SAIIE28 Proceedings, 25th - 27th of October 2017, Riverside Sun, South Africa © 2017 SAIIE

USING THE GTFS FORMAT TO IMPROVE PUBLIC TRANSPORT DATA ACCESSIBILITY IN GAUTENG

S. Koetsier[1*] & Q. van Heerden[2] & N.K. Maditse[3]

[1]Spatial Planning and Systems
Council for Scientific and Industrial Research, South Africa
skoetsier@csir.co.za

[2]Spatial Planning and Systems
Council for Scientific and Industrial Research, South Africa
qvheerden@csir.co.za

[3]Spatial Planning and Systems
Council for Scientific and Industrial Research, South Africa
kmaditse@csir.co.za

ABSTRACT

Public transport plays an important role in cities. It is a less expensive option than private transport and could reduce congestion and improve accessibility to jobs. However, ridership could be hampered when information regarding an operator's routes, schedules and fares is not readily available. Furthermore, public transport operators seem to maintain their data in different formats and not all data are updated frequently. From a planning perspective, this makes it difficult to determine the state of public transport at a provincial level. Converting public transport data into a standard format, such as the General Transit Feed Specification (GTFS) format, can be beneficial to operators as well as provincial officials, since it can be integrated in journey planning applications such as Google Maps. This increases accessibility of public transport information and can possibly increase ridership. In this paper we discuss the benefits of converting data into the GTFS format, benchmark the data of three public transport operators in Gauteng, and discuss the role that Industrial Engineers can play in using the data in the GTFS format to, for instance, optimise public transport networks.

*Corresponding author

1. INTRODUCTION

Public transport plays an important role in cities. It is a less expensive option than private transport and could reduce congestion and improve accessibility to jobs. However, ridership could be hampered when information regarding an operator's routes, schedules and fares is not readily available, outdated or incorrect. This is especially true in a developing country such as South Africa.

South Africa's public transport system consists of formal operators, such as municipal bus services, bus rapid transit services and commuter rail services, as well as informal operators in the form of mini-bus taxis. A large part of the population rely on public transport to get to work, therefore it is imperative that the public transport services be of high quality. Unfortunately, this is not the case, as the public transport industry in South Africa is plagued with problems ranging from unroadworthy vehicles, to overloading, to taxi associations fighting over territory.

Furthermore, public transport operators seem to maintain their data in different formats and not all data are updated frequently. From a planning perspective, this makes it difficult to determine the state of public transport at a provincial level, let alone improve or even optimise the network.

By making use of a data standard is useful in many ways. It removes ambiguity and allows data capturers to know exactly what data to capture and in which format to maintain these. If a standard is used across the industry, it allows much better control and analysis of the data, as generic analyses can be developed because the data format will always be the same for each operator. There are a number of public transport data formats used throughout the world. The *General Transit Feed Specification* (*GTFS*) [1] is a well-known example in the public transport space. *TransXChange* [2] is a United Kingdom national xml-based format used to maintain bus schedules and related information. Kauffman [3] explains that the *HAFAS Rohdatenformat* and *DIVA* are two proprietary public transport data formats that are widely used in Germany. In this paper, we focus on the GTFS format.

Many public transport operators in South Africa do not publish transit information and those that do, normally publish it in PDF format on their websites. This makes it extremely difficult for commuters to access the information. The use of the GTFS format can be beneficial to commuters, since it can be integrated in journey planning applications such as Google Maps or the recently released VayaMoja app [4] of the City of Johannesburg. This increases accessibility of public transport information and can possibly increase ridership.

Data standards are easier to implement in developed countries where data are more accessible. A standard such as GTFS could be extremely valuable in a developing country such as South Africa, both from a commuter perspective as well as a modelling perspective.

In this paper we discuss the benefits of converting data into the GTFS format, showcase a case study where two Gauteng Bus Rapid Transit (BRT) operators have been uploaded onto Google Maps, benchmark the data and operational situations of a number of public transport operators in Gauteng, and discuss the role that Industrial Engineers can play in using the data in the GTFS format to, for instance, optimise public transport networks or do transport modelling.

2. THE GENERAL TRANSIT FEED SPECIFICATION (GTFS)

GTFS, or General Transport Feed Specification, is an international format in which public transport agencies can maintain their data. GTFS was developed by Google in collaboration with TriMet, a public transport agency in Portland, Oregon [5]. The reference documentation is maintained and updated regularly by a number of transit agencies, developers and other stakeholders who use GTFS [6], as well as Google Developers [7]. Since its inception, there are over 400 GTFS feeds that have been made publicly available [8]. Thus far, Cape Town is the only city in South Africa that has adopted this trend by uploading their bus rapid transit system, MyCiTi, onto Google Maps [9].

When the GTFS format is used to maintain public transport data, it can be included in self-developed journey planning applications, or existing platforms such as Google Maps, the largest mapping site in the world. The route planner in Google Maps uses the existing road network and GTFS feeds to find directions between selected origins and destinations using public transport.

When public transport schedules and routes are published on Google Maps, visibility of the public transport service is increased, since the information is easily accessible by commuters. Seasoned riders may discover new

routes, private vehicle owners may be convinced to opt for using public transport instead, and tourists will be able to more easily travel in foreign countries. The contact details and websites of the operators are also published on Google Maps, ensuring that commuters have all the information regarding the service.

2.1 GTFS Static

A GTFS feed requires a collection of comma-separated text files, six mandatory and seven optional. For each file, there are required fields and optional fields. The required field names must always be present in the files, even if some entries have blank values for those fields. For detailed information on the fields, the reader is referred to the GTFS Reference website [1].

The required files contain details about the agency, routes, trips, stops, stop times, and calendar dates of service. The optional files include information about the fares, any calendar dates that deviate from the standard service, rules for transfers between routes, and the route shape information to ensure that routes displayed on Google Maps follow the curves of the road. A summary of the required files can be seen in Table 1 and the optional files are shown in Table 2. The relationships between the files can be seen in Figure 1.

Table 1: A summary of the required GTFS files.

File name	Purpose
agency.txt	Details of the public transport agencies providing the data for this GTFS feed, such as the name and contact details.
routes.txt	Details about the routes (also called lines) covered by each agency, such as the name of the route and mode of transport used on that route.
trips.txt	The trips for each route. A trip is a journey made by a public transport vehicle on a route. Each route generally has two trips, one in each direction. A trip can occur multiple times per day, at specific times of the day or at certain frequencies.
stops.txt	Geo-coded locations of stops where passengers board and alight. Stations (containing multiple stops) are also listed in this file.
stop_times.txt	The time that the vehicle arrives and departs at each stop for each trip.
calendar.txt	Periods during which the service is available. Some schedules may change on a monthly basis and the exact dates of operation can be specified in this file.

Table 2: A summary of the optional GTFS files.

File name	Purpose
calendar_dates.txt	Exceptions from the services specified in the calendar.txt file. For example, there may be public holidays that occur in the period specified in the calendar.txt file. If the service does not operate as usual on these public holidays, these dates can be specified in the calendar_dates.txt file.
fare_rules.txt	Fare rules are defined in this file, specifying the price and currency of each rule.
fare_attributes.txt	Fare rules are allocated to specific routes.
frequencies.txt	When trips do not have exact stop times, the frequencies file can be used to specify the frequency of service.
transfers.txt	Rules for making connections at transfer stations between routes.
shapes.txt	Rules for drawing lines on a map to represent routes.
feed_info.txt	Additional information about the feed, such as the organisation responsible for maintaining the feed and the feed version.

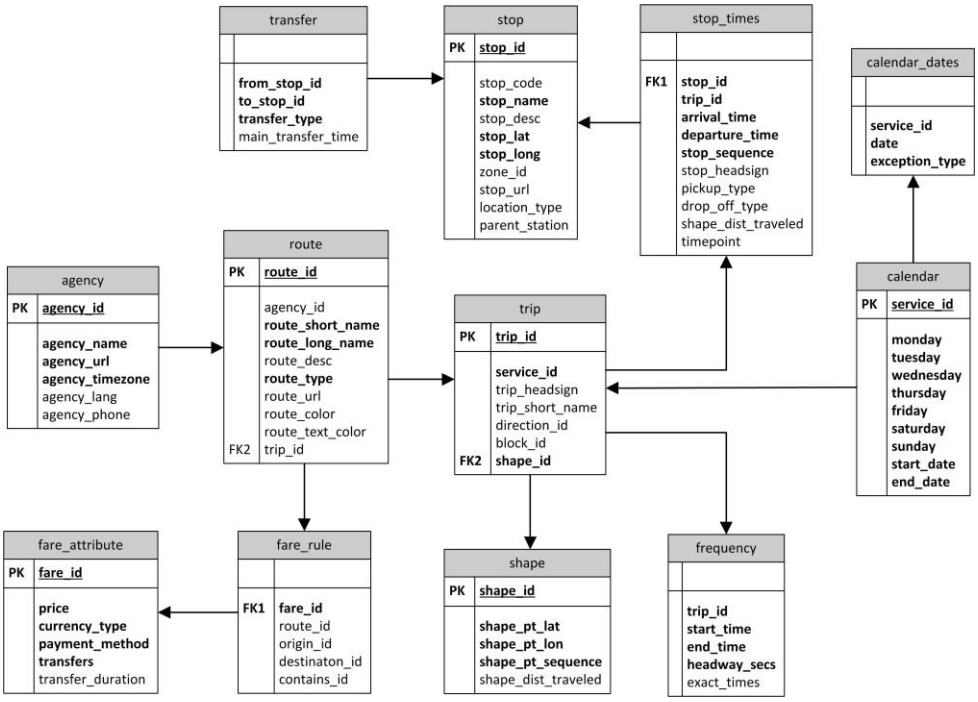

Figure 1: Entity relationship diagram showing the relationships between the different GTFS files, adapted from [1].

The logic of the required files is as follows. A public transport *agency* can have one or many routes. A *route*, also known as a line, is the path followed by a public transport vehicle when picking up and dropping off passengers. A route is made up of one or many *trips*. A trip is a single journey taken by a public transport vehicle on a specific route at a specific time and for a specific service, such as a weekday or weekend service. Different services are captured in the *calendar* file. A trip is further defined by a sequence of *stop times*, which indicates the times at which each *stop* along the route is visited. All stops are geo-located using coordinates, and have names and descriptions that are recognisable to commuters.

The optional files provide extra information regarding the services offered by the agency. Stations contain multiple stops at the same location and allow for *transfers* between different routes. A typical example would be the Marlboro Gautrain station where commuters can transfer from the main line to travel to the O.R. Tambo International airport. A transfer is made up of two stops; the stop where a passenger must get off on the first route and the stop where the passenger must get onto the second route to continue the journey to the destination. These two stops can differ or it can be the same stop, depending on the service. The fare of a route is determined by *fare rules*. There are three different fare rules, namely: zone-based, which is a fare that is dependent on the zones through which the commuter passes; trip-based, which is a fare that is dependent on the route taken, irrespective of the origin and destination stops; and distance-based, which is a fare that is dependent on the combination of origin and destination stops. Each route has a fare rule which is defined by different *fare attributes*, such as the cost and currency of the fare. When there are deviations from the standard service specified in the *calendar* file, these exception dates are specified in the *calendar_dates* file. An example is when a weekday, which normally falls under the weekday service, now follows a Sunday service because it is a public holiday.

2.2 GTFS Realtime

The GTFS Static specification, per se, is already beneficial for numerous reasons, as already discussed. The GTFS Static feed can be used to develop and implement the GTFS Realtime extension [11], with the ultimate aim of enhancing user experience.

GTFS Realtime essentially allows operators to provide realtime updates regarding their operations and fleet, such as delays, route changes, cancellations, the impact of unforeseen events on the network, actual vehicle positions, and live congestion levels. These factors are beneficial to the commuter as the commuter can better plan actual travel time and any possible delays that could occur throughout the journey.

While the benefits of GTFS Realtime are plentiful, the focus of this paper is on GTFS Static as the first stepping stone to obtaining better quality public transport data and making these available to the public. GTFS Realtime will be the natural next step and will be taken into account throughout the design process of a central database for public transport data.

2.3 GTFS usage in developing countries

A World Bank team concluded a pilot program in 2015 where they assisted three Asian developing countries in establishing a GTFS database [11]. These cities were the capital city of the Philippines, Manila; Zhengzhou, a city in central China; and Haiphong, the largest city in Vietnam.

In Manila, there were no complete maps or databases for the transit system, which includes a wide variety of services, such as buses, jeepnies, minivans, and rail. While the different operators in Zhengzhou have information regarding the services they operate, there was no central database integrating all transit information. Haiphong used an outdated map to operate their bus service, and there were no maps capturing the local road network.

Many of the challenges faced by these countries are similar in South Africa. The major problem facing the Gauteng Province is a lack of integration between public transit services. While each operator maintains information to an extent, there is no central database available to get a snapshot of the state of public transport in the province.

Since Zhengzhou's transit data were captured across multiple platforms, the World Bank team used an experienced technical specialist to develop algorithms able to convert all the data into the correct format. For Haiphong and Manila, they had to commence their GTFS project by capturing the data from scratch. From their experience, they estimate that the time and cost of the data collection effort would be 25% more than the expected time and cost. This is due to the number of new illegal routes being operated of which the agencies and government may not be aware of. They suggest that the data cleaning required when using a mobile application to capture all routes takes approximately 15% person-hours more than it took to capture the data. When conventional GPS devices are used to map routes, they suggest adding 50% to the total person-hours.

Once the GTFS databases were completed, the World Bank team ended off their project by establishing a hand-over process. Central stakeholders were identified who are able to provide the required technical and financial resources to host and maintain the GTFS data. In Manila, the National Department of Transport and Communications was identified to lead the GTFS project, while the Department of Science and Technology was identified to host the database. Since China had a longer-term objective of uploading GTFS feeds on a national scale, the National Ministry of Transport took over the GTFS project, while the local city Department of Transport took over responsibility of this task in Haiphong.

It was also critical to consider how the cities would be able to continually maintain and update their GTFS databases. Two open-source packages, developed as a result of the World Bank's pilot study, were made available to the agencies. This included a mobile application TransitWand [12], which can be used to capture route information, and GTFS Editor, a web-platform that can be used to edit data and export into the GTFS format.

Lesson learnt from the World Bank project can be applied to a South African context, since there are similarities between the operators. One concern of pursuing usage of the GTFS format in South Africa is that of people's access to the internet. However, a Pew Research Group study [13] conducted in 2015 indicates that 37% of South African adults report to owning a smartphone, and 42% report to using the internet. Millenials, aged 18 – 34, are more likely to own a smartphone or use the internet than those above the age of 35.

While this study indicates that more than half of the South African adult population does not own smartphones or have access to the internet, there is still a large part of the population who could benefit from accessing transit information on Google Maps. It is also believed that recent campaigns by municipalities to increase access to free wi-fi has allowed more South Africans to access to the internet. Using a standard, though, would allow for the development of software that can query the database and integrate alternative options such as SMS or USSD for non smartphone users.

2.4 Using GTFS for Transport Planning

The value of GTFS is not only found in increased accessibility to public transport information. It can also be leveraged by transport modellers to build more accurate transport models by including up-to-date transit information in a format that can easily be used by the software.

One example is MATSim [14], the Multi-Agent Transport Simulation Toolkit. MATSim is used to model transport vehicles as individual agents with decision-making capabilities. Agents are able to score their transport-related decisions and try out alternative options in the next iteration of the model. Other than private vehicles, MATSim has incorporated freight and public transport vehicles into their models in order to simulate real-life traffic conditions. The GTFS format is used to specify the routes and schedules used by public transport vehicles.

There are a number of route planning software packages available. Scotty [15] is a widely-used Austrian commercial route planning tool to plan rail and bus routes in Europe. OpenTripPlanner [16] is an open source multi-modal journey planning application which is able to determine the shortest path between origins and destinations, given the available road network and public transport services. The GTFS format is also used by OpenTripPlanner to specify transit routes and schedules. The CSIR has used OpenTripPlanner to determine least-cost paths between zones in municipalities, which feed into urban growth models [17].

If South Africa's transit operators would more readily capture and maintain their data in the GTFS format, transport modellers would have access to high quality data that can feed into models that are able to assist the government in transport planning.

While there are many challenges in South Africa inhibiting immediate adoption of the GTFS format, the literature shows that it is possible to capture and maintain transit data in this format in developing countries. In the next section, we elaborate on the methodology followed to capture and upload two GTFS feeds onto Google Maps.

3. GTFS GENERATION FOR GAUTENG'S BUS RAPID TRANSIT (BRT) SYSTEMS

To show the value of using the GTFS format and having information readily available online, a proof of concept (POC) exercise was undertaken to convert the data of Gauteng's Bus Rapid Transit (BRT) systems to the GTFS format and to upload these to Google Maps. This also served as a means to develop a methodology to follow to convert and upload data of any public transport operator in future.

There are currently two operational BRT systems in Gauteng: the City of Johannesburg's *Rea Vaya* and the City of Tshwane's *A Re Yeng*. Ekurhuleni is still in the planning phase of their BRT system, *Harambee*, thus they do not currently have data to convert and upload to Google Maps.

Rea Vaya was launched in 2009 and has three trunk routes, 12 feeder routes and five complementary routes. The trunk routes operate between Thokoza Park and Ellis Park East, between Thokoza Park and Braamfontein, and between Thokoza Park and Library Gardens East. A Re Yeng has been operational since 2014 and has two trunk routes, of which the second one was launched recently on 1 May 2017. The trunk routes cover Pretoria CBD to Hatfield and Pretoria CBD to Wonderboom. There are five feeder routes, and there are plans to extend the operations to the East of Pretoria.

3.1 Collecting raw data

The Gauteng Department of Roads and Transport provided the initial contact with each metropolitan's Head of Transport, who referred us to the appropriate persons in the BRT offices.

Face-to-face meetings were held with the agencies as these enable them to better understand the purpose and value of the project. When there is buy-in from the agencies, it greatly improves communication and willingness to provide the required assistance. Further to this, it is very important to make contact with the person with the most appropriate designation. Unfortunately this is difficult to determine from initial contact, since one initially does not have a full understanding of the operations at the agency. If the designated contact person is one who does not have direct access to the data, or does not have the appropriate authority to get access to the data, it introduces unnecessary delays in the conversion process.

There were also delays when the data had to be collected from multiple sources. It often occurs that agencies have separate departments working with route, schedule, and fare information. Each department captures

their data differently and there is no entity responsible for merging all data into a central database. Conversely, when all route, schedule, and fare information are collected and maintained in a central database and a single person can be contacted to collect all required information, it greatly reduces response times. When data are maintained in a database, it is also easier to convert the data into the GTFS format, since the relationships in the database can be leveraged by a script that can automatically write the data into the correct GTFS files.

Often, while the conversion was taking place, we found that queries arose regarding the raw data collected. When there is a single contact person who is knowledgeable about the data, it is easier to gain clarification about the data.

3.2 Manually converting raw data

Data collected from different sources often have to be consolidated before the GTFS conversion can take place. Additionally, if some data are captured in Word of PDF documents, it prolongs the conversion process. Manually capturing and converting raw data also requires extra validation efforts, since human error can occur when reading in text data.

3.3 Automatically converting raw data

When data are captured in a centralised relational database, a script can be used to read in all the data and write it to the appropriate GTFS file. Unfortunately, when data are incorrectly or inconsistently captured into the database, the script needs to be adjusted to allow for these deviations. Again, this illustrates the need for operators to capture and maintain their data consistently.

3.4 Validating GTFS feeds

Before the GTFS feeds were uploaded onto Google Maps, an online GTFS Validator [18] was used to ensure that the data were captured in the correct manner. The validator runs through a variety of checks such as whether required fields or files have been excluded, whether invalid values are entered into specific fields, whether stop times between stops make sense in terms of the distance between stops and most likely vehicle speed, etc. While this tool cannot determine whether the information captured is correct, it does help to ensure that validation errors are resolved before uploading it onto Google's Partner Dashboard. Once it is uploaded, Google performs a variety of validation checks which pick up finer nuances in the data that may hamper the end user experience.

3.5 Uploading GTFS feeds onto Google Maps

Limited information is available about the process that needs to be followed to upload GTFS data to Google Maps. Only when the process is undertaken does one really understand the effort that is required to publish a high quality GTFS feed. The steps that were followed are listed in

Table 3. The timeframes listed in the table specify the actual time it took to complete each step. These timeframe will differ for each agency, since the timeframes are heavily dependent on the responsiveness of the agency to react to the requests. Also note that there was a publisher involved in this process, who acted as the middleman on behalf of the agencies to publish their data onto Google Maps. In many instances time delays occurred due to the back and forth communication between Google, the publisher, and the agencies. Therefore the timeframes could be reduced if the agencies had direct contact with Google and could immediately respond to their requests.

Table 3 : Summary of steps followed and approximate timeframes to upload GTFS data onto Google Maps.

Step	Timeframe	Responsibility
Submit interest form	< 1 day	Publisher
Send letter of endorsement	2 – 11 days	Agency
Register designated signing authority	1 – 22 days	Publisher
Sign online agreement	1 – 2 days	Agency
Create Google Account	< 1 day	Publisher
Provide access to Transit Partner Dashboard	1 day	Google
Upload GTFS data	1 – 6 days	Publisher
Generate private Google Maps view	2 – 3 days	Google
Validate GTFS data	2 – 3 weeks	Agency, Publisher
Submit pre-launch checklist	1 day	Publisher
Run quality assurance tests	2 – 3 weeks	Google
Final review	1 week	Google
Decide on official launch date	1 week	Agency, Publisher
Launch GTFS data	Depends on launch date agreed upon	Google
Expected total time	**1-2 months**	

The first step was for the publisher to submit an interest form, providing the websites of the agencies to be uploaded. Google uses this information to evaluate whether the agency meets all the requirements to be uploaded onto Google Maps. The agency is required to follow a pre-determined schedule, be open to the public, and not allow reservation of seats.

Since the agencies did not upload the data themselves, they were required to send an endorsement letter to Google Transit specifying that they endorse the publisher to act on their behalf to upload their data. There were some delays in this step since the contact persons at the agencies had to get a higher authority to sign the endorsement letter.

Once the endorsement letter was received by Google, the publisher had to register a designated signing authority at each agency. The designated signing authority is the contact person at the agency that was required to sign an electronic form indicating their agreement to Google's terms and conditions. The online agreement form had to be accepted by each agency before the process could continue.

Once the online agreement was signed by the designated signing authority, a new Google Account had to be registered and sent to Google Transit. Only this Google Account has permission to upload GTFS data. Once the Google Account was created, access was granted to the Transit Partner Dashboard. The Transit Partner Dashboard is a platform where GTFS data can be uploaded to Google. There are two options to do this, namely a manual push of the GTFS data, or by specifying a URL where Google can fetch the data. Once the data was successfully uploaded, a validation report was generated on the dashboard reporting on any validation warnings and errors identified by Google.

The GTFS feed is then included by Google in the next build cycle and a private Google Maps view is generated, which usually takes 2 – 3 days. Google encourages the publisher and agencies to log on to Google Maps using the new Google account to validate the data. It is strongly suggested to include the agencies in this step, since they have a better understanding of the services offered. The private view displays all the information that would be visible to the public if the GTFS feed was launched.

The validation process also includes going through the validation report, addressing all issues to ensure that all routing information is correctly displayed to commuters. Any errors identified need to be addressed and the updated GTFS feed uploaded onto the dashboard again, repeating the process until the agency is satisfied with the quality of the data.

A pre-launch checklist has to be completed by the publisher before communicating to Google that the GTFS feed can be launched. Google then performs quality assurance tests to ensure that the data conforms to their standards. Once the GTFS feed successfully passes the quality review assurance tests, a final review is done on the feed, where additional concerns may be raised. Once these are all successfully resolved, the agencies and publisher decide on a launch date. Finally, the GTFS feed is launched to the public on the specified date.

At the time of writing, A Re Yeng's GTFS data successfully passed all review processes, and has been launched on Google Maps. An example of directions using A Re Yeng routes can be accessed here: https://goo.gl/maps/E89rtDRsNEn, and a screenshot of these directions are shown in Figure 2. Rea Vaya's data are still being validated, and concerns in the raw data are being addressed. The aim is to launch both feeds during South Africa's Transport Month in October.

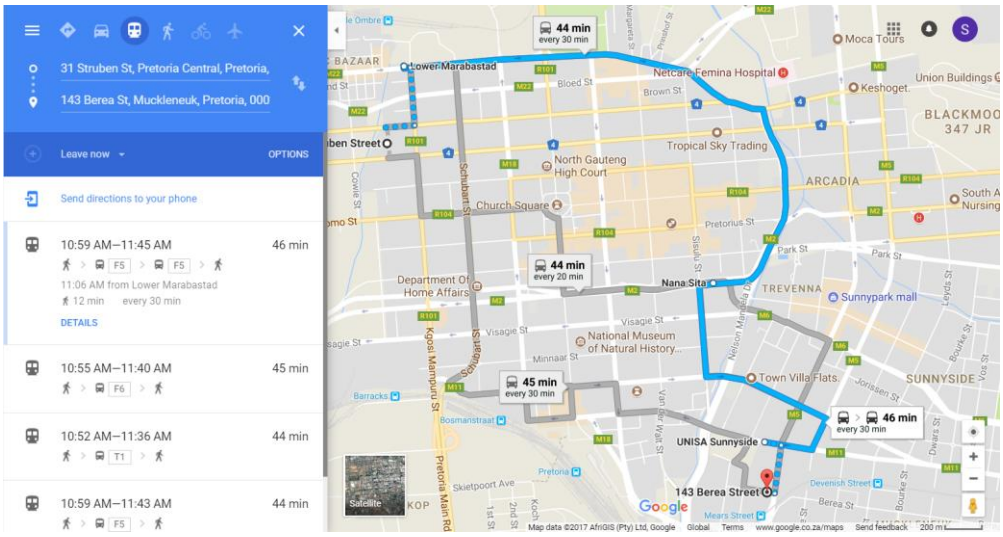

Figure 2 – Directions in Google Maps using A Re Yeng bus services.

4. BENCHMARK OF OTHER PUBLIC TRANSPORT OPERATORS

Six additional public transport operators were visited to understand their operations, data procedures, as well as to determine the quality of their data in order to assess how plausible it would be to convert their data to GTFS. There was a mix of full state-owned entities, semi-state operators, as well as municipal and provincial operators. Five of the operators visited provide bus services and one of them rail services. The taxi industry was not included in this exercise since they do not have fixed stop locations, as required by the GTFS format.

4.1 Operations

All operators who operate on a provincial level have to provide services according to their contractual agreements with the Gauteng Province. The contracts describe the routes that they service, when these routes are operated, and the subsidies earned per kilometre.

To extend operations, operators are required to apply for contract amendments to ensure that they receive subsidies for the additional services that they want to provide. This is a cumbersome process, often resulting in operators failing to apply to provide the additional services. Since municipal operators are funded by the municipalities themselves, they have more freedom in planning their routes.

4.2 Planning and Scheduling

Planning and scheduling are often done manually by the operators. Data collection methods for planning purposes differ for each operator. Some use surveys to determine where demand for new services arises, while other operators look at developments around the province to assist their planning decisions. Some operators require their bus drivers to log ridership information, which informs the planning department on which routes are over- and underserved. Other operators are using their electronic fare collection systems to capture information regarding ridership.

In most cases, a single employee is responsible for scheduling. This makes the operators vulnerable, since their service is dependent on one person who can retire or resign at any given moment.

Various methods are used by the operators to schedule their buses. In one instance, a complete manual process is followed using outdated planning boards. Although this process has been captured in a policy document, no knowledge transfer is taking place to ensure that the work can continue if something were to happen to the person responsible for scheduling.

Some operators use software to schedule their buses, however most of the software packages are extremely outdated and not user-friendly. Most of the operators stated that the software also does not comply with all their requirements, and therefore still need to rely on manual processes to perform their scheduling.

4.3 Fare collection

Some operators are still operating on out-of-date paper bus ticketing systems, where weekly or monthly bus tickets can be bought at specific stations around the province. Some of these operators still allow commuters to pay cash on the bus, although this is discouraged by the operators due to theft. Other operators have started migrating to electronic fare collection systems, which also allows them to capture more detailed information about their ridership.

4.4 Data formats

All operators capture and maintain their data in different ways. In most cases, the operators communicated that they have all data required by the GTFS format. The one thing they have in common is that none of them have a centralised database or server containing all the route, schedule and fare data. Route descriptions are often captured as text in Word documents or in outdated maps displaying the routes. Oftentimes, schedules are captured in Excel spreadsheets. There is also a lack of time-series data, which makes it impossible to obtain historic data of routes and changes to itineraries.

There were three exceptions where the operators' data were already captured and made available on Google Maps. The operators were approached by consultants working in the public transport domain, who are attempting to make public transport data more readily available to the general public.

4.5 Specific issues brought up by the operators

On two occasions, the issues surrounding subsidies and contract renewals were brought up. Some operators are still operating on subsidies that were awarded many years ago (as far back as the 1990s), without being adjusted. Because of this, operators are operating at very low subsidies making it difficult to improve and maintain their fleets. As a result, a lot of them continue to operate with old and unstable fleets making it difficult for them to provide quality services to their users.

4.6 Summary of the benchmarking exercise

The benchmarking exercise proved useful since it helped gain a clearer understanding of the state of operations at various public transport operators in Gauteng. From the data gathered, two categories of operators were identified: those with high operating cost and low ridership, and those with low operating cost and high ridership.

The issues faced by the operators in each category are similar, in that operators in the former category mostly struggle with planning and scheduling, and hence have low ridership values. Although operators in the latter category also struggle with their planning and scheduling, a greater struggle is that of funding and providing quality day-to-day operations. Since the routes covered by these operators have a high demand, these buses are often operating over capacity.

A recommendation that can be made to all operators, irrespective of the type of operator, is that the capturing and maintaining of data be standardised, and that all datasets be hosted on a central server or database.

For operators who are struggling with low ridership, it is recommended that their data be converted into the GTFS format and uploaded onto Google Maps. In this way, visibility of their services can be increased and ridership could possibly increase.

For operators that are struggling with day-to-day operations as a lack of funding, it is recommended to perform business process re-engineering in order to improve efficiencies in day-to-day operations. Studies can also be done to identify wastes in the system and decrease operating costs, freeing up a greater part of the budget for proper maintenance of their fleet.

5. CONCLUSION

As found by the Pew Research Group, more than half of South Africa's adults do not own smart phones or have access to the internet. This is a problem facing many developing countries, as new technologies cannot be readily applied to improve service delivery. Although these challenges exist, it has been shown in literature and by this case study that it is still possible to make public transport data more accessible to the general public in developing countries.

By uploading the two Gauteng BRT operators onto Google Maps the visibility to the services has been increased. It will, however, take a few months to determine whether this project has had a direct impact on ridership and awareness of the services. To determine the impact, A Re Yeng and Rea Vaya could include a question into their registration process where commuters are asked to indicate where they heard of the service, and if their presence on Google Maps has in any way influence them to use the service.

Even though most operators evaluated in the benchmarking exercise have all the data required to set up a GTFS feed, standardised transit data capturing and maintenance procedures are still missing in Gauteng. Standardisation could inform operators on what data to maintain and how often to update it. It also ensures that the provincial government has access to the same set of information for all operators, assisting in their transport planning.

There are many ways in which Industrial Engineers can take up stewardship and improve public transport in South Africa. Since the operators struggle with their day-to-day operations, Industrial Engineers could assist public transport operators in establishing processes to improve data standardisation across the public transport sector. Using database design, Industrial Engineers could also play an integral role in establishing a central GTFS database for the Province, and even the country. The CSIR could also play a strategic role by hosting the central GTFS database, and being the link between the operators and the government.

Another avenue for future work is the establishment of an open data GTFS Exchange platform, where South African public transport operators can upload their GTFS feeds, and the general public and researchers alike can access the data.

6. REFERENCES

[1] **Google Developers. 2017.** GTFS Reference. https://developers.google.com/transit/gtfs/reference/.
[2] **UK Government.** 2014. TransXChange. https://www.gov.uk/government/collections/ transxchange.
[3] **Kaufmann, S.** 2014. Opening public transport data in Germany. Diploma thesis, University of Ulm. http://dbis.eprints.uni-ulm.de/1054/.
[4] **Rea Vaya.** 2016. The Vaya Moja App: Travel planning at your fingertips. https://www.reavaya.org.za/news-archive/june-2016/1335-the-vaya-moja-app-travel-planning-at-your-fingertips.
[5] **McHugh, B.** 2013. *The GTFS Story* in *Beyond Transparency: Open data and the future of civic innovation*, American Press.
[6] **GTFS.org.** 2016. General Transit Feed Specification. http://gtfs.org/.
[7] **Google Developers.** 2016. GTFS Static Overview. https://developers.google.com/- transit/gtfs/.
[8] **Transit Wiki.** 2017. Publicly-accessible public transportation data. https://www.transit-wiki.org/TransitWiki/index.php/Publicly-accessible_public_transportation_data.
[9] **My Citi.** 2017. https://myciti.org.za/en/home/.
[10] **Google Developers.** 2016. GTFS Realtime Overview. https://developers.google.-com/transit/gtfs-realtime/.
[11] **Crambeck, H. and Qu, L.** 2015. Toward an Open Transit Service Data Standard in Developing Asian Countries, *Transportation Research Record: Journal of the Transport Research Board,* 2538, pp 30-36.
[12] **Conveyal**. 2017. TransitWand. http://transitwand.com/.
[13] **Poushter, J.** 2016. Pew Research Center: Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. http://www.pewglobal.org/2016/02/22/-smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/.
[14] **Senozon**. 2017. MATSim. http://www.matsim.org/.
[15] **OEBB**. 2017. SCOTTY - der Routenplaner fur Offis. http://fahrplan.oebb.at/bin/- query.exe/en.
[16] **OpenTripPlanner.** 2017. http://www.opentripplanner.org/.
[17] **Waldeck, L.J. and Van Heerden, Q.** 2017. Integrated Land-Use And Transportation Modelling In Developing Countries: Using Opentripplanner To Determine Lowest-Cost Commute Trips, in *Transportation, Land Use and Integration: Applications in Developing Countries, 100*, WIT Press.
[18] **TransitScreen.** 2017. GTFS Feed Validator. http://gtfsfeedvalidator.transitscreen.- com/.