

# Digital Matatus: Using Mobile Technology to Visualize Informality

“Big data” will not transform the world unless it is collated and synthesized into tools that can activate change.

## INTRODUCTION

Access to data is essential for generating urban plans for rapidly developing cities in the global south, however it is often challenging to acquire. This is because it either does not exist or is tightly controlled by those in power.[1] However, data does exist—locked away in records of everyday interactions such as cell phones and credit card transactions. My research team, Digital Matatus, set out to test whether it is possible to capture this data by leveraging the ubiquitous nature of mobile technologies in developing countries, to collect data on an essential infrastructure, and open that data up for anyone to use. The results show that developing strategies that allow the public to interact with data through the creation of civic tools and visualizations helps to improve civic life while also helping to generate new planning strategies.

Focusing on Nairobi’s semi-formal bus system, or Matatus as they are commonly known, the project was not only successful in generating data but also in transforming it into tools that could be acquired by civic actors to develop urban change. Distributed using the open data standard GTFS, the local technology community used the data to develop mobile phone routing applications. Non-governmental Organizations (NGOs) used it to develop Bus Rapid Transit Plans (BRT) plans. The downloadable paper transit map allowed Nairobi’s residents and the government to visualize, for the first time, the comprehensive system that serves their city, which generated a debate about its future amongst the various stakeholders. The implications of the research go beyond Kenya as it shows how we can leverage the power of “Big Data” to generate urban change through active data collection, data sharing and the development of visualizations and civic tools with that data.

**SARAH WILLIAMS**

MIT

**JACQUELINE KLOPP**

Columbia University

**DANIEL ORWA & PETER WAIGANJO**

University of Nairobi

**ADAM WHITE**

GROUPSHOT



1

Figure 1: Matatu bus in Nairobi. Matatu operators develop elaborate designs on their vehicles to attract customers. Image Credit: Sarah Williams

## THE NEEDS FOR DATA IN DEVELOPING COUNTRIES

Data is essential for developing infrastructure models, and one of the biggest challenges for transportation systems is acquiring data that accurately represents these systems. [2] Data for para-transit or semi-formal bus systems rarely exist. [3] This is particularly true in developing countries and makes it difficult for residents to know how to navigate their city, and, more importantly, makes it difficult for cities to know how to route and manage traffic flows.

Semi-formal transit, including Nairobi's Matatus, often constitute the backbone of mass transit for the majority of citizens in the rapidly growing cities of Africa, Asia, and Latin America. [4] This is true in Nairobi, where the Matatus system is the main form of transit [5] serving over 3.1 million people in Nairobi metropolitan area on 130 routes. The Matatu system is run by hundreds of privately-owned companies licensed by Nairobi's City Government. Matatus range from small vans that can hold up to 9-14 passengers to 32-seat vehicles. [6] Matatus often clog the already overstrained roadways and knowledge of the routes and schedules are essential for developing models of traffic flows as they represent a significant portion of the vehicles on the roadway.

Collecting data on semi-formal bus systems is challenging because hundreds of different operators individually set route and schedules. Governments are often unwilling to collect data on these systems as they find them too "chaotic" or do not want to sanction their existence by providing data. [7] Operators of semi-formal systems often collect data about the individual routes they provide, but the data is rarely shared with others since the information is primarily used to optimize their own business operations. [8] Often in analog form, these data sets are typically incomplete and hard to use for planning purposes.

At the onset of the project, data on Nairobi's Matatus system was incomplete and difficult to use for planning purposes. The City Council records only identified 90 of the 130 routes. These government files, only noted the start and end location of the Matatus. Which made it difficult to know how the Matatu navigates the city and where a passenger might be able to find a stop. The research team found a stylized paper based map created in 2010 by Kenya Buzz, a Nairobi-based media company. [9] While informative, the map was no longer publicly available, and the data used to create it was never released. While performing our work we found several incomplete sources of the data collected by private actors; however, none were usable for developing a comprehensive data set. [10]

## USING MOBILE TECHNOLOGY TO COLLECT TRANSPORT DATA

Research studies in which cell phone users actively collect data on their mobile devices have greater potential for use in planning urban infrastructure as the collector owns the data and can decide how it will be distributed. Investigations into the geo-locative capabilities of cell phone records to develop data on transit flows have been successful, but acquiring this data from telecommunication companies is extremely difficult because of privacy concerns. [11] Other research projects have used the geo-registration provided by mobile devices to capture transit data on formal systems, but the majority of these projects look at capacity and transit flows. [12] Often used to evaluate existing routes or identify unserved communities they rarely identify the location of transit infrastructure

Few projects use cell phones to capture the route location of the transit infrastructure by asking study participants to crowd source the location of vehicles in real-time on formal transit systems. [13] Some formal transit agencies have

2

modeled routes by attaching GPS devices to transit vehicles.[14] Several Matatu operators in Nairobi have tested this method. However, the data is limited to the routes they operate, is not consistent, and is rarely shared. [15] Active collection projects have been more successful for planning purposes but have not been applied to semi-formal systems in a comprehensive way, our research project set out to test the ability to collect data for an entirely unmapped system.

### THE UBIQUITOUS USE OF CELL PHONES IN NAIROBI

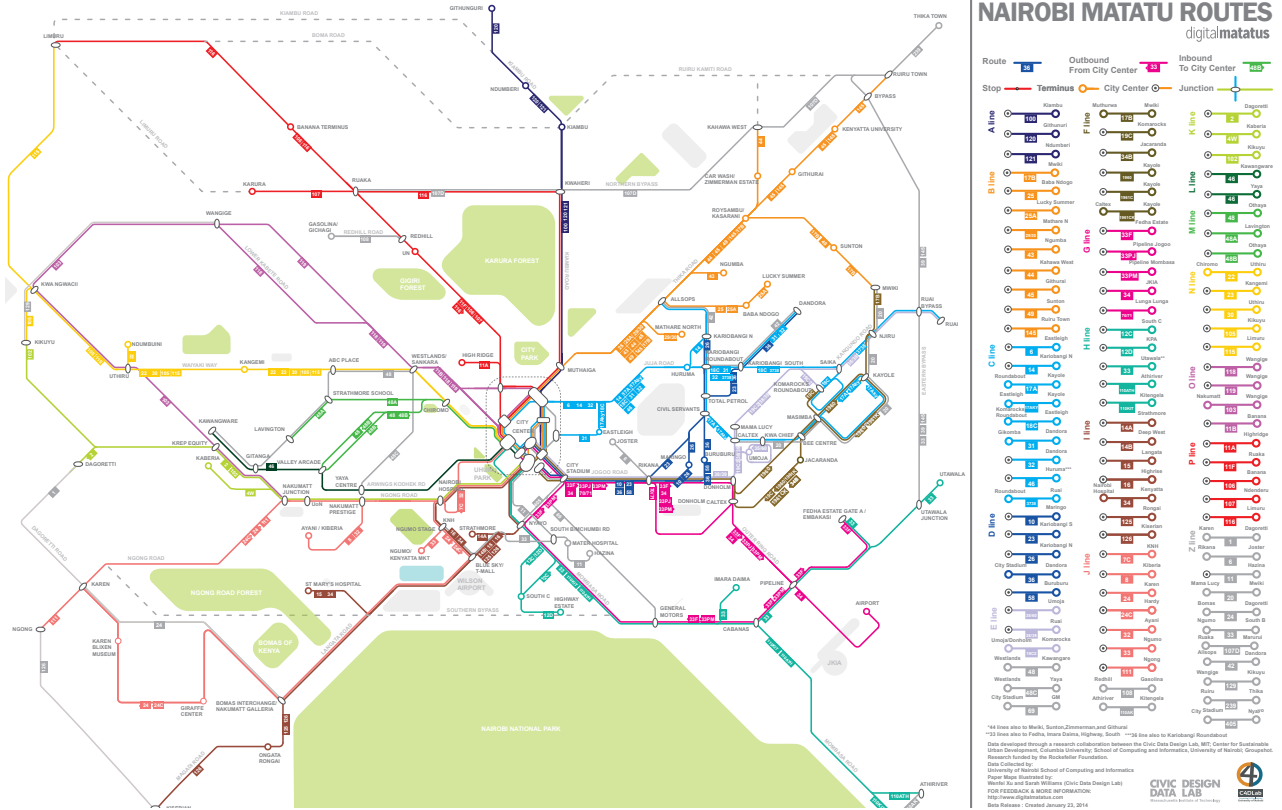
Cell phones are widely used in Nairobi. They provide essential services such as buying products, transferring funds, and providing health education. The development of these kinds of cell phone applications has exploded in Africa, and Kenya has become a center for some of these developments. [16] One of the biggest exports is the M-Pesa application, which allows users to send cell phone minutes for payment of everything from a cup of coffee to tickets to a movie. M-Pesa was started in 2007 by Safaricom, the local telecommunications network. After two years of operation, M-Pesa had adopted 8.5 million Kenyan users and transferred US \$3.7 billion in funds (equivalent to 10 percent of Kenyan's GDP) through the system. [17] In sub-Saharan Africa, there are ten times as mobile phones as mobile phones and between 2002 and 2007 cell phone subscriptions have increased by 49 percent annually while in Europe the increase has only been 17 percent. [18] Our research project sought to test whether we could leverage the widespread use of cell phones in cities like Nairobi to collect essential data for planning.



### COLLECTING THE DATA ON THE MATATU SYSTEM

While data collection is still ongoing, the initial process was performed during 2012-2013 in coordination with our research team, which was comprised of a partnership between the University of Nairobi School of Computing and Informatics, Columbia Universities Earth Institute, and Groupshot. The process involved identifying the current routes; developing technology to collect

Figure 2: Image of University of Nairobi student collecting data on a Matatu.  
Image Credit : Adam White



3

Figure 3: Front of the official Matatu map of Nairobi developed by the project team.  
Image Credit : Sarah Williams and Wenfei Xu

geo-registered data on routes and stops; translating the data in the General Transit Feed Specification (GTFS); developing unique numeric identifiers necessary for the specification; validating the data with the local transit community; and ultimately openly releasing the data by posting it on the GTFS Exchange.

Working with our team at the University of Nairobi, we tested and modified various mobile applications for data collection. Tool development was based in Nairobi to ensure that the technology would work on local mobile devices and cell networks and would remain a local resource. The team modified an open source mobile application to perform the data collection, developing a unique interface. [19] The application allowed University of Nairobi students riding the Matatus to collect latitude and longitude data trails of the Matatu routes. When the Matatus made a stop, they used the mobile tool to mark the stop and generate a unique identifier for use in the GTFS data protocol. Students collected the data by riding the Matatus, but in some of the more dangerous areas in Nairobi they followed the Matatus in a private car.

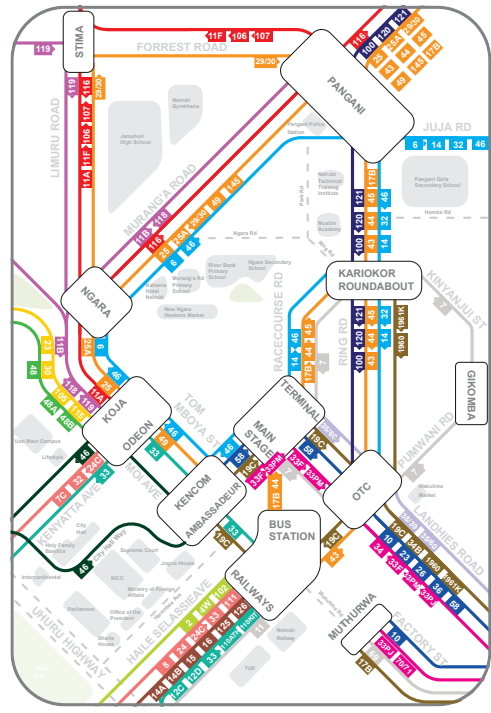
Once collected, the route data was translated into the GTFS data specification. This specification was chosen to help facilitate the development of mobile applications since many open-source transit applications already use this data specification as their base. The GTFS data specification was developed in 2005 by Portland's Tri-Met transit agency in partnership with Google. [20] This format is now widely used by transit agencies across the world to openly share their data and is the basis for transit routing directions provided by Google Maps. The openness with which GTFS data is distributed has encouraged its use for transit applications beyond trip planning by creating tools to improve transit operations and planning overall. [21] Once transformed into this common standard, the data was

4

## MATATU STATION LIST

ABC	59	53	55	56	57						Kariokor Roundabout	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522														
Adams	29	30	31	32	33	34					Kariokor Roundabout	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522														

## CITY CENTRE



uploaded to the GTFS exchange, a common repository for openly sharing GTFS data; anyone can use the site to download the data our team created for Nairobi. [22]

One of the biggest challenges of the project was modifying the GTFS specification to semi-formal transit since it was initially developed for more formalized systems with less chaotic scheduling and fare changes. However, storing the data in GTFS proved beneficial as it allowed the local technology community to easily generate mobile applications as detailed information was provided for the protocol.

### VISUALIZING THE DATA: PUTTING THE MATATU SYSTEM ON THE MAP

The GTFS data, when plotted in the map in its raw form, was too complex to read. The mess of latitude and longitude points representing overlapping routes and stops was hard to comprehend. Order was given to this complex network using the visual language of traditional subway maps seen in Paris, London, and New York. The map structured the Matatu data, and the visual cues provided by the style of organized transit systems made map readers contextualize Nairobi's Matatus as one organized system.

Maps are powerful as they can expose a system or idea by clearly representing it, thereby giving it a form of legitimacy. [23] Nairobi's Matatus mostly follow the few main road corridors heading out of the city's center with routes splitting off onto smaller tributaries as they move away from the city. In the map, each corridor's routes were grouped and color coded, as seen on more formal transit systems. Identifying major stops and points of interest including parks, airports, and landmarks allowed users to position themselves on the visualization.

Figure 4: Back of the official Matatu map of Nairobi developed by the project team. This image shows the location matrix which allowed users to more easily find the route they needed. Image Credit : Sarah Williams and Wenfei Xu

A legend, inset map, and matrix that matched routes with landmarks was included in the stylized map. The legend helped to associate colors with route numbers and to indicate the start and end location of all routes. The complexity of the Matatu stations in the city center made an inset map of this area essential. The matrix became the most useful addition to the composition as it mimicked the way people navigate the city—using landmarks instead of numeric addresses. The Matrix was constructed by listing the name of every landmark in the city and the Matatu routes that went to them. This allowed users to find their route in much the same way they would ask for directions.

#### **OPERATIONALIZING THE DATA:**

Transforming data into a meaningful tool for urban development needs civic actors that trust the data and can see potential for its use. Data development projects are useful, but it is not until the data is leveraged for change that it generates value. Throughout the data collection process, the research team engaged the local transit community through a series of workshops. These outreach activities sought advice from the various organizations, allowed them to participate in the data collection, provided information on how the data would be released, and ultimately created trust in the data and its creation so that it could be readily adopted.

Workshops were held in coordination with the Kenya Institute for Public Policy Analysis (KIPPRA), Kenya's primary government think tank tasked with transport modeling. Workshop participants included members of the government, academia, the head of the Matatus Owners Association, Matatu operators and drivers, NGOs, and the local technology community. The University of Nairobi also held a "hackathon" using the data to encourage the development of civic application by the private sector technology community. [24] This collaboration was successful in gaining support from the government, the Matatus owners and, perhaps more importantly, for engaging the technology community in developing applications that use the data.

One of the most powerful moments of the stakeholder collaboration was our team's presentation of the stylized transit map. The transit community was able to see the data we collected with their own eyes, and the visualization instantly generated a debate about its use. The Matatus operators used the maps as a planning tool. Conversations with them helped us to identify routes that we missed in the collection process. However, more time was spent on discussing under served neighborhoods, which were clearly identified in the visualization, and the potential for developing new routes in the areas.

When the Kenyan government saw the data visualized, they began to think of themselves as co-owners and made the visualization the official Matatu map of the city. Until the maps were presented, government officials appeared disinterested in the data collection process and primarily played a listeners' role in the workshops. The visualization allowed them to see the potential of the data to create positive interaction with the public. The Ministry of Transport held a press conference releasing the visualization as the official Matatu map of the city. The forum created a debate between the government, the transport community, and the public about the future of the Matatus system. Participants questioned the government about how they would respond to necessary services changes using the map as a tool. After the press event, the downloadable maps went viral on social media and large versions were printed in local newspapers allowing

everyone to access the data we created.

The local technology community saw the potential of the data before it was visualized and developed two mobile applications using the data before the map was officially released. Laban Okune developed Ma3Route, a mobile app that provided routing information using the data as a base. The tool allows users to share real-time, crowd-sourced data about Matatu route changes, accidents, and traffic congestion. Ma3Route is widely used in Nairobi and won the Kenya’s Vision 2030 ICT Innovation award in April of 2014. Sonar is another routing application developed by Jeremy Gordon of Flashcast.

UN-Habitat/Institute for Transportation and Development Policy (ITDP) used the data as the basis for creating the Bus Rapid Transit (BRT) Service Plans for the city. The data allowed them to identify the location of existing routes along potential BRT service lines proposed for the Thika highway corridor. It also allowed them to identify the location of overlapping routes, showing the potential for other BRT lines in the city.

The conversations generated during the engagement process allowed stakeholders to come together, edit the data, and generate a debate on how the data could be leveraged to create change in the city. These conversations were essential for operationalizing the data to develop important changes to an infrastructure most citizens depend upon.



5

### GIVING MEANING TO DATA

Leveraging “big data” to develop informed urban strategies often needs active data collection, data sharing, and ultimately the synthesis of data into visualizations and civic tools. The Digital Matatus project shows that anyone can leverage mobile technology to capture hard-to-find data, transform that data into representations that can expose informal systems, and use the data and its visualizations to help develop conversations for the future of infrastructure development

Figure 5: Identify locations for new Matatu routes with transit stakeholders in Nairobi.  
Image Credit : Adam White

in developing countries.

Collecting data outside formal channels is often essential for developing strategies for the creation of urban infrastructure. While data may exist in the hands of the government, cell phone operators, or private organizations, it is often hard to acquire this information. This is particularly true for informal systems. Semi-formal transit systems are the primary form of transit in most cities in the developing world, and the tools and methodologies developed through this research project can be used to capture data anywhere these systems exist.

Data visualizations are powerful vehicles for generating debate and evidence for planning strategies. The stylized transit maps developed by the project allowed the government to engage in conversations with the public. Nairobi's Matatu operators, who are currently the de-facto planners of the systems, used the map to identify and develop new routes for the system. Most importantly, the citizens of Nairobi now have vital information for navigating their city.

Development and dissemination of data needs to be inclusive and open in order to allow multiple civic actors to trust the data for their research. By engaging the transit community and opening the data up for anyone to use, the project was able to generate new products and tools for planning, including the development of mobile applications for citizens and the development of new transportation projects in the form of BRT lines.

This unprecedented growth of data has generated excitement about using it to reshape the way we live. [25] However, "Big Data" will not change the world unless it can be acquired and transformed into visualizations and civic tools that can be operationalized by governments, designers, and other civic actors to advocate and develop urban change. The Digital Matatus project provides an example of how we can use big data to develop necessary changes to the urban environment.

## ENDNOTES

1. Williams, S., Marcello, E., & Klopp, J. M. (2014). Toward Open Source Kenya: Creating and Sharing a GIS Database of Nairobi. *Annals of the Association of American Geographers*, 104(1), 114-130.
2. Herrera Juan, Daniel B. Work, Ryan Herring. Xuegang (Jeff) Ban, Quinn Jacobson and Alexandre Bayen. "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile century field experiment" *Transportation Research Part C* 18 (2010): 568-583.
3. Ching, Albert, et al. "A User-Flocksourced Bus Experiment in Dhaka: New Data Collection Technique with Smartphones." *Transportation Research Record: Journal of the Transportation Research Board* (2013).
4. Cervero, Robert, and Aaron Golub. "Informal Transport: A Global Perspective." *Transport Policy* 14.6 (2007): 445-57.
5. Salon, Deborah and Eric Aligula. "Urban Travel in Nairobi, Kenya: Analysis, Insights, and Opportunities." *Journal of Transport Geography* 22 (2012): 65-76.
6. Mutongi, K. "Thugs or entrepreneurs: Perceptions of 'Matatu' operators in Nairobi, 1970 to the present." *Africa: Journal of the International African Institute* 76.4 (2006): 549-568.
7. Cervero, Robert, and Aaron Golub. "Informal Transport: A Global Perspective." *Transport Policy* 14.6 (2007): 445-57.

## ACKNOWLEDGEMENTS:

Digital Matatus is a collaborative project. Jacqueline Klopp from Columbia University's Center for Sustainable Urban Development brought the team together. Adam White, of Groupshot, helped context the team to Nairobi's local technology community. Peter Waiganjo and Daniel Orwa Professors of the University of Nairobi's School of Computing and Informatics and Researcher at C4DLab directed the project in Nairobi. This work could not have been performed without the students who worked under their direction. Special thanks to student coordinators in Nairobi including Peterson Kamiri, Samuel Kariu, Jack Mutua. 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 Roualt, 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. Finally, I would like to thank Benjamin de la Peña, who was our project officer at the Rockefeller Foundation and supported our work finically but also intellectually. This project would not have been possible without the funding provided by the Rockefeller Foundation.



- Kemei, K. (2014) 'Move on night transport won't end accidents, Says Former TLB Chairman
- Hassan Kamwaro' East African Standard January 3, 2014. Retrieved from: <[http://www.standardmedia.co.ke/?articleID=2000101376&story\\_title=kamwaro-move-on-night-transport-won-t-end-accidents](http://www.standardmedia.co.ke/?articleID=2000101376&story_title=kamwaro-move-on-night-transport-won-t-end-accidents)>
- 8 Eros, Emily Jean. Transportation data as disruptive innovation in Mexico City. Diss. Massachusetts Institute of Technology, 2014.
  - 9 Last Accessed 10/9/2014 (<http://www.globalpost.com/dispatch/kenya/100127/kenyas-wild-matatus-captured-map>)
  - 10 The "Living in Nairobi" blog published a highly stylized route map in 2012 after we had begun our work, but they also did not publish any of the data collected to create the map, and have not maintained the map. (Last Accessed 10/9/2014 (<http://www.jambonairobi.co.ke/services/public-transport/nairobi-route-maps/nairobi-matatu-route-map-eastlands/>))
 

Panga Safari, formally Matatus Online, developed a private matatu route database that covered 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 been expanded, and can be searched through a web interface, but the back-end data was never made public. Last Accessed 10/9/2014 (<http://www.matatuonline.com/>)
  - 11 Ratti, Carlo, Dennis Frenchman, Riccardo Maria Pulselli, and Sarah Williams. "Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis" *Environment and Planning B* 33.5 (2006): 727-748.
 

Gonzalez, Marta C., Cesar A. Hidalgo, and Albert-Laszlo Barabasi. "Understanding individual human mobility patterns." *Nature* 453.7196 (2008): 779-782.

Caceres, Noelia, et al. "Traffic flow estimation models using cellular phone data." *Intelligent Transportation Systems, IEEE Transactions on* 13.3 (2012): 1430-1441.

Talbot, David. "African Bus Routes Redrawn Using Cell-Phone Data" *MIT Technology Review* 30 April (2013) Web. 6 October 6 2014. <<http://www.technologyreview.com/news/514211/african-bus-routes-redrawn-using-cell-phone-data/>>

Wakefield, Jane. "Mobile Phone Data Redraws Bus Routes in Africa." *BBC* 30 Apr. 2013. Web. 30 Sept. 2014. <<http://www.bbc.com/news/technology-22357748>>.
  - 12 Caceres, Noelia, et al. "Traffic flow estimation models using cellular phone data." *Intelligent Transportation Systems, IEEE Transactions on* 13.3 (2012): 1430-1441.
 

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

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): 318-323.
  - 13 Thiagarajan, Arvind, et al. "Cooperative transit tracking using smart-phones." *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. ACM (2010).
  - 14 Farzin, Janine M. "Constructing an automated bus origin-destination matrix using farecard and global positioning system data in Sao Paulo, Brazil." *Transportation Research Record: Journal of the Transportation Research Board* 2072.1 (2008): 30-37.
  - 15 Interviews with Matatu drivers.
  - 16 Aker, Jenny, and Isaac Mbiti. "Mobile Phones and Economic Development in Africa." *Journal of Economic Perspectives* 24.3 (2010): 207-232.
  - 17 Safaricom. Industry Update. 12 March 2009. Web. 5 Oct. 2014. [http://www.safaricom.co.ke/fileadmin/template/main/downloads/investor\\_relations\\_pdf/Industry%20Update%20120309.pdf](http://www.safaricom.co.ke/fileadmin/template/main/downloads/investor_relations_pdf/Industry%20Update%20120309.pdf)

Mbiti, Isaac, and David N. Weil. "Mobile banking: The impact of M-Pesa in Kenya." *National Bureau of Economic Research No. 17129* (2011).
  - 18 UITP (International Association of Public Transport). *Public Transport in Sub-Saharan Africa*. Trans-Africa Consortium. (2010)
  - 19 We tested the accuracy of this tool with standard GPS units. Towards the end of the study, we tested the application *Transit Wand*, which had greater accuracy but the data was harder to work with after it was collected. Modifications to Transit Wand, might make it a better resources for the project. We are currently working toward creating a new tool for data collecting that modifies this open source resource.
  - 20 McHugh, Bibiana. "Pioneering Open Data Standards: The GTFS Story." in *Beyond Transparency: Open Data and the Future of Civic Innovation*. Edited by Brett Goldstein with Lauren Dyson: (2013).
  - 21 Catala, Martin. 2011. Expanding the Google Transit Feed Specification to Support Operations and Planning. Retrieved from: <[http://www.dot.state.fl.us/research-center/Completed\\_Proj/Summary\\_PTO/FDOT\\_BDK85\\_977-15\\_rpt.pdf](http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_PTO/FDOT_BDK85_977-15_rpt.pdf)>
 

Lee-Gosselin, Martin and Ron N. Buliung. "The role of ICTs in the transformation and the experience of travel" *Transportation* 39 (2012): 873-876.
  - 22 We developed several tools to transform the data into GTFS. They are not detailed here as they are very technical in nature including information about the use of a unique coding structure we developed for the data itself.
  - 23 Kitchin, Rob, and Martin Dodge. "Rethinking maps." *Progress in Human Geography* 31.3 (2007): 331-344.
  - 24 These workshops should not be confused with collaborative design workshops, rather they asked the key stakeholder to be active participants in the data collection process.
  - 25 Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.