) constrain the potential for adjustments. Nevertheless, these findings suggest that real-time information can increase service reliability.

Further up the management structure, information (especially real-time information) can assist operators in managing and optimizing route locations, schedules, and frequencies (Ji et al., 2014). Regulators and local authorities can also benefit (Holdsworth, et al., 2007). Increased information could enhance regulators’ capacity to oversee transit services and enforce regulations (Eros, et al., 2014). Moreover, if information successfully increases ridership and public transit mode shares, it can help these officials to achieve local targets or larger-scale policy objectives (Basford, et al., 2003) – though this argument would benefit from empirical evidence. Moreover, open transit data may have wider implications for civic engagement. Rojas argues that it may enable a cycle of information availability, public feedback, and government response (Rojas, 2012). She notes that further research is necessary to investigate the outcomes of data disclosure initiatives, in terms of both citizen mobility and improved performance from transportation agencies. Nevertheless, the existing studies do suggest that open transit data may have significant impacts upon users, operators, regulators and transportation systems themselves – in places where these data exist.
Mind the global gap

While transit data became increasingly available and useful to individuals in rich countries, information remains scarce for those in the developing world. Fig. 1 shows the locations of 1,048\(^5\) public and private transit agencies around the world that release GTFS data through Google Transit (Google Maps, 2013). The map shows the concentration of GTFS feeds in the USA, Canada, Europe, China, and Japan – as well as a distinct absence in Latin America, Africa, and parts of Asia.

Figure 1: Google Transit GTFS feeds by location (as of November 2013)

Lack of official information presents one barrier to data availability. In developed countries, transportation services often include subways, streetcars, and buses that government agencies own or oversee. The same does not hold true in the Global South. Here, public transportation is often provided by non-scheduled, non-stop-based, private services provided by vehicles like tuk-tuks, jeepneys, matatus, colectivos, etc. Figure 2 shows the share of public transportation trips serviced by informal transit in cities described by peer-reviewed journals or World Bank studies (Cervero & Golub, 2007; Kumar & Barrett, 2008; Wilkinson, 2008). This figure demonstrates that many of the world’s large cities – and much of the global population – depend on informal structures of public transportation provision.

---

\(^5\) This figure does not include agencies that only release a visual layer of routes rather than a GTFS with stop and route information.
Figure 2: Global paratransit shares in selected cities, as a percentage of total public transportation trips

Key

City

City size scaled to reflect population
The characteristics of informal transit services vary widely by context, but their histories and operating structures exhibit many similarities, reflecting a legacy of institutional and financial capabilities, regulatory and operational ebbs and flows, and implicit or explicit policy decisions (e.g. laissez faire). In many cities –both industrialized and developing– public transportation arose when private vehicles started to operate along demand-based routes, filling a gap or absence of government-provided services. Vehicles often began operating as shared taxis and then grew to establish more fixed routes. They increased in number, expanded their coverage, and often supplanted other transportation modes to become the de facto public transportation systems in many cities.

Privately-owned informal transit fills an important service gap in many cities at a low direct cost to the government. It poses several challenges for urban environments, however, including violent on-street fare competition, crowded vehicles, traffic congestion from oversupplied routes, local air pollution from aging fleets, and law/regulation enforcement difficulties for local authorities (Cervero & Golub, 2007). In addition, privatized transit services often become too powerful or too difficult to oversee, and government agencies make attempts to consolidate and regularize the services. When the costs of service provision grow too large, governments privatize the services once again. Figure 3 depicts the phases of this cycle as described by Gómez-Ibáñez and Meyer (Gómez-Ibáñez & Meyer, 1993).

Gwilliam, having studied public transportation systems in various countries, finds that industrialized and post-colonial nations undergo slightly different cycles of regularization and privatization with respect to bus service in particular (Gwilliam, 2008). Figure 4 and Figure 5 depict these cycles for industrialized and post-colonial countries, respectively. These cycles can “stall” at different points. In post-colonial countries, operating dynamics can make major government intervention very difficult. Vehicles typically belong to owner-operators, who form route associations or cartels. Many city transit agencies hold some degree of official regulatory authority over these groups. In practice, however, route associations can (and sometimes do) protest government intervention by blocking intersections or suspending service, thus paralyzing the city. Because government agencies are unable or unwilling to break private sector cartels by force, the regularization process stalls at this stage, forming an atomized stasis depicted in Figure 6 (Zegras & Flores-Dewey, 2013). In cities with multiple types of public transportation, fixed-route modes (e.g. subway) may simultaneously be stalled at the public monopoly stage, leading to two sets of stasis (illustrated in Figure 6).

---

*C (Cervero & Golub, 2007) and (Iles, 2005) provide an extensive discussion of “informal” transportation in developing countries. See also (Sohail, et. al., 2006). For specific case studies, see: (Golub, et al., 2009; Salazar Ferro, et al., 2013), amongst many others.*
Mexico City offers an example of this phenomenon (Zegras & Flores-Dewey, 2013).

Figure 3: Gómez-Ibáñez and Meyer's public transportation privatization cycle

Reproduced from Zegras and Dewey, 2013
Figure 4: Regularization/privatization cycle in industrialized countries

Process: Return to public ownership in response to high perceived cost of service

Regulated public monopoly

Process: Direct regulation or quality agreements in response to local procedures

Regulated private local monopoly

Process: Free entry or franchising to combat regulatory capture and budget burden

Competitive private supply

Process: Consolidation by merger or success in franchise competitions

Private sector area monopoly

Process: Consolidation by merger or success in franchise competitions

Figure 5: Regularization/privatization cycle in post-colonial countries

Process: Withdrawal or nationalization of private suppliers due to stringent fare controls

Public/municipal monopoly

Regulated private monopoly

Process: Decline and failure of formal suppliers due to fare restraint; emergence of fragmented informal supply

Fragmented informal supply

Process: Self-regulation of informal sector suppliers to share revenue and avoid damaging conflict

Informal sector cartel

Process: Re-establishment of formal company supply by forced consolidation

Adapted from Gwilliam, 2008
In an atomized stasis, the government lacks the capacity to break private cartels but may successfully undertake less dramatic regulatory actions. Informal transit often operates in a grey space between legality and illegality. Operators may hold a government-granted concession to operate a certain number of vehicles along a particular route, for example. However, a concession holder may operate many more vehicles than he has registered. He may have purchased his concession informally from another owner. Or he may continue to operate under an expired concession that the regulatory agency has neglected to renew in a timely fashion. All of these actions place the informal operator in a space where he has some amount of legal right to operate, but is unwilling or unable (due to government inaction) to fully comply with the conditions. Illegality grants the government a certain amount of leverage to enforce service or safety regulations on operators during times when the government regulator has a high capacity to regulate services and to fund or otherwise expand formal modes of service. At the same time, operators can, to some extent, claim a legal right to their operations, and do provide a service that the government relies on, particularly during times in which government agencies are unable to provide an adequate supply or quality of formal
transit services. Thus, the government typically tolerates informal operations with varying and fluctuating degrees of opposition.

Information –or lack thereof– plays an important role in the dynamic relationship between informal transportation operators and regulators. As Gwilliam’s regularization cycle illustrates (see Figure 5), individual concession holders and owner-operators typically organize themselves into cartels in order to strengthen their hand against the government, set fares and operating areas, possibly share revenues, and mitigate conflict between operators. Cartels or vehicle fleet owners often possess data concerning their vehicle counts, route locations, frequencies, passenger counts, revenues, and profits. However, operators typically do not disclose this information to the government. Fluctuating power dynamics, tense relationships, and mistrust can leave little incentive for cooperation or coordination; the less the government knows about a service, the more difficult regularization becomes, and the more powerful operators become. This creates a landscape of "hidden information". In Gwilliam’s model, information is also "hidden". His cycle does not explicitly depict the role of information in movement from one stage to the next. However, the acquisition of such information may be a means of advancing along the cycle.

Until recently, for services operating in Gwilliam’s fragmented or cartelized informal supply scenarios, government agencies or the general public would have faced prohibitive difficulties generating or acquiring transportation route and schedule/frequency data without the assistance –or at least acquiescence– of transit operators. As a result, many government regulators lack detailed knowledge about where and when transit vehicles operate, and the data they do possess is often stored in the form of outdated, paper-based records rather than accurate digital maps and databases. This limits regulators’ capacity to monitor, regulate, assess, and plan its transportation services. A network of buses can be difficult enough to manage when cities know where and when the fleet operates. Absent even this basic information, regulatory enforcement and city-wide planning may become a prohibitively difficult processes.

The hidden information landscape also has important implications for transportation users. Information unavailability and inaccessibility means that travellers in many cities in the Global South lack access to transportation route maps and schedules, as well as other physical or digital information that would assist with trip planning (e.g. travel times, frequencies, fares, physical signage, etc.). Instead, users must generate information themselves in informal ways – either from personal experience, by speaking to other travellers, or by boarding a vehicle and asking its operator for information. These means of acquiring knowledge may produce incomplete information and cause difficulties for users, particularly when
navigating a new area of the city. Individuals may not know that they can reach a particular destination by public transit, or they may take an inefficient and time-consuming route, even for regular trips. Uncertain frequencies or schedules may cause passengers to wait in unsafe areas for long periods of time. Unexpected service irregularities could disrupt users' schedules and personal plans. All of these dimensions can constrain mobility, waste time, and limit the opportunities available to travellers. These impacts fall disproportionately upon a city's poorer residents, as users of informal transit almost always have lower incomes than users of other transportation modes. Moreover, in the long term, the negative aspects of unpredictable public transport escalate users' desire to escape dependency on the system – likely exacerbating motorization rates. In these respects, while many of their counterparts in, say, the USA can receive next-bus arrival times and live location information for individual vehicles, public transit users in the Global South may experience significant difficulties from a lack of official transit information.

Technological innovation, increased connectivity, and the growing popularity of smart city initiatives have begun to close the gap between transportation information availability in the Global North and Global South. Several cities around the world recently operationalized the mobile phone as a low-cost, portable, and discreet data collection tool. Local agencies or groups partnered with international institutions and, using various publicly-available or specifically-designed smartphone applications, they launched data collection initiatives in the Philippines (Krambeck, 2012), Dhaka (Zegras, et al., Forthcoming), Nairobi (Klopp, et al., 2014), China, and Mexico City (Eros, et al., 2014). The projects used similar methods; teams trained volunteers, students, or employees and sent them out to locate and ride every bus route in the city. The mobile application logged location points and stop locations along the way. Teams then cleaned, processed, and converted the data into GTFS feeds, GIS shapefiles, and sometimes physical bus maps (Badger, 2014; Burri, 2013). The project teams released their GTFS feeds on publically accessible websites and encouraged developers to begin working with the data.

Data collection and release represents the beginning of an important shift from hidden to open data; government regulators and teams of public users demonstrated the ability to generate digital route information, create public transit maps, and disseminate this information to the public in a standard format – without requiring the assistance or acquiescence of operators. In that sense, these projects began to democratize transportation information. This thesis investigates the potential impacts of newfound data availability on regulators, operators, and users. It asks whether transit data can act as a disruptive technology on public transportation structures and service provision.
Creative disruption

The term “disruptive technology” grew from a 1995 Harvard Business Review piece (Bower & Christensen, 1995). Intended for a management audience, the article describes how markets for goods and services often operate at an equilibrium stage marked by either stagnation or else movement along an upward trajectory of improving performance measures. Within this ecosystem, existing firms often fail to anticipate competition and the potential for a new customer base with different needs than the company currently meets. This leaves room for a new and different technology to enter the market and change the industry as a whole. For example, the advent of email increased the speed and ease with which individuals send messages and documents and, as a result, replaced much of the postal service’s volume. Plastics introduced a cheaper, lighter, and more customizable material for packaging and products. Digital media virtually replaced CDs and floppy discs. In these and many other sectors, disruptive technologies transformed a variety of commonplace activities. Table 3 lists examples and characteristics of several products that qualify as disruptive technology.

According to Christensen and Bower, disruption can create a new value network in two ways: either the disruptive technology enters the existing market and introduces an alternative that meets users’ needs at a lower cost, or else the disruptive technology outperforms incumbents through a new and different performance measure that becomes increasingly important. Typically, the disruptive technology first captures the market for low-end users, then improves its product to attract users with high performance needs (simplified in Figure 7). For example, early digital cameras offered a product that could instantly capture, display, and store hundreds of images without requiring the cost and inconvenience of using and developing film. These cameras offered a lower image quality than film, but could satisfy mainstream users’ needs. As technology developed, image quality improved such that digital cameras can now compete for users with high performance needs.

The emergence and impact of a particular early-stage technology may be difficult to predict. Based on Christensen’s description of market dynamics (represented in Figure 7), however, firms or individuals could anticipate innovation by examining a product’s current user base and searching for untapped, but related, user needs and market potential. In addition, firms could identify early-stage or “niche” disruptive technologies that have the potential to scale upwards and improve performance measures in order to disrupt the industry as a whole.
Figure 7: Disruptive technology in a market

derived from Bower and Christensen, 1995

Table 3: Disruptive technology examples and characteristics (products)

<table>
<thead>
<tr>
<th>Industry</th>
<th>Incumbent</th>
<th>Disruptive technology</th>
<th>Disruptive features</th>
<th>Enabler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communications</td>
<td>Postal mail</td>
<td>Email</td>
<td>Instant delivery, fast to compose, no delivery cost</td>
<td>Technology</td>
</tr>
<tr>
<td>Packaging, products</td>
<td>Natural materials (e.g. glass, metal, wood)</td>
<td>Plastics</td>
<td>Low-cost, light-weight, easily customizable</td>
<td>Technology</td>
</tr>
<tr>
<td>Memory</td>
<td>Physical media (e.g. CDs, floppy discs)</td>
<td>Digital media storage (e.g. the “cloud”)</td>
<td>Higher capacity, less subject to physical damage, easier to share and access and duplicate</td>
<td>Technology</td>
</tr>
<tr>
<td>Photography</td>
<td>Film camera</td>
<td>Digital camera</td>
<td>Instant ability to capture, view, and share images, does not require processing procedures costs</td>
<td>Technology</td>
</tr>
</tbody>
</table>

Derived from Bower and Christensen, 1995
Christensen initially focused his research on products themselves. Disruptions can involve services and organizational structures, however, and for this reason, Christensen soon revised his term in favour of “disruptive innovation” (Christensen & Raynor, 2003), a notion which considers not only the technology in question, but also the role of business models and consumer behaviour. These innovations may make use of new technologies (e.g. the internet) but their primary advance arises from their organizational or service-delivery model. Examples include Ryanair, easyJet, and Southwest’s “low cost, no frills” approach to air travel (Charitou & Merkides, 2003), Ford’s invention of the assembly line to produce affordable private automobiles, online open courseware’s entry into the for-profit education industry (Yuan, et al., 2013), Amazon.com’s online vending model (Charitou & Merkides, 2003), and Wikipedia’s crowdsourced knowledge database (Wikipedia, 2014). Table 4 describes the characteristics of these disruptions. All of these innovators offered a comparable product as incumbents, but did so in a way that was more affordable, more convenient, or more accessible to existing and potential users.

The examples given above all pertain to private firms. Disruptive innovation occurs less frequently in the public sector. Instead, productivity often falls over time, because bureaucratic structures and regulatory hurdles tend to dampen or prevent profit-based objectives, competitive pressures, and other innovation drivers (Eggers, et al., 2013). A Deloitte study argued that the opposite is possible, however. It argues that public agencies often hold significant buying power and can use it to shape product entry in new markets. In this way, agencies can steer sectors towards different approaches and methods of service delivery, enabling them to break stases or “deadlock” situations that sometimes stifle innovation (Eggers, et al., 2013). The report cites several examples (described in Table 5) where government bodies have adopted or created products or service models that can provide higher-quality services at lower costs: using digital tracking devices to release and monitor low-risk prison populations, considering alternative education models, etc.
Table 4: Disruptive technology examples and characteristics (services)

<table>
<thead>
<tr>
<th>Industry</th>
<th>Disruptive innovation</th>
<th>Incumbent</th>
<th>Disruptive features</th>
<th>Enabler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air travel</td>
<td>Low-cost carriers</td>
<td>Flagship carriers</td>
<td>Cheaper service with fewer “unnecessary” features</td>
<td>Service-delivery model</td>
</tr>
<tr>
<td>Automobile</td>
<td>Assembly line</td>
<td>Item-by-item production</td>
<td>Faster and cheaper</td>
<td>Production model</td>
</tr>
<tr>
<td>Vending</td>
<td>Amazon.com</td>
<td>Physical stores</td>
<td>Accessible by anyone at any time, eliminate the need for travel</td>
<td>Service-delivery model</td>
</tr>
<tr>
<td>For-profit education</td>
<td>Online courseware</td>
<td>Physical classrooms and classes</td>
<td>Low-cost or free, accessible by anyone at anytime</td>
<td>Service-delivery model</td>
</tr>
</tbody>
</table>

Derived from Christensen and Raynor, 2003; Charitou and Merkides, 2003; Yuan, et al., 2013; Wikipedia 2014

Table 5: Disruptive technology examples and characteristics (public sector)

<table>
<thead>
<tr>
<th>Industry</th>
<th>Disruptive innovation</th>
<th>Incumbent</th>
<th>Disruptive features</th>
<th>Enabler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrections</td>
<td>Digital tracking devices for low-risk inmates</td>
<td>Incarceration</td>
<td>Cheaper means to provide adequate surveillance</td>
<td>Technology</td>
</tr>
<tr>
<td>Public education</td>
<td>Charter schools, alternative learning methods</td>
<td>Traditional classrooms</td>
<td>Education tailored to specific populations</td>
<td>Service-delivery model</td>
</tr>
<tr>
<td>Defense</td>
<td>Unmanned aerial vehicles</td>
<td>Manned aircraft</td>
<td>Lower costs, no physical danger to pilots</td>
<td>Technology</td>
</tr>
</tbody>
</table>

Derived from Eggers, et al., 2013
Applied to public transportation (see Table 6 for detailed examples), disruptive innovation can occur through technological advances or modified organizational structures. These are not mutually exclusive categories, but innovations tend to fall more towards one category than another. Regarding technological change, new products or developments can affect supply or demand in a variety of ways. On the supply side, for example, bus rapid transit systems (BRT) carry high volumes of passengers on surface streets with low dwell times; when implemented properly, they can provide speed comparable to subway systems without necessitating heavy fixed-infrastructure investment costs (Levinson, et al., 2003). Additionally, intelligent transportation systems can reduce congestion, lower management costs, or provide information services to facilitate trip planning. On the demand side, many human interactions (both social and commercial) have shifted into the online realm, reducing a certain amount of individuals' need or desire for face-to-face interactions. Moreover, carsharing, the possibility of autonomous vehicles, transportation demand management efforts, and other innovations could reduce vehicle ownership and vehicle miles travelled (Shay & Khattak, 2010).

A second avenue for innovation concerns the organizational structure of transportation service provision. Governments hold a fundamental responsibility to ensure and promote quality of life for their constituents. Transportation agencies aim to fulfill this general mandate by providing planning and oversight for services that enable the city population to move around – ideally in an efficient manner, with smooth service operations. Thus, a disruptive innovation could change an agency’s organizational structure or operating methods in such a way that its capacity or public accessibility increases. BRT, for example, tends to spur organizational reform such as service consolidation, formalization, and regulation (Flores-Dewey & Zegras, 2012).

While the transportation literature discusses disruptive innovations enabled by technology (Shay & Khattak, 2010), it focuses less attention on those relating to organizational structures. One article does call attention to this avenue; Andersen (Andersen, 2013) classifies disruptive transportation innovation into three categories: (1) solutions that challenge conventional definitions of travel mode, (2) changes in mobile technology that undercut conventional notions of time or transportation’s value proposition, and (3) technologies that undercut the business model of urban transportation planning. The first two groupings concern technological change that changes transportation supply or demand, respectively. Andersen describes these at length. The final grouping relates to organizational structure, though with an emphasis on fiscal management. Andersen devotes only cursory attention to this type of innovation, devoting only one paragraph to a discussion of electric vehicles and their relationship to the gas-tax model of road funding.
### Table 6: Disruptive technology examples and characteristics (transportation)

<table>
<thead>
<tr>
<th>Disruptive innovation</th>
<th>Incumbent</th>
<th>Andersen’s category</th>
<th>Disruptive features</th>
<th>Enabler</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRT</td>
<td>Rail and bus</td>
<td>1</td>
<td>Easier to build and more flexible than rail; faster service and lower dwell times than conventional bus</td>
<td>Technology (vehicle type, dedicated lane design, boarding infrastructure)</td>
</tr>
<tr>
<td>BRT</td>
<td>Private, atomized transit services</td>
<td>3</td>
<td>Forces consolidation and formalization, enables regulation</td>
<td>Organizational model</td>
</tr>
<tr>
<td>Electronic tolling systems</td>
<td>Conventional (cash-based) toll booths</td>
<td>3</td>
<td>Speeds travel, reduces congestion, reduces fuel consumption, lowers management costs, increases information for planning purposes</td>
<td>Technology</td>
</tr>
<tr>
<td>Online vending and services (e.g. banking)</td>
<td>Physical retail locations</td>
<td>2</td>
<td>Lowers transportation demand by making services available online</td>
<td>Technology, service-delivery models</td>
</tr>
<tr>
<td>Carsharing</td>
<td>Private vehicles</td>
<td>1</td>
<td>Provides a lower-cost and convenient alternative to car rentals for individuals who need occasional vehicle access</td>
<td>Service-delivery model</td>
</tr>
<tr>
<td>Electric vehicles</td>
<td>Gas-tax model of road funding</td>
<td>3</td>
<td>Prevents users from having to “pay as they drive”; requires an alternative funding structure</td>
<td>Technology, organizational structure</td>
</tr>
</tbody>
</table>

Derived from Levinson, et al., 2003; Shay and Khattak, 2010; Flores-Dewey and Zegras, 2012; Andersen 2013

*Note: Andersen’s categories are coded as:
1: Solutions that challenge conventional definitions of travel mode
2: Changes in mobile technology that undercut conventional notions of time or transportation’s value proposition
3: Technologies that undercut the business model of urban transportation planning.

### Disruptive innovation in transportation innovation

This review of disruptive innovation demonstrates the power of innovation to create value for both existing and potential markets – either through technological developments or organizational models. While disruptive innovation represents
the norm for private-sector product development and service provision, examples are less common in the public sector.

Public transportation sorely needs innovation. In industrialized as well as developing countries, transit systems tend to stall at either a monopolized or atomized stasis in the privatization cycle. As a result, operators and service providers often lack the capacity or the incentive to improve services for existing and potential users.

Advances in information and communications technology (ICT) have produced conditions conducive to innovation in transportation information. Increasing mobile phone penetration and the availability of tracking applications now enables data collection efforts to take place in environments characterized by hidden information. Until recently, residents of these areas typically lacked maps, schedules, and other tools to access formal information about their transportation systems. Transit agencies themselves often lacked comprehensive route information, which made planning processes very difficult and exacerbated struggles to regulate informal transit services. By potentially democratizing data production and consumption, ICT enables transportation regulators (and the public at large) to generate, access, and share comprehensive, up-to-date transportation data in a digital and consumable format. Smart city and open data movements provide additional impetus for these initiatives.

In order for ICT to realize their disruptive potential, they must spur innovation that increases the accuracy and/or comprehensiveness of existing information in a manner that disrupts the current market for transportation planning, service provision, and/or service usage. These innovations would change dynamics amongst transit regulators, operators, and/or users and add value to transportation supply.

This thesis investigates the potential role of information as a disruptive innovation with respect to urban public transportation systems characterized by hidden information and the strong presence of informal services. I hypothesize that innovations stemming from new information could disrupt transportation supply in a few ways. First, the tools built from new data could assist users in trip planning tasks, increasing accessibility and mobility – particularly for users navigating new parts of a city. Information therefore disrupts transportation by adding value to current services; the service operates in the same manner, but offers a better quality of experience or a lower economic cost to users. This fits, roughly, into Andersen's second category of disruptive innovation since it changes the value of existing services. Second, data could enable planning professionals to better visualize and analyze accessibility within a city, raising their capacity for...
informed planning. This scenario leverages data and technological tools to disrupt the organizational model of transportation planning (Andersen’s third category). Finally, in a landscape characterized by hidden information, data could change relationships between transit agencies and informal transit operators; access to transit data may increase an agency’s relative power with respect to cartels and individual drivers, particularly in an atomized stasis. This could ameliorate some of a transit agency’s regulatory difficulties and perhaps enable it to improve service quality for passengers. In this scenario, information enables an agency to resolve an impasse by disrupting its organizational structure of service provision and perhaps its service delivery model (Andersen’s third category).

I explore these hypotheses and situate my analysis within the Federal District (DF) of Mexico City. In 2013, the DF’s Ministry of Transportation and Roads (SETRAVI) conducted a public transportation data collection project, funded by the World Bank. I participated in these efforts as a short-term consultant to the World Bank. I use this project as a case study to examine the role that information could play as a disruptive innovation in a particular context, focusing on the potential for innovation to create disruption amongst and between users, operators, and regulators.

Methods

The analytical method in this thesis focuses on three interlinked cases – users, regulators, and operators – examined in a prospective fashion, based on the data collected during fieldwork and follow-up efforts. The cases represent the different actors of relevance. Rather than formally testing my hypotheses, the approach is necessarily speculative. The research can be considered as an extension of hypothesis-generation; more like hypothesis validation, by formalizing the understanding of possible pathways of effects.

I base my analysis on existing academic literature, observations and informal discussions that took place over two project-related trips to Mexico, and semi-structured interviews with 16 microbus operators and route association leaders.

Project-related fieldwork occurred in April and June 2013. During this time, I participated in meetings and training sessions with representatives and staff from all seven of the city’s public transportation agencies.

Interviews took place over the course of one week, in June 2013, and were recorded to facilitate later analysis. These discussions lasted anywhere from half an hour to two hours. They took place at route offices or private homes and were conducted in Spanish. Staff at EMBARQ’s Mexico City office were instrumental in contacting operators and making arrangements for these meetings. 
Interview questions probed for details regarding four main areas:

(1) Characteristics of microbus operations and management
(2) Experiences and perceptions regarding microbus formalization
(3) Opinions about sharing bus operations information with the government
(4) Opinions about sharing real-time information with users

Responses were translated into English and analyzed for content.
THREE | EMPIRICAL CONTEXT

Public transport in the DF: A cyclical century

Mexico City’s Federal District (DF) and its metropolitan area (MCMA, shown in Figure 8) epitomize today’s megalopolitan challenges. The DF, itself, represents essentially a single jurisdiction with one mayor and an estimated population of 8.9 million, yet the broader MCMA encompasses another 12 million people spread over some 40 additional local jurisdictions in two states. This arrangement poses institutional and operational challenges for transport and other sectors. This case focuses almost exclusively on services in the DF, where since 1975, the Ministry of Transportation and Roads (SETRAVI) has regulated both technical and non-technical aspects of public transportation planning and policy.

Figure 8: Mexico City Metropolitan Area (MCMA)
Public transportation in the DF is improving but has historically been characterized as disorganized and poorly integrated (Loftus-Otway, et al., 2009). A historic lack of integrated planning, along with the government’s inability to enforce regulations, led to this disorganization. Throughout the 20th century, Mexico City’s public transportation experienced several periods of privatization and public incorporation. The sector moved through the stages of Gómez-Ibáñez and Meyer’s privatization model (see Figure 3) with cycles beginning anew roughly every twenty years. Within this evolving political landscape, transit modes were developed independently and without sufficient attention to network interconnections (Loftus-Otway, et al., 2009). This left the Mexico City government with discontinuous policies and a “weak institutional capacity to plan, build, and regulate public transport” (Flores-Dewey & Zegras, 2012).

Mexico City’s public transportation services began in an entrepreneurial fashion in the early 1900s, when a British company began operating the city’s first trolleys. Rapid urban growth occurred over the next few decades; Mexico City’s population expanded exponentially from 350,000 in 1900 to two million in 1940 (Wirth, 1997). Faced with increasing transit needs, the Mexican government purchased the trolleybus company in 1946. It added electric trolleybuses and expanded the service, renaming it Servicio de Transportes Eléctricos del Distrito Federal (STE). As the city’s population continued to grow, the DF created a diesel bus service to accommodate passenger demand without necessitating fixed infrastructure.

From 1960 to 1981, Mexico City’s diesel buses were run privately, typically by owner-operators. To organize themselves, bus drivers formed a union, the Alianza de Camioneros, which grew over time into a politically powerful and highly corrupt association (Wirth, 1997). The group undercut trolley budgets by pressuring the government for subsidies. As more and more buses entered the roads, they increasingly dominated the transportation mode share. In addition, fixed-route taxis (peseros) began serving service gaps, particularly on the outskirts of the city. Trolley services lost prominence as a result (Ward, 1990).

Other forms of transit also proliferated during these decades. A growing middle class fostered increased demand for personal vehicles, and private vehicles became more common. The rise of the automobile brought with it considerable traffic congestion, contributing to growing air pollution levels and an increasingly crowded urban centre (Legorreta, 1988; Wirth, 1997). The Mexican government responded by creating an underground subway system, Sistema de Transporte Colectivo (STC), whose first line opened in 1967. Shortly afterward, the DF government began to license and regulate privately-owned and operated minibuses and vans. At this time, public transport in the DF was dominated by high-capacity vehicles: large buses, trolleybuses, and Metro. However, as Wirth
argues, these high-capacity modes increased too slowly to keep pace with rapid population growth and urbanization (Wirth, 1997). Moreover, the government became increasingly wary of the Alianza de Camioneros’s growing power and influence. The city therefore attempted to downsize the bus system and willingly granted licenses necessary for microbuses to multiply. By 1980, approximately 41,000 microbuses (or colectivos) operated in the DF (Wirth, 1997). These vehicles filled service gaps left by higher-capacity vehicles and reduced the relative power of the Alianza.

In 1981, the mayor of Mexico City revoked all 7,000 bus drivers’ existing concessions and again transformed its diesel bus system into a publicly run entity, incorporated as the Autotransportes Urbanos de Pasajeros Ruta-100, or simply Ruta-100 (Flores-Dewey & Zegras, 2012). Financial troubles and internal strife marked this period. The bus drivers’ union expressed opposition to the ruling party through periodic, disruptive protests; Ruta-100’s leadership became known for internal corruption; and expanding subsidies and other fiscal difficulties left the organization struggling to manage its operational costs (Wirth, 1997). Service levels deteriorated accordingly (Flores-Dewey & Zegras, 2012). The government reacted by continuing to permit the microbus sector’s rapid growth, and by tolerating microbus drivers’ unregistered operations. By 1995, microbuses in the DF numbered 72,000 and were approaching 55-60% of motorized trips in the city (Flores-Dewey & Zegras, 2012). In this year, Ruta-100 was declared bankrupt after a strike. Bus services ceased and the association spurred violent confrontations and protests for a year afterwards. In 2000, the government re-instated bus services under a newly-created public entity, the Red de Transporte de Pasajeros, which currently operates the diesel bus service.

Over the last two decades, privately owned, low capacity vehicles grew increasingly prominent – despite SETRAVI’s desire to disrupt the atomized stasis by formalizing, restructuring, or removing microbus services. The microbuses provide extensive transit services to the majority of the city’s population and require relatively low expenditures and operational burdens. These conveniences fail to compensate for larger issues, however. An oversupply of vehicles increases traffic congestion and air pollution, prompts on-street competition for passengers, and spurs dangerous driving and collisions (Cervero & Golub, 2007; Dodero, 2013; Flores-Dewey & Zegras, 2012; Wirth, 1997). The organization structure of the microbus system also renders regulation and planning exceedingly difficult for SETRAVI and its sub-office for microbus oversight, the Dirección General de Transporte (DGT). This difficulty stems partly from logistics. Over 100

---

2 Cervero estimates that about half of the city’s paratransit operators lacked licenses and adequate insurance (Cervero, 1998).
independent associations represent over 22,000 microbus owners in the DF; The DGT spends considerable time and focus mediating conflicts between these bus operator organizations, and between microbuses and other transit modes, at the cost of system planning efforts (Flores-Dewey & Zegras, 2012). It also must engage with these organizations, typically in ad hoc negotiations, in order to determine and enforce rules and regulations. The associations accumulated significant political power over the years, which they have wielded in myriad ways: blocking the routes of bicitaxis and Ruta-100 vehicles, threatening to cut off main arteries if they are fined for stopping at undesigned points, committing fraud in vehicle sales and registrations, and generally leveraging power for financial gain (Wirth, 1997).

This situation and history features enough complications to keep any transportation agency busy. On top of all this, Mexico City faces added jurisdictional challenges because transportation services are not integrated between the DF and the greater MCMA. “Mexico City” functionally encompasses both the DF and the states of Mexico and Hidalgo (see Figure 8); 17% of trips originating in the DF end in the State of Mexico, and 25% of trips that originate in the State of Mexico end in the DF (SETRAVI, 2007). The DF border acts as an impermeable barrier for public transportation services and planning, however. Passengers making these trips must take one form of transit (likely a microbus) to the DF boundary, then transfer to another vehicle to transport them into the DF. They first pay the (higher, less subsidized) microbus fare in the State of Mexico, then pay a second time once they cross into the DF. This creates financial and economic costs for passengers that can restrict their accessibility. In addition, it poses a challenge for regulators seeking to provide effective services for the megalopolis.

In the midst of this context, the DF began efforts to shift to a more formal system organized around a new, high-capacity Metrobús bus rapid transit (BRT) service, launched in 2000. Metrobús recently launched a fifth line and SETRAVI plans to expand this service further to add high capacity trunk routes to the DF’s public transportation system. However, SETRAVI’s inability to regulate the microbuses slows its attempts to reform and restructure transportation services. Protracted and expensive negotiations, lasting upwards of one year, must take place with opposed microbus route associations before a microbus route can be displaced to make room for BRT services. A lack of data and coordination exacerbates these difficulties.

In addition to BRT-related expansions, the DGT has also been interacting with microbus route associations in order to facilitate an incremental formalization process that would transform microbus routes into a system like the RTP, with
higher-capacity vehicles and driver remuneration that is decoupled from passenger fares.

This entire narrative can be encapsulated by Gómez-Ibáñez and Meyer's privatization cycle (Figure 3). As the model predicts, transit services emerge in an entrepreneurial fashion, undergo consolidation, and fall under government regulation until declining profits force service reductions and require buttressing from public subsidies until rising costs force prompt the government to de-regulate the service back to its original, private state (Gómez-Ibáñez & Meyer, 1993). In fact, Mexico City today arguably more closely represents the situation depicted in Figure 6, with both atomized private and public (e.g., Metro) stases simultaneously strong.

**Current state of public transportation operations**

At the present moment, in 2014, all public transport in the DF falls, to some extent, under SETRAVI's oversight, either as operator (e.g., STE) or regulator (e.g., DGT). This includes six relevant agencies (not including taxis or the Ecobici bike share system). Figure 9 describes the fleet size and estimated ridership of each of these agencies, illustrating the overwhelming importance of microbuses as a means of transit for the bulk of public transportation trips in the DF.

Technical and non-technical capacity varies between the transit agencies, as does their access to accurate, comprehensive, formal data about services provided. Table 7 summarizes the type of data held by each agency, its source, and its consumers. Figure 10 attempts to visualize this by taking each group of stakeholders and situating transit modes along a continuum of data accuracy (i.e. correctness of the information) and data comprehensiveness (i.e. completeness and precision of the data) in order to demonstrate variance in data quantity and quality that may be accessed by different stakeholder groups. On the x-axis, comprehensiveness moves from very basic knowledge about a minority of routes, to knowledge about all routes, to knowledge about all routes and their frequencies or other aspects, to real-time knowledge of fleet patterns and locations. On the y-axis, accuracy moves from very low to very high. Metrobús (BRT) belongs at one end of the spectrum (high accuracy and comprehensiveness) because it keeps digital records of its routes, stops, frequencies, and travel times, and the agency can track passenger trips through electronic fare cards. Its vehicles also have on-board AVL data capable of transmitting real-time location information back to a new data and operations centre. Slightly lower on the accuracy and comprehensiveness axis, the STC possesses significant data concerning subway operations, but does not have live-tracking capabilities for the fleet. At the other end of the spectrum, microbus users may only know about a small selection of routes, but may have
reasonably accurate knowledge from travelling on these vehicles regularly. Up until the data collection project, regulators had knowledge of a greater variety of routes and some additional aspects, but these data were outdated and somewhat inaccurate. Data are generally more difficult for the DGT to obtain, since this sub-office regulates services that do not operate on fixed infrastructure. Microbus route associations, on the other hand, collectively possess quite complete and accurate knowledge of the system's routes and frequencies, though they rely on drivers to collect fares and monitor compliance\(^8\), and do not have a digital means to track ridership or vehicle locations.

Figure 9: Public transportation in the DF\(^9\)

\(\text{TREN SUBURBANO} \quad 1 \text{ rail line} \quad 134,000 \text{ daily pass. trips}\

\(\text{METROBUS} \quad 4 \text{ BRT lines} \quad 700,000 \text{ daily pass. trips}\

\(\text{STC (METRO)} \quad 12 \text{ subway lines} \quad 4,200,000 \text{ daily pass. trips}\

\(\text{STE (TROLLEYBUS)} \quad 8 \text{ electric trolleybus lines} \quad 241,000 \text{ daily pass. trips}\

\(\text{RTP (DIESEL BUSES)} \quad 100 \text{ bus routes} \quad 750,000 \text{ daily pass. trips}\

\(\text{DGT (MICROBUSES)} \quad 1,200 \text{ bus routes} \quad 8,700,000 \text{ daily pass. trips}\

\(^8\) Some microbuses are now equipped with motion sensors to track passengers but drivers have strong incentives to tamper with the devices (or otherwise block them from operating correctly) in order to report lower ridership and retain greater revenue for themselves.

\(^9\) Data sources: Personal communications with agency staff; academic literature and news media (Francke, Macias, & Schmid, 2012; McConville, 2010; Webb, 2009)
<table>
<thead>
<tr>
<th>Agency</th>
<th>Information available</th>
<th>Information source</th>
<th>Information consumers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STE</strong></td>
<td>Routes and stops</td>
<td>Fixed route info stored as ArcGIS shapefiles or KML files, available as web-based or printed maps</td>
<td>Users, operators, regulators, developers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Users, operators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ticket sales</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td><strong>STC</strong></td>
<td>Routes and stops</td>
<td>Fixed route info stored as ArcGIS shapefiles or KML files, available as web-based or printed maps</td>
<td>Users, operators, regulators, developers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Users, operators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Electronic fare cards and turnstiles</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td><strong>Metrobús (BRT)</strong></td>
<td>Routes and stops</td>
<td>Fixed route info stored as ArcGIS shapefiles or KML files, available as web-based or printed maps</td>
<td>Users, operators, regulators, developers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Users, operators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Electronic fare cards and turnstiles</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>On-board vehicle AVL systems that relay information to a new operations centre</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td><strong>Suburbanos</strong></td>
<td>Routes and stops</td>
<td>Fixed route info stored as ArcGIS shapefiles or KML files, available as web-based or printed maps</td>
<td>Users, operators, regulators, developers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Users, operators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Electronic fare cards and turnstiles</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Operator records</td>
<td>Operators, regulators</td>
</tr>
<tr>
<td>Data Source</td>
<td>Information Type</td>
<td>Data Provider</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>------------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td><strong>RTP</strong></td>
<td>Routes and stops</td>
<td>Users (though not in apps), operators, regulators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fixed route info stored as ArcGIS shapefiles or KML files, available as web-based or printed maps</td>
<td>Users (though not in apps), operators, regulators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Approximate schedules and frequencies</td>
<td>Operator records</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Passenger counts</td>
<td>Ticket sales</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revenues</td>
<td>Operator records</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vehicle counts</td>
<td>Operator records</td>
<td></td>
</tr>
<tr>
<td><strong>DGT</strong></td>
<td>Routes and stops</td>
<td>Regulators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frequencies</td>
<td>Operators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Revenues</td>
<td>Operators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vehicle counts</td>
<td>Operators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vehicle counts (less accurate)</td>
<td>Operators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vague route descriptions (outdated)</td>
<td>Operators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vehicle routes, stops, frequencies for some studied routes</td>
<td>Operators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vehicle routes (out of date)</td>
<td>Operators</td>
<td></td>
</tr>
<tr>
<td><strong>SETRA VI</strong></td>
<td>Mode share, origin-destination information, other demographics and mobility metrics</td>
<td>Regulators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Origin-destination survey results stored in digital databases</td>
<td>Regulators</td>
<td></td>
</tr>
</tbody>
</table>

Data source: Agency websites, contact and meetings with agency staff
In terms of external data availability, STE, STC, Metrobús (BRT), RTP, and the suburbanos individually publish route maps and schedule information on their websites. Software developers have been using this information for years in order to create mobile apps with transit maps and route planning capabilities. Before the open data project began, 28 transit-related apps were already available for Mexico City on Android or iOS platforms. Most of these apps simply showed the locations of the Metro stations and trip-planning directions from the user’s location – not integrated with, but simply overlaid upon, a map. A few apps also included this information for the electric trolleybuses, suburbanos, Metrobús, and STC. The apps remained limited, with no reliable travel time information and oft-outdated
maps. None of the apps included information for the RTP or microbus services, and microbus maps are nonexistent in general.

SETRAVI experiences the most difficulty in collecting and maintaining data about microbus services due to a lack of regulation and access to accurate, comprehensive route and operations data. That does not imply that the service is disorganized or “informal”, however. Rather, information exists as hidden data that route associations do not share with the government. Likewise, organizational structure exists, but is largely outside the scope of government regulation.

Interviews with microbus operators, researchers, and route association officials revealed that microbuses in the DF operate in an organizational structure that can be likened to the form of a pyramid (see Figure 11). Bus operators lie at the bottom of this pyramid. While many drivers own and operate their own vehicle, others earn a portion of the fares they collect. Transportation legislation states that an individual can own and operate no more than five microbuses. However, concessions are often illegally bought and sold, such that roughly 15% of microbuses are owned by large fleet owners, or flotillas. The remainder belong to individuals possessing a single or small number of buses. Each bus owner belongs to a route association (also referred to as a union, though technically functioning as a cartel), to whom he pays concessionary fees in exchange for operating privileges, legal assistance in the event of an accident, organizational structure, and access to route-owned garages. Larger unions also purchase fuel at wholesale prices and sell it at private petrol stations. The route association in turn pays licensing fees and other charges to the DGT, a sub-office of SETRAVI that holds responsibility for regulating the microbus services. The DGT grants concessions to route associations, giving them the right to operate along specific ramales, or branches, belonging to a ruta, or route. At a macro level, the rutas organize themselves into cupulas which act as cartels and represent the rutas in negotiations with the DGT and SETRAVI. This structure transfers most of the organizational burden and financial risk from the DGT onto microbus unions and operators, but creates strong regulatory challenges in return. Opacity characterizes the system; the DGT possesses scant information about the specific location of each ramal, which vehicles and drivers operate along it, and how many passengers they typically transport.

A sub-office of the DGT, staffed by approximately 25 employees, collects and handles information regarding microbuses. Its existing data are outdated and managed inefficiently. The DGT last surveyed the routes in 2001 by requesting vehicle and route information from each operator. It stores the results as a collection of binders containing 35,000 paper records that list each ruta’s ramales, their origin and destination, the distance of the route, the number of vehicles
operating along it, and the passenger load per day. Shortcomings abound in this information system. First, the records often do not describe the routes themselves, they simply list the approximate start and finish of the ramal without necessarily describing intermediary points. Second, the records are twelve years out of date. Moreover, the data are self-reported, so it is difficult to determine their initial accuracy. Finally, this information is not stored digitally or linked to spatial records. The DGT does have a shapefile showing about 100 of the microbus routes, but this file is likewise incomplete, outdated, and of questionable accuracy; it was created by a consulting team hired under the previous administration and disappeared after the last administration change; SETRAVI had to request a copy from a third-party. This highlights another problem regarding data management in Mexico City: lack of continuity following political change.

Figure 11: Microbus organizational structure

To address some of these issues, the DGT began an attempt to update its information in March 2013 by sending a form to each route association asking for route and fleet details and by meeting with the route association leaders. Several aspects limit this approach's effectiveness. The data are once again self-reported by route association leaders who have incentives to misrepresent their routes. Furthermore, the DGT did not specify a consistent format (e.g., Excel or KML) in which operators should provide information, so employees will have to manually process responses from each of the 121 rutas. Finally, response rates may be very low; two months after the DGT's request, only one route association had supplied the requested information.
DGT may not possess information about the microbuses, but that does not imply that this information does not exist. Again, the microbuses operate under a highly-structured system of branch and route leadership. Access to, and control of, information forms a major source of route associations’ political power. Leaders in the associations generally know exactly where their services run, and at what frequencies. Some of this information is collected with modern technology; during fieldwork in Mexico City, the author met with one association that employed its own dedicated GIS staffer. Other data are collected through more analog or labour-intensive means. For example, other operators explained that their dispatch services are performed by young men (and sometimes underage boys) who stand on the street and use radios to perform their duties. These individuals are sometimes tasked with keeping records of bus frequencies during their shift. Through these methods, bus owners and route association leaders keep a close watch on where, when, and how frequently their vehicles travel from origin to destination. Operators are unwilling to share this information with SETRAVI, however, because of their pervasive distrust and suspicion of the government in general and the Ministry of Transport in specific – or because they prefer to prevent SETRAVI from enforcing regulations or otherwise using information to the operators’ disadvantage.

Lack of data for the microbuses poses a hurdle for broader service improvements, such as BRT expansions. Discussions with microbus operators revealed that closed-door displacement negotiations take place between cupula and the DGT when a proposed new BRT route requires the expulsion or modification of an existing microbus ramal. These negotiations often last between 12 and 18 months. During the process, the DGT often contracts external consultants to conduct expensive field studies of the ramal, involving field observation and vehicle counting, which can take upwards of three months. Many of the data collection efforts are undertaken to gain basic information about route characteristics, vehicle counts, and passenger loads; these data are used to assess the financial value of the ramal so that DGT and the cupula can determine adequate compensation for the ramal’s displaced operators.

Even though operators generally collect their own frequency, route, and revenue data, information asymmetries necessitate external studies. Displacement negotiations require the DGT to offer compensation for an asset (i.e. the right to operate along a certain route) whose value it does not know. Route associations have strong incentives to over-report passenger loads and trip frequencies or otherwise misconstrue their route information to exaggerate the concession’s value

---

10 Previous research confirms this (Dodero, 2013). Similar dispatch approaches are common throughout such service types in Latin America.
and gain increased displacement compensation from the DGT. Meanwhile, the DGT has incentives to underestimate a concession’s value, and nearly all interview respondents described a general lack of trust in the government and reported a strong reluctance to rely on government estimates or figures.

The DGT typically relocates *ramales* and microbus operators that become displaced by Metrobús lines, since the majority of operators do not become Metrobús drivers. Rather than solving the problem, this approach causes it to migrate to a different area – a phenomenon sometimes referred to in Mexico as the “cockroach effect” (The Economist Blog, 2011). Shifting routes may remove tensions around the Metrobús area but simply transfers the stresses elsewhere. Territorial conflicts can ensue and even the relocations may only serve in the short-term; for example, during the construction of Metrobús Line 3, the DGT moved one *ramal* from *Ruta 1*, only to relocate it a second time for another Metrobús expansion. The lack of comprehensive and accessible route information may prevent the DGT from identifying a more acceptable place to relocate displaced *ramales*.

Overall, approximately nine million passenger trips take place aboard microbuses in the Distrito Federal each day, and SETRAVI has very little information about where these trips occur and how its citizens get around, other than the latest origin-destination survey (administered in 2007). It is difficult to regulate a service with so little information. SETRAVI’s regulatory and planning abilities are therefore severely hindered by the opacity of the transport industry.

### 2013 data collection and dissemination project

Within this context, the global open data movement reached Mexico City and began to influence politics. In December 2012, Miguel Ángel Mancera won the race for the DF’s mayoral office. His election platform was based partly on open and responsive governance and included specific reference to open data. After his election, SETRAVI’s new leadership followed through by enrolling relevant agencies in the process of transit data collection and spurring the creation of the city’s first consolidated transportation data database. These efforts received financial support from the World Bank and technical assistance from a team of consultants from Integrated Transport Planning (ITP) and Conveyal.

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 the 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 a static GTFS.

GTFS creation involved two processes: existing data conversion, and new data collection. In April 2013, the team held a series of workshop sessions with government officials and NGOs to discuss the role and potential uses of open transit data and introduce the Android app and GTFS editor. After the workshop, transit agencies with existing route and stop information used the GTFS editor to convert their data into GTFS format. The converted data includes information for about 125 lines, 260 route variations, and over 5,000 stops, covering the entire Metro, Metrobús, RTP, STE, and suburban rail systems.

The team also oversaw an extensive data collection effort focusing on the microbuses. Since existing data were unavailable, the team contracted staff from a local office of a global transportation NGO (EMBARQ) to use TransitWand to collect route and stop location information for all microbus routes within the DF. Operators were not informed that their vehicles were being tracked. By October 2013, the staff had spent roughly 2,000 person-hours tracking nearly 1,100 *ramales* – more or less every route in the city. Though highly labour-intensive, this brute-force method, combined with the existing data conversion, resulted in the city’s first comprehensive and accurate database of public transportation information.

From the converted and collected datasets, SETRAVI created a GTFS feed that includes route, stop, and estimated frequencies and trip times for each responsible transit agency in the DF. The GTFS data feed is currently limited to static information since real-time positioning was not available for most modes of transportation. Moreover, SETRAVI has so far excluded microbuses from the publically available feed. Microbus operations are not easily GTFS-compatible since they do not consistently adhere to fixed stop locations, routes, schedules, consistent headways, or trip times. A workaround was developed for this project in order to include the microbuses in the GTFS feed, but SETRAVI still chose to exclude them. This is an important point to note; the Mexico City data collection project was intended as an open data initiative, yet SETRAVI has chosen to leave potentially the most valuable set of data as “hidden information”. This could stem from several motivations. If the data are incomplete or inaccurate, then a sponsored data release would reflect poorly upon the agency. Moreover, releasing the data could have political significance that SETRAVI prefers to avoid, since data availability could highlight the importance of the microbus sector and the lack of high-capacity, fixed-route services serving much of the city. Finally, data release could make the service look more attractive, in contrast to SETRAVI’s demonstrated interest in consolidation and restructuring into higher-capacity
modes. In any case, the SETRAVI has so far chosen not to include microbuses in the public GTFS feed.

After creating the initial GTFS feed, SETRAVI held a second series of meetings and workshops in June 2013. The team shared results with agencies, introduced a live service-disruption-tracker tool for internal use, released the GTFS feed on SETRAVI’s website, met with Google to discuss feed inclusion on Google Transit, promoted the launch of an open trip planner, and showcased the data at a meeting with software and mobile-app developers. Additionally, non-profit organizations and SETRAVI jointly hosted a “hackathon” challenge in Mexico City to spur mobility-related solutions to specific problems.

The feed gives SETRAVI a single data stream (for internal use) that describes all of the multimodal transit services under its authority. It also allows the agency to move towards a more formal, comprehensive, and consumable set of internal data on the microbus sector. However, these data could still be improved; the feed faces certain limitations because it does not include real-time information and relies on estimated trip times and frequencies. Nonetheless, the data may act as an important and useful first step in Mexico City’s move towards open transportation data. Completion of SETRAVI’s data collection efforts (including the microbus data) marks the first time that the government of the DF has had a complete and accurate snapshot of transportation within the city – all in a standard format primed for data analysis and utilization. While SETRAVI previously held data describing its more fixed transportation services, it lacked virtually any information concerning the routes or operations of microbuses and the nine million passenger-trips they accommodate each day.

The transportation history overviewed in this chapter indicates the microbus network’s growing importance over time, the DF’s current dependence upon these vehicles, and the deliberate information asymmetry between route associations and the DGT. New data’s entrance into a previously opaque information landscape could have impacts upon various microbus system stakeholders and the relationships between them. This thesis’ next chapter identifies and describes the relevant stakeholder categories, their potential uses for the transportation data, and the possible ways in which the data may affect relationships between stakeholder categories for Mexico City’s predominant form of public transportation. I assess potential scenarios moving forward, and the ways in which transportation information may act as a disruptive innovation in transportation service provision and regulation.

---

11 In the following months, the project team released a tool for alerting passengers of service disruptions
FOUR | POTENTIAL IMPACTS OF INFORMATION

Transportation data availability could have impacts on at least three categories of stakeholders: users, regulators, and owners/operators. This chapter examines how recent data collection possibilities change the data quality and availability within and between each stakeholder category, and uses this analysis to assess the types of disruptive innovation that could occur as a result. It then presents three potential scenarios going forward and examines the feasibility and role of information in each in order to determine how, and to what extent, disruptive transportation innovation is likely to take place in Mexico City.

The scenarios consider “transportation information availability” to mean an accurate, comprehensive, digital database of transit routes that are stored in a consumable way (e.g. a static GTFS feed). Frequency information and real-time data are considered and discussed specifically, but are not assumed. Data collection methods could involve methods similar to the EMBARQ-facilitated data collection that took place in Mexico City, a crowdsourced method (e.g. a similar, smartphone-based data collection initiative, but involving a large group of users to track the individual microbus routes they already ride, rather than a handful of paid staff to locate and track the DF’s entire microbus network), or a top-down approach (e.g. installing sensors on vehicles themselves). For the purpose of assessing disruptive innovation, this analysis considers methods to be less important than resulting actions and potential. For example, while a crowdsourced data collection effort would represent a greater democratization of data, the resulting dataset itself would have the same disruptive potential as a top-down data collection effort that yielded the same type and quality of data and was similarly released for use by third parties and transit users. The critical point of analysis is assessing how the information can/would be used to improve existing aspects of transportation’s accessibility, value proposition, organizational model, and service-delivery model.

Data changes and potential disruption

Transportation users

Transit users in Mexico City constitute a very large and diverse set of individuals. Roughly two thirds of all motorized trips in the MCMA take place on public transportation, indicating that the majority of Mexico City’s residents rely on transit to navigate the city (Guerra, 2014). Socioeconomics do affect mode choice, however. Anecdotal evidence suggests that wealthier individuals favour the Metrobús (along with taxis and private cars) because they generally perceive the
BRT system to be safer than other public transport modes, both in terms of personal security and driving behaviours. In contrast, lower-income populations rely more heavily on the Metro and microbus networks in order to travel throughout the city. Table 8 shows mode choice by income level for microbuses and cars, demonstrating how microbus service appears to be considered an inferior good in mode choice decisions.

Table 8: Mode choice by income level

<table>
<thead>
<tr>
<th>Income quintile</th>
<th>Estimated average monthly income in 2007 dollars</th>
<th>At least one microbus segment</th>
<th>A private car</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorest</td>
<td>$206</td>
<td>62%</td>
<td>14%</td>
</tr>
<tr>
<td>Second</td>
<td>$381</td>
<td>60%</td>
<td>19%</td>
</tr>
<tr>
<td>Third</td>
<td>$603</td>
<td>57%</td>
<td>22%</td>
</tr>
<tr>
<td>Fourth</td>
<td>$943</td>
<td>50%</td>
<td>31%</td>
</tr>
<tr>
<td>Fifth</td>
<td>$2,511</td>
<td>31%</td>
<td>53%</td>
</tr>
</tbody>
</table>

Data source: Erick Guerra, calculated from 2007 Origin-destination survey data (INEGI)

Figure 12 depicts the accuracy and comprehensiveness of data accessible to users before the 2013 open data project (grey), after the project (black), and with future steps that could occur regarding the microbuses (red). A major shift on this graph could indicate potential for disruptive innovation, since higher data quality can increase the value of existing transportation services with respect to users.

Before SETRAVI released the GTFS feed, users of fixed routes or more structured transit services (e.g. Metrobús, Metro) already had access to information about these modes in the form of posted system maps, online map and route information, and several Android and Apple iOS apps that could direct users to the nearest transit station. Data accuracy and comprehensiveness were therefore fairly high. Users had little formal or official information about microbuses, however, and generally relied on personal experience, word-of-mouth, and other informal methods to determine how to get from one point to another. While the accuracy of personal experience may be quite high, the comprehensiveness of most users’ information was likely very low, covering a limited number of relevant routes.
Since SETRAVI released the GTFS feed, users can access the public transit information by downloading the feed online individually (unlikely), using an app or tool that uses the feed to give routing directions (depending on internet/smartphone access), or searching for transit directions on Google Transit (pending official release). This data access may serve as a convenience to individuals travelling to a new part of the city, and could be useful for tourists visiting the city. However, results of the recent data collection project may raise accuracy and increase data accessibility slightly, but would not produce a major shift for the user population. This slight shift will likely have very low impacts on transit users' mobility or accessibility due to the nature of transportation information that SETRAVI has released. Because the GTFS feed does not include microbuses, it simply releases already-available information in a more convenient format for developers to use. This does not add significant value to transportation services and is unlikely to disrupt current dynamics.

If SETRAVI were to include its microbus data in the publically-available GTFS feed, this would increase the comprehensiveness of transit information available to the public (see Figure 12); data release would enable mapping and app development that would give users access to route and stop information for all
microbus routes within the DF and enable travellers to use trip planning services that integrate all modes of public transit within the DF. This would enable citizens to reach current destinations more efficiently and potentially venture into new areas that were previously considered inaccessible. Thus, data release would increase macro-level accessibility for public transit users in the city. The result would qualify as Andersen’s second form of disruptive innovation in public transportation; it changes the value proposition of transit such that current services operate in the same manner, but offer users a greater level of accessibility and a higher-quality experience.

Despite this potential, several factors limit the usefulness of the current data held by SETRAVI. First, the data collection method used in the 2013 project creates information that may not be completely accurate and are not comprehensive. Data collection for the microbuses represents a snapshot from a single moment in time of a system of some largely unknown dynamism; labour constraints limited fieldwork efforts to one complete ride on each microbus ruta. With approximately 1,100 ramales mapped, there is a high likelihood that some of these routes represent deviations from typical paths. Second, even if all route information are fully accurate, a one-pass vehicle tracking effort prevents the GTFS data from being used to accurately predict travel times, estimate frequencies, or determine the current location of a vehicle. The feed will also become incorrect over time as routes change. Finally, even a completely accurate and comprehensive GTFS feed for the microbus system in the DF would not fulfill many travellers’ needs because it would not include transit data beyond the DF, in the greater MCMA.

Means of information access further restricts usability. Microbuses could be considered an inferior good; transit users tend to shift away from this mode as their incomes rise. Thus, microbus users, on average, have lower incomes and lower access to smartphones and web-based tools than users of other transit modes. A mobile data collection pilot test in Buenos Aires supports this finding. Researchers conducted 74 onboard passenger surveys along a BRT route and 34 along a conventional bus route. Results suggested lower smartphone penetration amongst users of less desirable (non-rapid) transit (Butts, et al., 2013). Thus, smartphone apps and high-tech means of dissemination may not reach the majority of users. Smartphone penetration is rising quickly in many parts of the world. However, for the time being, printed maps, schedules, and signs may be more useful to passengers than a software-based tool.

Based on this analysis, or in areas in which personal safety is a concern. In these ways, real-time location data could add additional value to the microbus data that SETRAVI has already collected, increasing its disruptive potential shows potential shifts in the quality of data accessible to transportation users. The 2013 open data
project results in no change for microbus users and only a small improvement for fixed-route users. The figure also illustrates the increases in data accuracy and comprehensiveness that could arise from releasing the microbus data to the public, or from a new intervention: developing a live-tracking system to release real-time microbus information.

To further increase comprehensiveness and accuracy and maximize potential impacts, the microbuses would need to be included in a publicly-available GTFS-RT feed.\textsuperscript{12} Real-time location information would enable SETRAVI to generate accurate, up-to-date frequency, trip time, and current position information for the microbuses, which would enable two key outputs for users: comprehensive routing and trip planning, and next-bus arrival time information. The impact of trip planning information depends on users’ access to internet-capable devices that could relay this information. The impact of next-bus arrival time information may be negligible during periods with short headways; traffic congestion causes travel time to vary greatly between days and times. If a user is about to spend over two hours in a bus to the city centre, saving a few minutes in waiting time likely does not offer significant benefits. However, arrival times could be useful during periods of infrequent service or in areas in which personal safety is a concern. In these ways, real-time location data could add additional value to the microbus data that SETRAVI has already collected, increasing its disruptive potential with respect to users.

\textit{Transportation regulators}

Transportation regulators include the sub-offices responsible for each of the DF’s transit services: STE, STC, RTP, Metrobús, \textit{suburbanos}, and microbuses. The sub-offices for fixed-route transportation (STE, STC, RTP, Metrobús, and \textit{suburbanos}) did not undertake new data collection efforts as part of the 2013 open data project, but simply converted existing data into a GTFS format for public release. The accuracy and comprehensiveness of data available to regulators did not change for these modes and does not have disruptive potential. This section therefore focuses on microbus data changes and potential impacts.

The DGT, another sub-office of SETRAVI, acts as the regulatory agency for microbuses. As discussed, this office conducts relatively little planning; its main efforts focus on (1) smoothing conflict between route associations or individual drivers, and (2) managing relationships between the microbuses and other modes of transit, particularly during BRT expansion efforts. As discussed, the DGT possesses scant information about the microbus services that transport nine million passengers around the DF over the course of a day. The 2013 open data

\textsuperscript{12} Metrobús and certain RTP buses already have this capability, but microbuses generally do not.
project therefore offers a significant improvement in data accuracy and comprehensiveness (depicted in Figure 13).

Figure 13: Accuracy and comprehensiveness of data accessible to transit regulators

Data availability could raise SETRAVI’s capacity in two areas: regulation and planning, both of which add value to the “business model of transportation planning” that Andersen describes as his third form of disruptive innovation.

Regarding regulation, the DGT would benefit from any information that could enhance its position with respect to the microbus route associations and potentially alter dynamics in the existing, atomized stasis. A GTFS feed for the microbuses has some potential to change existing dynamics. Data from the 2013 project significantly increases the accuracy and comprehensiveness of the DGT’s transportation information (see Figure 13). The current, static GTFS feed may assist the city to determine whether vehicles on the street are operating outside their ramal’s concessions, either because of short-term deviations or long-term shifts in routes – though this potential is limited by the need for a regulator to observe on-street operations and, presuming that the static GTFS feed is correct,
then using the vehicle’s route number (displayed on the bus) to validate whether
the driver is operating along his permitted route. For maximum effectiveness,
however, the DGT would require ready access to trip frequencies, vehicle counts,
and travel times. Gaining this information would require live-tracking devices
placed on board the buses to generate a GTFS-RT feed. While there is some
disruptive potential for the existing data, information would have a much greater
capacity to improve regulatory capacity if it included data aspects that could be
generated from real-time location sensors.

The GTFS data, as a whole, have yet to play a significant role in SETRAVI’s
internal data management and analysis, but SETRAVI has been exploring
potential avenues with respect to planning. First, the microbus data can be used to
visualize transport network coverage within the city, and to identify areas lacking
in service. Second, The OTPA tool offers significant promise for a forward-
thinking administration like SETRAVI. However, it faces limitations with respect
to the microbus services. Travel time is central to OTPA’s functionality; the tool
essentially examines the locations an individual can reach in a given amount of
time. The microbus trip times in the DF’s current GTFS feed may not be reliable,
however, because each ramal was measured only once (sometimes in peak-hour
traffic, other times in off-peak), frequencies are merely estimates, and the data will
grow outdated over time. Moreover, OTPA itself cannot perfectly model a
transportation network, since it does not consider vehicle capacity – a key
limitation in Mexico City’s crowded systems. It would, however, benefit from a
GTFS-RT feed to better measure and model the city’s microbus network.

In addition to macro-level analysis enabled by OTPA, the 2013 microbus data
could also facilitate longer-term planning and strategies for managing ramal
displacement during BRT expansion. The increased accuracy of the data for all
microbus routes could help the DGT to quickly determine the location of ramales
and number of microbus operators affected by a Metrobús expansion. It could
then determine appropriate compensation¹³ and devise a long-term relocation
strategy for affected operators. This would expedite negotiations, save costs, and
enable SETRAVI as a whole to implement its planning decisions more quickly –
thus adding substantial value to its organizational model of transportation service
provision.

Overall, the new information contained in the GTFS feed can serve as a tool for
SETRAVI to visualize and analyze its transportation network; this can facilitate
more comprehensive planning endeavours. The feed can also reduce time and cost
barriers to cooperation regarding Metrobús expansion. Both of these outcomes

¹³ Respondents gave conflicting answers about the type of compensation offered to displaced ramales; some
reported financial compensation while others stated that they were given new concessions only.
qualify as Andersen’s third form of disruptive innovation. Both are possible, to varying extents, with the static GTFS feed. However, GTFS-RT data would enable larger impacts and enable SETRAVI to “do more with less” – a hallmark of disruptive innovation.

Transportation operators

The “operator” category includes a variety of stakeholders with different interests and levels of agency: fixed-route operators, microbus owner-operators, microbus operators who do not own their vehicles, small-scale microbus owners with a more vested interest, larger-scale flotillera owners, and members of the route association leadership or cupula. Since the open data project does not affect data available to fixed-route operators, this section considers open data concerns and impacts that could affect various stakeholders in the microbus industry. It considers potential disruptions that would benefit these stakeholders, as well as those that would impact them negatively.

The accuracy and comprehensiveness of data available to operators (see Figure 14) generally has not changed as a result of the 2013 open data project since SETRAVI has not released the microbus data and because other sub-offices supplied the information contained in the GTFS feed.

Figure 14: Accuracy and comprehensiveness of data accessible to transit operators
Route association leadership typically holds detailed information about their *ramales*' routes, stops, frequencies, and headways. Therefore, a static GTFS may increase their information regarding other routes, but probably would not confer benefits to the association since the routes are generally fixed to a relative geography and could not be easily moved to a new area. In contrast, by increasing SETRAVI's information access, the route association may lose power with respect to Metrobús negotiations. Route information could also enable SETRAVI to more easily modify or overhaul the route structures, or to increase regulatory enforcement for existing concessions. In all of these scenarios, an organizational disruption occurs as a result of data availability, but it incurs some degree of loss for the route associations. This would be exacerbated with information from a GTFS-RT feed (see Figure 14). While data accuracy and availability would increase for bus operators, SETRAVI would gain even more information about frequencies and route characteristics, which would enhance its ability to negotiate fair compensation for displaced microbus routes and perhaps enable the agency to identify high-demand *ramales* that would benefit from a higher capacity transportation service like *Metrobús*.

On the other hand, real-time microbus information would confer certain benefits to microbus operators. Route association members and bus owners responded positively when asked how real-time information would impact their operations. Interview respondents indicated that they would be willing to share location data with the government, provided they were not required to disclose any financial information regarding earnings and operating costs. In general, the greater an individual's position of authority and the closer his connection with the government, the more he expressed willingness to generate and share information. High-level officials in formalized routes expressed the strongest support for open data; they explained that “everything must be open ... It’s all for a better service” and that information benefits everyone — the government, the operators, and transit riders. Respondents indicated that GTFS-RT information would enable route associations and owners to better track their vehicles and ensure driver compliance with respect to routes and stops. They also suggested that real-time information might prompt passengers to choose one route or service over another, increasing profits for data-oriented route associations. Bus owners were particularly enthusiastic about vehicle location information, stating that it would enable them to better respond to vehicle theft and identify bus drivers who use the vehicles for personal purposes. These answers indicate that real-time location data can improve the organizational model of transit service delivery and increase the value of transit services in such a manner that causes disruption in favour of different types of microbus operator.
Real-time microbus location data could also play a role in dispatch services. Interview respondents cited dispatch as one of a route association’s most significant responsibilities. They described the service as extremely necessary to keep the bus service functioning effectively. However, route associations conduct dispatch in a very informal manner, often hiring several young boys or teenagers to perform the service. These individuals typically stand on street corners with radios, timing and counting the buses, and informing bus drivers when to begin their route and whether they should linger at a stop to increase the headway. Some routes require dispatchers to record bus counts and times on paper; other routes do not collect data for future reference. Route association officials suggested that access to live data would enable agency staff to replace labour with technology and better coordinate the fleet movements. It would also create historic records that could inform future planning and could be used to track accidents and other metrics. These improvements could reduce traffic congestion through more optimized dispatch systems. Overall, this represents a positive disruption for the organizational model of transportation service delivery.

As the organizational and service provision models of transportation change, certain members of the microbus industry may incur losses from disruption. While route associations and bus owners stand to benefit from real-time information, drivers seem to lose more as information opacity decreases. Drivers receive earnings as a share of the fares they collect; their incentive is to maximize profits by transporting as many passengers as possible, potentially by competing for boardings or pocketing fares. Increased information could mean that route associations and owners can better regulate operators, likely cutting into their profits.

Bus drivers also stand to lose the most from Metrobús expansion and, potentially, from franchising agreements. When the BRT system adds or expands a line, it generally pushes drivers out to a new area that (by argument of revealed preference) is likely less convenient for them to access. While the route associations receive compensation for affected ramales, these payments may or may not trickle down to the operator level. In theory, displaced microbus drivers could seek employment with Metrobús, but the BRT’s high capacity vehicles require fewer drivers per route, implying that there would be an oversupply of candidates for new jobs. Furthermore, interview responses indicated that bus operators prefer to work in the microbus system since it offers opportunities for advancement; drivers can aim to earn enough to purchase a bus and move upwards through the system. They receive remuneration daily, have a greater sense of independence, and prefer driving a microbus for other cultural reasons (e.g. more flexible schedules, not required to wear a uniform). These factors also affect driver perceptions regarding the initial shift towards franchising microbus
routes; formalization, in general, can curtail independence. In the long term, however, drivers may benefit from health and insurance benefits, stable wages, and regular hours.

Bus drivers’ potential loss from formalization processes implies that they may prefer to resist or drag out negotiations and franchising processes. They have little capacity to do so, however, because negotiations take place at a higher level. Thus, drivers’ main tool for opposition would be strikes. These measures would require a high degree of organization and significant levels of motivation.

Driver strikes may be unlikely or extreme responses to formalization, but drivers can resist information availability through other means. In the past, microbus owners on different ramales installed passenger-counting devices on their vehicles in order to ensure that drivers were reporting all fares. Drivers immediately tampered with the devices and prevented them from functioning properly. Similar reactions could prevent the creation of a GTFS-RT feed, unless this were done through surreptitious crowdsourcing of passengers with smartphones. Crowdsourcing is currently too difficult, however, because smartphones are not yet ubiquitous in Mexico City, particularly amongst lower-income populations that rely on microbuses. Generating real-time information would thus require support and participation from bus operators, since location-sensing equipment would need to be placed in the vehicles and allowed to operate without interference. The feasibility of a GTFS-RT feed therefore depends upon driver perceptions.

When asked how they would feel about real-time data collection, leaders of non-formalized ramales expressed some suspicion about how the data would be used, as did individual operators. Lack of trust in the government (falta de confianza en el gobierno) presented a recurring theme in the interviews. Every operator from a non-franchised route raised this topic and expressed his or her frustrations and suspicions at length. Key themes centered upon corruption, opacity, and lack of accountability; one individual mentioned that a 20-year legacy of under-the-table payments made him reluctant to trust transportation officials. Even respondents from franchised corridors expressed some dissatisfaction with SETRAVI and eagerly shared anecdotes. Operators’ perceptions of these experiences contribute to less willingness to cooperate in future interactions.

Despite these concerns, operators declared that they may be willing to share location information with the government if they believed it would benefit their route. Respondents cited financial motives for these sentiments. SETRAVI recently increased microbus fares from 3 pesos to 4.5 pesos. As a result, passengers expect commensurately better service. Bus operators explained that better information
services could increase transit riders’ satisfaction with the bus service. Research reviewed in Chapter Two supports these assertions. Real-time data would be most powerful for these purposes; operators explained that accurate frequency and arrival times (accessible by text message services) would be especially useful on the outskirts of the DF during off-peak hours when frequencies are lower, and within the city center during peak hours when travel times are highest. Since non-formalized bus operators’ only revenue comes from passenger fares, expectations for higher ridership may motivate these operators to allow location tracking to take place on the vehicles they operate.

Overall, the current, static GTFS feed (whether or not it is released) may have little impact on microbus operations but could slightly increase SETRAVI’s power with respect to microbus associations. A real-time location service appears to offer a larger net gain to route associations and bus owners. Bus operators may oppose this development and could interfere with location-sensing technology because of their mistrust towards the government and opposition to route association regulation. However, drivers may allow location-sensing to take place if they felt that increased information would raise passenger trips and generate higher earnings.

**Paths forward**

This analysis suggests that transportation information could cause disruption and impact stakeholders differently depending on circumstances and relationships between actors. Next, I review three potential scenarios to explore how microbus information could play out in the context of the DF.

**Scenario 1: No additional major intervention**

With the current results from the 2013 data collection project (a public GTFS feed that does not include microbus data), SETRAVI gains access to route information and estimated frequencies for the microbuses and can reap some gains with respect to more efficient negotiations over Metrobús expansions. SETRAVI’s planning endeavours would also benefit from route location information. Route associations may lose some degree of power as information opacity decreases, but impacts would be mild because the static GTFS offers limited insight into microbus operations. Users experience insignificant direct impacts since the GTFS feed is not publically available to them, does not incorporate data for services outside the DF, does not include real-time information or accurate frequency information, and would need to be disseminated through a text-message-based service or physical maps in order to reach a broad audience. Indirectly, they would benefit from any service improvements induced by SETRAVI’s potentially
enhanced planning and regulatory powers. As discussed, there are some possibilities for disruptive innovation. However, if SETRAVI takes no additional major interventions, then the GTFS feed will have relatively minor impacts on transportation dynamics in Mexico City and would not noticeably disrupt transportation supply. Table 9 summarizes these impacts.

This scenario seems moderately likely for the city since it requires no explicit action on the part of SETRAVI or other stakeholders. It would not result in a noticeable, disruptive innovation with respect to transportation supply.

Table 9: Stakeholder impacts under Scenario 1

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Impact</th>
<th>Degree</th>
<th>Implementation Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulators</td>
<td>+</td>
<td>Minor</td>
<td>n/a</td>
</tr>
<tr>
<td>Operators</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Route associations</td>
<td>-</td>
<td>Minor</td>
<td>n/a</td>
</tr>
<tr>
<td>Bus owners</td>
<td>-</td>
<td>Minor</td>
<td>n/a</td>
</tr>
<tr>
<td>Bus operators</td>
<td>-</td>
<td>Negligible</td>
<td>n/a</td>
</tr>
<tr>
<td>Users</td>
<td>+</td>
<td>Minor</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Scenario 2: Government intervention for data collection

My analysis indicates that both SETRAVI and the route associations could gain significant benefits from a GTFS-RT feed. Creating a real-time dataset would require costly and challenging inputs. Since smartphone penetration is currently too low to make crowdsourcing a likely option, location-sensing devices would need to be installed on a sufficient number (if not all) of the city’s microbuses – either by the government or by route associations. Advances in digital technology have lowered the cost of the necessary devices. However, even if devices could be procured for as little as $50 each, the cost (and logistics) of equipping 30,000 microbuses would still be a substantial expense for the government or a route association to bear. Second, SETRAVI and the route associations would require the approval or, at minimum, the acquiescence of bus operators in order for location-sensing equipment to be left to function properly. In addition, these passive devices would not easily generate other information, especially stop locations, which, in the location data, could be easily confused with stop light locations or consistently congested points.

Under current dynamics, bus operators’ distrust towards the government may prompt skepticism regarding data collection efforts. Some operators may view location-sensing equipment as an unwelcome intrusion and tamper with the devices. Interviews indicate that drivers could be persuaded to support data
collection efforts if they believed that resulting information would increase public transit ridership and, ultimately, driver incomes. SETRAVI would have to approach this situation carefully, however; the agency should be very explicit about its purposes and processes in order to assuage operators’ sense of distrust, and it must clearly explain the benefits of transit data from an operator’s business standpoint. The successfulness of the data collection initiative would depend on the level of driver buy-in. In the long-term, however, SETRAVI may wish to decouple driver remuneration from passenger fares to prevent on-street fare competition and decrease congestions. This could remove drivers’ main motivation for permitting location sensors. The new data sources considered in this thesis cannot on their own decouple fares from remuneration, although new fare collection media in combination with real-time information could pose as a powerful reformatory combination.

This scenario results in a very substantial disruptive innovation with respect to the ways and degrees in which transportation is regulated, planned, and supervised. It also causes some disruptive innovation by increasing the value of existing transit services for users. Ultimately, however, the expense of data collection equipment and the riskiness of placement may make a top-down intervention too risky for SETRAVI or the route associations to undertake. This scenario is therefore somewhat unlikely. Table 10 summarizes these impacts.

Table 10: Stakeholder impacts under Scenario 2

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Impact</th>
<th>Degree</th>
<th>Implementation Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulators</td>
<td>+</td>
<td>Major</td>
<td>Cost, driver compliance</td>
</tr>
<tr>
<td>Operators</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Route associations</td>
<td>+</td>
<td>Major</td>
<td>Driver compliance</td>
</tr>
<tr>
<td>Bus owners</td>
<td>+/-</td>
<td>Moderate</td>
<td>Driver compliance</td>
</tr>
<tr>
<td>Bus operators</td>
<td>+/-</td>
<td>Moderate</td>
<td>n/a</td>
</tr>
<tr>
<td>Users</td>
<td>+</td>
<td>Moderate</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Scenario 3: Data collection as part of incremental formalization

SETRAVI’s recent move towards incremental microbus formalization may present an ideal situation for launching a real-time data collection effort. When microbus routes become franchises, they typically purchase new, larger vehicles. These vehicles could be equipped with location-sensing technology (either standard GPS equipment or cheaper, lower-tech sensors) for a relatively small cost, compared to the vehicles’ total purchase value, and would preclude the logistical hassle of retrofitting existing vehicles.
The labour dynamics of a franchise system would be much more conducive to data collection. Because drivers earn a regular, fixed wage, they face lower incentives to underreport fares, deviate from fixed routes, and interfere with regulatory processes. These drivers may be less independent in general and much less likely to tamper with onboard equipment. On the other hand, fixed wages imply that drivers would be indifferent to whether or not the public found location data to be useful; increased revenues from passenger fares would make little difference to drivers and could not be used to incentivize compliance.

Under this scenario, a GTFS-RT feed would increase SETRAVI’s institutional capacity to plan and regulate transit, as well as route associations’ ability to conduct monitoring and dispatch for their routes. Both of these are a planning and service delivery form of disruptive innovation. Bus drivers, having already undergone the franchising process, would neither gain nor lose from data collection efforts. Moreover, users may benefit from additional bus information. Table 11 summarizes these impacts.

Table 11: Stakeholder impacts under Scenario 3

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Impact</th>
<th>Degree</th>
<th>Implementation Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulators</td>
<td>+</td>
<td>Major</td>
<td>Time required to franchise routes throughout the city</td>
</tr>
<tr>
<td>Operators</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Route associations</td>
<td>+</td>
<td>Major</td>
<td>Additional vehicle costs</td>
</tr>
<tr>
<td>Bus operators</td>
<td>None</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Users</td>
<td>+</td>
<td>Variable</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Overall, this scenario seems feasible for SETRAVI to undertake (particularly since the static GTFS data could speed the franchising process) and would confer benefits to regulators, users, and route associations without stirring significant resistance from microbus drivers. Thus, Scenario 3 suggests that state capacity to regulate is necessary before the information revolution may spur disruptive innovation; by exercising its regulatory capacity and strength, a transportation agency can allow a particular technology to become effective.

**Limitations**

This analysis is subject to limitations due to the relatively short time spent in fieldwork and in direct contact with agency staff and microbus operators; a lengthier period in Mexico could have added additional insight into the project. Moreover, my analysis of disruptive innovation faces limitations because it is extremely difficult to predict a disruptive innovation or measure the impacts of
data availability, particularly in a city like Mexico. The type of data collection initiative described in this study represents one of only a handful of initiatives in the world, all of which occurred within the last two years. Therefore, there has not been sufficient time to observe data release efforts elsewhere in order to identify factors that could be important in the Mexico case, or to learn from other researchers' methods. In order to follow up on this research and investigate actual outcomes, years would have to pass in order to allow the data time to make an impact on various actors. Even then, a method to measure and, ideally, quantify outcomes would be very difficult to develop in such a way that the study engages a representative sample of individuals from various stakeholder categories.
Like many megacities around the world, Mexico City currently suffers from congestion, air pollution, oversupplied vehicles, dangerous road conditions, and inefficient public transportation service. Some form of transportation innovation, particularly with respect to the microbus system, may improve this dynamic. However, recurring cycles of privatization and regularization have led to a dynamic in which SETRAVI lacks the power and institutional capacity to regulate and restructure its microbus system. This creates an atomized stasis equivalent to the “deadlock” often described in disruptive innovation literature.

In this context, this research investigates whether, and how, newfound transportation data availability has the power to become a disruptive innovation with respect to public transportation in the DF. I find that two categories of disruptive innovation can occur: that which changes the value proposition of transportation and increases the utility of existing services, and that which alters the organizational model of transportation planning and service delivery. Within each of these categories, I analyze the change in data accuracy and comprehensiveness for various modes, then examine how these changes might create specific disruptions to transportation regulators, operators, and users. I then apply these possible disruptions to different paths of government and stakeholder action.

Many of these scenarios consider and depend on whether SETRAVI chooses to release its microbus data to the public, and whether the government (or route associations) would consider implementing a live-tracking system. The current, static GTFS feed may have some effects on government capacity and operations. However, transportation data would be most useful if they included real-time information; this would enable frequency calculations and live vehicle tracking that would enhance the disruptive potential of several outcomes.

While various sets of outcomes are possible, the magnitude of their impacts and the likelihood of each outcome occurring would vary. SETRAVI may continue to choose not to release their microbus data to the public. This scenario is likely to occur but unlikely to produce a disruption with respect to transportation supply; its effects would be limited amongst all stakeholders.

Another possibility is that a top-down or bottom-up data collection initiative could enable real-time data collection to take place on the microbuses. In this case, the value of existing transportation services increases; passengers become increasingly able to access different parts of the city in the most efficient way, and live location information could improve public safety and increase user
satisfaction. In addition, disruptive innovation could have significant impacts on transportation regulation, planning, and operations management. More accurate and comprehensive public transportation information could increase SETRAVI's regulatory capacity with respect to the microbuses, facilitate faster expansion of the BRT system, and enable the government to undertake more informed transportation and land use planning. A real-time data collection intervention is unlikely to occur as a top-down approach, and a crowdsourced approach would require much higher smartphone penetration rates than likely among current system users, as well as the motivation to participate. Therefore, the disruptive potential of a successful intervention would be very high, but the likelihood would be quite low.

Despite the unlikelihood of a standalone initiative, disruptive innovation may still occur. If SETRAVI continues to formalize microbus routes into franchises and launches a data collection initiative as part of the franchise process, then the agency could feasibly begin collecting location data. In this way, institutional reforms undertaken by SETRAVI could activate information in a manner than enables disruptive innovation, spurring improved transportation service provision within the DF. Even so, institutional barriers prevent data-driven innovation from causing disruption beyond the jurisdictional boundary that surrounds the DF. The innovations described in this thesis do not address the need for metropolitan-wide service integration. Until integration (or at least institutional coordination) begins, even real-time data can have only limited impacts on public transportation in the Mexico City Metropolitan Area.

Findings from this case study have relevance for a number of cities around the world. Mexico City experiences its own particular challenges, but its transportation dynamics bear many similarities to developing cities across the world: tensions between more and less formal services, cycles of privatization and regularization, government difficulty in regulatory and planning capacities, and a general lack of data to describe transportation systems. Thus, while the DF case study has location-specific findings, it also yields generalizable results – that transportation data availability can change relationships within and between groups of stakeholders, potentially eliminating “deadlock” situations and improving the government’s capacity to provide transportation services for its citizens and, more generally, improve their quality of life. As city governments contemplate the costs of data collection efforts, this case study suggests that open data projects can be beneficial to regulators, operators, and users alike. It also underscores the importance of considering how data may be operationalized in a specific context, given stakeholder relations and other constraints.
Much of the analysis and conclusions in this thesis rely on speculation and predictions because the newness of the data and the current unavailability of real-time information limited efforts to measure or quantify data impacts. Therefore, this research would benefit from future studies to examine actions amongst stakeholder groups and measure the outcomes of data availability over time.

Future research could take a multi-city case study approach to measure data impacts in different settings. The Mexico City project represents an early example of a data collection effort that will become more and more common due to advances in ICT the open data movement. Therefore, research studying data impacts could select case studies, bearing in mind the differences in cities’ political structures, regulatory norms, geography, transit modes, and institutional/colonial legacies. In order to measure disruptive innovation, these case studies would examine dynamics amongst stakeholders and attempt to identify “deadlock” situations in which information improvements with respect to accessibility, accuracy, and comprehensiveness may be able to effect change. Performance measures would be developed. For example, in Mexico City, the inefficiencies of BRT displacement negotiations represent a potential “deadlock” situation. Therefore, changes in the average length of displacement negotiations could serve as an indicator of disruption with respect to BRT expansion and planning capacity. More qualitative approaches (e.g. semi-structured interviews with different stakeholders) could also be included to discuss potential data impacts. These methods could help future studies attempt to investigate and measure the benefits and impacts made possible by newly-available information.


