Open Transport Data Assessment in Mysore

WORLD BANK GROUP

Contents

Acknowledgments

This study was conducted by the World Bank's team of experts led by Alla Morrison, Digital Development Specialist, in close collaboration with the experts from the Karnataka State Road Transport Corporation (KSRTC), the Republic of India. The principal author of this report is Daniel Rudmark.

The team thanks the Karnataka State Road Transport Corporation (KSRTC) and Mr. R. K. Kataria, former Managing Director, for their participation in this study. The team also expresses gratitude to the Directorate of Urban Land Transport, Mysore City Transport Division, Bangalore Metropolitan Transport Corporation, Mysore City Corporation, Mysore Urban Development Authority, Mysore City Police, Karnataka State Remote Sensing Applications Centre, and Sri Jayachamarajendra College of Engineering for their participation in the interviews. The team is also grateful to Lepton Software, YoRide, Marata technology, and TakeUz for their participation in the roundtable and contributions to the dialogue.

The team also thanks colleagues from the World Bank and external experts who have made a valuable contribution to this study, particularly Nupur Gupta, Andrew Stott, Oleg Petrov, Raman Krishnan, and Tatiana Peralta.

The team expresses its appreciation and gratitude to Olga Ryzhuk and Genevieve Maria Dutta for their assistance in the preparation of this report.

Standard Disclaimer:

Copyright Statement:

Abbreviations

API	Application Program Interface
AVL	Automated Vehicle Location
BMTC	Bangalore Metropolitan Transport Corporation
CC0	Creative Commons License
CC-BY	Creative Commons Attribution License
CCTV	Closed-Circuit Television
DULT	Directorate of Urban Land Transport
EU	European Union
GBFS	General Bikeshare Feed Specification
GIS	Geographical Information System
GPS	Global Positioning System
GSM	Global System for Mobile Communications
GTFS	General Transit Feed Specification
ICT	Information and Communication Technology
IT	Information Technology
ITS	Intelligent Transport System
KSRSAC	Karnataka State Remote Sensing Applications Centre
KSRTC	Karnataka State Road Transport Corporation
MCP	Mysore City Police
MCTD	Mysore City Transport Division
MUDA	Mysore City Development Authority
OTP	Open Trip Planner
PDF	Portable Document Format
SJCE	Sri Jayachamarajendra College of Engineering
SMS	Short-Message Service
SIRI	Service Interface for Real Time Information
SUTP	Sustainable Urban Transport Project
TCS	Tata Consultancy Services
XML	Extensible Markup Language

Executive Summary

The World Bank conducted this study of the potential for Open Transport Data in Mysore, India as part of the Sustainable Urban Transport Project (SUTP). It is principally intended for use by Karnataka State Road Transport Corporation (KSRTC), as it contains detailed recommendations and actions to enable KSRTC to open their data for third-party developers. However, in parallel to the in-depth study of the specific conditions of KSRTC, the World Bank team undertook a limited Transport Data Assessment. Thus, the latter part of this report contains our findings and recommendations for local transportation authorities as they consider and implement an open data program in line with the National Data Sharing and Accessibility Policy.[1]

Following the implementation of an Intelligent Transport System (ITS) by KSRTC as part of the SUTP, valuable data are now being generated. The data are not currently available for external re-use, but in the cases where it has been made available—for instance through the appathon held in Mysore in the spring of 2016—it has been welcomed by citizens and authorities. To this end, KSRTC is seeking recommendations on how to advance an open data program in a gradual way to minimize negative effects on the operations of KSRTC.

During the assessment, four broad areas related to third-party development were identified for further investigation. These were:

- the terms of use that third-party developers needed to comply with;
- static public transport data;
- real-time public transport data; and
- the launching strategy.

The report makes the following recommendations for the consideration of KSRTC:

1. To *utilize customized terms of use*, encompassing clauses to specify intermediate developer server architecture, tiered rate limits, no liability, no use of trademarks, no endorsements, and to allow termination of access if KSRTC finds that the terms of use have not been fulfilled by a third-party developer. While standard international open licenses, such as Creative Commons Attribution (CC-BY), are generally preferable, customized terms of use are proposed to KSRTC in order to enable their preferred approach for a gradual, closely managed opening of their data.
2. To *publish static public transport data as General Transit Feed Specifications (GTFS)*. In addition, the recommendation in this report is to provide the GTFS feed openly, but the terms of use should state that if the data are used, then the user is deemed to have accepted the license conditions.

[1] http://ogpl.gov.in/NDSAP/NDSAP-30Jan2012.pdf

3. To *publish real-time public transport data as GTFS-RealTime*. To cater to third-party developers there are also recommendations to perform load testing of the existing server and to compare results to benchmarks and inquire with Bangalore Metropolitan Transport Corporation (BMTC) on a potential cooperation on authorization/authentication mechanisms and caching technologies.
4. To *launch the data to third-party developers nonexclusively*.

In the limited transport data assessment, evidence pointed toward readiness of the city of Mysore in regard to leadership and strategy, the actual and potential internal use of data, and the re-use ecosystem. However, because the policy framework lacks a statewide policy, the collection and management of transport and related data are still hindering publication of highly valuable datasets, and the actual and potential internal and external use of data are yet to reach their potential.

There were several existing high value datasets that could be released in a rapid manner:

- routes, fares, stops, and timetables as well as real-time running data for Mysore City buses;
- historical individual accident data;
- parking data; and
- rental bike facilities and real-time availability data.

We also found data that through additional collection and management could be released as open transport data:

- road network and conditions, and
- traffic counts/traffic speed measurements.

Background

Like many other Indian cities, Mysore faces problems with road congestion, including arrival delays of buses at bus stops, lack of traveler information about different bus routes and stops, as well as time and frequency of service. In its efforts to support the overall public transport system, Karnataka State Road Transport Corporation (KSRTC) introduced an Intelligent Transport System (ITS) to deliver better services and make the transport system more passenger friendly by drawing on ITS capabilities. While the implementation of the ITS in Mysore is overseen by KSRTC, it is the Mysore City Transport Division (MCTD) that operates the buses in Mysore. MCTD oversees a fleet of about 400 buses out of three depots. The implemented ITS system at MCTD provides the following applications:

- Global Positioning System (GPS)-based Automatic Vehicle Location
- Global System for Mobile communications (GSM)-based data and voice communication
- Central Control Room applications
- Real-time passenger information in vehicles, at bus stops and bus stations
- Real-time passenger information on the Internet and via Short Message Service (SMS)

One of the key events triggering the assessment was that several actors had approached KSRTC and requested access to their data to build *third-party applications*. These actors ranged from global software developers to local experts and students. As the new ITS entailed a range of new highly valuable data that could be turned into useful services for travelers, KSRTC requested the World Bank to produce a set of recommendations in order to potentially engage with the third-party developers.

During the assessment, four broad areas were identified for further investigation. These were:

- the terms of use that third-party developers needed to comply with;
- static public transport data;
- real-time public transport data; and
- the launching strategy.

These four areas are described in detail below.

Area 1: Terms of Use

A sustainable program for third-party development requires that the data owner publish clear and fair terms of use of the data. There are several reasons for data owners to approach the task of formulating terms of use mindfully:

1. The terms explicate the responsibilities of both the data owner and the third-party developer. For instance, a data owner may waive the ownership over data once published—as in no copyright reserved (CC0) (see below)—or renounce liability.
2. The terms explicate whether the intended uses by third-party developers are within the boundaries of the allowed use by the data owners. One example is whether a certain business model is feasible or whether the data can be imported into open source platforms such as OpenStreetMap. This means that data owners should be aware that too strict terms of use might hinder third-party developers, and thus they should apply any restrictions with caution.

For a data owner considering terms of use there are two principally different approaches. First, the data owner may use a standard open data license. The main benefit of this approach is interoperability as a standard open data license removes ambiguity in respect to what a third-party developer is allowed to do with the data. Moreover, by choosing a standard license, agencies do not have to engage in formulating terms of use that are both legally compliant and attractive to third-party developers. To this end, many transport agencies use standard licenses such as CC-BY[2] and CC0.[3] Second, there are cases when additional restrictions are seen as necessary, especially when agencies are commercial or semicommercial. In such cases, there may be a need for additional restrictions and a custom license may be used.[4]

It is important to understand that standard licenses address the *use of data*. This includes, e.g., to what extent the data may be modified, allowed usage scenarios, e.g., commercial, and whether attribution is mandatory, optional, or disallowed. However, in cases where limitations are necessary, such as for real-time data, the terms for *access to the data* can be specified separately and in a way that fitts the idiosyncratic conditions of the publishing organization. For instance, use of the UK real-time Weather Forecast and Observation[5] data is governed by the UK Open Government License, meaning that users can modify it, publish it freely themselves, and re-distribute it. However, to obtain the data one needs to use one of three application program interfaces (APIs): (1) a bulk file download from data.gov.uk—a conventional free on-demand best endeavor download without registration, (2) a free best endeavors API

---

[2] C.f. http://www.liikennevirasto.fi/web/en/open-data/terms-of-use#.V3E8mZOLRZI and https://opendata.transport.nsw.gov.au/content/dam/apiportal/transportnsw/docs/TfNSW-Open-Data-Portal-Terms.pdf

[3] https://data.public.lu/en/datasets/horaires-et-arrets-des-transport-publics/

[4] It is not uncommon that agencies start with a more restrictive license and gradually move toward more open and standardized licenses once the risks identified at the outset of a third-party developer program have been experienced are not realized.

[5] https://data.gov.uk/dataset/metoffice_uklocs3hr_fc

for which one needs a developer key and which has some basic rate limiting, and (3) a paid for API which has a guaranteed level of service in terms of availability and response time up to a contractually agreed API usage rate. When an organization is instead using a custom license, it typically covers both the use of data and the access to data.

Standard Open Data Licenses

While there are a number of standard open data licenses, the most used standard open data license by transport agencies is CC-BY. However, many agencies are currently considering moving towards the less restrictive CC0.

CC-BY

The Creative Commons Attribution (CC-BY) is currently at version 4 and is used extensively by open data publishers. The license is permissive and non-revocable, as long as the re-user follows the license terms. The license stipulates that a re-user may both *share* and *adapt* the licensed material in any way, including commercial uses. The license has two important restrictions. First, the re-user must *attribute* from where the data was obtained and if any changes were made in respect to original data. The data owner may stipulate in the license how such attribution should be done, and CC-BY states that such attribution may not be formulated such that it would appear as an endorsement from the data owner (claiming endorsement is not allowed under CC-BY). Second, the license *removes any liability* on behalf of the data owner. This means that no warranties for the data are given and that a re-user must use the data as-is and at their own risk, to the fullest extent of applicable laws.

CC0

More recently there have been arguments raised to move from CC-BY to CC0. This is due to both "attribution stacking" that may be a result of using many different data sources[6] and/or that it creates incompatibility with other licenses and thus hinders uptake into platforms such as OpenStreetMap. In such cases, CC0 may be used to deliberately dedicate copyright protected works to the public domain. Once CC0 is in effect, the work belongs to the public domain and can be used by anyone without any restrictions or obligations. Just as for CC-BY the license removes liability and the re-user must not imply endorsement from the data owner.

National Licenses

Some countries use a national license for their open data.[7] This can be a way to cater for special national jurisdiction that standard licenses do not consider. Further it may be a way to control the terms under

[6] http://web.archive.org/web/20160426155458/http://wir.okfn.org/2012/01/27/attribution-stacking-as-a-barrier-to-reuse/

[7] C.f. http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/ and http://open.canada.ca/en/open-government-licence-canada

which the data are shared. However, a national license may create incompatibility issues that may hinder international uptake.[8] Currently in India, a new national license is undergoing public consultation.[9]

Custom Terms of Use

When agencies use custom terms of use, they must cater for covering all relevant aspects and make some strategic decisions on what to allow and what to constrain. Custom terms of use are more prone to mistakes, especially for agencies with limited legal capacity and resources, and should be used only when standard licenses cannot achieve their objectives or address all context specific concerns. In the sections below the most important aspects are accounted for.

Architecture

Some agencies choose to put constraints on application developers' systems architecture as a measure to decrease the load on agency servers. This constraint typically specifies that the clients of a third-party application are not allowed to connect directly to the agency's server but must access agency data through an intermediate server hosted by the third-party developer. However, while this measure decreases the number of requests, it also may hamper developer adoption since it requires the developer to host their own server.

Rate limits

Rate limits are the number of calls a third-party developer is allowed to make within a specified time frame. Rate limits are a response to a common dilemma for transport data publishers: significant take-up and re-use of (real-time) transport data that may infer high hosting costs and problematic scaling issues for the legacy systems producing the data. To this end, transport data publishers may need to impose restrictions regarding the frequency at which data are retrieved.

This can be done in several ways:

1. A *flat rate* for all developers.[10] This model affords any developer a fixed number of calls regardless of the purpose or need of a third-party application. The upside of this model is less administration than the tiered model but if offers little flexibility if the flat rate doesn't suffice.

2. A *tiered model* based on application attributes.[11] Several agencies have implemented a tiered model where a third-party application gets the number of API calls depending on its needs. Typically, there are two to three tiers where all developers are granted the basic level but must, e.g., prove traveler utility before being allowed to make additional requests.

---

[8] http://opendefinition.org/licenses/ogl-canada-subnational/

[9] https://www.mygov.in/sites/default/files/mygov_1466767582190667.pdf

[10] https://tfl.gov.uk/corporate/terms-and-conditions/transport-data-service#on-this-page-2

[11] https://trafiklab.se/

3. A *general exhortation* from agencies to access data sparsely.[12] Some agencies have not formalized the number of requests a third-party developer is allowed to make but rather make a general request to third-party developers not to query agency servers excessively. While this model allows the agency to avoid the resources necessary to create a data access infrastructure, it is contingent on developers' goodwill to follow the guidelines.

Liability

A necessary clause in terms of use for third-party developers concerns specifying liability. This typically means stating that the agency cannot be held liable for any damages the data may have caused the third-party developer and that the agency typically does not provide any warranties regarding the quality, security, reliability, availability, or performance of the API or the API content.

Trademarks

Many agencies are careful about how third parties may use their trademarks as intellectual property rights and typically protect these. Moreover, the use of trademarks of an agency may be interpreted as being an official app on behalf of the agency or that the third-party developer is endorsed by the agency. To this end, the use of trademarks is typically forbidden,[13] unless an explicit permit by the agency has been granted in some cases.[14]

Endorsements

When agencies engage in third-party development, there emerges a need to explicate the boundaries between the agency as data owner and producer and the application developer. One such important boundary is that the third-party developer must live on its own merit and not draw on the agency's reputation or goodwill. For this reason, agencies typically disallow any use that suggests that a third-party application is endorsed by the agency.

Termination

A delicate clause in the terms of use concerns the rights for the agency to terminate the access for a third-party developer. A termination clause in terms of use typically specifies that termination may occur when a third-party developer is in breach of the terms of use. However, some agencies reserve the right to terminate access for any reason, not just breaching the terms of use. Typically, the terms of use specify that it is the agency that is the sole decision-maker on when a termination shall be done.

---

[12] http://www.bart.gov/schedules/developers/gtfs

[13] http://www.transitchicago.com/developers/terms.aspx

[14] http://web.mta.info/developers/developer-data-terms.html

Usage

Agencies may put a special clause on the actual use of the data.[15] Below are the most common—and noncontroversial—items regarding usage:

- Prohibit disclosing any assigned API keys to outside parties;
- Prohibit use that is in conflict with local jurisdiction; and
- Stating that by using the data, the user is deemed to have accepted the license conditions.

In addition, some agencies implement constructs in the terms of use that will disallow applications that would be in conflict with agencies' business interests. While such a clause may seem as a valid safeguard against potential loss of revenue from third-party applications, it may also raise concerns among third-party developers. A third-party developer must mobilize necessary resources to cover for the investment necessary, and when there are ambiguities whether the current third-party developer business model is compliant with terms of use, it is likely that the third-party developer will not make that investment.

Recommendations

Disclaimer

The preliminary analysis and recommendations in this section are based on information and opinions collected from interviews undertaken and materials provided by the government and other local stakeholders during this study. This section is not based on detailed, legal due diligence and does not constitute legal advice. Accordingly, no inference should be drawn as to the completeness, adequacy, accuracy or suitability of the underlying assessment, or recommendations, or any actions that might be undertaken resulting therefrom, regarding the enabling policy, or legal or regulatory framework for Open Data in the country. It is therefore recommended that, prior to undertaking any legal action to address any legal assessment issue raised herein, a formal legal due diligence be performed by competent, locally qualified legal counsel, preferably assisted by international legal experts with relevant experience and knowledge of these areas.

Current state

Currently, public transport information in the city of Mysore is not shared with third-party developers—except for the current mobile app developed in a KSRTC-held hackathon—and hence no existing license terms exist. In regards to related open data initiatives, there is a thorough initiative on the national level.[16] However, KSRTC falls under the jurisdiction of Karnataka state which is still to adopt an open data policy.

[15] Some agencies may have specific sections on each item under usage, but for the purpose of this report all such clauses are grouped under "usage."

[16] https://data.gov.in/

In such an event, it is possible that KSRTC may need to revise its terms of use to align with such a policy once it is in place in the future.

Custom License

We recommend KSRTC to craft a custom terms of use prior to publishing transport data for third-party developers.

Motivation

To the best of our knowledge no transport agencies in India are currently publishing their transport data for third-party developers. Hence, in case KSRTC were to decide to proceed with implementing such a program, they would be a first-mover and as such have room for making decisions matching the more specific conditions of KSRTC. However, and more importantly, during our assessment several concerns were raised as nonnegotiable preconditions for data to be released at all, such as the need to revoke the license if the result was not beneficial, or not allowing third-party applications hurting the Transport Corporation's revenue. Finally, while there is an existing open data policy on a national level, KSTRC falls under the jurisdiction of Karnataka state where such a policy, including license terms, has not yet been implemented. However, as stated earlier in this report, crafting a custom terms of use requires substantial legal competence and effort, and the risks associated with a custom terms of use must be weighed toward the business risks associated with a standard open data license. Also, once the policy on the state level is implemented, KSRTC should migrate to the license recommended in the policy.

Short-term Actions (1–3 months)

1. Assemble tentative terms of use using existing license terms in other cities as inspiration. Based on the information provided by KSRTC during the assessment, we recommend to consider the following clauses:

    - *Architecture*. Require third-party developers to use an intermediate server. This will reduce the load that KSRTC's servers need to handle, an issue that has been identified as important. The terms of use from New York's Metropolitan Transportation Authority (MTA) in the U.S. may be used as inspiration.[17]

    - *Rate limits*. Following the need to maintain server load within an acceptable range, we recommend that the terms of use include rate limits. Here we find that using a tiered model would be appropriate as there exists a large tentative re-use community given existing interest from businesses and universities. By using a utility-based tiered model, KSTRC could grant any developer a rate limit for experimentation and testing purposes, a production targeting rate limit for applications that may demonstrate some traveler value, and a partner-level rate limit for applications of strategic importance. (For more information on a similar approach see

[17] http://web.mta.info/developers/developer-data-terms.html

Annex 1: The Tiered Model of Stockholm Public Transport). Also, this section should include a clause specifying that KSRTC reserves the right to reduce the limit if the system is under stress, yet it should make a best effort to inform its third-party developers in such a case.

- *Liability*. We strongly recommend KSTRC to explicitly not be liable for neither the content nor the uptime for its transport data. In addition, the terms of use should clearly state that no warranties are given. The terms of use from Transport for London may be used for inspiration.[18]

- *Trademarks*. We recommend that third-party developers are not allowed to use the trademarks of KSRTC. Adding a clause that any use would require written consent from KSRTC could be included, but this could infer additional administration on behalf of KSRTC. The "Terms of Agreement", §6, from the MTA can be used as an inspiration.[19]

- *Endorsement*. We recommend KSRTC to restrict third-party developers from claims to be endorsed by KSRTC. The terms of use from Transport for London may be used for inspiration.[20]

- *Termination*. We recommend KSRTC to have the right to terminate access for third-party developers and that such a decision is to be made solely by KSRTC in the case that KSRTC may find that the terms of use have not been met by the third-party developer. The terms of use from Chicago Transit Authority may be used for inspiration.[21]

- *Usage*. Given the information that was provided during the assessment, we recommend KSRTC to craft a specific clause on usage. Such a clause could include the following items:

    - Prohibit disclosing any assigned API key to outside parties;
    - Prohibit use that is in conflict with local jurisdictions; and
    - Explicate that the application should promote the use of public transport. However, such a clause may be perceived as a barrier by some third-party developers. It is therefore important that such a clause is used with great caution and that KSRTC in due time would consider revising it (see below). For inspiration, the terms of use from Trafiklab.se may be used.[22]

2. Once the tentative terms of use are assembled, it is of utmost importance that the terms are validated by local legal expertise. The examples and inspirations given in this report are best practices, but it is necessary that they are adapted to local jurisdictions.

---

[18] https://tfl.gov.uk/corporate/terms-and-conditions/transport-data-service (§1–§2)

[19] http://web.mta.info/developers/developer-data-terms.html)

[20] https://tfl.gov.uk/corporate/terms-and-conditions/transport-data-service#on-this-page-4

[21] http://www.transitchicago.com/developers/terms.aspx (section V)

[22] https://www.trafiklab.se/api/sl-realtidsinformation-3/licens ("Requirements on use")

Long-term Actions (6–12 months)

1. Follow the development regarding the work on policy and implementation of an open data program in Karnataka state. In the case that an open data state policy is published, including recommendations for data licensing, KSRTC should work toward harmonizing with such a policy for compatibility reasons.
2. Evaluate to what degree the terms of use are aligned with the business models of third-party developers. If many concerns are raised which have led to non-adoption, KSRTC should consider revising the terms and adopt a standard open data license.

Area 2: Static Public Transport Data

One of the most valuable datasets that is published by many agencies around the globe concerns static data on agency operations. These datasets include timetables, stops, routes, and fares and can create much value for the agency, third-party developers, and travelers as they provide the foundation for creating a large array of traveler applications, integration of public transport information into new contexts, and agency analytics.

Choosing a Format for Static Public Transport Data

When an agency decides to publish static data on its operations, they have to make a decision on the technical format through which this information is conveyed. The most common format to publish this information is the General Transit Feed Specification (GTFS).[23] In the EU, the agreed-upon standard for sharing such data is an implementation of Transmodel[24] called NeTEx.[25] These datasets may then be run through a travel planning algorithm, such as the OpenTripPlanner (OTP),[26] to present options to go from point A to point B. Other agencies have not published the data but rather, have published programmable services to search for trips, stops within a given geographical range, or fares for a certain trip.

GTFS

GTFS has emerged as a de facto standard for publishing transit data around the globe, and at least 1,000 agencies are currently publishing feeds in this format.[27] In addition, there exists a large community around this standard in the form of e-mail lists,[28] software tools, and frameworks. GTFS is designed around providing a standard with the end user experience in mind. However, given its wide adoption it has come to be used in additional use cases such as public transport network analysis.[29] More information on GTFS can be found in Annex 2: General Transit Feed Specifications (GTFS).

NeTEx

NeTEx is a standard for exchanging public transport information as extensible markup language (XML) documents. It is based on an XML schema using the Transmodel representation of common public transport concepts and data structures and can be used to exchange many kinds of data between passenger information systems, including for stops, facilities, timetabling, and fares. The data can be used by both

[23] https://en.wikipedia.org/wiki/General_Transit_Feed_Specification
[24] https://en.wikipedia.org/wiki/Transmodel
[25] https://en.wikipedia.org/wiki/NeTEx
[26] https://github.com/opentripplanner
[27] http://www.gtfs-data-exchange.com/agencies
[28] https://groups.google.com/forum/#!forum/transit-developers
[29] http://www.opentripplanner.org/analyst/

operational management systems and customer facing systems for journey planning. NeTEx is the selected standard for the Commission Delegated Regulation[30] on harmonizing open transport data among EU member states. The drawback of using NeTEx over GTFS is adoption, and the surrounding ecosystem is still to emerge and does not compare to GTFS—for instance, there is no accepted open source route planning algorithm published to date. However, NeTEx can convey more complex data structures than GTFS, for instance on fare designs and interagency information exchange.

Programmable Services

Some agencies choose to serve third-party developers, not by publishing the data itself but rather through programmable services. For instance, Transport for London does not publish static data files but serves its third-party developers through an API instead. The benefit of this solution is the ability to provide a coherent user experience, as the agency controls the route planning algorithm and different algorithms may yield different results. However, this also creates a significant extra load on agency servers as each travel planning request generated by a user must be served by the API. Developers may also find that this solution is less flexible than working with other agencies that provide the data themselves.

Choosing a Distribution Mechanism for GTFS

Agencies also need to decide *how* the static public transport data are to be shared. In this section, such mechanisms for GTFS are accounted for. There are three principally different ways of sharing GTFS files with third-party developers:

1. *Provide an open feed*. This means that the URL to the feed is published publicly and anyone with the link may access the data. By using the data feed, the user has accepted the terms of use. In the United States and many other countries this distribution mechanism is the most common for GTFS. One of the benefits of this model is compliance with guidelines of open data in that no registration should be required to gain access to the data. Another benefit is that agencies do not have to spend resources on creating an authentication mechanism and manage its use. However, by having an open feed the agency will not know who is using their GTFS feeds.
2. *Require registration before access is granted to the GTFS feed*. When agencies find that is important to know what developers are accessing their GTFS feeds they may require registration before gaining access to the feed. In addition, by requiring registration they may also force the developer to accept the terms of use and collect e-mail addresses to inform about upcoming changes. The drawbacks of this approach are mainly that a required registration may be perceived negatively by third-party developers, that it takes additional resources from the agency to build and maintain the registration mechanism, and that it conflicts with the principles of open data, which should not require registration.

[30] http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32013R0886&from=EN

3. *Require that terms of use are accepted before access is granted to the GTFS feed*. Some agencies[31] have struck a middle ground—they do not require registration but still enforce that third-party developers have accepted the terms of use.

Recommendations

Current state

Currently the data necessary to create a GTFS feed is available in a digital format within KSRTC, and according to our interviews the process of converting the current data into GTFS was not assessed as requiring substantial work. We also found that that there is consensus that GTFS was the preferred format to publish static public transport data.

GTFS and providing an open feed

We strongly recommend that KSRTC use the GTFS standard to publish static data to third-party developers. We also recommend KSRTC to use Option 1 as a distribution mechanism: providing the feed openly. However, the terms of use should state that if the data is used, then the user is deemed to have accepted the license conditions.

Motivation

GTFS has emerged as a de facto standard among transit agencies around the globe. As a consequence, there currently exists a large pool of readily available software developers that easily can integrate their existing application into a new setting such as Mysore. Moreover, for new and more local developers there is a large array of existing open source software tools and frameworks that new developers could draw upon and enter the market rapidly. Finally, because of the large uptake of GTFS, more agency-oriented tools, such as the OTP Analyst, are being developed and could be utilized if the static transit data were available as GTFS. The NeTEx standard is interesting and offers more flexibility, especially in scenarios with information exchanges among agencies, but has insofar not reached a satisfying market penetration to yield a recommendation. Programmable services may provide a way of ensuring a consistent user experience but require additional server resources – an issue that has been identified as crucial during the assessment.

Concerning distribution mechanisms, agencies typically do not require registration for GTFS feeds, and choosing Option 3 above would ensure KSRTC that third-party developers have accepted the terms of use without a need for KSRTC having to build and maintain the registration mechanism. In addition, since these files change relatively infrequently—compared, for example, to real-time files—they do not pose a threat to excessive server load and hence do not require registration to manage rate limiting.

[31] https://www.data.vic.gov.au/data/dataset/ptv-timetable-and-geographic-information-2015-gtfs

Short-term Actions (1–3 months)

1. *Convert static information required into GTFS*. The most likely actor to perform this conversion is TCS[32] given the information that was collected during our assessment. However, conversion to GTFS is deemed a relatively simple task and could thus be performed by other actors such as KSRTC's information technology (IT) department or a consultant.

2. *Include GTFS into existing timetable and fare change processes*. It is important that the GTFS reflects the currently used timetables, routes, and fares. Thus, updates to such information should render a new set of GTFS files, and to achieve consistency in GTFS updates it is important that KSRTC adds GTFS updates into its existing processes.

3. *Prepare a section on the KSRTC's website where GTFS feeds can be distributed*. Currently, there is no open data platform in the state of Karnataka. To this end, KSRTC needs to publish this data on its own web page.

4. *Implement functionality that requires third-party developers to accept terms of use before accessing the feed*. A good inspiration for this task is the state of Victoria in Australia.[33] If building this functionality is technically challenging, simply stating in the terms of use that if the data are used then the user is deemed to have accepted the license conditions, is also a viable option.

5. *Publish the feed in relevant GTFS feed directories*. Once the feed is in place it is important that it becomes known that the feed is available. To this end KSRTC should publish the location of its GTFS feeds into directories such as Transitland[34] and TransitFeeds.[35]

[32] TCS stands for Tata Consultancy Services, current vendor that provides the ITS system and operational support to KSRTC.

[33] See https://www.data.vic.gov.au/data/dataset/ptv-timetable-and-geographic-information-2015-gtfs

[34] https://transit.land/feed-registry/

[35] https://github.com/TransitFeeds/TransitFeeds-Public/issues

Area 3: Real-time Public Transport Data

Real-time information about public transport is a highly-valued dataset. Application developers typically urge operators to release the public transport data as it provides a means to calculate more accurate arrival and departure times, rather than relying on predefined timetables. For instance, real-time running data give travelers a new decision-making basis as by knowing the actual location of public transport they can decide whether a certain departure is a feasible option.

Choosing a Format for Real-time Public Transport Data

There are two main approaches to distributing public transport data. First, agencies may employ a "*fire hose*" model where the entire dataset is continuously streamed to its clients. The primary benefit of this model is that it takes the load off agency servers. However, it requires more bandwidth and that third-party developers store the data locally before distributing the data to its clients. Second, agencies may provide a "*faucet*" API where third-party developers may, for example, query the location of an individual vehicle on a particular route. This model is suitable for mobile application usage where third-party applications may query agency servers directly; however, this mode creates many more agency server requests than the "*fire hose*" model.

When agencies choose to publish real-time public transport information, they need to decide the format or standard through which this information should be conveyed. The three most common formats are GTFS-RealTime, SIRI, or a custom API. These are briefly presented below.

GTFS-RealTime

GTFS-RealTime is a feed specification that allows public transportation agencies to provide real- time updates about their fleet to application developers. According to Google, "it is an extension to GTFS, an open data format for public transportation schedules and associated geographic information. GTFS-RealTime was designed around ease of implementation, good GTFS interoperability and a focus on passenger information."[36] Given the large uptake of GTFS for static data, many agencies have started to use the real-time extension as it affords easy integration of static and real-time data. However, the uptake of GTFS-RealTime is lower than of static GTFS feeds because many agencies can or will not provide real-time information and agencies have chosen to publish real-time information through other standards such as SIRI (which offers more flexibility) or custom APIs. GTFS-RealTime is based on the "fire hose" model.

[36] https://developers.google.com/transit/GTFS-RealTime/

SIRI

The Service Interface for Real Time Information (SIRI) is a protocol that allows exchange of real-time information about public transport services and vehicles. Just like NeTEx, SIRI is based on Transmodel for public transport information and comprises a general-purpose model and an XML schema for public transport information. Given its history, SIRI has mainly seen uptake in Europe, but some U.S. agencies, such as MTA, offer real-time information about their services through SIRI.[37] The main benefit of SIRI is that it may convey more details about public transport than GTFS-RealTime and that it can employ both a "*fire hose*" and "*faucet*" approach.

Custom API

A final approach is custom APIs where agencies move an internally used API—or its portions—up through the design hierarchy and expose it to third-party developers. The advantages of this model are that it does not require any conversion to a standard and that typically the API is already tested and proven. In addition, the agency may choose exactly what information is released to third-party developers. Typically, a custom API uses a "*faucet*" approach.

Managing High Server Load

While real-time data offers exciting opportunities for third-party developers, it puts additional requirements on agency server infrastructure. Due to its high value, the number of requests for real-time data may exceed what agency servers can manage to complete. To this end most agencies implement a policy on the maximum number of allowed requests per third-party developer (see Rate limits) and an authentication mechanism to materialize this policy.

The key to managing high server load is understanding two issues:

1. The number of requests per time unit that the server infrastructure can handle, and
2. The anticipated load given the specified rate limits.

The first issue requires that a load test is performed onto the servers that are to serve third-party developers. This way the agency is able to assess actual capacity of the exiting servers. The primary reason to perform a realistic load test is that the performance of the system, especially under stress, is difficult to anticipate. Hence, through a load test—preferably with some sorts of instrumentation or logging of the individual components—the agency can determine the "weak link" that may need to be optimized for increased performance.

The second issue is an assessment of the anticipated load from third-party developers. This is much more difficult to assess, as it is highly dependent on the type of services being developed, how much data they require, the number of third-party developers, and user adoption. A cost-efficient way of dealing with

[37] http://bustime.mta.info/wiki/Developers/SIRIIntro

uncertain developer demand under cost constraints is to initially only allow a basic developer service (typically the entry category in a tiered model).

In addition, there is a need to enforce the rate limits. This is typically performed by issuing API keys[38] to developers during registration and have all API calls make requests through an API proxy that checks whether the key is valid and if the request is allowed given its rate limits. There are several ways of implementing the enforcement of rate limits:

1. *Custom development* where the agency (or its IT suppliers) complements its existing applications by writing API key issuing, authentication, and authorization software;
2. *Commercial packages* where the agency procures an API management platform from a vendor of such a platform.[39] This platform typically comes with a range of additional tools such as developer consoles and analytics tools; and
3. *Open source packages*[40] where the agency (or its IT suppliers) may get appropriate software without additional costs but need to invest to understand and configure it appropriately.

Recommendations

Current state

Regarding the format of public transport real-time information, it currently is available as a custom API. This custom API is currently used by the mobile app recently released by KSRTC. In terms of converting the custom API there was a discussion around this matter during the assessment. Currently, the underlying ITS system does not have support for publishing real-time data as GTFS-RealTime and according to TCS such a conversion would be a nontrivial task.

During the assessment, it was also found that there are currently two web/app servers that could be used to serve third-party developers. First, there is a cluster of web servers that hosts both the internal applications for the control room and the public website. This server cluster is fully load tested by TCS. Second, there is a cluster of application servers serving an SMS application and the newly launched mobile app. This cluster has yet to be load tested as up until now it has not been as strategically important, although with the launch of the mobile app its importance has increased. There has been a concern about acquiring new server clusters, and the preferable solution is to use one of these existing server clusters.

In terms of rate limit, enforcement does not exist today and thus needs to be acquired. During our assessment, we also found that BMTC was preparing for launching its data for third-party developers and was considering the same need to enforce rate limits.

---

[38] https://en.wikipedia.org/wiki/Application_programming_interface_key

[39] There are many actors in this space ranging from more general purpose actors such as Microsoft Azure and Amazon Web Services to specialized firms such as Mashery, ApiGee, Modern Age, and 3Scale.

[40] See http://apievangelist.com/2014/10/05/taking-a-fresh-look-at-what-open-source-api-management-architecture-is-available/ for an overview.

Recommendations

In terms of a real-time information format the recommendations are contingent on the launching strategy that KSRTC is seeking to implement (see Area 4: Launching Strategy).

- If KSRTC is to make a staged launching process with selected third-party developers, it is recommended to use the internal API for real-time information.
- However, our long-term recommendation is for KSRTC to use the GTFS-RealTime format, given that such a conversion carries a reasonable cost. The reason to choose GTFS-RealTime is both its integration possibilities, the existing ecosystem, and the anticipated decreased server load capabilities.

In terms of managing high server load we strongly recommend that KSRTC undertakes a load test on the app server cluster and uses this as the third-party developer server, at least initially. Once the capacity of this server cluster is established, it is possible to make a preliminary assessment of how many additional third-party applications the server could serve on its lowest rate limit tier. To reduce initial investments by KSRTC, the program could start by only offering a basic service and work with developers on how to optimize the use of the limited capacity.

In the long term, KSRTC should either itself increase its server capacity, utilize data management capacity of Karnataka State, or outsource production quality serving of data at scale to a private sector. KSRTC could fund this with paid for access with service quality guarantees, as well as data conversion services for the developers after taking the data from KSRTC's own systems—such as caching or "firehose to faucet" conversion. Private sector companies can offer value-added opportunities—for instance Place/Transport API in the UK takes data feeds from different transport agencies and turns them into a guaranteed service for business users of transport data with a common API, irrespective of the original API used by each transport authority.

In terms of enforcing the rate limit policy, we recommend that KSRTC initially inquire with BMTC in regards to a potential cooperation around this issue. This way the two companies may share the development cost on this issue. As a fallback option, we recommend KSRTC inquire with TCS on the cost of implementing such a rate limit enforcement system, potentially in parallel to BMTC discussions.

Short-term Actions (1–3 months)

1. Prepare a change procurement order for TCS for custom API–GTFS-RealTime conversion.
2. Perform a load test on the app server cluster.
3. Assess whether the current app server cluster would be able to handle third-party development.
4. Inquire with BMTC on a possible cooperation on rate limit enforcing software.
5. Inquire with TCS on an estimate of developing rate limit enforcing software.

Medium-term Actions (4–6 months)

1. Implement a custom API–GTFS-RealTime conversion.
2. Implement rate limit enforcing software.
3. If a staged approach has been undertaken, assess whether the existing server infrastructure needs to be expanded.

Long-term Actions (6–12 months)

1. Launch key-based GTFS-RealTime to third-party developers.
2. Evaluate options for serving production quality APIs by increasing own server capacity, leveraging state resources, or outsourcing to the private sector.

Area 4: Launching Strategy

Background

During the assessment KSRTC displayed substantial interest in spurring third-party development in a rapid manner. However, being a first-mover among public transport agencies in India, several concerns were raised on how such a launch could be implemented in a controlled manner. For instance, there were concerns about whether third-party development would hurt revenue, to what extent the existing infrastructure would be able to handle the perceived web request load, and whether it was feasible to develop the necessary interfaces within a short time frame.

To this end two launching strategies—in addition to those actions proposed earlier in this report—were discussed during the assessment. They are presented below, together with pros, cons, and necessary actions.

Recommended: Full-access Approach

In this scenario KSRTC would release all its data nonexclusively when launching. The rationale for such a full-access approach would primarily be to enforce the first-mover position and quickly draw on third-party development on a larger scale. This way KSRTC is likely to reap the benefits of open transport more rapidly with larger benefits for travelers, leveraging the deep community of all GTFS developers and applications in the market.

Table 1: Characteristics of the Full Access Approach

	Static Data Format	Static Data Availability	Real-time Data Format	Real-time Data Availability	Real-time Data Authentication
Trial period	—	—	—	—	—
Full-access period	GTFS	Nonexclusive	GTFS-RealTime	Nonexclusive	API keys

Pros

- In the case that a full-access period could be achieved rapidly, KSRTC would manifest itself as the first-mover and likely attract attention from many third-party developers.
- A full-access approach would eliminate the need for selecting third-party developers to a potential trial period, something that was identified as potentially problematic.

Cons

- There would be less room for KSRTC to test-drive third-party development in regards to the concerns identified by KSRTC. However, through the suggested terms of use, KSRTC would still need to terminate its third-party developer program.
- Launching with full access would entail unpredictable server loads for which KSRTC would need to be ready to scale up quickly.

Alternative full-access approaches

While the above is the recommended approach, there is an alternative version of a full-access approach:

- *Same as above but initially release real-time data as Custom API during the full-access period*. This situation could be considered if the conversion and implementation of GTFS-RealTime would be lengthy. The pros of this approach are that KSRTC still would be able to publish its real-time data to third-party developers in a timely manner. The cons are that KSRTC would not be able to fully draw on the GTFS community (including international third-party developers), a projected increased load on its web servers due to its "*faucet*" approach, and that all developers will need to redesign their applications once GTFS-RealTime is launched.

Necessary actions

This launching does not require any additional actions other than those brought forward in previous sections of this report.

Alternative Strategy: Staged Approach

In this scenario, the first release of KSRTC data would be exclusive toward a selected group of third-party developers. This group would get access to both static and real-time data over a fixed period of time (around 6 months). The rationale for such a staged approach would primarily be that it would be a way for KSRTC to "test-drive" third-party development, albeit not withholding the long-term strategy of releasing the data nonexclusively.

Table 2: Characteristics of the Staged Approach

Period	Static Data Format	Static Data Availability	Real-time Data Format	Real-time Data Availability	Real-time Data Authentication
Trial period	GTFS	Exclusive	Custom API	Exclusive	IP address
Full-access period	GTFS	Nonexclusive	GTFS-RealTime	Nonexclusive	API keys

Pros

- One concern is that third-party development would hurt KSRTC's revenue and during a trial period this could be monitored closely.
- Assessing the necessary server capacity can be difficult without real data—however such data could be generated during a trial period.
- During a trial period third parties may develop new and useful applications that can be used as showcases and motivate additional investments in third-party development.
- The trial period would generate feedback from third-party developers and potential deficiencies in both the legal and technical infrastructure could be adjusted before the full access period.

Cons

- The exclusive access inherently means that some third-party developers will not be able to participate in the trial period which may reflect negatively on KSRTC.
- With a longer period to full launch there is a risk that currently interested developers invest elsewhere and/or that other agencies are publishing its data openly in the meantime, creating more competition for the third-party developers. This risks affecting developer adoption once KSRTC offers nonexclusive access.
- Since developers would have to be selected, there is a risk that the selection process and criteria may be questioned.
- Developers using the custom API would need to subsequently convert their solution into GTFS-RealTime.

Alternative staged approaches

If KSRTC were to employ a staged approach, there are two alternative versions of such a strategy:

- *Same as above but release static data as GTFS nonexclusively during the trial period*. The pros of this approach are that KSRTC would likely be the first-mover and in this way attract third-party developers, but also that KSRTC is likely to avoid third-party development based on scraping (which otherwise is likely to occur). The cons are that the more controlled test-driving aspect of the trial period is attenuated and that KSRTC must implement the recommendations for static public transport data (see Recommendations) more rapidly.
- *Same as above but release real-time data as a Custom API during the full-access period*. This situation could be considered if the conversion and implementation of GTFS-RealTime is perceived as too lengthy to maintain interest among third-party developers. The pros of this approach are that KSRTC still would be able to quickly publish its real-time data to third-party developers. The cons are that KSRTC would not be able to fully draw on the GTFS community

(including international third-party developers) and there would be a projected increased load on its web servers due to its "*faucet*" approach.

Short-term actions (1–3 months)

- *Develop selection criteria and perform a selection process*. This is a challenging task since the idea is to move rapidly, but designing a comprehensive selection criteria and a nondiscriminatory selection process would work against the objective of a trial period. The ideal candidates should be a mix of local developers (e.g., the developers that have brought proposals to KSRTC) and international firms with proven track records. Such candidates would be likely to both address local issues and bring proven and popular technology into Mysore.

Medium-term actions (4–6 months)

- *Prepare a launching event for nonexclusive access to data*. To get additional traction when the data are made available to third-party developers, it is often a good idea to use an event such as an appathon. This way public relations for the launch will be generated and a suite of different prototypes will be developed in a rapid manner, providing immediate traveler utility.

Transport Data Assessment

The main purpose of the World Bank mission to Mysore and Bangalore concerned identifying data-driven solutions for KSRTC. In particular, the mission sought to provide recommendations and help identify actions for KSRTC to commence publishing transport data to third-party developers.

However, an additional purpose was conducting partial Transport Data Assessment. Although the Bank has a methodology for identifying key actions for clients that decide to proceed with an open transport data program, the technical visit did not allow the time or the resources to conduct a full formal Assessment. It was however possible to use the Assessment framework to benchmark Mysore's current level of achievement against global best practice.

The assessment has two main outputs. First is the team's rating of Mysore's readiness to unlock full potential of transport data. Second is that the high value datasets with a proven ability to provide utility have been identified and rated in regards to when they could be released.

During the visit the team met with the following organizations:

- The Directorate of Urban Land Transport (DULT)
- Karnataka State Road Transport Corporation (KSRTC)
- Mysore City Transport Division (MCTD)
- Bangalore Metropolitan Transport Corporation (BMTC)
- Mysore City Corporation (MCC)
- Mysore Urban Development Authority (MUDA)
- Mysore City Police (MCP)
- Karnataka State Remote Sensing Applications Centre (KSRSAC)
- Sri Jayachamarajendra College of Engineering (SJCE)
- A roundtable discussion with potential re-users comprising
    - Lepton Software
    - YoRide
    - Marata Technology
    - TakeUz

The results of this assessment are shown in the following table:

Table 3: Transport Data Assessment Results

Dimension	Importance	Findings for Transport Data in Mysore	Rating[41]
Leadership and strategy	Using transport data to maximum effect requires the implementation of change—often including legal, institutional, technological, and cultural changes—and may affect stakeholders both inside and outside government. Focused, strong, and sustained political/senior leadership and a clear strategy for improvement in strategy are therefore critical to helping a government overcome resistance and inertia of all kinds, to helping incentivize actors to make the necessary changes in a timely and effective manner, and to achieving the desired objectives and benefits of a city's transport strategy.	There was evidence for solid transport leadership in Mysore. The transport planning is based on the Mysore Master Plan[42] that runs up until 2031, and all transport investments are aligned with this plan. The green rating is also based on several far-reaching projects that have been implemented according to the plan. This includes road up-gradation, new road networks, and ring roads. In addition, Mysore became among the first cities in India to implement an ITS system for its city buses and is currently sharing their experiences with other interested Indian cities and regions. Another Mysore example includes a public-private partnership around rental electric bikes (that is claimed to be the first in India). As Mysore attracts more than 3 million tourists each year, rental electric bikes provide a sustainable alternative to motorized transportation for tourists.	**GREEN**

[41] Green means there is clear evidence of readiness. Yellow means that evidence of readiness is less clear.

[42] http://mudamysore.gov.in/master_plan.htm

Dimension	Importance	Findings for Transport Data in Mysore	Rating[41]
Policy framework	The long-term success and sustainability of a program for realizing maximum value from Transport Data, including through Open Data, depends greatly on the enabling policy and legal framework. Wider use of data requires that a range of policy and legal issues be addressed—for example, with respect to the licensing and re-use of data, ensuring privacy and data protection, and anonymizing personal and personally identifiable data. It is important to identify at an early stage the existing policies, laws, and regulations with respect to a core set of issues, and to identify actual or perceived obstacles in order that policy or legal change can be initiated early if needed.	On a national level there currently exists an implemented data policy,[43] and the national open data portal[44] have some 27,000 resources published. The policy mandates that all shareable and nonsensitive data should be published through the national data portal. However, the Mysore agencies interviewed in this assessment do not fall under the responsibility of the central government but answer to Karnataka state. In Karnataka state, there is not currently an operational open data program. However, an open data policy is under preparation by the Centre for e-Governance at the Government of Karnataka and is yet to take effect. Also, as evidenced by our visit with KSRSAC, there are currently legal barriers for sharing map data, and especially so for commercial re-use.	**YELLOW**

[43] http://ogpl.gov.in/NDSAP/NDSAP-30Jan2012.pdf

[44] https://data.gov.in/

Dimension	Importance	Findings for Transport Data in Mysore	Rating[41]
Transport and related institutions and companies	The organization of transport in many cities is complex: often there are separate responsibilities for private road transport and public transport services; in some cases, there are separate responsibilities for infrastructure and for operations; and some transport services may be operated by private sector actors who may or may not be regulated. It is important to understand these structures and how the relevant agency/agencies use data to carry out their strategic planning, operational planning, tactical management, and reporting responsibilities, and where appropriate how they work together and share data to carry out their responsibilities.	While all interviewed agencies were using data for planning their operations, there was awareness that additional data could be used for decision making (such as bus occupancy or tickets data from Depot Computerization System). There was multiple evidence on interagency data exchange for planning purposes. One notable example is MUDA that draws data from a multitude of actors to produce the master plan.[45] However such data exchange requests were currently handled manually on an as need basis.	**YELLOW**

[45] http://justmysuru.com.md-in-23.webhostbox.net/mudamysore.gov.in/MasterPlan/MP_Reports/VOL%201_DATA%20COLLECTION,%20ANALYSIS%20&%20PROJECTIONS.pdf

Dimension	Importance	Findings for Transport Data in Mysore	Rating[41]
Collection and management of transport and related data	Programs to derive value from data can build on established digital data sources and information management procedures within transport and other agencies where they already exist. Where data are only available in paper form, it will be hard to make data available in a re-usable format quickly and cheaply. Conversely, good existing information management practices within transport and other agencies can make it much easier to put processes in place that make the release of data a sustainable, business-as-usual, downstream process as part of day-to-day information management.	The visited agencies showed a good awareness of the data they held and how data was collected. While most of the high value datasets found during the assessment were available in a digital format, some crucial datasets were not readily available to allow a green rating. This includes real-time traffic information (such as traffic counts and traffic flows). However, we found evidence of an awareness that new transport projects should include a data component (e.g., in a bike rental project currently under implementation).	**YELLOW**

Dimension	Importance	Findings for Transport Data in Mysore	Rating[41]
Actual and potential internal use of data	The extent to which transport agencies themselves use data to drive decisions and to report performance is a key indicator of whether data is regarded as important by the transport agencies in improving transport performance. Where there is potential for greater use of data and for additional data to be obtained for transport management, there may also be synergies with making that data available to others.	There was multiple evidence for the use of data for strategic, tactical, and operational purposes in Mysore. MUDA relied on extensive data collection from a number of Mysore agencies for a variety of planning objectives, including the Masterplan. These data included parking spaces and accidents from the police, bus routes from KSRTC, and 3D mapping of geospatial data from KSRSAC. KSRTC revised all their schedules as a result of the data produced by the ITS (where actual and perceived arrival differed significantly). KSRTC is also using the ITS system to do daily follow-ups on its operations. Finally, plans are in progress for a shared control facility between Mysore police and KSRTC. MCP has recently installed some 40 CCTV cameras around the city that are used both for traffic monitoring purposes and also to identify traffic violations. Accidents are collected manually and later entered into a computerized system. The data is subsequently used by MUDA for planning purposes.	**GREEN**

Dimension	Importance	Findings for Transport Data in Mysore	Rating[41]
Actual and potential external use of data	The value of data is in its use. A strong demand-side "pull" of data is important not only in creating and maintaining pressure on transport agencies to release data but also in ensuring that the wider Data Ecosystem develops and that data is turned into economically or socially valuable services for citizens.	Evidence found for limited information available for citizens and businesses. MUDA has recently begun sharing its forthcoming projects to the citizens through its website. However, these plans are only shared in formats (PDFs) that hamper value-adding re-use. KSRTC has just recently released a smartphone app with travel planning functionality, but its uptake is still to be evaluated given its recent launch. KSRTC also offers similar functionality through its website. There is currently no official real-time traffic information released to the public (however crowdsourced data is available through Google Maps). No information was found that is currently distributed to third-party developers. MCP currently shares accident reports with newspapers and other agencies on request. In addition, accidents are published on both the MCP web page and Facebook page.	**YELLOW**

Dimension	Importance	Findings for Transport Data in Mysore	Rating[41]
Data re-use community	Experience among leading agencies has demonstrated that Open Data initiatives are more sustainable and high impact when Open Data efforts use an "community" approach—meaning agencies invest not only in supplying data but also address the policy/legal framework, institutional readiness, capacity building (for government and infomediaries), citizen engagement, innovation financing, and technology infrastructure. Transport agencies need to play a multidimensional role in the transport data ecosystem and create new types of partnerships with a wide range of stakeholders.	There was multiple and strong evidence of a vibrant re-user community. Both KSRTC and BMTC have been approached multiple times by third-party developers making formal and informal requests for access to public transport data. These requests include a wide range of actors, from local students to international market leaders. KSRTC organized a well-attended appathon and recently launched the winning application. A roundtable discussion in Bangalore attracted participation of several third-party developers (on a short notice), where they expressed great interest in re-using transport data. Finally, Mysore and Karnataka have strong traditions of having a world-class ICT sector. Mysore is hosting both the Infosys Campus[46] and SJCE, educating more than 3,000 engineers (many in computer science and software engineering). In addition, SJCE is starting a Big Data Analytics research center together with IBM and has expressed great interest in working with transport data within the center.	**GREEN**

[46] https://en.wikipedia.org/wiki/Infosys#Training_centre_in_Mysore

Identified High-value Datasets

Identified high-value datasets are assessed on a three-point scale:

Quick win

The data are already held in a form suitable for publication as open data or from which it could reasonably be easily extracted with the skills available, and publication of the data would not raise new issues of policy (for instance because the data are already in the public domain but in non-reusable form) and would not involve a significant loss of revenue. These datasets are good candidates for an accelerated program of release to create momentum and to stimulate re-use of data within, say, the first six months of a transport data initiative.

Priority

The data is already held in a form from which it could be extracted for publication reasonably easily, but publication may require policy decisions and/or technical work such as anonymization or budgetary adjustment. These datasets may not be possible to release at the time of the launch, but the potential value of the data is such that priority should be given to resolving the issues and releasing the data within around a year after the launch of the transport data initiative.

Longer term

The data are important, but are not yet held in a form from which they could be extracted for publication, perhaps because a new data collection system would need to be developed or because there are difficult policy issues associated with the release of the dataset which would need to be resolved after extensive study and consultation

Table 4: Identified High-Value Datasets

Ref.	Dataset	Description	Key Benefits and Risks	Recommend Action
D01	Road network and conditions	Data on the road network and its conditions in Mysore are collected by MCC. MCC collects data on trunk and regional roads. The data being collected include position, surface, road type, length width, and condition. The data are collected through surveying. The data collection is headed by a zone commissioner that oversees the field staff that performs the actual data collection. The data are not stored in a geocoded format within MCC. KSRSAC has geocoded the collected information (using ArcGis) and is thus able to provide the basis for decision making for, e.g., MUDA. However, the data access is firmly restricted beyond other governmental agencies.	MCC expressed that they infrequently are asked by re-users for access to data on the road network and typically allow it. The requests are made by mail and MCC responds typically within one week. If the data do not exist, they will consider collecting it. Since the data are not available in a geocoded format, the delivery is typically performed through a word file. By publishing these datasets as is in an open manner, the manual work of handling these requests may stop. However, the datasets could be further enriched to provide additional value. As these maps are not currently geocoded, it hinders most value-adding re-use. Investing in GIS tools that would make such re-use possible would also provide new opportunities for data-driven decisions for MCC. Also, since the data are available within KSRCAC there would be no need for an initial conversion. This geocoded information could hence be shared more openly with re-users.	**QUICK WIN (as is)** **LONGER TERM (geocoded)**

Ref.	Dataset	Description	Key Benefits and Risks	Recommend Action
D02	Routes, fares, stops and timetables—Mysore city buses	MCTD operates some 400 buses within the city of Mysore. There is (semi) static information on routes, fares, stops, and timetables available, currently stored in a relational database. During the visit the assessment by local technical experts was that conversion of this static information to GTFS static format would not require substantial work.	This dataset converted to GTFS would present new data-driven opportunities for, e.g., application development, accessibility, and planning analytics. There are several specific recommendations and actions in this report to bring such publication forward.	**QUICK WIN**
D03	Real-time running data Mysore city buses	With the implementation of ITS systems, it is currently feasible to publish current vehicle locations (and possibly vehicle occupancy) as open data by using the installed in-vehicle GPS devices and smart-card data. These data are currently held in a relational database within MCTD.	Real-time information about public transport is a highly-valued dataset. Application developers typically urge operators to release these data, as they provide a means to calculate more accurate arrival and departure times, rather than planned ones (using predefined timetables). For instance, real-time running data give travelers a new decision-making basis (as passengers through the actual location can decide whether a certain departure is a feasible option). The preferred standard for these types of datasets are GTFS-RealTime. There are several specific recommendations and actions in this report to bring such publication forward.	**QUICK WIN**

Ref.	Dataset	Description	Key Benefits and Risks	Recommend Action
D04	Individual accident data (historical)	MCP records information on accidents such as potential fatalities and description of the accident. The data are currently collected manually but are subsequently entered into a data analysis system the same day as the accident occurs and the data are stored in a relational database. There is an ongoing project to move data collection onto handheld computers. Currently, the data are shared with newspapers and other agencies on request and published on the web page and Facebook page of the Mysore police.	Publishing a georeferenced and anonymized accident database entails several benefits. First, it reduces the administrative work currently performed to distribute this dataset to interested governmental and nongovernmental stakeholders. Second, making it public in a machine-readable way opens up various visualizations, safety planning measures, and meshing with other datasets (e.g., weather) to reveal insofar unaddressed patterns. As this dataset in its raw form contains sensitive privacy dimensions, it is important that the procedures for anonymization are in place.	**QUICK WIN**

Ref.	Dataset	Description	Key Benefits and Risks	Recommend Action
D05	Parking	Mysore City Police records data on the parking spaces in Mysore. While the visit did not allow for a deeper investigation of the collection, management, and sharing of parking spaces we found evidence that the data are available and appear to be sufficiently detailed for publication.[47] The assessment team was not able to determine in what format the data were currently stored.	Parking availability is a highly-valued dataset and typically one of the most re-used datasets when cities publish open data. Typically, parking data are transformed into consumer apps that utilize the smartphone GPS to locate nearby parking spaces. In addition, if the parking data specify additional details around the parking such as type of vehicle, accessibility options, and payment types, the data adds significant value to the dataset. While this dataset carries significant value, it is important that any changes to parking availability are properly collected and managed. If the data are not updated in a timely manner, the confidence in applications using the open data is at stake with the value of this dataset being depreciated.	**PRIORITY**[48]
D06	Traffic counts/traffic speed measurements	Currently there are no real-time traffic counts or speed measurements available at Mysore. However, the traffic police have 40 operational CCTV cameras being used by the police control room to detect traffic violations. These cameras could be used to collect real-time traffic counts. In addition, the AVLs from KSRTC's ITS could be shared and utilized for real-time traffic speed measurements.	Traffic counts can generate applications that indicate the current traffic situation in a city but also serve as a basis for historical analysis and insights. To this end, traffic counts and traffic speed measurements are often part of a city's open data program.	**LONGER TERM**

[47] http://justmysuru.com.md-in-23.webhostbox.net/mudamysore.gov.in/MasterPlan/MP_Reports/VOL 1_DATA COLLECTION, ANALYSIS & PROJECTIONS.pdf (p. 78)

[48] This recommendation is based on the limited evidence currently at hand. However, a closer investigation of the data available and the data collection and management procedures could yield a different recommendation.

Ref.	Dataset	Description	Key Benefits and Risks	Recommend Action
D07	Rental bike facilities and real-time availability	Mysore was the first city in India to launch an electric bike rental system[49] that commenced operations in July 2016. The system is owned by MCC with support from DULT, and the bike rentals are operated by Green Wheel Ride (on a 6-year contract). The system comprises both the location of the cycle hire facilities and real-time availability, and the data are owned by MCC.	Publishing rental bike station locations and real-time availability has several benefits. First, it enables integration into existing information outlets for tourists and city citizens such as apps, websites, and signage. In addition, it provides a useful component in producing multi-modal travel applications. As these types of data have a low degree of standardization, it may lead to low uptake of non-local third-party developers. One way to mitigate this is joining the system around pyBikes[50] that currently acts as an aggregator on more than 400 cities rental bike data. In addition, the publishing agency should consider investing in the emerging standard GBFS[51] (a rental bike version of GTFS).	**PRIORITY**

[49] http://greenwheelride.com/index.php/rent-an-ebee

[50] See http://citybik.es/ for more information

[51] https://github.com/NABSA/gbfs

Annex 1: The Tiered Model of Stockholm Public Transport

The tiered-based model used in Stockholm, Sweden, concerns a common dilemma for transport data publishers: significant take-up and re-use of (real-time) transport data that may infer high hosting costs and problematic scaling issues for legacy systems producing the data. To this end, transport data publishers may need to impose restrictions regarding the frequency at which data are retrieved.

Stockholm Public Transport is using a tiered model for re-users (bronze, silver, gold).[52] Every interested re-user is granted a Bronze level license. The Bronze level (max 30 requests/minute and 10,000 requests/month) is sufficient for development and experimental purposes (and in some cases also for a deployed application). Once an app or other digital service is operational and the re-users deem the Bronze level as insufficient, they may apply for the Silver level (60 requests/minute and 500,000 requests/month). This is a very streamlined process where the re-user is required to argue for the value the service creates for travelers and where Stockholm Public Transport subsequently assesses the application. In practice, however, the absolute majority of applicants is granted the Silver level. In that case, the Silver level also proves insufficient, the re-user may apply for a Gold level license where no formal limits on requests exist. Instead, the integrations are tailored in close collaboration between Stockholm Public Transport and the re-user to optimize the integrations. The Gold level license to date is only granted to a handful of re-users.

The main benefit of this model is that API requests are considered a finite resource that are (roughly) distributed in accordance with user value. A potential downside of this more pragmatic approach is that re-users are left to the discretion of Stockholm Public Transport in case they need additional API requests (compared to a "freemium" model where any re-user may pay for additional API requests beyond the free level). In this sense, this model might not be acceptable in more ideologically driven approaches to open transport data.

The main reason for Stockholm Public Transport to use this model is that they have been keen on restricting the overall number of API calls while not imposing additional costs onto successful re-users (as they provide much traveler value yet often suffer from weak business cases). To this end, they have instead chosen traveler value as an API request distribution mechanism. (On a side note, Stockholm Public Transport is currently working to strengthen the business case of re-users by allowing ticket sales through APIs where re-users are given a commission for each sold ticket.)

[52] https://translate.google.com/translate?sl=auto&tl=en&js=y&prev=_t&hl=en&ie=UTF-8&u=https://www.trafiklab.se/api/sl-realtidsinformation-3&edit-text=&act=url

Annex 2: General Transit Feed Specifications (GTFS)

The simple GTFS file structure has prompted high levels of adoption and made it the standard format for transit data. As of June 2016, an estimated 1,048 transit operators (including 326 agencies) had released official GTFS feeds.[53] Most of the feeds are from operators in the United States, Canada, Europe, Australia, New Zealand, and Japan, but experience with GTFS is also growing in Bank client countries.

GTFS describes a standardized format for public transit data, which requires data to be released as a package of comma-delimited text files, each containing one aspect of the transit information and a set of rules on how to record it. Six files are mandatory (agency, stops, routes, trips, stop times, and calendar) and seven optional (calendar dates, fare attributes, fare rules, shapes, frequencies, transfers, and feed info). To accommodate the varied and changing nature of transit services, Google continuously updates the GTFS specification by adding extensions, optional fields, and additional valid responses, which then can be used for different applications (see Table 5). While initially the GTFS file format only managed static transit information, such as routes, stops, and schedules, with the release of GTFS-RealTime (or GTFS-RT), dynamic information on real-time bus locations and service disruptions (using data from on-board Automated Vehicle Location (AVL) systems) can also be shared. Although the GTFS standard is not owned or managed by Google, its active involvement and ongoing support for GTFS modifications adds to the institutionalization of the data format. In Europe, real-time transport data have been managed already for some time using a different standard, SIRI (Service Interface for Real-time Information), a real-time transport data protocol initially developed by France, Germany, the United Kingdom, and the Scandinavian countries.

From a user perspective, the emergence of GTFS coincides and aligns with the emergence of a broader open data/open governance movement; in this broader context, the data can be used to not only access transit data for personal use, but also monitor and evaluate the provided services. The open and standardized reporting format, for example, allows citizens and transit users to compare scheduled services with their own experiences. With GTFS-RealTime, users can even monitor actual performance of service delivery. This ability for citizen engagement, enabled by the GTFS protocol, is in itself an important development and one that governments are paying increasing attention to. In Brazil, for example, citizen dissatisfaction with public transport service quality and tariffs was one of the catalysts for the protests that paralyzed many Brazilian metropolitan areas in June of 2013. In the wake of the protests, cities across Brazil have focused on opening up transport service data, and in recent months both the municipalities of Sao Paulo and the State of Rio, which manage inter-municipal services in the Rio metropolitan area, have

[53] GoogleTransitDataFeed. List of Publically Accessible Transit Data Feeds. https://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds. Accessed May 2016.

promised open access to service data. In the case of Rio, supported by a World Bank policy loan, data will be released using the GTFS protocol.

A number of tools have been created that advantage the GTFS standard to provide better information about the public transportation network, data analytics on the performance of the transportation network, timetables, and accessibility analytics, etc. Some of these applications are highlighted below.

Table 5: Uses of GTFS Data[54]

Category	Description	Examples
Trip planning and maps	Applications that assist a transit customer in planning a trip from one location to another using public transportation	Google Maps, OpenTripPlanner, Bing Maps, Hopstop, Transit App for iOS, Nokia transport, RouteShout, Tiramisu
Ride sharing	Applications that assist people in connecting with potential ride sharing matches	Parkio, Avego
Timetable creation	Applications that create a printed list of the agency's schedule in a timetable format	TimeTablePublisher
Data visualization	Applications that provide graphic visualizations of transit routes, stops, and schedule data	Walk Score, Apartment Search, Mapnificent, Analyst extension for OpenTripPlanner
Accessibility	Applications that assist transit riders with disabilities in using public transportation	Sendero Group BrailleNote GPS, Travel Assistant Device
Planning analysis	Applications that assist transit professionals in assessing the current or planned transit network	OpenTripPlanner analyst extension, Graphserver, Transit Boardings Estimation and Simulator Tool, TransCAD 6.0

[54] Modified from A. Antrim and S. Barbeau, "The Many Uses of GTFS Data: Opening the Door to Transit and Multimodal Applications," http://www.locationaware.usf.edu/wp-content/uploads/2010/02/The-Many-Uses-of-GTFS-Data-%E2%80%93-ITS-America-submission-abbreviated.pdf

Interactive Voice Response (IVR)	Applications that provide transit information over the phone via an automated speech recognition system	BusLine, TransitSpeak, TravelSpea
Real-time transit information	Applications that use GTFS data along with a real-time information source to provide estimated arrival information to transit riders	OneBusAway, NextBus, TransLoc, Moovit, next bus arrival signs at bus stops (e.g., in Santiago, São Paulo)
Dedicated SMS applications	Applications designed for feature phones without data capabilities	RouteShout, Transantiago's SMS Bus