Journal of Transport Geography 49 (2015) 39–51

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrg

The digital matatu project: Using cell phones to create an open source data for Nairobi's semi-formal bus system

Sarah Williams [a,*], Adam White [b], Peter Waiganjo [c], Daniel Orwa [c], Jacqueline Klopp [d]

[a] MIT, Department of Urban Studies and Planning, 77 Massachusetts Avenue (Building 10-485), Cambridge, MA 02139-4307, United States
[b] Groupshot, 157 Pearl Street, Cambridge, MA 02139, United States
[c] School of Computing and Informatics, University of Nairobi, P. O. Box 30197-00100 GPO Nairobi, Kenya
[d] Center for Sustainable Development, Columbia University, 475 Riverside Dr. Suite 520, New York, NY 10115, United States

ARTICLE INFO

Article history:
Received 14 October 2014
Received in revised form 27 September 2015
Accepted 2 October 2015
Available online 30 October 2015

Keywords:
Mobile data collection
Mobile phones
Open data
General Transit Feed Specification (GTFS)
African transit
Informality
Semi-formal transit
Paratransit
Nairobi
Kenya
Informal transit
Data standards
International
Matatu

ABSTRACT

In many of the world's growing cities, semi-formal buses form the basis of public transit systems. However, little open and standardized data exist on these systems. The Digital Matatus project in Nairobi, Kenya set out to test whether the geo-locative capabilities of mobile technology could be used to collect data on a semi-formal transit system and whether that data could be translated into the General Transit Feed Specification (GTFS) data standard for wider use. The results of this work show that mobile technologies, particularly mobile phones, which are increasingly prevalent in developing countries, can indeed be used effectively to collect and deliver data in a modified GTFS format for semi-formal transit. Perhaps more importantly, through our work in Nairobi, we were able to identify the benefits and technical needs for developing data on semi-formal transit. Overall, the work illustrates (1) how the GTFS can be adapted to semi-formal systems and used by other cities with such transit systems, (2) that there is demand from technologists as well as transport communities for comprehensive data on semi-formal transit, (3) that releasing the data openly in the GTFS standard format can help to encourage the development of transportation applications, and (4) that including the entire transit community during the data development can create a community of users and mechanisms for institutionalizing a process of data updating and sharing. The engagement strategies our research team developed around the data collection process in Nairobi became just as important as the resulting data it produced.

1. Introduction

For millions in the developing world, citywide transportation options are often limited to semi-formal networks of buses and minibuses run by hundreds of diverse operators. Often referred to as paratransit, these systems constitute the backbone of mass transit for the majority of citizens in the rapidly growing cities of Africa, Asia, and Latin America (Cervero and Golub, 2007; Behrens et al., 2012; Guillen and Cordova, 2012). System-wide maps of station locations, routes, fares, schedules, operating calendars, and other key information are simply not available to the public for the majority of paratransit routes around the world. Lack of data makes it hard for users to know how to navigate these systems and creates limitations for transit planners when developing transit models (Thakur and Sharma, 2009; Barcelo et al., 2010). This is a stark contrast to cities with formal, planned transport systems where such information is expected of operators and increasingly being integrated with new technology to allow better planning decisions in real time (Catala, 2011; Lee-Gosselin and Buliung, 2012; Sussman, 2005; Kramers, 2014).

Semi-formal bus networks are composed of many private actors that, like taxis, operate for profit and are owned either by the drivers themselves or by businesses of varying size (Cervero and Golub, 2007; Guillen and Cordova, 2012). Vehicle size and capacity can vary widely, from small cars to full-size buses (Zhang et al., 2013). Unlike regular taxis, these paratransit bus systems often follow set routes with designated stops, much like formal transit systems (Cervero and Golub, 2007). They deliver an essential transportation infrastructure to developing cities by providing mobility to residents, especially the urban poor and lower middle class who often cannot afford other means of transport (Zhang et al., 2013). While they help to fill a transportation gap, paratransit systems have some drawbacks including contributions to traffic congestion, crashes, and environmental pollution (Cervero, 2000) as well as unreliability and safety concerns (Klopp and Mitullah,

* Corresponding author.
E-mail addresses: sew@mit.edu (S. Williams), adam@groupshot.org (A. White), waiganjo@uonbi.ac.ke (P. Waiganjo), dorwa@uonbi.ac.ke (D. Orwa), jk2002@columbia.edu (J. Klopp).

http://dx.doi.org/10.1016/j.jtrangeo.2015.10.005

2015). However, the ability to provide transit where the government has not offers some advantages including demand responsiveness and flexibility as well as local ownership (Mutongi, 2006; Woolf and Joubert, 2013).

With the spread of mobile phones with geo-locative abilities and improvements in information and communication technologies, new possibilities are emerging to collect paratransit data by individuals at a dramatically lower cost. Key questions emerge out of these new technological developments: Can data be captured on paratransit systems outside formal institutional frameworks? What is the best way to collect such data? What data structure makes the most sense for the storage and distribution of paratransit data? Can the data be distributed using the General Transit Feed Specification (GTFS), which is largely used by more formal transit agencies to facilitate transit routing applications? Does a new data standard need to be developed that better captures the informal aspects of paratransit systems? Who will use the data and for what purposes?

Our research team set out to answer these questions by testing whether the geo-locative capabilities of mobile technology could be used to collect a comprehensive data set on a semi-formal transit system and whether that data could be translated to the GTFS data standard allowing it to be more widely used by the larger transit and technology communities[1]. We used Nairobi, Kenya as our case study and set out to analyze the city's semi-formal bus system called matatus. Our research team collected basic route data using mobile devices for all Nairobi's 135 matatu routes. The data was then converted into GTFS, a standard widely used by transportation routing applications. The team worked with the GTFS community to develop changes in the GTFS format that accommodate the differences in the way the matatus and other semi-formal systems operate. The research team also worked with Nairobi's transit and technology community to inform them about the data, which was ultimately made open for anyone to use. This allowed local technology entrepreneurs to extend the value of the data by creating mobile routing applications, and the data was eventually uploaded to Google Maps, a first for an informal transit system. The transit community and the government used the data to discuss and develop transport plans for the city. By sharing the work with a broader set of actors as it was developed, we helped to generate a local and global community around using GTFS data for semi-formal transit.

2. Theory literature framing

2.1. Leveraging mobile devices to collect transit data

One of the biggest issues for studying and modeling transport is acquiring data to accurately represent these systems (Herrera et al., 2010). The prevalence of mobile devices with GPS positioning has produced research on the possibility of using the data generated by these devices to collect critical transport data. Many of these studies have shown cell phones can help to model transit flows by actively collecting GPS data (Caceres et al., 2012; Choi and Jang, 2000; Herrera et al., 2010; Wang et al., 2010). Other studies have looked at how the GPS data stored by cell phone providers can be used to model traffic flows in both developed and developing countries (Ratti et al., 2006; Gonzalez et al., 2008; Caceres et al.;, 2012; Talbot, 2013; Wakefield, 2013). Other projects look at how transit riders can crowd source transit vehicle locations in real time (Thiagarajan et al., 2010). Many formal transit agencies globally are actively collecting GPS data from devices they install on their vehicles (Farzin, 2008). However, studies that look at public transportation data collection often focus on formal systems rather than semi-formal ones (Farzin, 2008). Acquiring cell phone records from telecommunications companies is one key way to access mobility data but it is often extremely difficult to obtain (Gonzalez et al., 2008). Experiments in which cell phone users actively collect and contribute data through their mobile devices are more successful as the data is owned by the collector and can be shared. Our research team wanted to see if we could apply this type of methodology to semi-formal transit.

2.2. Data availability: semi-formal bus systems

When our research team started the project in 2012, we did not know of any organizations using mobile devices to generate data on semi-formal bus systems. However, as our work progressed, we discovered a handful of initiatives working in parallel to ours. A team at the World Bank, with support from the Australia Agency for International Development (AusAID), worked with the Philippines Department of Transport and Communications and other transport-related agencies in Manila to set up a transportation information system. This system includes an open database containing basic service information for the myriad of public transport modes in the city (World Bank and AusAID, 2014). The World Bank also supported a project in Mexico City with the Department of Transport (Secretaría de Transportes y Vialidad del Distrito Federal [SETRAVI]) and is conducting similar work in three Chinese cities (World Bank and AusAID, 2014; Eros et al., 2014). The MIT-based team, Urban Launchpad, has collected data, although not initially in GTFS format, for the bus system in Dhaka (Ching et al., 2013; Zegras et al., 2014). In each case, the groups involved in these projects created mobile tools to collect routing and stop data.

The informal and flexible nature of paratransit systems make them highly variable and erratic, which presents a serious challenge to data collection (Guillen and Cordova, 2012). Governments are often reluctant to collect data on these systems as they find them too “chaotic” or complex to address. Some government and industry actors collude and mutually benefit from the lack of transparency of data in these systems (Cervero and Golub, 2007; Kemei, 2014; Klopp and Mitullah, forthcoming; Klopp, 2012; Republic of Kenya, 2009). When government agencies do in fact collect data, they often hire consultants who do not always share the data (Williams et al., 2014). Furthermore, governments are sometimes hesitant to share data they have on semi-formal transit systems because they often do not want acknowledge these systems for political reasons. This is the case for Mexico City, which recently collected data on the formal and semi-formal bus system but so far has only released data on the formal bus system to the public (Eros, 2014).

Semi-formal transit operators sometimes collect analog data on their systems to help manage their services, but this data is rarely standardized or shared across transit operators or with the public. Many semi-formal transit operators do not see an immediate benefit to creating and sharing data or, alternatively, do not have the means to collect it. The informal, and often unsanctioned, nature of these operations may lead some owners to keep their activities hidden from government oversight. The operators who do collect data on their systems do so to maximize profit (Eros, 2014), and the data is usually incomplete, unstandardized, private and, therefore, unavailable for comprehensive transportation planning or the development of user-centered transit information.

While the recent initiatives to collect data on semi-formal bus systems marks a change from the past, few cities in the developing world are currently generating or sharing transit data in a standardized format, such as GTFS. A review of the GTFS Exchange, a widely-used web-based platform for sharing GTFS transit data[2], shows that only four of around 766 agencies producing feeds were in Africa, including

[1] In Nairobi, a number of technology entrepreneurs were starting to develop transit applications before we started our work but were generally not perceived as part of the “transit community” of planners, regulators, operators, insurers, mechanics, drivers, and passengers. That is now changing, and we from now on include technology entrepreneurs in our category of “transit community.”

[2] GTFS exchange was started and is maintained by Jehiah Czebotar.

our own, and two were mostly rail (not bus) data. [3] There are also no semi-formal transit systems on the feed.

2.3. General transit feed specification (GTFS) and open data for transit

GTFS was first developed in 2005 for Portland's TriMet transit agency in conjunction with Google to provide transit agencies a way to standardize their data for use with trip routing software, such as Open Trip Planner and Google Maps (McHugh, 2013). The standard was implemented in Google Maps in 2006 and adopted by transit agencies across the United States that wanted to provide their users with better access to route and schedule information. 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, shapes, routes, trips, stop times, and other schedule data[4] (Google Developers, 2014). By 2007, many formal transit agencies had adopted the GTFS standard to share their data, even if this transit data was originally collected in a different standard, so their transit routing information could be viewed in Google Maps (McHugh, 2013; Wong, 2013). This process has created a worldwide standard for openly sharing transit data, which is often posted on the GTFS Exchange.

The openness and sharing of GTFS data has encouraged its use for transit applications beyond simple trip planning. Other tools have been developed to improve transit operations and planning overall (Catala, 2011; Lee-Gosselin and Buliung, 2012). A Brisbane study used GTFS data, along with *go* card[5] data, to identify the travel paths of passengers on their transit system (Tao et al., 2014). Another study used GTFS data from Auckland (New Zealand), Vancouver (Canada), and Portland, OR (United States) to develop a model that would allow public transport agencies to assess and benchmark different services (Hadas, 2013). Before the release of GTFS data, this type of analysis and assessment was hard to achieve because of the varying data standards across rail, bus, and subway routes (Hadas, 2013). *Open Trip Planner*, a tool originally developed for GTFS routing, created a plug-in to allow users to determine the accessibility of transit. The plug-in has been widely employed and was used for determining transit accessibility in New York City directly after Hurricane Sandy (Byrd et al., 2012; Wong, 2013). Overall, transit planners are beginning to realize that GTFS can be used for applications beyond trip planning and are starting to use this data to analyze transit in new ways (Catala, 2011).

3. Nairobi context

3.1. Nairobi's technology community

Nairobi, Kenya provides a good case study for how mobile phones can be used to collect transit data. Over the last decade, mobile technology use has exploded in developing countries, and Kenya, particularly Nairobi, has become a center for some of these developments (Aker and Mbiti, 2010). The number of mobile connections in Kenya rose from 30.4 million in 2012 to 31.2 million in 2013, and Kenya's current mobile phone penetration rate is 74.9%, above the average for Sub-Saharan Africa (Kenya National Bureau of Statistics, 2015). The low cost of handsets and texting plans facilitates the rapid spread of mobile phone use. This rapid expansion of mobile use in Kenya is evident in the success of the M-Pesa, a mobile banking service. Nearly two years after starting in 2007, M-Pesa has 8.5 million Kenyan users, and US$3.7 billion (equivalent to 10% of Kenya's GDP) has been transferred through the system (Safaricom, 2009; Mbiti and Weil, 2011).

Nairobi has a thriving technology community and higher mobile phone use than the rest of the country. It is home to the iHub, an innovation and technology space developed to encourage and support technology entrepreneurs by creating a shared community of learning (Hersman, 2012). *Ushahidi*, a crisis mapping tool now used worldwide, was developed in Nairobi as a response to the 2007 election crisis. In 2013, IBM launched a research lab in Nairobi in collaboration with the Ministry of Information, Communications and Technology (ICT) through the Kenya ICT Board. The lab focuses on applied research and solving problems "relevant to Africa and [that] contribute to the building of a science and technology base for the continent" (McLeod, 2013). Much interest and experimentation in the use of mobile technology has focused on health, economic development, and humanitarian response. The application of mobile technology to the many problems in transportation appears to be just beginning.

3.2. Nairobi's semi-formal transit (matatu) system

Nairobi's matatu network comprises over 135 routes that, according to the 2009 census, serve a population of well over 3.1 million within the metropolitan area. Matatus act as the main motorized public transport for the majority of city inhabitants even though they are privately run and operated (Salon and Aligula, 2012). In Nairobi, the matatu network developed in reaction to the gap in service left by poor funding and management of the municipal public transport systems (Mutongi, 2006; Klopp and Mitullah, forthcoming). In contrast to other infrastructure, the vehicles are locally owned and involve large numbers of small businesses and independent workers, from the operators (who often own large matatu fleets), to the drivers, touts, and mechanics (Mutongi, 2006). Matatus largely run on "official" routes, usually remnants of the former bus network. However, as the city expands and new roadways are constructed, additional unsanctioned routes are developed by the operators. Service does not always have fixed schedules and fares, and drivers often take detours to avoid traffic or police and sometimes take the liberty of improvising stops. Currently, approximately 9554 matatus and buses serve the Nairobi region (Transport Licensing Board, 2012).

4. Data Collection Team and Methodology

Over the course of 2012–2013, our research team from three universities (University of Nairobi, MIT, and Columbia University) and one US design firm (Groupshot) successfully collected data on 135 routes that comprise Nairobi's matatu system. The University of Nairobi led the data collection process with a team of five students who performed most of the field work. Students rode on the buses and collected route and stop names as well as physical characteristics of the stops. In the few areas where the matatus were too dangerous for the student to ride, students followed the matatus in cars. Data collection occurred from September 2012 to September 2013; however, the process was periodically stopped to test data quality and retool our collection software. Once our tools and methods were functioning well, final collection took roughly six months. Routes often needed to be surveyed multiple times to ensure we obtained the most consistent route. Routes can change because of construction, avoidance of police, and school opening and closing times. Once collected, the data was validated using the Google GTFS validator.

The data collection process involved identifying existing routes, developing and testing mobile GPS-enabled tools to collect the data, creating a unique coding structure to allow the data to be formatted in GTFS, generating a methodology for data collection in the field, translating the data into GTFS, interfacing with Nairobi's transit community, and releasing the data by posting it on GTFS Exchange website in conjunction with a public launch. This is the first time these routes

[3] The other African GTFS feeds included railway data from: the Tunisia Society National Des Chemins De Fer (or Tunisian railways); Gautrain Management Agency which has oversight for rail in Pretoria and Johannesburg South African rail; MyCiti, which was in Cape Town and had some bus Bus Rapid Transit data. From GTFS Exchange last accessed, August 24, 2014.

[4] The full the guidelines can be found on the Google transit developer site. https://developers.google.com/transit/, Last accessed September 26, 2015.

[5] TransLink's South East Queensland electronic ticket.

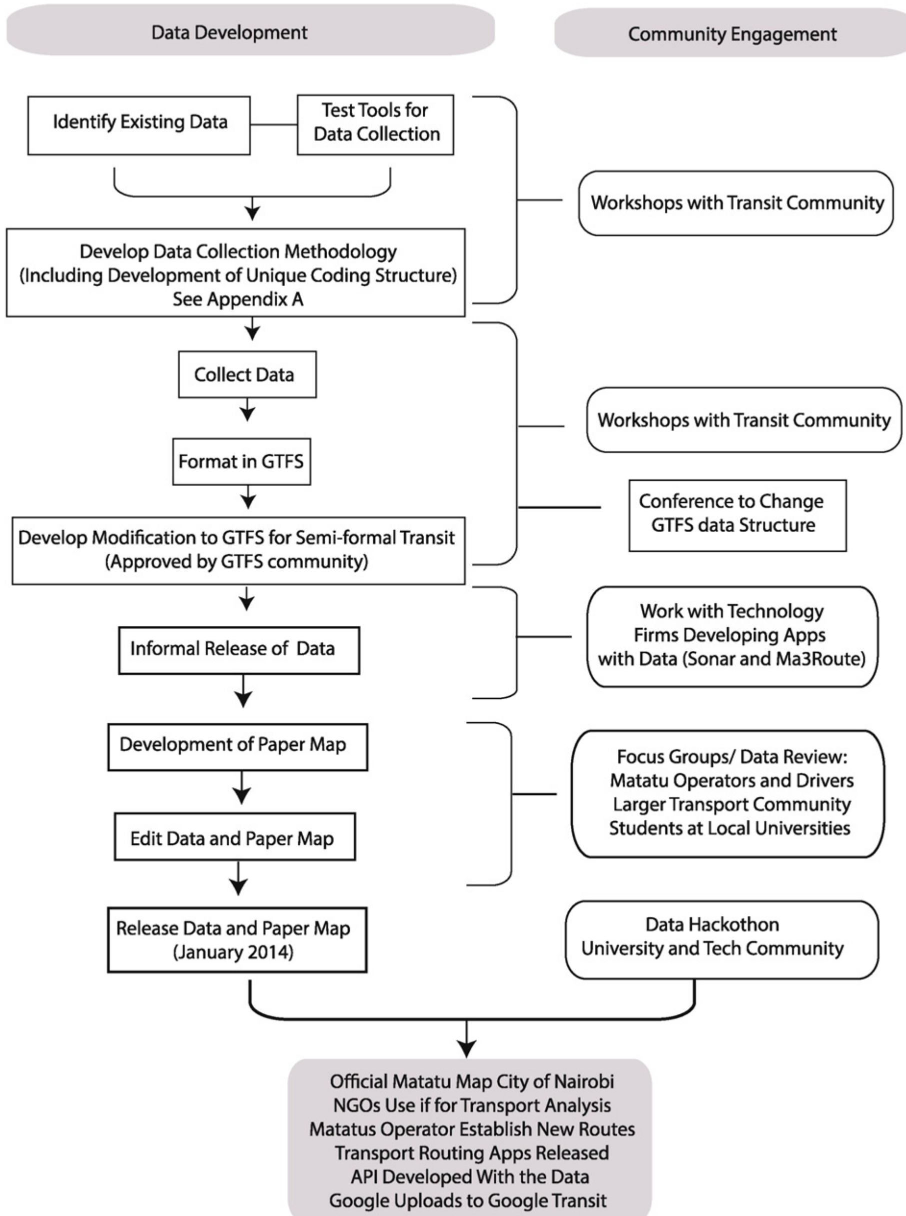


Fig. 1. Flowchart showing research strategy and process. One can see that engaging the community with the data was just as important as developing the data itself. This process helped the transit community trust and understand the data and, therefore, use it.

have been mapped in a comprehensive manner, as well as the first time paratransit data in Nairobi has been fully integrated into the GTFS standard and later uploaded to Google and Open Street maps. Details of the methodological process are below. Fig. 1 provides a flowchart of the overall research methods, data collection, and development process.

4.1. Identifying the current routes

The first step in this work involved finding and collecting existing data on routes. We obtained government data in the form of Microsoft Word document files but found it to be incomplete, outdated, and inaccurate. Route changes are often developed by the matatu industry, not the government, in response to demand. These changes are usually not recorded in the government files. It should be noted that the Kenya National Transport and Safety Authority recently started moving towards publishing matatu route changes as well as information about new matatu licenses in the *Kenya Gazette*, the official government publication. This publication may help in updating the data moving forward.

The research team discovered a paper-based map created in 2010 by Kenya Buzz, a Nairobi-based media company, for commuters.[6] However, the map had a small print release and was not available at the time of the study. The data used to develop the map was never released and was not digital. "Living in Nairobi" published a highly stylized route map in 2012 after we had started our work but did not publish any of the data collected to create the map and has not maintained it.[7] *Panga Safari*, formally *Matatus Online*, developed a private matatu route database covering some parts of the city but did not include standardized routing information or consistent stop documentation, making it difficult to upgrade this data to a standardized format such as GTFS. The database has since been expanded and can be searched through a

[6] Last accessed 10/9/2014 (http://www.globalpost.com/dispatch/kenya/100127/kenyas-wild-matatus-captured-map).

[7] Last accessed 10/9/2014 (http://www.jambonairobi.co.ke/services/public-transport/nairobi-route-maps/nairobi-matatu-route-map-eastlands/).

web interface, but the backend data was never made public. [8] We also identified and reviewed existing data collection projects performed by entrepreneurs for business or social reasons. Many were incomplete, included major errors, or employed inconsistent methodologies and data structures that made the data impossible to combine or format into GTFS. Also, only a few were willing to share their data. Given the limitations of the existing data on the matatu system it was clear that in order to develop a comprehensive, consistent, and standardized data set that could be accurately used for transit modeling and provide accurate routing information our research team would need develop the data ourselves.

4.2. Tools used to collect data and method of collection

The team began by testing various Android smartphone-based data collection tools, eventually focusing on using *MyTracks*, a basic GPS tracking system for mobile devices developed by Google. GPS units were used as back-up so we could compare the accuracy of the multiple forms of data collection (See Figs. 2 and 3). Through testing, we found that standard GPS units and the mobile applications on Android phones had similar accuracy. However, mobile phones sometimes took longer to lock in on GPS satellite signal and could lose connection more regularly. The *Mytracks* app allowed for the easier digital collection of meta-data (for example, the name of a stop and current passenger counts could easily be recorded). Collection with GPS units needed a paper recording to cross reference waypoint numbers, which was then digitized and joined to the GPS data later on.

As we were engaged in this data collection process, we discovered an open source web and mobile app *TransitWand* created by the consulting company Conveyal for a similar transit data collection project in Mexico City, involving the World Bank and the Department of Transport for Mexico City. We tested *TransitWand* in Nairobi and found it resolved some of the drawbacks of earlier apps. When compared to data produced by *MyTracks* and GPS units, *TransitWand* generated cleaner data because the software automatically snapped location data to roads. However, because the tool was still in beta development at the time, the ability to directly export to GTFS was not operational. This made post-production work of *TransitWand* data more time consuming than for other applications. While *TransitWand* will be very useful if it is developed further, the team decided to complete the data set using *MyTracks*. We provide a summary of our findings on the various data collection tools in Fig. 4 below.

The biggest challenges in using the mobile data collection applications included extremely limited battery life, the slow speeds of affordable Android phones, phone theft, and small screens size and frequent stops, both of which made digital data entry more time consuming. Still, we found that mobile phones were the most effective tools for data collection and determined that there is potential for the development of new phone applications to streamline information processing while in the field and automate conversion to GTFS.

While data was largely collected onboard matatus, on particularly dangerous routes data was also collected in private vehicles that followed matatus. Data collection onboard the matatu vehicles themselves was found to be the most optimal method because it allows data collectors the opportunity to engage with drivers and passengers about stop names and route information. It is also more affordable and scalable than the alternative of employing a tracking car. While private cars allowed the data collector to observe multiple vehicles at the same time and provided extra time to take notes, the information obtained from talking with passengers on the matatus proved more valuable.

After testing several different tools and processes used for the data collection on well-known routes, we devised a standard protocol and methodology for creating route, stop, and shape data to fit the GTFS coding structure (See Appendix A). In all cases, data collectors would ride a route (either in the matatu or following in a separate car), use the data collection tool to generate latitude and longitude points along the route, and record all of the stops as well as specific coding information we developed for each route, stop, and shape, which was essential for the GTFS protocol (See Appendix A).

While many paratransit systems involve some stopping at varied locations based on customer demand, regular and central stops and large terminals exist. Students identified stops based on their personal knowledge, information from frequent users of these routes, visual notation (e.g., signs, shelters), and, if necessary, confirmation from discussion with matatu crews or a group of commuters on the route. In many cases, stops were identified as either designated (established by a government agency) or undesignated (established by matatu operators based on user demand and not officially sanctioned)[9] (See Appendix A). Adding this additional data to the GTFS file could be a useful tool for the city should it move to formalize many of the more heavily used undesignated stops (see Fig. 5).

5. GTFS Formatting for Semi-Formal Transit

Once the essential data on the routes shapes and stops were collected, we started the work of translating the data into the GTFS standard. The GTFS data format assumes that the system is part of a formal transit agency and that the transit agency has developed a unique identification system for routes and stops. Therefore, we needed to develop a unique identification system (See Appendix A). GTFS also assumes there are standard schedules and fares, standard vehicle types, scheduled service outages, and that transit agencies are maintaining the data. Given that matatus have loosely-set schedules, we had to generate rough estimates for departure and frequency of trips from the main terminus at peak and off-peak periods as well as the stop times (a matatu generally leaves a stop every two minutes during peak hours). Matatus do not have standard fares, as the fares are largely demand driven. For instance, when it rains in Nairobi, fares can triple. There are also cases of predatory fares—fares that are artificially lowered to lock out competitors (Salon and Gulyani, 2010; UITP (International Association of Public Transport), 2010). Fare information is optional in the GTFS format; therefore, we decided not to populate this field since it would be difficult to develop it in a standardized way.

GTFS requires an Agency file, usually a transit agency. Given that the data was developed for the hundreds of “agencies” operating matatus, the research team is listed in that field. The matatu system is fragmented and complex. Therefore, a neutral and technically capable institution should collect the data can ensure quality and uniformity. Ideally, this function should eventually be taken over by a government agency, such as Kenya Institute for Public Policy Analysis (KIPPRA) or the National Transport and Safety Authority, with a steady budget allocation for updating the data along with a strong mandate to make it openly available. KIPPRA has expressed interest in maintaining the data and the methodology, which will be refined in a next phase focusing on streamlined and user-friendly systems and tools for updating. More recently, the government has made moves to create a Nairobi Metropolitan Area Transport Authority, which will have clear responsibilities that include data gathering. The data, methodologies, and tools developed through this work, along with the expertise KIPPRA has gained through our collaboration, will be a helpful in kick-starting the data and transit planning work of this new agency.

[8] Last accessed 10/9/2014 (http://www.matatuonline.com/).

[9] We used three criteria to identify designated stops: 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 of approval by being noted in official road maps. However, as the city government has not been actively planning and designating official stops, the majority of stops remain informal and undesignated. Therefore, we collected both the designated and undesignated matatu stops and coded them in the stop ID data file. (See Appendix A).

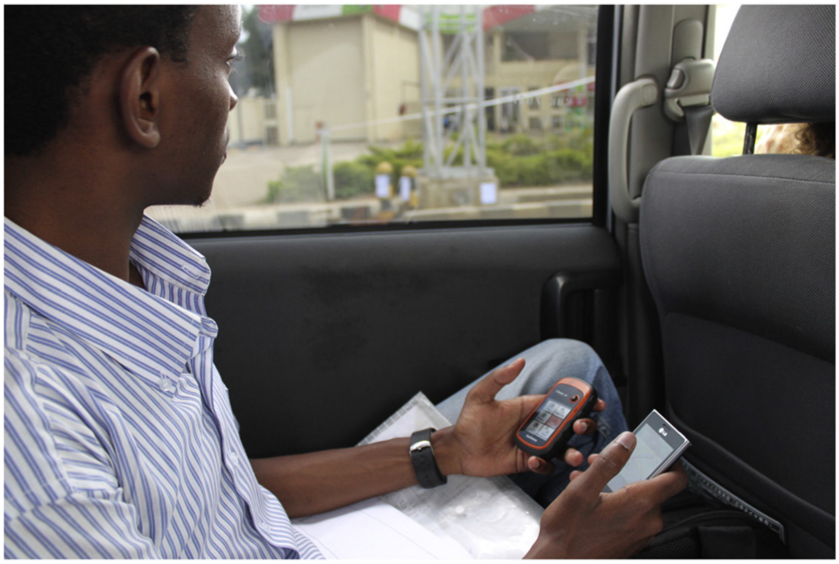


Fig. 2. Image of student collecting data using an android cell phone and a GPS unit as back-up. Image Credit Adam White.

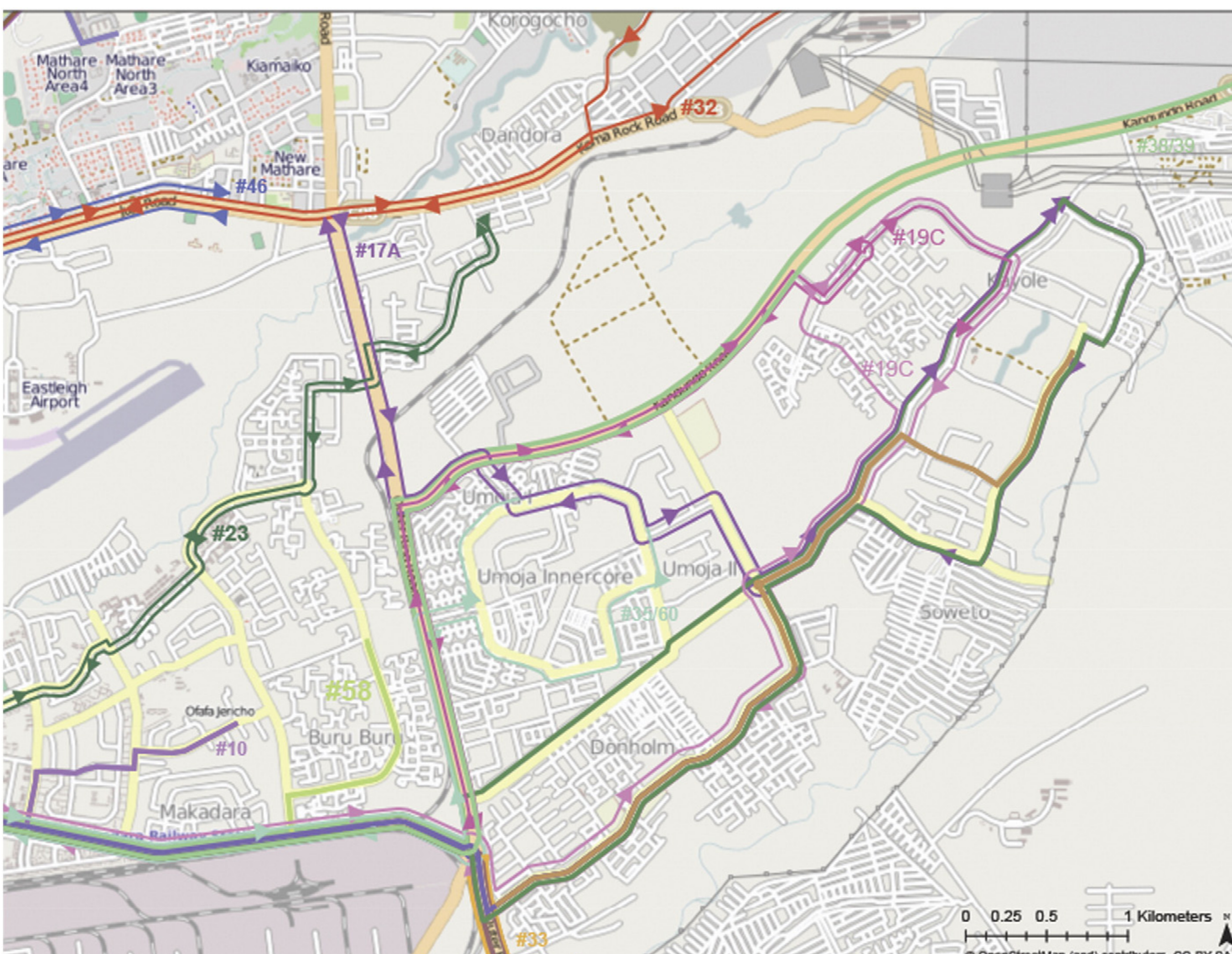


Fig. 3. Image of data mapped in Open Street Maps.

Tool Tested	Pro	Con
MyTracks	• Snapped latitude and longitude data to the roadways allowing for cleaner data. • Easy to modify for additional data elements. • Easy to visualize the data captured immediately.	• Needed to develop tool to post process data collected in the field to GTFS data standard.
TransitWand	• Snapped latitude and longitude data to the roadways allowing for cleaner data. • Developed specifically for the collection of GTFS data. • Allowed for modification of questions asked in the field. • Good user interface.	• Needed programming modifications for use in the field. Had this tool been found earlier in the process these modifications might have been very similar to what was needed to be performed by MyTracks. • GTFS export was not working properly making post processing to GTFS necessary and cumbersome.
FlockTracker	• Allowed for latitude and longitude data collection on phones. • Could be modified to allow for survey questions.	• Needed many modifications to be used in Nairobi. A new interface would need to be developed. • As this tool was in the early phase of development when we used it, programming communication errors causes it to lose connection with GPS satellites. We understand some of these bugs have been fixed in more recent versions of the software. • Did not snap data to the roadways - which meant that much post-processing would be needed.
Fulcrum	• Allowed for latitude and longitude data collection on phones. • Could be modified to allow for survey questions.	• Designed for development on iphones, which are not common in Kenya as they are too expensive. Most smartphone users in Kenya use Android phones.
GPS Surveyor	• Allowed for latitude and longitude data collection on phones.	• Works much like a GPS unit and does not allow for additional data (such as stop names to be collected in the field) no advantage over GPS units.
Open Data Kit	• Allowed for latitude and longitude data collection on phones.	• Needed significant programming development to allow for use in the field.
App Inventor	• Allowed for latitude and longitude data collection on phones.	• Needed significant programming development to allow for use in the field. Was the basis for the FlockTraker program previously mentioned.
• It should be noted that for all the products we tested, the interaction between the phone and GPS satellites needed for data collection drained the battery life of the phones, making it difficult to complete whole routes without extra battery packs for the phones themselves.		

Fig. 4. A table comparing the different data collection tools.

5.1. Changing the GTFS standard for semi-formal transit

As the previous discussion shows, semi-formal transit systems operate differently from traditional buses. The research team wanted a way to indicate this difference in the GTFS data format. Modifying GTFS is particularly important for hybrid transit systems made up of both formal and semi-formal systems, because it would allow for more accurate transfer and routing between the two systems and would also allow planners to analyze the dynamics between the two transit types. Our team sought to actively address the changes needed to GTFS for use with semi-formal transit. With support from the Rockefeller Foundation and the World Bank Open Transport Initiative, we convened a conference of groups involved in developing the GTFS standard. We included research teams focusing on developing GTFS for semi-formal transit (this included members of the team in Mexico City, Manila, and Dhaka), and members of the paratransit community in the United States who are struggling with similar issues with using the GTFS standard.[10] The GTFS standard is particularly interesting in that it has never been formalized by any agency or multi-lateral body but has become a de facto standard through adoption by growing numbers of users globally who want their data to appear on Google maps. Modifications to GTFS to make it more user friendly for paratransit might encourage increased adoption of this standard as well as increased information to users of these systems.

Conference participants proposed and approved a change to the GTFS format. The group added a “continuous stops” field to the stop times and routes table to indicate that a route and its stops do not follow normal bus transit behavior but rather that it is possible to board or debark from a transit vehicle at any point along the vehicle's path of travel. The field can have the following non-negative integer values: 0 or blank = normal stop behavior along entire route (default), 1 =

[10] We acknowledge the role of Holly Krambeck who leads an Open Transit Initiative at the World Bank in convening and facilitating this conference.

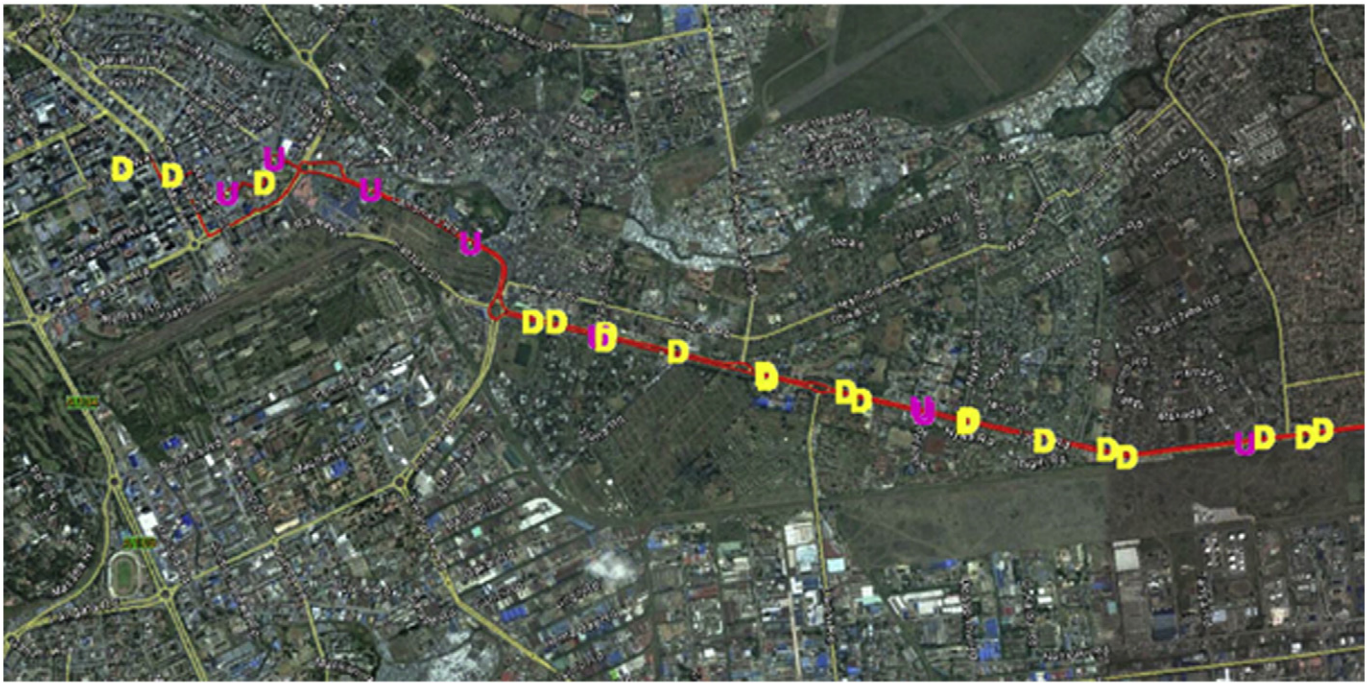


Fig. 5. Image of the designated and undesignated stops along one matatu route in Nairobi. “D” marks stops designated by the city. “U” marks undesignated stops. The majority of stops are undesignated. This can contribute significantly to the traffic congestion problem in Nairobi, although the problem is also of poor traffic management and road design that does not account for the needs of matatus and their riders.

continuous stopping behavior along entire route. If 1 is specified, a valid shape file must be identified for the route to indicate the complete path of travel for each trip. This makes shape files, which are optional for formal agencies, more essential for the development of GTFS for semi-formal transit.

6. Working with the transport community in Nairobi

As we proceeded with 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 data analytics and modeling. The research team held two workshops for technologists, various government transportation offices, policy analysts, and transport operators to discuss the project. The workshops were held to obtain early feedback from potential users and to inform members of the transit and technology community about the data collection process so they could trust the data we collected. In the process, the team assisted Laban Okune, who used the data to improve his award-winning *Ma3Route* mobile app, which shares real time, crowd sourced matatu and traffic data between users. We worked with Jeremy Gordon of Flashcast who developed a routing program called *Sonar* using the data and who also shared data with us. We also facilitated the use of the data by UN-Habitat/Institute for Transportation and Development Policy (ITDP) who found the data useful as they began a Bus Rapid Transit Service Plan for the city (ITDP and UN-Habitat, 2014). In turn, they shared further data with us, creating more circulation of valuable data. The transparency in which we discussed the development of the data at these workshops helped to create users of the data and an ethos of sharing (Williams et al., 2014). Since the release of the data in January, 2014 there have been more than 5000 downloads of the data and paper maps, showing that it has been used by a broad group of people.

The research team also developed workshops with university students and matatu drivers and operators to obtain feedback on the maps we developed from the data. The matatu drivers provided helpful information about routes and stop names missing from our data set. They were also excited by the development of the paper map (See Fig. 6), which allowed them for the first time to see the extent of the transport system. After seeing the map, the matatu drivers were able to illustrate new ideas for routing to reduce congestion. University students, who are frequent matatu users, helped with landmarks for the map and also gave feedback on the utility of the map for wayfinding and journey planning (Klopp et al., 2015). A similar conversation with officials from the National Transport and Safety Authority (NTSA) helped them recognize newly established routes as well as generate an overdue conversation on transit routing and planning within Nairobi's transport community.

To further disseminate information about the data, the research team hosted a transit hackathon at the University of Nairobi. Over the weekend of January 25–26, 2014, eight teams of up to four university students each participated in the collaborative programming competition. The teams came up with a number of ideas for mobile apps including a trip-planning application and another that estimated fares for different routes. One team developed an application that functioned as the backend for a group ridesharing program, enabling partygoers to “crowdsource” a matatu ride home late at night. Another would alert drivers of notoriously accident-prone areas, or “blackspots.” The winner of the hackathon was Paul Mutie who devised *Ktransit*, a program that created an Application Program Interface (API) to access the GTFS data more efficiently by translating the series of comma-delimited data into a data structure that is more accessible by other mobile applications.

Finally, our data was accepted by Google for uploading onto Google Maps. This means that anyone using Google Maps can plan a trip not only by car, which was previously available, but also by matatu. The new transit feature, made possible through our data, was launched on August 26, 2015 in Nairobi and provides different route options based on different user-specified origins and destinations (Fig. 7). It is the first semi-formal transit system to be included in the Google Maps transit routing feature. The data's inclusion in Google Maps provides us with a valuable opportunity to study whether having better information about one's transit system changes ridership behavior. The research

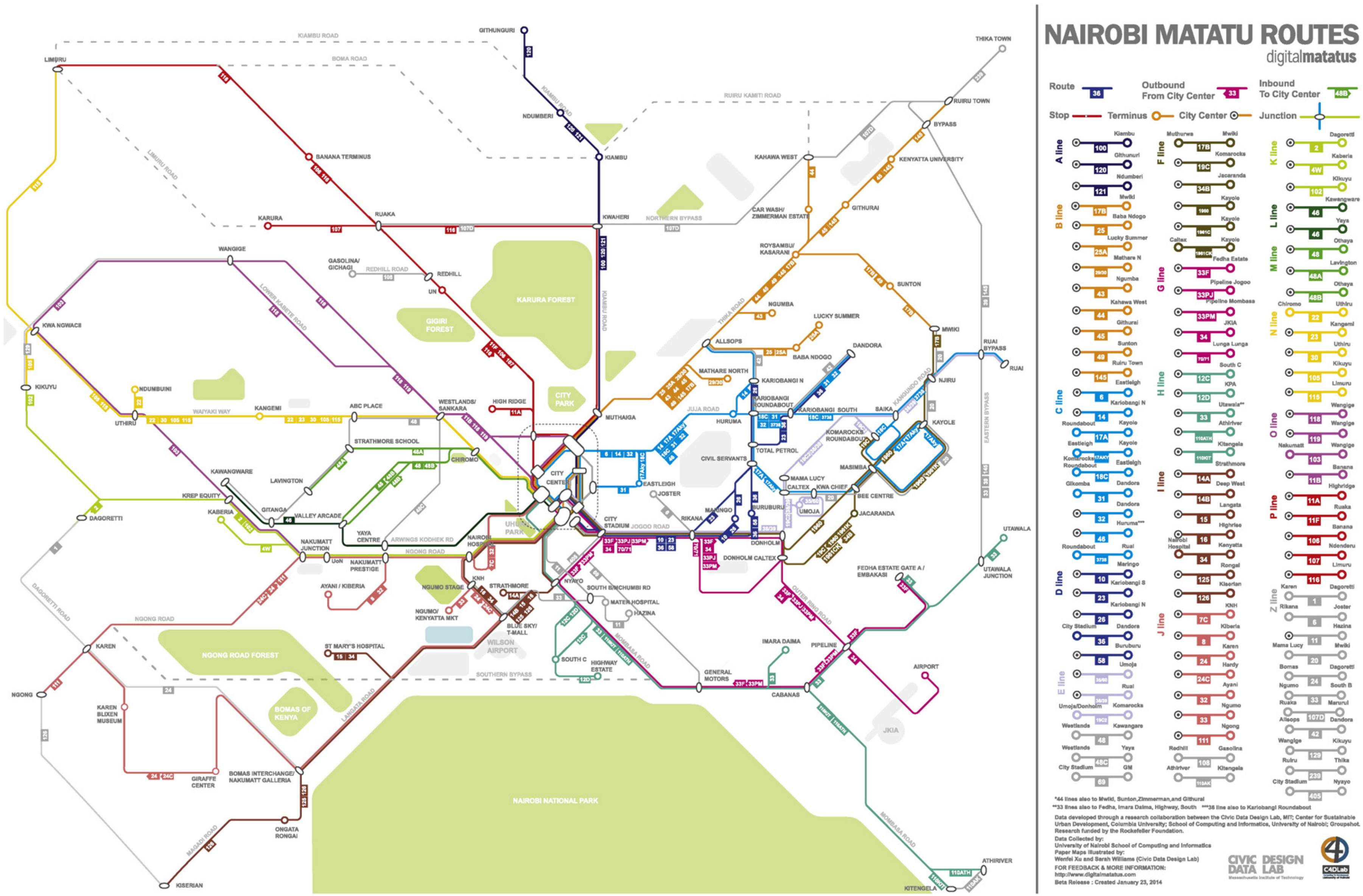


Fig. 6. Matatu map used during our focus group with matatu drivers and owners. Nairobi's city government made it the official matatu map when the final edited version was released in January 2014.

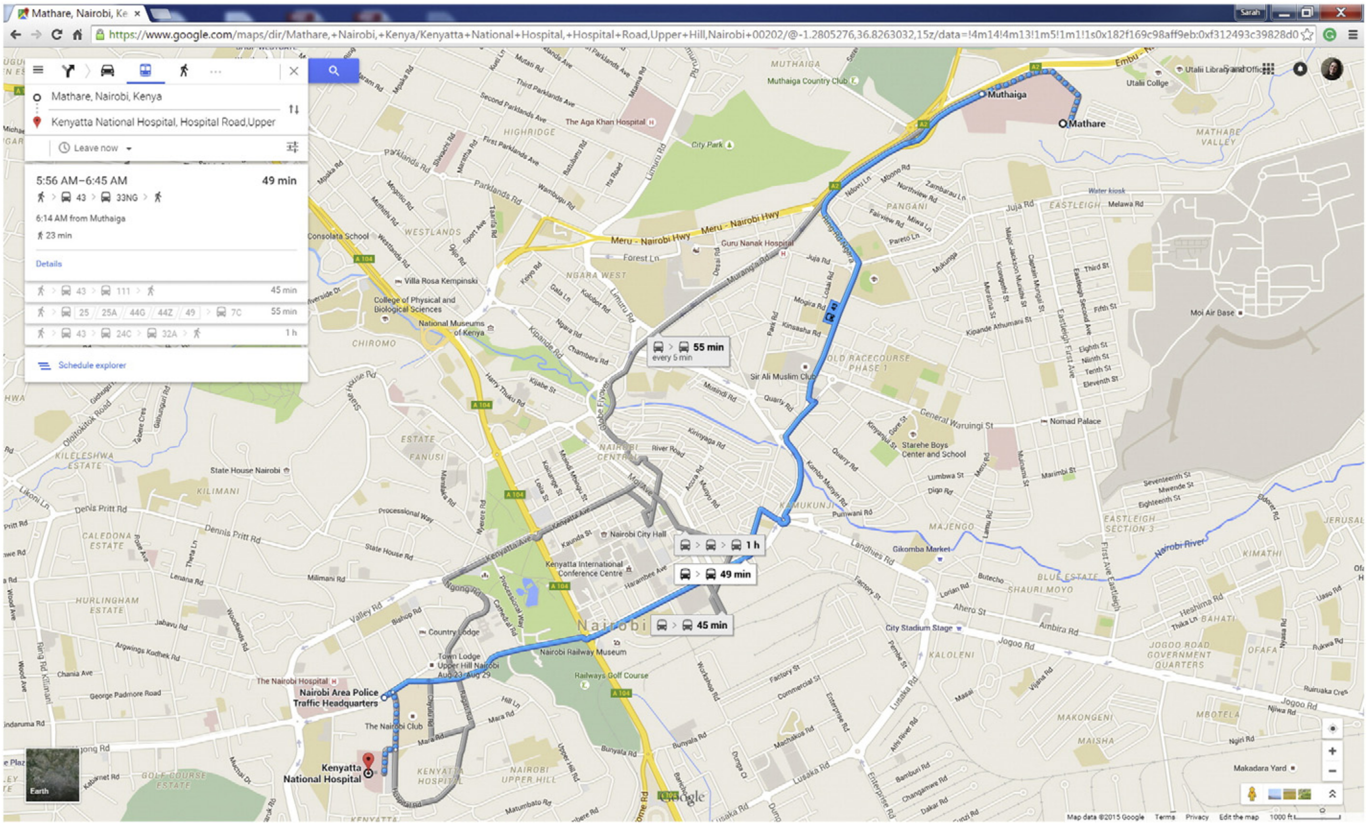


Fig. 7. The above map shows a screen shot of a Google Maps transit search for direction from Mathare, an informal settlement in the northwest to Nairobi, to Kenyatta National Hospital. The direction now includes options for taking a matatu, which would be the only option for many people living in Mathare.

team has planned a series of surveys, one implemented before the launch and several to be administered after, to determine how the access to the data through Google Maps changes passenger behavior.

7. Discussion and Conclusion

In many cities with paratransit, basic transport data often does not exist or is inaccessible. This project demonstrates that with a dedicated team and by using mobile technology, it is possible to create valuable data for semi-formal bus systems. In addition, we showed how to transform this type of data into a GTFS format useful for planning, research, operations, and transit routing applications. Further, we pinpointed specific changes needed to the GTFS standard to accommodate the nature of paratransit. Overall, we found that the GTFS format is a very helpful framework for paratransit data collection because of its integration with emerging planning software developed for the format and its requirements for more detailed and structured analysis of key features of these transit systems.

We also discovered that the GTFS format allows for the inclusion of additional data points that are not part of its core. This feature can be helpful for future modeling and planning of paratransit systems. For example, we used this feature to develop additional information on whether stops were designated or undesignated. Other data, such as ridership statistics or vehicle safety, could also be collected and would help with transit planning. More importantly, the standardized nature of this data has created the possibility of using plug-ins and programs developed for GTFS to measure transit accessibility and transit flows among other planning applications (Byrd et al., 2012; Hadas, 2013; Wong, 2013).

Our tests of existing technology for mobile geographic data collection, including *My Tracks* and *Transit Wand*, also show that many of these tools can be adapted for GTFS data collection. With some modifications, these tools could better facilitate the collection of GTFS data for paratransit systems. Data collectors found it cumbersome to enter in the metadata necessary for the GTFS format while in the field. Future research should address changes to the tools to assist with the data collection process. Data storage and export from the tools made it difficult to translate the raw data into the formatted text files GTFS requires. Future work should also look into developing data collection tools to automate GTFS formatting. *Transit Wand* developers hoped to pursue this, but as part of a consulting company, they would need a project specifically tasking them to create those tools. This points to the need to find a mechanism for more public investment in some of these tools and also the open data, innovation and research they enable. Future research should also look at the possibility of developing crowd sourced data collection tools for paratransit; so far, these tools have largely been applied to more formal transit systems (Thiagarajan et al., 2010). Creating new tools that facilitate data collection processes in the field and the ability to generate GTFS data on the fly would help the needed development of comparable data on semi-formal transit systems.

Interestingly, our team discovered other researchers in different parts of the world who are simultaneously working on similar projects to develop data on semi-formal transit in the GTFS format. The Digital Matatus project helped bring this group together through a “GTFS for the Rest of Us” conference convened with the World Bank. Continued development and expansion of this community and sharing of insights, data, and tools could help support a new paratransit inclusive GTFS format and encourage the development of transit planning tools for semi-formal transit that use the format. This work can help spread the use of emerging GTFS data for analyzing networks and systems and facilitate cross-city comparative studies on how these systems function and perform.

The Digital Matatus project also illustrates that there is demand for comprehensive data on informal transit, which is stored in a standardized format, such as GTFS. This is evidenced by groups in Nairobi that took the data we openly shared to develop over five mobile matatu routing applications, *Ma3Route*, *Sonar Flashcast*, *Matatu Maps*, *Digital Matatus* and *Transit App*, which are now in use in Nairobi. Furthermore, it is not only the technology community that benefited from the GTFS data. NGOs, such as Institute for Transportation and Development Policy (ITDP), multilaterals such UN-Habitat and the World Bank have used the data in their project work for Nairobi, and UN-Habitat and ITDP more recently have tried to replicate this work in Kampala, Uganda.

The Kenyan government is increasingly seeing the benefit of developing this kind of data. The Nairobi City County Government has designated the map as the official transit map for the city (see Fig. 6). The government's acceptance of the data was in large part because of its inclusion in workshops about the data collection and the open sharing of data. By engaging Nairobi's transit community during the data development, we created trust in the accuracy of the data, demand for its use, and a stronger data sharing ethos (Williams et al., 2014).

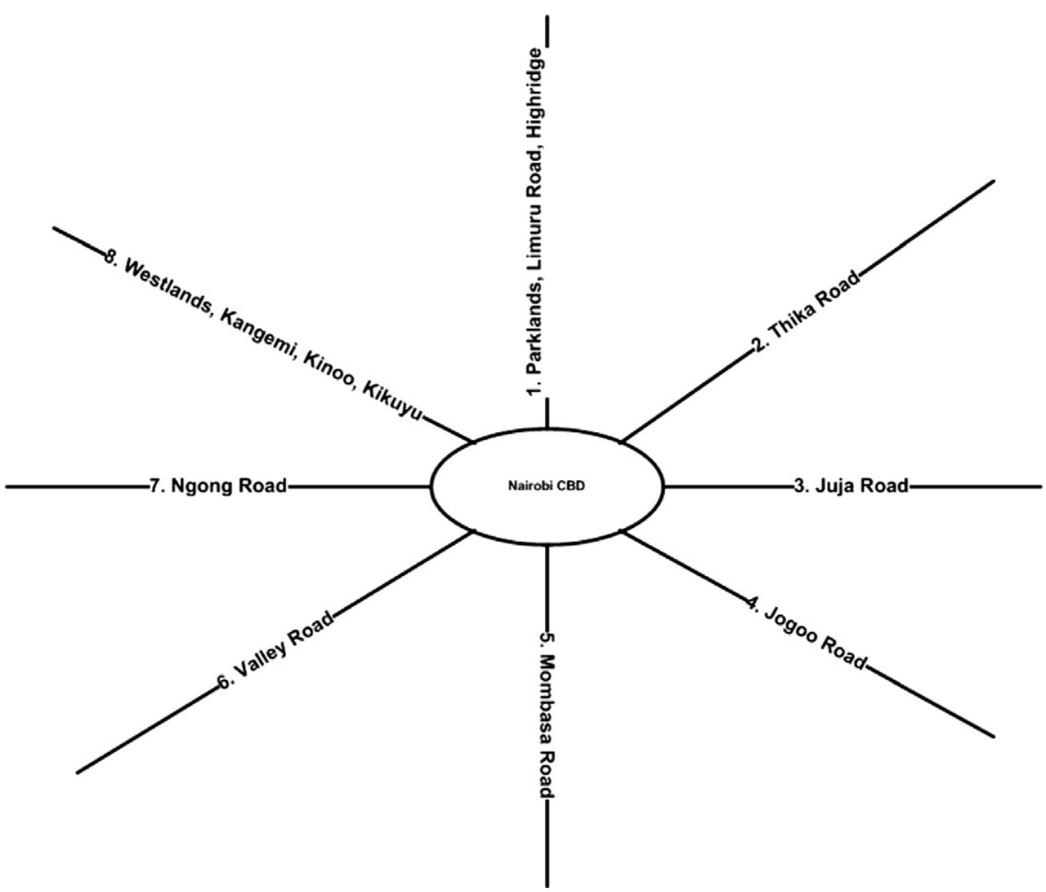
In brief, leveraging technologies, such as mobile phones, that are ubiquitous in cities in developing countries, to create data and then linking this data to open-data architecture, such as GTFS, has the potential to fundamentally transform what is often a closed data-deficient transport planning process in many cities. This is especially the case if the tools and techniques of data creation are anchored locally allowing for updating of data over time. Overall, this kind of work ensures cities that depend on paratransit will be included in and benefit from the growing technology revolution in transportation (Townsend, 2013).

Acknowledgments

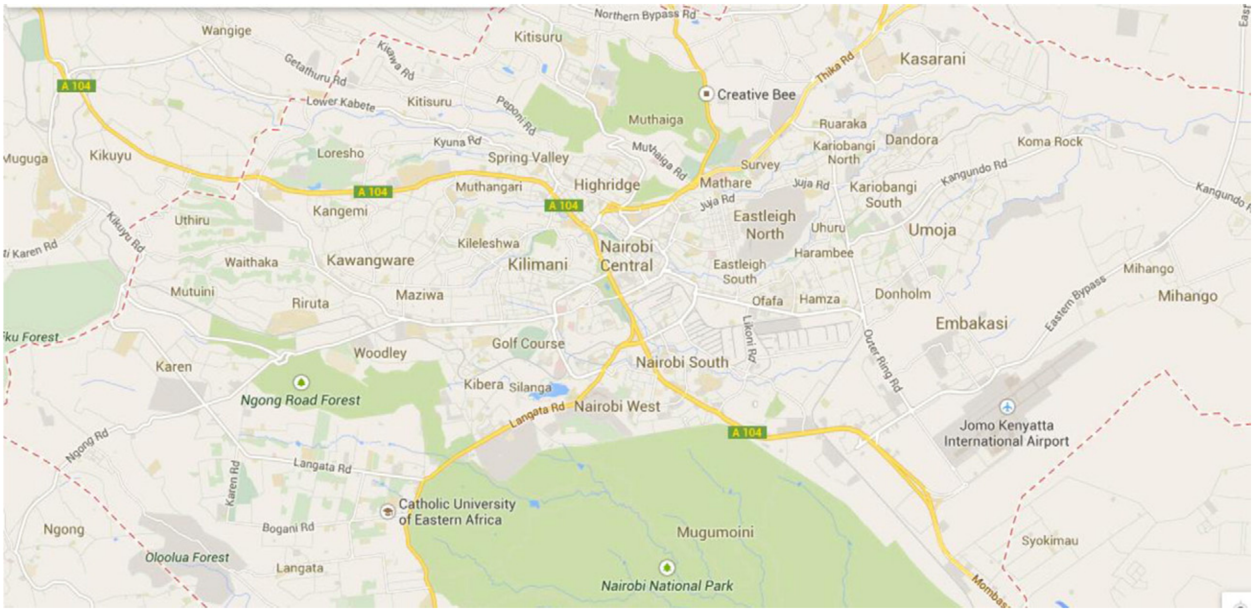
This work would not have been possible without the critical thinking and hard work of the following University of Nairobi, School of Computing & Informatics students: Ikamar Ekessa, Peter Kamiri, Samuel Kariu, Maureen Mbinya, Mureri Ntwiga. Researchers at MIT's Civic Data Design Lab also contributed significantly to the work including; Jonathan Andrew Campbell, Emily Eros, Alexis Howland, Lindiwe Rennert, Alicia Rouault, Christopher Van Alstyne, Catherine Vanderwaart. Special thanks to Wenfei Xu from the Civic Data Design lab who was instrumental in the development of the Matatu Map. 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 into this work. We also greatly benefitted from conversations that compared projects with Prof. Chris Zegras, Albert Ching, Stephen Kennedy, Neil Taylor, and Kevin Webb. Last, but not least, we thank James Gachanja and Dr. Zachary Gariy at KIPPRA for hosting the workshops and the Kenya Alliance of Resident Association for their work on the launch of the data and map for this work helped our thinking, made connections, and showed us a way forward in sustaining data collection work.

Appendix A. Route coding structure developed for GTFS

Nairobi's transit routes largely fall along the major road corridors. The team gave each corridor a numeric identifier and used that as the basis for the unique identification system developed for the GTFS data (see Fig. 1). A different alpha-numeric identification code was then developed for the routes, stops, schedules, and shape files that are part of the GTFS data structure. The codes included metadata about each data point collected, to help maintain knowledge gained about the system during the data collection process. The coding structure methodology was developed in a way that would easily allow new stops and routes to be added to the data over time. The development of the identification

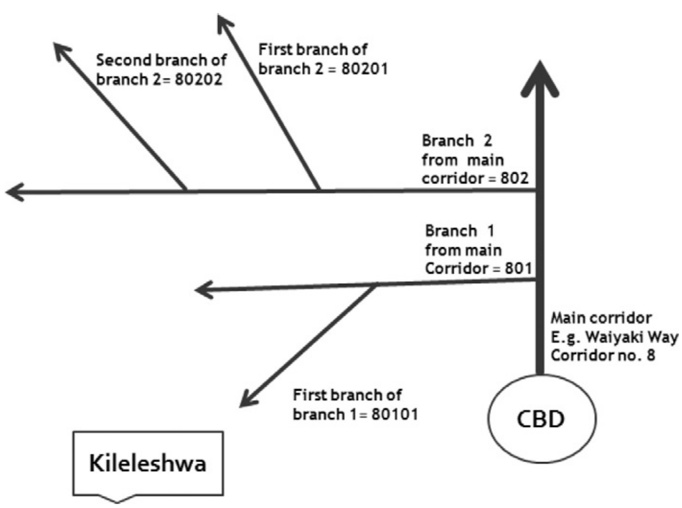


Appendix Fig. 1. The above image shows how we broke down Nairobi's matatu system into a series of corridors.



Appendix Fig. 2. Here you can see Nairobi's main corridors.

system is detailed below.Appendix Fig. 3 Route coding: The figure below illustrates the route coding based on our protocol.



Routes coding structure

In the GTFS file structure, the routes file includes a unique identifier for the route. It also includes route short and long name as well as a description of the type of route (e.g., bus, rail, subway). Each matatu route falls along a major corridor. The first digit in the route unique identifier is the number representing that corridor (see Fig. 1). Then as the matatu route branches off that corridor, it is given a series of numeric identifiers representing its place along the branch. Therefore, the second two digits in the identifier represent the second level branch. The next four digits represent the alpha-numeric characters developed for unique route numbers. The next digit represents whether it was a designated (1) or non-designated (0) route. The final digit represented within it was an inbound (1) or outbound (0) route. See Fig. 3 for how the route branching structure works.

Route coding: corridor|1st level branch|2nd level branch|route no.|gazetted|direction

E.g. route code for route 48 is: 8|01|01|0048|1|1.

Operates from Odeon terminus (in CBD) through Riverside Drive to Kileleshwa (along Waiyaki Way — Corridor 8)

Stops coding structure

The GTFS data structure for the stops includes a stop unique identifier, the stop name, latitude and longitude information for each stop, along with the stop type and a determination of whether it had a parent location.

The first digit in the stop unique identifier represented the name of the main corridor. When the stop was designated (1) or undesignated (0), the next digit represented within it was an inbound (1) or outbound (0) route. The next three digits were character abbreviations of the stops. For example "WST" for Westlands.

Stops coding simplified:

Bus stop coding: corridor|designation|direction|stop name

Example of a bus stop code: 08|1|1|AAA

Shape coding structure

In the GTFS data structure, the shape file recreates the path of the route. It includes a numeric identifier for the route, and a series of latitude and longitude points and a sequence numbers so the routes can be drawn in various software packages.

The first digit of the shape unique code is the corridor number. The next four digits represent the alpha-numeric characters for the route or the route's short name. The next digit represents the origin, or what we called Level: 1 — for matatus originating from the main terminus, 2–9 — for matatus not originating from the main terminus. The next number represents the route variation. Many routes vary slightly at the end, and this would indicate that variation. Examples 2 and 3 below show different variations on the same route.

(c) The shape file coding

Shape coding is made up of:

corridor|route no.|route level|route variation no.

Using two examples to illustrate this coding.

Example 1 – Karen route 5|0024|1|1 – originates at Ambassador to Karen through Langata Road

Example 2 – Karen route 5|0024|2|1 – originates at Bomas (Galleria) to Karen through Karen "C"

Example 3 – Karen route 5|0024|2|2 would represent a route originating at Bomas (galleria) to Karen through Hardy Shopping Centre (a variation on the original route).

Appendix B. Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.jtrangeo.2015.10.005. These data include the Google map of the most important areas described in this article.

References

Aker, Jenny C., Mbiti, Isaac M., 2010. Mobile phones and economic development in Africa. Center for Global Development Working Paper 211.

Barcelo, Jaume, Kuwahara, Masao, Miska, Marc, 2010. Traffic data collection and its standardization. Int. Ser. Oper. Res. Manag. Sci. 44, 144.

Behrens, Roger, McCormick, D., Mfinanga, D.A., 2012. An evaluation of policy approaches to upgrading and integrating paratransit in African urban public transport systems: results of the first round of a Delphi survey. *CODATU XV*: *The role of urban mobility in* (*re*) *shaping cities*.

Byrd, A., Emory, D., Webb, K., 2012. Open trip planner analyst: open source, open data driven transportation network analysis. Poster P13 - 6039, Session 616 (Open Data in Transportation). Presented at the Transportation Research Board Annual Meeting 2012, Washington, DC.

Caceres, Noelia, Romero, Luis M., Benitez, Francisco G., del Castillo, Jose M., 2012. Traffic flow estimation models using cellular phone data. Int. Transp. Syst. IEEE Trans 13 (3), 1430–1441.

Catala, Martin, 2011. Expanding the Google Transit Feed Specification to Support Operations and Planning. Final Report Contract Number BDK85 977–15. University of South Florida (November 15, Last accessed 6/11.2015 from: http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_PTO/FDOT_BDK85_977-15_rpt.pdf).

Cervero, Robert, 2000. Informal transport in the developing world. UN-HABITAT.

Cervero, Robert, Golub, Aaron, 2007. Informal transport: a global perspective. Transp. Policy 14 (6), 445–457.

Ching, Albert C., Zegras, Stephen Kennedy, Mamun, M., 2013. A user-flocksourced bus experiment in Dhaka: New data collection technique with smartphones. Transp. Res. Rec. J. Transp. Res, Board.

Choi, Keechoo, Jang, Wonjae, 2000. Development of a transit network from a street map database with spatial analysis and dynamic segmentation. Transp. Res. Part C Emerg. Technol. 8 (1), 129–146.

Eros, Emily Jean, 2014. Transportation data as disruptive innovation in Mexico City (Ph.D. diss.) Massachusetts Institute of Technology.

Eros, Emily, Mehndiratta, Shomik, Zegras, Chris, Webb, Kevin, Ochoa, Maria, 2014. Applying the general transit feed specification to the Global South: experiences in Mexico City, Mexico—and beyond. Transp. Res. Rec. J. Transp. Res. Board 2442, 44–52.

Farzin, Janine M., 2008. Constructing an automated bus origin–destination matrix using farecard and global positioning system data in Sao Paulo, Brazil. Transp. Res. Rec. J. Transp. Res. Board 1, 30–37 (2072).

Gonzalez, Marta C., Hidalgo, Cesar A., Barabasi, Albert-Laszlo, 2008. Understanding individual human mobility patterns. Nature 453 (7196), 779–782.

Google Developers, 2014. What is GTFS? Last accessed 6/11/2015 https://developers.google.com/transit/gtfs/

Guillen, Marie Danielle, Cordova, Lorenzo, 2012. Mapping informal public transport terminals: the case of tricycles of Brgy. 176, Bagong Silang, Caloocan City. In: CODATU XV The role of urban mobility in (re) shaping cities.

Hadas, Yuval, 2013. Assessing public transport systems connectivity based on Google Transit data. J. Transp. Geogr. 33, 105–116.

Herrera, Juan C., Work, Daniel B., Herring, Ryan, Xuegang Jeff Ban, Quinn Jacobson, Bayen, Alexandre M., 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. Transp. Res. C. Emerg. Technol. 18 (4), 568–583.

Hersman, Erik, 2012. Mobilizing tech entrepreneurs in africa (innovations case narrative: iHub). Innovations 7 (4), 59–67.

ITDP, UN-Habitat, 2014. Nairobi Ndovu/A104 Bus Service Plan Last accessed 22/9/2015 https://www.itdp.org/wp-content/uploads/2015/02/Nairobi-Ndovu-A104-BRT-Service-Plan.pdf.

Kemei, K., 2014. Move on night transport won't end accidents, says former TLB Chairman Hassan Kamwaro. East African Standard (January 3, Last accessed 6/11/2015 http://www.standardmedia.co.ke/?articleID=2000101376&story_title=kamwaro-move-on-night-transport-won-t-end-accidents).

Kenya National Bureau of Statistics, 2015. Economic Survey of Kenya. Last Accessed 6/11/15 https://www.opendata.go.ke/Economy-and-Finance/Economic-Survey-2015/yvq8-75uz.

Klopp, Jacqueline M., 2012. Towards a political economy of transportation policy and practice in Nairobi. Urban Forum 23 (1), 1–21.

Klopp', Jacqueline M., Mitullah, Winnie, 2015. Politics, policy and paratransit: a view from Nairobi. In: Behrens, Roger, McCormick, Diane, Mfinanga, David (Eds.), Paratransit for African cities (forthcoming Routledge).

Klopp, Jacqueline M., Sarah, Williams, Waiganjo, Peter, Orwa, Dan, White, Adam, 2015. Leveraging cellphones for wayfinding and journey planning in semi-formal bus systems: lessons from Digital Matatus in Nairobi. Planning Support Systems and Smart Cities. Springer.

Kramers, Anna, 2014. Designing next generation multimodal traveler information systems to support sustainability-oriented decisions. Environ. Model Softw. 56, 83–93.

Lee-Gosselin, Martin E.H., Buliung, Ron N., 2012. The role of ICTs in the transformation and the experience of travel. Transportation 39 (5), 873–876.

Mbiti, Isaac, Weil, David N., 2011. Mobile banking: the impact of M-Pesa in Kenya No. w17129 National Bureau of Economic Research.

McHugh, Bibiana, 2013. Pioneering open data standards: The GTFS Story. In: Goldstein, Brett, Dyson, Lauren (Eds.), Beyond transparency: open data and the future of civic innovation.

McLeod, David, 2013. SA losing to Kenya in tech race. TechCentral (9 June. Last accessed 6/1/2015 http://www.techcentral.co.za/sa-losing-to-kenya-in-tech-race/40955/).

Mutongi, K., 2006. Thugs or entrepreneurs: perceptions of 'Matatu' operators in Nairobi, 1970 to the present. Afr. J. Int. Afr. Inst. 76 (4), 549–568.

Ratti, Carlo, Frenchman, Dennis, Pulselli, Riccardo Maria, Williams, Sarah, 2006. Mobile landscapes: using location data from cell phones for urban analysis. Environ. Plan. B 33 (5), 727–748.

Republic of Kenya. Report of the national task force on police reforms, 2009. Safaricom. Industry Update.12 March 2009. Last accessed 6/5/2015 http://www.safaricom.co.ke/fileadmin/template/main/downloads/investor_relations_pdf/Industry % 20Update%20120309.pdf.

Safaricom, 2009. Industry Update (12 March, Last accessed 6/5/2015).

Salon, Deborah, Aligula, Eric, 2012. Urban travel in Nairobi, Kenya: analysis, insights, and opportunities. J. Transp. Geogr. 22, 65–76.

Salon, Deborah, Gulyani, S., 2010. Mobility, poverty, and gender: travel 'choices' of slum residents in Nairobi, Kenya. Transp. Rev. Transnatl. Transdiscipl. J. 30.5, 1464–5327.

Sussman, J.M., 2005. Perspectives on intelligent transportation systems (ITS). Springer, Cambridge, MA.

Talbot, David, 2013. African Bus routes redrawn using cell-phone data. MIT technology review (30 April, Last accessed October 6, 2014. http://www.technologyreview.com/news/514211/african-bus-routes-redrawn-using-cell-phone-data/).

Tao, Sui, Rohde, David, Corcoran, Jonathan, 2014. Examining the spatial–temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap. J. Transp. Geogr. 41, 21–36.

Thakur, R., Sharma, M., 2009. GIS and challenges to planning and development applications in peripheral regions. In: Gatrell, J.D., Jensen, R.R. (Eds.), Planning and socioeconomic applications. Springer, pp. 125–137.

Thiagarajan, Arvind, Biagioni, James, Gerlich, Tomas, Eriksson, Jakob, 2010. Cooperative transit tracking using smart-phones. Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems. ACM, pp. 85–98.

Townsend, Anthony M., 2013. Smart cities: big data, civic hackers, and the quest for a new utopia. WW Norton & Company.

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 (Last accessed 6/11/2015 (http://www.uitp.org/sites/default/files/cck-focus-papers-files/Transafrica_UITP_UATP_PublicTransport_in_SubSaharan_Africa_2008.pdf)).

Wakefield, Jane, 30 Apr. 2013. Mobile phone data redraws Bus routes in Africa. BBC (Last Accessed 9/30/2014: http://www.bbc.com/news/technology-22357748).

Wang, Huayong, Calabrese, Francesco, Di Lorenzo, Giusy, Ratti, Carlo, 2010. Transportation mode inference from anonymized and aggregated mobile phone call detail records. In Intelligent Transportation Systems (ITSC), 13th International IEEE Conference on pp. 318–323 IEEE.

Williams, Sarah, Marcello, Elizabeth, Klopp, Jacqueline, 2014. Open source Nairobi: creating and sharing a GIS database of Nairobi for community empowerment. Ann. Assoc. Am. Geogr. 104.

Wong, James, 2013. Leveraging the general transit feed specification for efficient transit analysis. Transp. Res. Rec. J. Transp. Res. Board 2338 (1), 11–19.

Woolf, S.E., Joubert, Johan W., 2013. A people-centred view on paratransit in South Africa. Cities 35, 284–293.

World Bank and AusAID, 2014. *An Overview of Open Transit Data in East and Southeast Asia* Washington DC.

Zegras, P. Christopher, Eros, Emily, Butts, Kuan, Resor, Elizabeth, Kennedy, Stephen, Ching, Albert, Mamun, Muntasir, 2014. Tracing a path to knowledge? Indicative user impacts of introducing a public transport map in Dhaka, Bangladesh. Camb. J. Reg. Econ. Soc. rsu028.

Zhang, Junyi, Li, Gang, Nugroho, S.B., Fujiwara, Akimasa, 2013. Paratransit-adaptive transportation policies for transition to sustainability in developing countries. Sustainable Transport Studies in Asia. Springer, Japan, pp. 137–166.