Towards a Standard for Paratransit Data?: Lessons from Developing GTFS Data for Nairobi's Matatu System

Submission Date: August 1, 2013

Word Count: 6726 (3 figures)

Authors:
Klopp Jacqueline
Center for Sustainable Development, Columbia University
475 Riverside Dr. Suite 520
New York, NY 10115
(tel) 212-851-2979 (fax) 212-851-1780
jk2002@columbia.edu

Jackson Mutua, Daniel Orwa, Peter Waiganjo
School of Computing and Informatics, University of Nairobi
P. O. Box 30197 - 00100 GPO Nairobi
tel +254-4447870
email: mutuajack@gmail.com (Jackson Mutua)
tel (254) 0733788598 (Peter Waiganjo)
email waiganjo@uonbi.ac.ke (Peter Waiganjo)
tel (254) 0722831793 (Dan Orwa)
email: dorwa@uonbi.ac.ke (Dan Orwa)

Adam White
157 Pearl Street
Cambridge, MA 02139
Tel (818) 414-5903
adam.e.white@gmail.com

Sarah Williams[1]
MIT, Department of Urban Studies and Planning
77 Massachusetts Avenue
Cambridge, MA 02139-4307
tel (617) 258-0461
sew@mit.edu

---

[1] All authors contributed to this work in important ways. Thus, the ordering of the authors is strictly alphabetical. Jacqueline Klopp will serve as corresponding author.

Abstract

Despite the significance of paratransit for public transport in many cities across the globe, little data exists on these flexible, often informal systems, let alone any standard for such data. With the spread of mobile phones, ICT and open data efforts, new possibilities are emerging to collect paratransit data at a low cost and to put it into GTFS to allow for app and tool development. This paper describes and analyzes an experiment in collecting basic paratransit data in Nairobi, converting it into GTFS format, opening it up for anyone to use, and exploring its utility. We present the methodology, results and major challenges involved in fitting paratransit data into GTFS. Overall, we found that modifications do need to be made to GTFS to account for key differences between paratransit and more formal, planned systems. Further, we discovered that space exists for improving or editing applications for transit data collection and adapting them to the challenges of collecting data on paratransit. Finally, we see it is critical to start more structured collaborations across the current experiments in paratransit data collection occurring in various cities across the globe. By comparing our approaches and findings, we will be better placed to know whether a move towards a new data standard for paratransit is needed and if so, what it might look like and who it might serve in the cities that so far are excluded from the benefits of the open transit data movement.

Introduction

“Paratransit” and “standards” are two words that do not easily go together. Flexible, demand responsive and seemingly “chaotic” paratransit is often the backbone of mass transit for the majority of citizens in the rapidly growing cities of Africa, Asia and Latin America. Despite the significance of paratransit for public transport in many cities, little data exists on these flexible, often informal systems, let alone any standard for such data. In cities with formal, planned transport systems, operators publish their schedules, routes, trips and fares. When this data is made available to the the public in a standard form such as GTFS (General Transit Feed Specification), this allows developers to create applications that provide better information for users and planners. However, in cities where paratransit systems are dominant, the lack of basic open and standardized transport data means many potential benefits are out of reach. Collecting key data for paratransit and developing a standard or modifying existing standards such as GTFS could thus have potentially far reaching impacts on improving these systems and empowering users, operators and planners with new forms of information and tools.

A number of reasons explain why good data for paratransit does not exist or is inaccessible. First, many operators in paratransit are small businesses that may not see the utility of data or do not have the time and money to collect it. Sometimes, the fact that some of what these businesses do is informal means that they also may also wish to stay under the government or public radar. Another reason for lack of data is that governments, used to seeing these system as “chaotic” or too complex to address, often do not bother to require operators-including the larger businesses-to collect data. Worse, some government and industry actors collude and mutually benefit from the lack of transparency. Also, when government agencies collect data, they most often hire consultants and do not always make the data collection methodology and data that is collected by these consultants open. Finally, the informal and flexible nature of these systems that make them highly variable is an intrinsic challenge to paratransit data collection. In sum, at the core of the data problem for paratransit is both the top down and non-transparent nature of current data gathering efforts and the lack of a useful and adaptable framework for collecting, sharing, and utilizing paratransit data across organizations, operators, citizens, and cities. In the end, basic data on paratransit logistics, infrastructure, and performance, needed to better regulate, manage and improve the transit system in many cities across the world, either does not exist or is inaccessible.

With the spread of mobile phones, ICT and open data efforts, new possibilities are emerging to collect data at a low cost, and a growing interest exists in applying new technologies to transportation. Key questions then become: what kind of data structure makes most sense for paratransit? What is the best way to collect it? Who will use it and for what purposes? Should basic route, stop, trips and fare data be adapted to existing standards such as GTFS or does a new standard need to be developed that better captures the unique aspects of paratransit systems?

Currently, this is an emerging area of discussion within separate experiments occurring across diverse cities including Nairobi, Manila, Dhaka and Mexico City. More specifically, the World Bank, with support from AusAid, is working with the Philippines Department of Transport and Communications and other transport-related agencies in Manila to set up a GTFS database that contains basic service information for the myriad of public transport modes in the city. The World Bank is also starting a project in Mexico City with the Department of Transport (SETRAVI or Secretaría de Transportes y Vialidad del Distrito Federal) and is also starting work in China. USAID has a project that involves collecting paratransit data in Honduras, and an MIT based team (Urban Launchpad) has collected data for the bus system in Dhaka (Ching et al 2012).

The aim of this paper is to describe and analyze one of the first experiments in Africa to collect GTFS paratransit data. We use this experiment to explore lessons for the wider discussion on paratransit standards emerging from these initial experiments. Our work focuses on Nairobi's "matatu" transit system. Similar to other paratransit systems in the world, Nairobi's "matatu" transit consists of numerous operators running large buses to small fourteen and even seven seaters. These vehicles are called by the popular name of "matatus". Matatus run on over 100 routes in a city of 3 million and more, serving the wider metropolitan area. This system provides the main motorised public transport for the majority of people in the Nairobi metropolitan area (Salon and Aligula 2012).

A team at the School of Computing and Informatics at the University of Nairobi conducted the data collection and conversion process to GTFS in collaboration with the Center for Sustainable Urban Development at Columbia University, the Civic Data Design Lab at MIT and Groupshot, a small Boston-based design firm that specializes in informality. This project involved collecting basic route data from scratch using handheld devices (GPS units and a mobile phone app) and exploring how to put it into GTFS format. In collaboration with the Kenya Institute for Public Policy Analysis (KIPPRA), this data and the data collection process was also shared in two workshops with potential and current users and posted on the GTFS exchange. Finally, the Civic Data Design Lab has been working on visualizations of the matatu transit system that will also be released to the public and city planners.

Overall, we argue that this work in Nairobi demonstrates that it is quite possible using simple hand held devices and riding the system to develop useful basic paratransit data. It is also possible with some basic modifications to put this data into GTFS format. We also have some preliminary evidence that, as in other parts of the world, such data once made open is useful to app developers and also those working in transportation planning. However, we also discovered a number of basic modifications that need to be made to GTFS to account for key differences between paratransit and more formal systems. In the end, it is clear from this work that determining a way to collect and maintain paratransit data, standardize it and make it it open access could help cities all over the globe struggling to build and operate the kind of public transit that are fundamental to making cities work.

2. Theoretical Overview

Data is critical for transport modelling and traffic simulation which in turn are critical for developing and organizing better transport systems. Increasingly, transit authorities in the United States and Europe are also releasing transit data to facilitate the development of applications to help transit users make more informed choices, which in turn influences the system as a whole (Roth 2010). Overall, given how critical transport data is for ensuring mobility and supporting well functioning transportation systems, it is hard to believe that the availability and standardization of this data is not more advanced (Bareclo et al 2010: 1). This is particularly true for paratransit systems.

A promising development in the effort to collect and create critical transport data is the growing use of mobile phones and the shrinking cost and increased accuracy of GPS. A number of recent experiments have shown how using cellphone signals can assist in mapping traffic flows (Caceres et al 2007, Herrara et al 2010, Wang 2010, Ratti et al 2006). This includes the recent analysis of data from 2.5 billion call records of 5 million mobile phone users in the Ivory Coast to look at traffic movement in and around Abidjan (Talbot 2013, Wakefield 2013). In practice, however, in many cases telecommunications companies are not willing to release cell phone data or are not willing to do so for free. Fortunately, other ways to collect data exist.

Besides the big data generated by mobile phone use, basic route, trip, fare, stop and schedule information are often produced by transit operators directly. When made open access this data can be profoundly useful for developing applications for trip and other forms of planning. A movement to create such open transit data feeds was spearheaded in 2005 by Bibiana McHugh, IT manager at Portland’s TriMet transit agency. She approached Mapquest, Yahoo! and Google, to partner on uploading TriMet transit data onto maps to provide trip planning information. Google became the partner, launching the free online Google Transit Trip Planner in December 2005. This involved formatting the data using a basic specification that resembled the form that transit operators usually use to keep basic information (Roth 2010). Thus the Google Transit Feed Specification was born, which has since been renamed the General Transit Feed Specification. This simple standardized data format consists of a series of text files collected in a ZIP file. Each file models a particular aspect of transit information, much of which is relational: stops, routes, trips, and other schedule data. Once the data is in this basic format, this allows for easy uploading onto google maps and use in diverse trip planning applications such as the open source application TimeTable Publisher developed by Tri-Met.

Since 2005, many cities and transit authorities have opened transit data in GTFS, making this a widely used standard particularly in the United States. The GTFS exchange, a web based platform for sharing GTFS transit data, started and maintained by Jehiah Czebotar, is one way that data is made accessible. Currently, most major transit agencies in the United States make their GTFS data public, and interest is growing in going beyond using this data for trip planning to creating tools to improve transit operations and planning (NCTR 2011, Lee et al. 2012). The most profound aspect of these developments is the open data architecture which allows for application development and use across the globe, provided good GTFS data exists for a location. It would be deeply problematic and in fact, unjust if the large number of cities in the world that rely on paratransit are left out of these new opportunities to develop data and tools.

3. Background to the Nairobi (Para) Transit System

Paratransit forms the core of public transit in many cities in Africa including Nairobi (UITP 2010). In almost all cases paratransit has arisen because of poor funding and management of municipal public transport systems, rapid urbanization and hence escalating demand for transport services and a poor regulatory and institutional environment which provides opportunities for organized interests to make untaxed and unofficial money from the system as it stands. These organized interests often form an oppositional block to change. The large numbers of paratransit operators, which form the backbone of urban transport, also have substantial political clout through their organisations and the ability to go on strike and withdraw their vehicles from the streets, bringing the cities to a standstill.

Many cities also have persisting colonial planning traditions which tend to ignore more consultative processes and the concerns of the majority of citizens. For Nairobi, this is clearly the case (Klopp 2012). These planning traditions focus on services for the car owning minority (often in the form of roads without facilities for public transit) and limited or no interest in upgrading services for the majorities including paratransit, especially as this would impinge on entrenched interests. At the same time, paratransit is also seen as form of African empowerment as the vehicles are locally owned and involve large numbers of small businesses from the operators to the mechanics tasked with repairing the vehicles (Mutongi 2006). Thus, informal forms of transportation, which emerged during the colonial period to address the unmet public transport needs of the majority, have expanded and continue to bear the brunt of increasing demand, predatory practices on the part of some government officers and also cartels who sometimes work in collusion and the neglect of planners and policymakers.

Currently, approximately 9,554 matatus and buses serve the Nairobi region on approximately 138 routes (Transport Licensing Board 2012). Over the last decade, responding in part to public concern, the government has made sporadic attempts to improve regulation, usually through punitive measures that inadvertently allow for more extraction in the form of bribes. Few attempts have been made to work with existing paratransit organisations to actually upgrade and plan for the existing system and integrate it into new transport projects. Matatu operators also do not collect data on their operations. Those government agencies tasked with planning for routes and stops-the transport and planning units of the former Nairobi city council and the Transport Licensing Board appear to have not fully executed their responsibilities even as they take fees for using some terminals (the city) and getting licences (Transport Licensing Board). In addition, little or no information is accessible from them.

New legislation puts the functions of the Transport Licensing Board under a National Transport and Safety Board. The new devolved Nairobi City County is also engaging in a new planning process. The Kenya Institute for Public Policy and Research Analysis (KIPPRA) is also exploring with us the concept of an open access transport data portal to improve modelling and planning. Thus, currently the institutional context is changing with new opportunities to engage in more bottom up data collection and approaches to planning and advocacy for improvements in service.

Finally, Nairobi, has a vibrant tech community that has already been trying to create maps of the matatu system (with limited success since it requires substantial labor and some resources) and applications for mobile phones, which are now ubiquitous in the city. Young developers are building new apps and showcasing them in the numerous tech hubs and incubators such as the iHub and mLab. Well structured, high quality basic route, trips and stops data can help in these efforts to develop ways to get better information to travellers.

One such example is the android app and web platform ma3route[2] that gives people directions and crowdsourced information on traffic conditions and rates the quality of matatu driving.

4. The Data Collection Experiment

Left is a picture of a University of Nairobi student collecting data on a Matatu route. Right is the raw data of Matatu routes generated for Nairobi, Kenya. At the time of this article in 2013 we were 70% complete. Photo taken by Adam White. Image of Route Map generated by the Civic Data Design Lab, MIT

The first step in this work involved collecting existing data on routes; very little data existed at the former Nairobi City Council (now Nairobi City County) except a dated document listing of some of the "official routes". After querying all of the relevant agencies including the Transport Licensing Board, it became clear that the existing government and agency data was incomplete, outdated, and often inaccurate especially given the large number of road projects and new real estate developments in and around the city. These were prompting the matatu industry to alter existing routes or devise new ones.[3] We also reviewed existing data collection projects done by entrepreneurs for business or social reasons. Many of these projects are now defunct because the cost of initial data collection (and continuous updates) and the clear uses of their tool did not align. While many of these projects successfully mapped reasonable percentages of the city, many had low accuracy, major errors, or inconsistent methodologies and data structures making the data impossible to combine. In addition, only a few were willing to share their data. Once we realized that the existing data was not consistent, reliable, or comprehensive, it became clear that the bus and matatu route data for the city would need to be collected from scratch.

[2] See http://www.ma3route.com/. See also prototypes by University of Nairobi students Sam Karui http://nairobi-routes.appspot.com/ and Kichitaro Shiojiri http://www.youtube.com/watch?v=4xfEs-HJRDY

[3] The Transport Licensing Board had commissioned a consultant to conduct a study but the methodology was not very rigorous (see Transport Licensing Board 2012).

As the team began data collection, we wanted to test different tools for transit data collection. After working with several different tools and processes, we devised a protocol and clear methodology for creating a route map and a GTFS compatible data structure. In all cases student team members from the University of Nairobi would ride a Matatu route, either in the car or matatu, use the data collection tool to generate latitude and longitude points along the route, and record all of the stops as well as specific meta-data about each stop, such as the stop name which is important for the GTFS protocol. While many paratransit systems involve stopping at varied locations depending on customer demand, regular and central stops and large terminals also exist. Students identified a stop based on their knowledge, information from frequent users of many routes, visual notation (signs, shelters etc), and if necessary, confirmation from discussion with an operator or group of commuters on a route.

The team knew early on that they wanted to experiment with using cell phones for data collection, but would also need to compare these with more traditional mobile mapping devices such as GPS units. Initially, due to availability, the team began using handheld GPS units and hand taken notes to document the first few routes. The team also began testing Android smartphone based data collection tools, focussing in on using *MyTracks* a basic GPS tracking system for cell phones developed by Google. Much later in the collection cycle we tested *TransitWand* which was released near the completion of the project.[4] GPS units were used as back-up so that we could compare the accuracy of the two forms of data collection.

We conducted experiments to compare the viability of each tool and found that the GPS units and mobile application on Android phones had similar accuracy but that mobile phones sometime took longer to lock in on a satellite and could lose a GPS signal more regularly. *Mytracks* app allowed for easier directly digitized entry of meta-data (for example, the name of a stop and current passenger counts could easily be recorded), while with GPS units a paper list was kept to cross reference waypoint numbers with meta-data that was then digitized and joined to the GPS data later on. The biggest challenges in using the mobile based GPS applications was the extremely limited battery life, slow speeds of affordable android phones, the risk of losing a more high value android phones to theft in a matatu where security is a constant problem (which did happen unfortunately), and the small screen size making digital data entry more time consuming particularly with frequent stops. Ultimately we believe mobile phones could be used more

[4] Some of the other tools we developed were; Paperless Tracker, an MIT student development, Fulcrum, GPS Surveyor, Open Data Kit, making our own tool in App Inventor.

for data collection, but a new application needs to be developed to make it easier to process the information while in the field and after.

As we were engaged in this data collection process, we discovered that an open source web and mobile app *TransitWand* was created by the consulting company Conveyal. *TransitWand* was developed for a similar GTFS data collection project in Mexico City, involving the World Bank and the Department of Transport for Mexico City.[5] We tested *TransitWand* in Nairobi and found that this tool, which is purpose built for transit data collection, resolved some of the issues of using an android phone. It was also more accurate than the GPS units as it snapped the data to the roads. However, because the tool was still in Beta development, the export to GTFS features were not fully working.[6] This made post-production work of the files collected with *TransitWand* more time consuming and cumbersome than the other applications we had been using. While *TransitWand* might be useful in the future, the team decided to complete the data set using *MyTracks*. GPS units remained as a fallback option. We believe future improvements in Mobile applications, such as *MyTr*acks and *TransitWand,* might make them more suitable for field data collection as they make including meta-data such as stop names and timing a lot easier.

Another issue in the methodology involved whether data would be collected from inside matatus or in a private car following the vehicles. Inside the vehicles is more affordable and scalable. Also being in the vehicle allowed the data collector the opportunity to engage with a matatu conductor and passengers in real time. Unfortunately, because the vehicles travel quickly and frequently alter course, the data collector can have problems of perspective (For example, it is hard to see whether most vehicles are travelling on the route or just the one that you happen to be in). This method also had the added problem of exposing the data collectors to risks of theft, a serious problem in Nairobi's matatu transit system. Using a private car allowed the data collector to observe multiple vehicles and to identify stops based on patterns of vehicles more easily than from within one vehicle. The control of the vehicle also allowed extra time to take notes, capture more meta data and make sure that a stop or notation is not missed. While The data was largely collected on matatus, but on particularly dangerous routes data was collected in private vehicles.

Once the data was collected it need to be cleaned and formatted in GTFS. This means

---

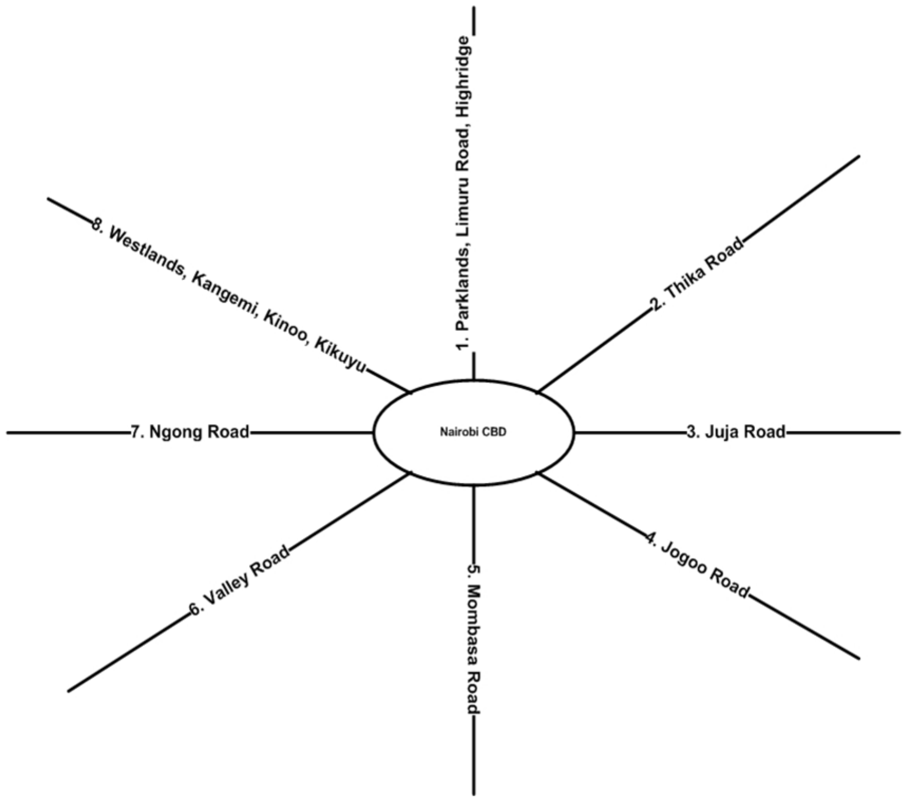
[5] Conveyal is a consultancy specializing in open data and open source technology for the transport sector. See http://www.conveyal.com/. Interestingly Conveyal was born out of the OpenPlans transportation technology group and involves the core developers behind the open source app OpenTripPlanner (OTP) and the creators of OTP Analyst, a research platform for generation of transport accessibility metrics.

[6] We are grateful to Kevin Webb of Conveyal for talking to our team about *TransitWand*.

creating the series of csv files that detail the transit system. In the formatting phase many interesting questions arose when deciding how to collect data to fit into the specific categories required for GTFS. For example, clear calendars, service frequencies and schedules, required for GTFS, do not exist for the matatu system. In addition, fares are often unstable, actual routes can change depending on traffic patterns, police checks, commuter demands or the prevailing weather. (For instance, when it rains in Nairobi, fares can triple and routes are adjusted to avoid traffic jams).

While the format may not be perfect for paratransit, trying to fit the data into GTFS was a useful framework for ensuring we had the correct data elements needed for developing a routing system and providing basic data to users online. As an example of a challenge, the team noted figuring out what a “designated stop” might mean was not always straightforward and in the end, three different features were captured in data collection that might connote “designation”. This includes 1) physical infrastructure (pullout from the road, bus shed or bus stop, a sign that the stop is “Matatu and bus crew organized”) 2) evidence of approval from Nairobi City Council (now Nairobi City County) or 3) evidence approval by being noted in official road maps.

The GTFS format does allow one to extend data to include additional information that is not part of its core. We also developed a protocol to embed limited meta-data in the unique ID of routes and stops. This was to help make the data easy to maintain. We also added notation for designated and undesignated stops. The team mapped the bus/matatu routes in Nairobi into eight major corridors in a kind of star structure (see below) and used this to codify clear routes and bus stops. We used this structure to develop a coding system based on branching and rotation to give each stop and route a unique and logical identifier. This coding allowed us to have a unique identifier for each route and stop. This was important as previously no way existed to identify the locations of a stop with similiar stop and route names in different parts of the system. This type of coding in more formal systems happens for data collected and is maintained by formal transit agencies for a number of service reason beyond routing. We developed a new coding system for that Matatu system itself that could be useful beyond the GTFS data collection.



Star structure of Nairobi's radial matatu transit system used to code routes and stops

So far, we collected data for 102 matatu routes and the 6 rail routes and posted them on the GTFS data exchange.

While we were going through the process of data collection, we also partnered with the Kenya Institute for Public Policy Analysis (KIPPRA), Kenya's primary government think tank tasked with transport modelling. We held two workshops where we invited technologists, policy analysts and transport operators to discuss the project and provide feedback. The idea was to get early feedback from potential users before a release. In the process the University of Nairobi team started working with a promising young app developer, Laban Okune, who used the data to improve his ma3route app. The app has now been on the market for ten months and recently won a Pivot East award. We also linked up with the UN-Habitat/Institute for Transportation and Development Policy (ITDP) who found the data useful as they begin a Bus Rapid Transit Service Plan for Nairobi. Overall, the conclusion of the workshops was that good, open access transit data is badly needed in Nairobi. We also started to discuss with KIPPRA a possible plan for hosting an open transport data portal that would include continually updated GTFS data among other needed data.

4. Findings

Overall, we found that by using handheld devices and devising a number of adaptations, it is possible to develop a GTFS data set for paratransit that has some real utility to developers, planners and policy analysts. Changes were clearly required in attributes that were “flexible” such as schedules, fares and stops. Below, we summarize by specification how we addressed this variability and the differences between paratransit and formal planned transit in the data required for GTFS files.

Agencies

Usually transit operators collect and publish GTFS data. Currently, there are no matatu and bus transit agencies collecting the this data. For the purposes of our project, the School of Computing and Informatics at the University of Nairobi acted as the agency producing the data. With the matatu system fragmented and complex, it in fact makes sense to have a neutral and technically capable institution collect the data and ensure quality and uniformity in quality and formatting. Ideally, this function is taken over by a government agency like KIPPRA that has a steady budget allocation for updating the data and a strong mandate to make it openly available.

Stops

In the matatu system there are both fixed terminals and stops and more “demand responsive” stops that are made to adjust to passenger needs, sometimes stopping at a particular location that has no resemblance to a stop. For the matatu system since many stops do have a relative fixity we thought it important to capture these. Since the city government has not been actively planning and designating official stops, a large number of undesignated stops exist. We thus collected both the designated and undesignated matatu stops and coded them in the stop ID data file. We also captured critical meta-data on designated or undesignated and if designated, why we considered it designated. This could help if the city wanted to formalize many of the high use undesignated stops.

Routes

Routes can also be flexible to some degree but overall they tended to have relative stability. It would be possible and interesting to study over time just how stable different routes actually are. We noted, for example, that when schools are in session, some routes adjust to accommodate students. Another reason for route change is traffic police spot checks. Matatu drivers try to avoid these spots, because usually they have to pay the traffic police a bribe and waste time in the process. As in more formal systems, routes often change when road construction is occurring and in one matatu we found a polite notice explaining the route changes. Thus, some of the flexibility is in fact desirable. We were able to determine the common relatively stable routes to put into GTFS files.

One required field in the route file is the the **route_type** which describes the type of transportation used on a route. (Currently, these include 0 Tram, Streetcar, Light rail. Any light rail or street level system within a metropolitan area **1** - Subway, Metro. Any underground rail system within a metropolitan area. **2** - Rail. Used for intercity or long-distance travel.**3** - Bus. Used for short- and long-distance bus routes.**4** - Ferry. Used for short- and long-distance boat service.**5** - Cable car. Used for street-level cable cars where the cable runs beneath the car **6** - Gondola, Suspended cable car. Typically used for aerial cable cars where the car is suspended from the cable **7** - Funicular. Any rail system designed for steep inclines). It might make sense, especially in hybrid transit systems that include both more formal, planned forms of transit and paratransit to modify this field to be able to distinguish between formal bus service and paratransit service. This could add flexibility in the data structure. It would allow planners to see changes as systems become more planned, sometimes unevenly based on bottom up efforts of operators or for example, the introduction of Bus Rapid Transit on some corridors.

Schedules

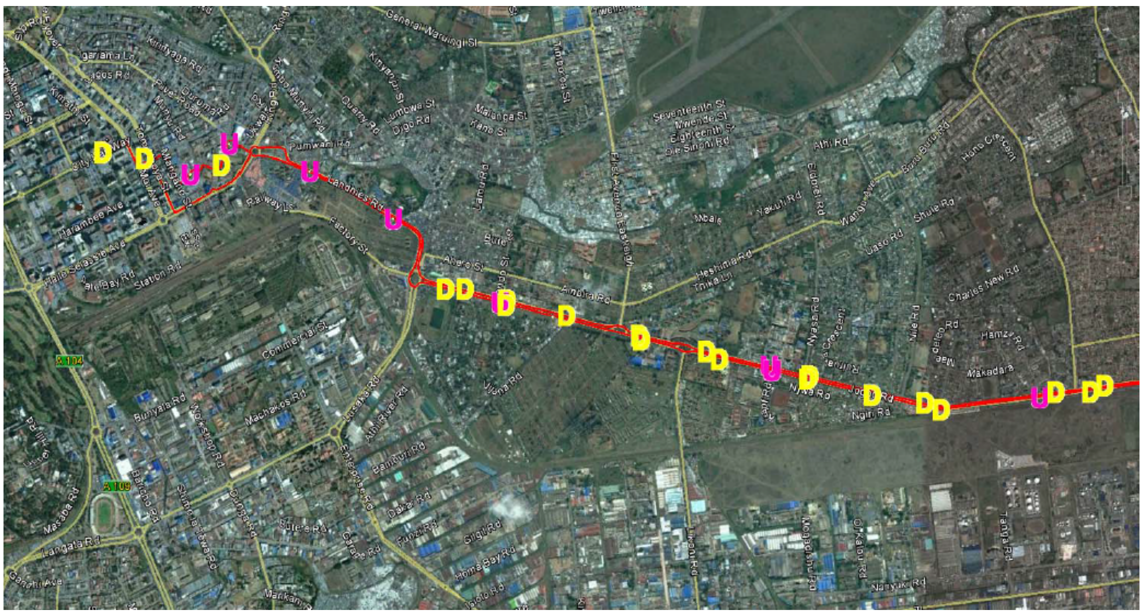
Ask a matatu crew, and it will say that no apparent need exists for scheduling since a vehicle only departs from a terminus when full. However, commuters would very much like to have better scheduling to be able to plan their trips better. To be able to create GTFS data, we in essence created a schedule to provide this required information. We did this by estimating the departure rate from the main terminus at both peak and off-peak periods. This is in no way accurate. However, It might be noted that formal planned systems are also often off schedule and standard GTFS has static information. More recently, GTFS realtime has been developed to allow agencies to upload realtime data . It would be possible to explore how this would work with matatus.

Fares

It is not very easy to follow fare information, largely because it is demand responsive and unregulated. Depending on the actual commuter demand at a terminus, fares can go up at will. This is even more pronounced when there is a downpour in Nairobi and passengers are less likely to want to walk. There are also cases of predatory fares that are lower than usual to lock out competitors. Note that besides this variability, fares are not cheap, and poorer residents can not afford them (Salon and Gulyani 2010). Fares are not a required file for GTFS so we avoided this problem but simply not creating a fare file. However, it should be possible to generate data of average fares on peak/off peak rain./non-rain days.

Applications

The GTFS data created so far has found users in the world of app developers (ma3route) and transport planners (ITDP and KIPPRA). With the GTFS data that we have so far, we have also been able to explore some data visualizations and their uses. Having data in this basic format allowed for visualizations of routes and stops and allowed for some basic analysis. For example, we could start to see where non-designated were, and the routes that had more undesignated stops than designated ones for some routes.



Designated and Undesignated Stops on a Matatu Route in Nairobi (Pink Undesignated, Yellow Designated)

This kind of analysis could in principle lead to more data-driven policy-making and empowered citizens. One major area to explore is the impact of making such data and the tools it enables open to the public. Simply making data freely available does not mean it will be transformative (Williams et al *forthcoming*) which is why we are specifically working with the technology community and reform minded policy analysts and transport operators.

5. Conclusions

In many of the cities in the world which rely heavily on paratransit even basic transport data does not exist. If it does exist, it is often not easily accessible. This makes it extremely difficult for governmental officials, policy experts, urban planners and the public to make informed decisions even as cities grow and change at unprecedented rates around them creating ever more challenges for public transit. Foreign NGOs and consultants hired by governments often create transport data to inform critical urban developments including large scale infrastructure projects that are reshaping the destinies of cities in fundamental ways. However, this data and the methodology for creating it are often not shared (Williams et al *forthcoming*).[7] This makes it extremely difficult to benchmark changes in transit use and do any substantive modelling or more data-driven policy-making. Further, in such a system, the public is not involved or engaged and the technology community, which has demonstrated its critical value to improving transit, is shut out or has to create its own data if it wishes to create useful applications and tools.

What is profoundly important about nurturing the open transit data movement in cities with paratransit is precisely the possibilities of providing critical data to the public, technologists and policy-makers and also supporting better interactions between them in the transport arena. Leveraging technologies such as cellphones that are ubiquitous in rapidly urbanizing countries to create data and linking this data to an open data architecture in a standardized way thus has the potential to fundamentally transform what, in effect, has been a closed and undemocratic transport planning process that often ignores the need to improve and include existing paratransit operators and users (Klopp 2012, Behrens et al. 2012).

In our experiment in Nairobi, we discovered that using a GTFS format is a very useful framework for initial data collection for paratransit and thinking more carefully about how these systems differ from formal, planned systems. We discovered that GTFS can be adapted and allows incorporation of rich and useful metadata in a structured way. This facilitated creating some of the first comprehensive visualizations of the paratransit system for the public and planners by using the GTFS data with Open Street Maps. It also allowed for the creation of and improvement of apps that provide better transit information to the public. However, by trying to fit aspects of the paratransit system into a GTFS format, it also became more clear precisely where the fit is often hard to make because of the differences between a planned formal transport system with fixed stops and schedules (even if they are not always strictly adhered too) and the demand responsive and flexible paratransit system.

Overall, it appears that modifications need to be made to GTFS to account for key differences between paratransit and more formal, planned systems. Further, we

[7] This is also a source of frustration for consultants who usually have to generate data each time they do any work. Thanks to Neil Taylor for pointing this out. We also discovered this when ITDP contacted us for our data to start to develop a Bus rapid Transit service plan for Nairobi. They usually have to develop the data from scratch.

discovered that there is space for improving or editing applications for transit data collection and adapting them to the challenges of collecting paratransit data. Finally, we see it is critical to start more structured collaborations across the current experiments in paratransit data collection across cities. By combining forces we will be better placed to know whether we need to move towards a new data standard for paratransit and if so, what it might look like and also who it might serve in the cities that so far have been excluded from the vast possibilities of the open transit data movement.

Acknowledgements
This work would also not have been possible with the critical thinking and hard work of the following University of Nairobi, School of Computing & Informatics students: Ikamar Ekessa, Peter Kamiri, Samuel Kariu, Maureen Mbinya, Mureri Ntwiga. At MIT, we would also like to thank students Alicia Roualt, Catherine Vanderwaart, and Alexis Howland for their work on this project. We also gratefully acknowledge the support of the Rockefeller Foundation which provided a grant for this work. Special thanks to Benjamin de la Pena at the Rockefeller Foundation for his support of and critical insights on this work. We also benefitted greatly from conversations that compared projects with Prof. Chris Zegras, Albert Ching, Stephen Kennedy, Neil Taylor, Kevin Webb and Emily Jean Eros. Last but not least, we thank James Gachanja, Dr. Zachary Gariy and Dr. Eric Aligula at KIPPRA for hosting the workshops on this work that helped our thinking, made connections and showed us a way forward in sustaining data collection work.

References

Barcelo, Jaume, Masao Kuwahara and Marc Miska. 2010. “Traffic Data Collection and Its Standardization” in J. Barcelos and M. Kuwahara (eds.) *Traffic Data Collection and Its Standardization* International Series in Operations Research and Management Science 144.

Behrens, Roger, Dorothy McCormick and David Mfinanga. 2012. “An evaluation of policy approaches to upgrading and integrating paratransit in African urban public transport systems: Results of the first round of the Delphi survey” Paper presented to CODATU XV: The role of urban mobility in (re) shaping cities. October 22-26, Addis Ababa.

Ching, Albert. P. Christopher Zegras, Stephen Kennedy and Muntasir Imran Mamum. 2012. “ A user-Flocksourced Bus Experiment in Dhaka: New Data Collection Technique with Smartphones” Available at http://web.mit.edu/czegras/www/Flocksource_JUT.pdf.

Herrera Juan, Daniel B. Work, Ryan Herring. Xuegang (Jeff) Ban, Quinn Jacobson and Alexandre Bayen. 2010. "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile century field experiment" Transportation Research Part C 18: 568-583.

Klopp, Jacqueline 2012. "Towards a Political Economy of Transport Policy and Practices in Nairobi", *Urban Forum*23: 1-21.

Mutongi, K. 2006. Thugs or entrepreneurs: Perceptions of 'Matatu' operators in Nairobi, 1970 to the present. *Africa: Journal of the International African Institute, 76,* 4: 549-568.

National Center for Transit Research. 2011. *Expanding the Google Transit Feed Specification to Support Operations and Planning.* Available at http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_PTO/FDOT_BDK85_977-15_rpt.pdf.

Ratti, Carlo, Dennis Frenchman, Riccardo Maria Pulselli and Sarah Williams. 2006. "Mobile Landscapes: using location data from cell phones for urban analysis" *Environment and Planning B* 33: 727-748.

Salon, Deborah and Eric Aligula (2012) Urban Travel in Nairobi, Kenya: Analysis, Insights, and Opportunities. *Journal of Transport Geography* 22 (May 2012), 65 – 76.

Salon, D., & Gulyani, S. 2010. Mobility, poverty and gender: travel "Choices" of slum residents *Transport reviews. A Transnational Transdisciplinary Journa*l, 30(5), 1464–5327.

Transport Licensing Board (Government of Kenya) 2012. PSV *Demand, Termani Capacities and Compliance Level with TLB Regulations in Nairobi metropolitan Area.*

UITP(International Association of Public Transport) 2010.*Public Transport in Sub-Saharan Africa.* Trans-Africa Consortium.

Wakefield, Jane. 2013. 'Mobile phone data redraws bus routes in Africa" BBC News Available at http://www.bbc.co.uk/news/technology-22357748.

Wang, F. Calabrese, G. Di Lorenzo and C. Ratti. Transportation mode inference from anonymized and aggregated mobile phone call detail records. In *Proc. 13th Int. IEEE Conf. Intelligent Transportation Syst.* (ITSC 2010), Madeira Island, Portugal, Sep 19-22, 2010, pp. 318-323.

Williams, Sarah, Elizabeth Marcello, Jacqueline Klopp. *Forthcoming.* "Open Source Nairobi: Creating and Sharing a GIS Database of Nairobi for Community Empowerment, *Annals of the Association of American Geographers.*