015

Transport

Pieter Colpaert and Julián Andrés Rojas Meléndez

Key points

- Public transport has been a poster-child of the open data movement with a variety of route planning applications used by millions of people every day. Transport data can also be used to analyse policy and advocate for service improvements.

- Tensions exist between centralised route planning services and distributed, open data-driven approaches to transport data. Only a fraction of the data used to drive mobility apps is truly open, and current technical architectures risk holding back a next wave of innovation.

- Data-driven transport tools have been developed worldwide; however, established standards need to be more flexible in order to accommodate semi-structured and informal transport networks in the developing world.

- The future success of "Mobility as a Service" will depend on a much greater range of open transport data and application programming interfaces (APIs).

Introduction

How far do you live from your place of work? Was your answer a distance or was it a duration dependent upon a specific mode of transport? The question of how far you can go, and how long it takes to go from one location to another, is key to identifying the opportunities you and your family can take advantage of. The amount of data an application could use to support an answer to this question is beyond imagination. Details of road networks, live public transport timetables, and even wheelchair accessibility of public buildings, are just a few of the applicable datasets.

Urban planners, real estate developers, travel application developers, and even manufacturers of autonomous vehicles, all need this kind of information to make their services better. For some, the availability of this data is even a primary condition for operation. Take the Dutch

company, GoOV,[1] for example, which aids people with a mental disability to get home safely and autonomously using public transport. Without access to live transport tables, they would not be able to offer these services.

Transport apps have served as a poster child for the open data movement, with route planning apps, such as CityMapper, Transit App, or Google Maps often appearing in presentations on the benefits of open data. In 2014, the International Association for Public Transport (UITP) made open data the main subject of a focus paper,[2] and the association featured open data talks in its IT-Trans conference. A year later, the American Public Transport Association (APTA) published a Policy Development and Research Paper on embracing open data.[3] Although these developments indicate significant traction to date on open transport data, gaining the disclosure of transit data has not been straightforward. As pointed out in studies by Rojas[4] and Colpaert et al.,[5] many cultural, technical, and legal obstacles have had to be overcome.

While transport data is hard to define, this chapter will focus on data that can be used by route planners and on three main challenges:

1. Route planning – determining who does what and how transport data is licensed.

2. The accessibility and availability of datasets.

3. Emerging technologies such as Mobility as a Service (MaaS) and autonomous driving.

From schedule data to advice on route planning

Route or trip planner apps advise consumers on how to get to a specific destination. Travel information is displayed using a plethora of interfaces from in-car navigation systems to the website of a local bus company or a third-party travel app. In-car navigation systems may weight data elements differently when providing route planning advice when compared with an application from a municipal transit agency, yet both of them need access to the same data.

The data needed to create these types of applications resides inside the organisations that manage and operate public transport networks, and, due to the high degree of heterogeneity that can be found from one organisation to another in terms of how they manage data, opening and using this data can be a big challenge. To address this issue, several standardisation efforts have arisen around the world to support public transport operators in openly sharing their data in an interoperable fashion. Standards, such as the General Transit Feed Specification (GTFS),[6] the European Network Timetable Exchange (NeTEx),[7] the Standard Interface for Real-time Information (SIRI),[8] or the American Transit Communications Interface Profiles (TCIP),[9] provide mechanisms to model and describe scheduled services and real-time updates from transport networks, including arrival predictions, vehicle positions, and service advisories in machine-readable formats.

Some operators also offer route planning application programming interfaces (APIs), which function as open innovation tools, encouraging the creativity of partners who need access to route planning information quickly. However, offering a public route planning API comes at a cost. API providers need to consider whether they are able to respond to all queries with the consequent server bill for each request. As a consequence, route planning APIs are often only available via registration, API keys, and rate limiting. Open data advocates would not call this truly open data as data users are not in control of the algorithms that modify, filter, and operate on the data that is finally exchanged via the API. In a truly open transport data ecosystem, everyone would be able to create their own specific route planning API based on all data being published as open data first.

Transport for London

Transport for London is the local government body responsible for the transport system in Greater London, England, and is commonly cited as a source of open data success stories. When Transport for London began opening up data and offering public APIs, the economic growth potential was estimated at GBP 130 million,[10] with more than 600 apps created and more than 500 people directly employed in the reuse of public transport data. Transport for London now focuses primarily on publishing the data and not on building their own route planning apps.[11] However, they still publish both the raw timetable data as well as a unified API. Read more at https://tfl.gov.uk/info-for/open-data-users/.

The manner in which data is published also reflects legal constraints. In 2016, Scassa and Diebel[12] published a paper in which they, from a legal perspective, argued that publishing real-time data as open data is troublesome. Indeed, when a route planning API is offered, a Service Level Agreement (SLA) is needed, guaranteeing the up-times of an expensive but free service. When, however, the raw data is published via downloads or file updates, the effort required by the publisher is lower and, thus, easier to guarantee.

Public authorities and non-governmental organisations (NGOs) also play a key role regarding open data in the public transport ecosystem. Public authorities provide the legal framework and regulations that drive public transport organisations to pursue open data strategies. They also may provide technical and standardisation guidelines for data publishing that help to achieve

greater interoperability. NGOs are avid users of public transport open data, which they use for different kinds of studies and data analysis that aim to shed light on social issues and potential solutions. For example, a study[13] conducted by the non-profit organisation, Despacio, on the current status of, and trends in, bike mobility in Bogotá (Colombia) relied on open data provided by the Secretaría de Movilidad of Bogotá to highlight the main challenges and gaps in terms of security and infrastructure for the growing number of bike users in the city. They also used this data, together with the public transportation routes information, to generate a mobility coverage map of the city. Another example is the study performed by the Public Knowledge Workshop, an Israeli NGO that facilitates open data initiatives, which used schedule and live update data from Israeli railway and bus companies to verify their operational synchronisation.[14] They revealed that despite the presence of an official government plan requiring joined-up scheduling, there was little synchronisation in practice between the trains and the corresponding buses that were supposed to deliver and pick up passengers to and from their trains. These open data-based studies provide a vital resource for urban planners to better design and plan the development of cities and for social organisations that work toward improving living conditions in cities.

Emerging technology

In 2015, Linked Connections was put forward as a middle-ground route planning solution, moving beyond the false dichotomy between data dumps and route planning APIs.[15] With Linked Connections, route planning happens on the infrastructure of the data user, but data is already prepared for the purpose of route planning by the provider. At the basis of the technology lies the same idea as behind Content Delivery Networks (CDN). By creating small fragments of data about the departures of public transport vehicles in cacheable documents, the raw data needed by users is published cost efficiently. The goal of the framework is to enable a new open source route planning ecosystem based on web querying. Further information is available at https://linkedconnections.org.

Toward global coverage: The need for accessible data

Although there have been major steps in opening up transit data in the last decade, building a global route planner that includes all public transport modes in the world remains close to impossible. The amount of effort and money required for such an endeavour exceeds what governments and companies are willing to invest. The obstacles are diverse, including technical, legal, and financial barriers, but the availability and accessibility of the required data is paramount.

The majority of public transport companies in the world still do not provide their schedules as open data, and even fewer publish live transit updates in machine-readable formats. Therefore, it is not possible to automatically include such data in a global route planning application. One approach to tackling this data gap might be the use of applications that crawl through transport provider websites and scrape schedule information. This kind of approach demands a high

effort, as for every company, there must be an ad hoc implementation of the scraper to extract data. Furthermore, there are often legal uncertainties as to whether scraping transport websites is legal in a particular jurisdiction.

Despite the relatively low availability of data and legal uncertainties around scraped data, there are still some entrepreneurs and established businesses that have been addressing this titanic challenge. The most famous, and notorious, is Google Maps. Google uses the GTFS specification that they maintain, together with a global community of developers in order to import data on different transport modes and networks into their route planner. They encourage public transport companies to generate and deliver their data in this format, but Google does not require the data to be openly published. Sometimes they will work out a direct arrangement with the public transport operators as is the case for the urban bus company, Transmilenio in Bogotá (Colombia), where the operator hires an external company to generate and deliver the GTFS feed to Google without publishing it for public access. According to the Google Transit website, they currently support 5 640[16] different transport companies within their route planning application that covers over 18 000[17] different cities around the world.

There are several other examples of applications and services that reuse transport open data and that seek to provide a global route planner, such as CityMapper, Transit App, Ally, Moovit, among others. Some of them even try to generate their own data to include cities and transport networks that do not publish their own data (e.g. CityMapper and their work on Mexico City and Istanbul).[18] Navitia[19] makes an API available that currently contains 434 transport datasets from around the world from which developers can use route planning features, generate maps of time/distances, and access timetables. They take advantage of publicly available open data and encourage users to provide new data sources. However, Transitland is potentially the largest catalogue of open transit data,[20] which reports 945 open GTFS feeds, covering 2 377 different public transit operators at the time of writing.

Mexico City

In Mexico, a GTFS feed was introduced to take advantage of the collection of GPS data throughout its transit systems. In a matter of weeks, this mega-city with several different transit providers was able to introduce a fully functional GTFS feed and obtain the benefits of work done on route planning tools elsewhere. A range of free or low-cost customer-facing applications and planning tools were able to immediately capitalise on this data.

Problematic, however, is the fact that part of the public transit system in Mexico is only semi-structured, meaning that some services do not have fixed stops, nor a defined timetable. The project revealed an important limitation of GTFS in its current form as it is unable to easily accommodate the kind of semi-structured public transit services that operate in many developing world cities.

Eros et. al (2014) have detailed the experience in a full paper for the Transportation Research Board.[21]

By providing a standardised way to model and describe public transport time schedules in machine-readable formats, GTFS has become one of the most important tools to increase the amount of available open data in the transport sector. However, it has some notable limitations when working toward global coverage of transport data. It was originally designed to model structured networks that define a set of fixed stops for vehicles and that run on predefined time schedules that are often specified down to the second. But this is not the case for most of the public transportation services offered in the major cities of the Global South, where operators may define a set of routes that are followed by a set of vehicles but without predefined fixed stops. This type of limitation in the modelling capabilities of the available standards adds difficulty to both standards and open data adoption in these parts of the world. Moreover, public transport operators in developing countries often have few incentives to provide data about their operation, and public authorities may lack the necessary regulatory framework and resources to drive or support these organisations in publishing open data.

To address these shortcomings, and to promote the wider implementation of open data initiatives, a number of different approaches have arisen. For instance, the GTFS-flex[22] specification, created and maintained by the independent developer community, is a proposed extension for GTFS that aims to provide the capabilities for modelling semi-structured public transport and demand-responsive transportation services. In Kenya, the Digital Matatus project[23] has made use of mobile communication and geolocation technologies to map and generate a GTFS data source for the semi-structured public transport service in Nairobi, which has proven to be a feasible mechanism to fill the gap when data on these types of transport networks is not available from official sources. Following this initiative, the Digital Transport for Africa community was created, which has supported open data generation projects for public transport services in Cairo, Maputo, Accra, and Abidjan.[24] Similarly, the World Bank began offering a course to empower participants to create, manage, and use GTFS feeds in resource-constrained environments.[25] It is important to note that these types of initiatives help to increase available open data for the transport sector, but they still require significant investment and political will from the public authorities in the developing world.

Today, there is evidence that disclosing public transport data can generate many benefits for different actors, including developers, entrepreneurs, users, and transport companies, and the discussion is no longer centred on whether data should or should not be openly published. The resistance still encountered around the world to engaging with open data is attributed more to a matter of the political will of organisations. Policies promoted at the national, regional, and local levels can play an important role in increasing the implementation of open transport data initiatives. One clear example of such promotion is the Intelligent Transport System (ITS) Directive[26] of the European Union. The directive aims to accelerate the deployment of innovative transport technologies across Europe, and the public accessibility of data is one of its main requirements, indicating that both policy and research discussions about open data in the public transport sector have now moved to a technical and a legal level. The key questions to address in scaling coverage relate to how transport data should be published to improve interoperability, while keeping costs to a minimum, as well as how to address legal considerations to protect the interests of involved parties, without limiting open data benefits.

World-wide and open source – Transportr

Open source route planning software exists today, such as Open Trip Planner, OSRM, Navitia, or RRRR, and many companies, like Plannerstack, Conveyal, Digitransit, and Kisio Digital, make use of this open source software to provide services to their clients. Navitia.io, based on the Navitia code-base, is a freemium SaaS solution for route planning. Transportr reuses this service to create a fully open source and free app with the data available via the web-services. Read more at https://transportr.grobox.de/.

Mobility as a service: An emerging challenge

Mobility is always a core point of discussion in urban planning. Ever since its introduction in the early 20th century, cities have been adapting to, or have been "taken hostage by", as some would proclaim, the car as the primary means of transport. The continued dominance and density of cars, and their negative environmental and social impacts within urban environments, has created a sense of urgency around the need to diversify the way we move from one place to another. Yet statistics on car use will not trigger a worldwide change by themselves. In order to change dominant behaviour, mobility activists and entrepreneurs have coined the term Mobility as a Service (MaaS). This new idea tries to activate people to leave their cars behind and diversify their mobility choices by means of an app. Instead of having to use multiple apps to find routes and buy tickets for each different mode of transport, an ecosystem for all-in-one solutions must be built.

In order to grow a MaaS ecosystem in a certain region, three requirements need to be fulfilled. The first is that the data needs to be available on where and when specific services can be used. Given the low availability of open transport datasets today, the MaaS movement is also an important advocacy force for open data, arguing that every mobility player, whether public or private, needs to publish their data in order to create a truly level playing field for MaaS.

In Belgium, for example, an Open Data Charter was created in 2018 by local governments and regional governmental institutions[27] that lays out 20 principles for open data, including the 19th principle stating that data resulting from a government concession should be open as well. Local governments adopting such a principle may push forward the agenda of open data and MaaS worldwide.

The second requirement for MaaS is that an open ticketing API must be in place. The more you allow third parties to sell your tickets, the more integration can happen with other mobility solutions. An open ticketing API may allow tickets to be granted to users in various ways (e.g. per hour, per km, etc.). As evidenced by the low availability of fare data in general, it is certainly early days, yet this is an area that is currently rapidly evolving. In Finland, for example, an API for ticketing has been created that can be used by anyone to buy tickets without signing complicated contracts. This allows apps created by vendors, such as MaaS global,[28] to start selling tickets as a third party.

Finally, open data and open ticketing alone are not going to create a seamless travel experience for end-users. As a third condition, a city needs to prepare itself for multimodality. Infrastructures

need to be better aligned with public and private transport offerings. Different enablers exist for brainstorming solutions in this area, such as Open Transport Camp in Australia[29] and TransportationCamp in the US,[30] or initiatives such as the MaaS alliance,[31] Fabrique Mobilité,[32] or Mobihubs[33] in Europe. Ultimately, it will take multiple data communities to output the policy, planning, and programmes of action that will truly reshape public space and mobility.

Conclusion

There is evidence worldwide that transport data is being released as open data, whether it is through crowdsourcing initiatives as in Mexico City or through official public transport or governmental organisations. In the US, thanks to the APTA, and in European countries, thanks to the ITS, Public Sector Information (PSI), and INSPIRE directives, policies are pushing the agenda for open transport data forward.

Now that the benefits of sharing public transport data openly are becoming visible through apps that can immediately turn these datasets into route planners, the way data is shared needs to evolve technically. The current de facto standard for sharing data via GTFS still requires a big investment from users before the data can be used in a route planner, and only a fraction of the data that exists in GTFS format is publicly available as open data. The true potential of open transport data is yet to be unlocked, although as the integration costs of transport data decrease and more data is made available, there is scope for substantial progress to be made.

Open data alone is not going to create a big change in how people move from one place to another. Advancement of MaaS will need to combine concepts of open data, support for open ticketing, and work on infrastructure investments in order to diversify the availability of transport options. It is up to policy-makers to create the right environment and infrastructure to properly prepare cities for the mobility of the future.

Further reading

Colpaert, P., Compernolle, M.V., Walravns, N., Mechant, P., Adriaenssens, J., Ongenae, F., Verborgh, R., & Mannens, E. (2017). Open transport data for maximising reuse in multimodal route planners: A study in Flanders. *IET Intelligent Transport Systems*, *11(7)*, 397–402. https://ieeexplore.ieee.org/abstract/document/8061184

Eros, E., Mehndiratta, S., Zegras, C., Webb, K., & Ochoa, M.C. (2014). Applying the General Transit Feed Specification to the Global South: Experiences in Mexico City, Mexico – and beyond. *Transportation Research Record*, *2442(1)*, 44–52. https://doi.org/10.3141/2442-06

Hogge, B. (2016). *Transport for London: Get set, go! Open data's impact*. GovLab and Omidyar Network. http://odimpact.org/files/case-studies-transport-for-london.pdf

Rojas, F.M. (2012). *Transit transparency: Effective disclosure through open data*. Transparency Policy Project. Cambridge, MA: Ash Center for Democratic Governance and Innovation. http://www.transparencypolicy.net/assets/FINAL_UTC_TransitTransparency_8%2028%202012.pdf

About the authors

Pieter Colpaert is a Researcher at Ghent University's Internet and Data Research Lab. His research focuses on enabling route planning at a large scale, using linked data. He is a board member of the Belgian chapter of Open Knowledge and a community coordinator of the open transport working group at Open Knowledge International. You can learn more about Pieter's work at http://pietercolpaert.be.

Julián Andrés Rojas Meléndez is a Researcher at the University of Ghent working on interoperable open data publishing strategies on the Web and decentralised route planning with open data. You can follow Julián at https://www.twitter.com/julianr1987.

How to cite this chapter

Colpaert, P. & Rojas Meléndez, J.A. (2019). Open data and transportation. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The state of open data: Histories and horizons* (pp. 215–224). Cape Town and Ottawa: African Minds and International Development Research Centre. http://stateofopendata.od4d.net.

Endnotes

1   http://go-ov.nl/

2   UITP. (2014). *Action points for the public transport sector. The benefits of open data*. Brussels: Union Internationale des Transports Publics. https://www.uitp.org/sites/default/files/cck-focus-papers-files/AP%20-%20Benefits%20of%20open%20data%20EN.pdf

3   APTA. (2015). *Public transportation embracing open data*. Washington, DC: American Public Transport Association. https://www.apta.com/resources/reportsandpublications/Documents/APTA-Embracing-Open-Data.pdf

4   Rojas, F.M. (2012). *Transit transparency: Effective disclosure through open data*. Transparency Policy Project. Cambridge, MA: Ash Center for Democratic Governance and Innovation. http://www.transparencypolicy.net/assets/FINAL_UTC_TransitTransparency_8%2028%202012.pdf

5   Colpaert, P., Van Compernolle, M., Walravens, N., Mechant. P., Adriaenssens, J., Ongenae, F., Verborgh, R., & Mannens, E. (2017). Open transport data for maximising reuse in multimodal route planners: A study in Flanders. *IET Intelligent Transport Systems*, *11(7)*, 397–402. https://ieeexplore.ieee.org/abstract/document/8061184

6   http://gtfs.org

7   http://netex-cen.eu/

8   http://www.transmodel-cen.eu/standards/siri/

9   http://www.aptatcip.com/

10 Deloitte. (2017). *Assessing the value of TfL's open data and digital partnerships*. London: Deloitte. http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf

11 Hogge, B. (2016). *Transport for London: Get set, go! Open Data's Impact*. GovLab and Omidyar Network. http://odimpact.org/files/case-studies-transport-for-london.pdf

12   Scassa, T. & Diebel, A. (2016). Open or closed? Open licensing of real-time public sector transit data. *JeDEM – EJournal of EDemocracy and Open Government*, *8(2)*, 1–20. https://jedem.org/index.php/jedem/article/view/414

13   Verma, P., López, J.S., & Pardo, C. (2015). *Bicycle account: Bogota 2014*. Bogotá: Despacio. http://www.despacio.org/wp-content/uploads/2015/01/Bicycle-Account-BOG-2014-20150109-LR.pdf

14   See the OpenTrain project at http://otrain.org/ and Open Bus project details at http://www.hasadna.org.il/en/projects/

15   Colpaert, P. (2017). Publishing transport data for maximum reuse. PhD Thesis, Ghent University, Belgium. https://phd.pietercolpaert.be

16   https://maps.google.com/landing/transit/cities/

17   https://maps.google.com/help/maps/mapcontent/transit/participate.html

18   Citymapper. (2015). Building a city without open data. https://medium.com/citymapper/building-a-city-without-open-data-124356672deb

19   https://www.navitia.io/

20   https://transit.land/feed-registry/

21   Eros, E., Mehndiratta, S., Zegras, C., Webb, K., & Ochoa, M.C. (2014). Applying the General Transit Feed Specification to the Global South: Experiences in Mexico City, Mexico – and beyond. *Transportation Research Record*, *2442(1)*, 44–52. https://doi.org/10.3141/2442-06

22   https://github.com/MobilityData/gtfs-flex

23   Williams, S., White, A., Waiganjo, P., Orwa, D., & Klopp, J. (2015). The Digital Matatu Project: Using cell phones to create an open source data for Nairobi's semi-formal bus system. *Journal of Transport Geography*, *49*, 39–51. http://www.sciencedirect.com/science/article/pii/S0966692315001878

24   http://digitaltransport4africa.org/

25   https://olc.worldbank.org/content/introduction-general-transit-feed-specification-gtfs-and-informal-transit-system-mapping

26   https://ec.europa.eu/transport/themes/its/road/action_plan_en

27   Smart Flanders. (2018). Open Data Charter. https://smart.flanders.be/

28   https://maas.global/

29   http://www.transportcamp.org.au/

30   http://transportationcamp.org/

31   https://maas-alliance.eu/

32   http://wiki.lafabriquedesmobilites.fr/wiki/Ev%C3%A9nements

33   https://mobihubs.eu/