The Role of Open Data in Sustainable Transport
Content

1. What is Open Data? ........................................................................................................................................................... 4
  1.1 Definition................................................................................................................................................................................................... 4

2. Open Data and Transport .................................................................................................................................................. 5
  2.1 Transit Stops (stops.txt) ................................................................................................................................................................. 6
  2.2 Transit Routes (routes.txt) ............................................................................................................................................................ 7
  2.3 What can it be used for? ............................................................................................................................................................. 8
  2.4 Case Studies........................................................................................................................................................................................... 9
    London ................................................................................................................................................................................................... 9
  2.5 Open Data in Developing Countries ...................................................................................................................................... 10
    Nairobi, Kenya .................................................................................................................................................................................. 11
    Mexico City, Mexico .................................................................................................................................................................... 11

3. Why Use Open Data in Transport? ............................................................................................................................ 12

4. Further Resources ............................................................................................................................................................. 13

5. Sources ................................................................................................................................................................................... 13
1. What is Open Data?

What does it mean for a city or agency to open its data? How does Open Data influence people in their everyday life? Through digitalization and new technologies, vast amounts of data are already being collected by governments and although much of that data is public by law, it is often not yet available in a form that is easy to use. The concept of “open data,” or free access to information, has been emerging in different fields and has started to become mainstreamed since 2009 when various governments announced initiatives towards opening up their public information. Still, this trend is still new to many people. A lack of knowledge and clarity can create hesitation and wariness towards opening data when, in fact, open data may help to make governments more transparent and efficient and could unlock the potential of official data to enable new services that, in turn, improve life of citizens.

1.1 Definition

Governments and organisations generate data in the form of documents, databases, records, transcripts, etc., but historically this information has not always been available or open to the broader public. Opening data to the public can drive internal efficiency, spark community engagement and fuel a civic tech ecosystem. Open data can be freely used, reused and redistributed by anyone, generally only requiring reference to the original source. In that way, opening up data can also inspire interoperability between various actors of different levels, for instance inter- and intra-governmentally, regionally, nationally or globally.

In order for data to be considered open, it must be:

1. **Complete** – all public data is made available, and is not subject to valid privacy, security or privilege limitations.

2. **Primary** – data is collected at the source, with the highest possible level of granularity.

3. **Timely** – it is made available as quickly as necessary to preserve the value of the data.

4. **Accessible** – it is available to the widest range of users for the widest range of purposes. It must be available on the Internet.

5. **Machine processable** – it is structured to allow automated processing.

6. **Non-discriminatory** – it must be available to anyone, with no registration requirement.

7. **Non-proprietary** – it is available in a format over which no entity has exclusive control.

8. **License-free** – it is not subject to any copyright, patent, trademark or trade secret regulation.

---

1 Civic Technology encourages citizens to participate in public good development, enhances citizen communications and improves government infrastructure. Developers create matching apps using the data provided by the government on open APIs in non-proprietary formats. Later these new solutions can be included in an open government platform to be used by both, citizens and governments.

2 Definition taken from Open Government Working Group (http://opengovdata.org/)

3 http://www.webopedia.com/TERM/B/big_data.html
2. Open Data and Transport

Public transport users increasingly demand smarter cities that provide digital information, particularly real-time updates on their journeys. In the context of transport, the concept of smart cities revolves around a more integrated approach of data and urban transport provision. Open data can save both time and money in collecting such information about routes. Real-time service information including route and stop locations, passenger volumes by location and time of day, planned schedules, service disruptions, pricing and fare products, and average travel and dwell times can all be collected automatically or manually by staff. This information can then be automatically uploaded to an accessible, central server. In comparison, traditional methods require far more staff hours, manual work to record and upload information and advanced computer knowledge of programs such as TransCAD or GIS.

Opening data can empower resource-constrained transport agencies to collect high-quality transport data with minimal effort and cost, as well as to conduct robust data analyses with minimal formal training in transport engineering and planning. In addition, providing access to official data for developers can support employment, entrepreneurship and user input in the development of new and innovative technology applications for users – benefits that government agencies may not have the time, resources or expertise to develop. Overall, open data can thus improve the efficiency and effectiveness of government services and create new products and services that incorporate knowledge from combined data sources and patterns. Open data does not only benefits individual transport users through the provision of personalised and relevant travel information, but also the economy, as the economic value of open data is estimated at several tens of billions of Euro annually in the EU alone. In the long term, Open Data is expected to create a virtuous cycle, where new products further increase the demand for Open Data which in turn catalyses the release of more data, services, and applications.

The General Transit Feed Specification (GTFS) has emerged as the standard to release public transit data around the world. GTFS was co-developed by Google and TriMet, the transit agency in Portland, Oregon.

Types of Data Collection

- Traditional: household surveys, field traffic counts
  - time-consuming and labour intensive
- Intelligent Transportation Systems (ITS): dedicated, sensor-based systems
  - costly, high technical requirements
- Open Transport: crowd-sourcing, existing ICT systems (cellular networks, Internet connectivity), open-source, license-free applications

*Open Knowledge Foundation, 2012*
What is the General Transit Feed Specification (GTFS)?

GTFS is a standardised free and open template for entering data related to basic transit system services. It consists of a package of comma-delimited text files for use in a spreadsheet application, each of which contains one aspect of the transit information and a set of rules on how to record it. There are six mandatory files (agency, stops, routes, trips, stops times, and calendar) that every agency must provide and seven optional files (calendar dates, fare attributes, fare rules, shapes, frequencies, transfers and feed info).

Example GTFS Feed (source Google Transit)

Each aspect of transit information has a required format that allows developers and applications to work compatibly across cities and datasets. These templates are provided by Google Transit. An example of the information is below:

<table>
<thead>
<tr>
<th>Transport Instrument</th>
<th>Traditional Method</th>
<th>Open Transport Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIS Route and Station/Stop Locations</td>
<td>• Collect data using dedicated GPS device</td>
<td>• Staff ride transit route using mobile app</td>
</tr>
<tr>
<td></td>
<td>• Manually upload data to desktop computer</td>
<td>• Enter route and stop details using the app as they ride</td>
</tr>
<tr>
<td></td>
<td>• Use specialised GIS software to relate collected data to city's road network, enter details about the route</td>
<td>• Data and meta-data automatically uploaded to accessible, central server</td>
</tr>
<tr>
<td></td>
<td>• Manually enter route meta-data</td>
<td>• Can be updated via a web-based graphical user interface</td>
</tr>
<tr>
<td></td>
<td>• Can only be updated by a GIS specialist</td>
<td></td>
</tr>
<tr>
<td>Passenger Volumes by Location and Time of Day</td>
<td>• Stop locations manually marked on map, as well as alightings andboardings</td>
<td>• Survey staff can record boardings and alightings along entire route using mobile app. Data is saved with route information and automatically updated</td>
</tr>
<tr>
<td></td>
<td>• Stop locations plotted in GIS platform</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Passenger counts manually updated in GIS for each surveyed location</td>
<td></td>
</tr>
<tr>
<td>Average Travel and Dwell Times</td>
<td>• Staff ride transit routes and measure travel time between pre-determined points on route map</td>
<td>• Travel time automatically recorded and linked to route file</td>
</tr>
<tr>
<td></td>
<td>• Travel time data manually entered on each route segment</td>
<td></td>
</tr>
<tr>
<td>Total Time Needed</td>
<td>On Route: 2 hours x 3 = 6 hours Additional data entry: 1 hour x 3 = 3 Total: 9 hours</td>
<td>On Route: 2 hours Additional data entry: 0 hours Total: 2 hours</td>
</tr>
<tr>
<td>Total Time and Cost for Metro Manila</td>
<td>9 hours x 900 routes = 8,100 hours</td>
<td>2 hours x 900 routes = 1,800 hours</td>
</tr>
</tbody>
</table>

Source: World Bank Open Transport Team

2.1 Transit Stops (stops.txt)

stop_id,stop_name,stop_desc,stop_lat,stop_lon,stop_url,location_type,parent_station
S1,Mission St. & Silver Ave.,The stop is located at the southwest corner of the intersection.,37.728631,-122.431282,,,

An example of stop information is shown above.
The required information is:

- Stop ID (stop_id)
- Stop Name (stop_name)
- Stop Description (stop_desc)
- Stop Latitude (stop_lat)
- Stop Longitude (stop_lon)
- Stop URL (stop_url)
- Location Type (location_type)
- Parent Station (parent_station)
However, if a piece of information is not known, it can be left blank. Providing as much information as possible, however, will make the GTFS data more useful by allowing for a broader set of applications. The example above describes a stop that has been labelled as S1 and is located at Mission St. and Silver Avenue. The description gives more detailed information about where it is located at the intersection, and the latitude and longitude allow for precise map locations.

2.2 Transit Routes (routes.txt)

The routes file asks for:

- Route ID (route_id)
- Route Short Name (route_short_name)
- Route Long name (route_long_name)
- Route Description (route_desc)
- Route Type (route_type)

The file above describes a route given the ID “A”. This ID is unique and no other route in this dataset can be given this name. The description notes that the “A” route travels from the lower Mission to Downtown. The short name is 17, while the long name “Mission” more logically describes the route. Finally, the route type describes the type of transportation used on the route. In this case “3” refers to bus. A more complete list can be found in the Google GTFS Reference.
2.3 What can it be used for?

GTFS feeds allow public transit agencies to publish their transit data in a format that is accessible to developers to access and write applications that consume the data.

GTFS data can be used for trip planners, timetable publishers and a slew of other applications that use public transit information in some way.

Because GTFS is an open-standard, applications that are designed for one city's GTFS data can be used with any other set of GTFS data. This means that applications or analyses performed for one city's data can easily be performed and adapted for another city.

It can be used not only to manage static transit information such as routes, stops and schedules, but GTFS-realtime (GTFS-RT) data feed specifications can provide live updates on transit fleets using Automated Vehicle Location (AVL) systems and static GTFS feeds.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trip planning and maps</td>
<td>Applications that assist a transit customer in planning a trip from one location to another using public transportation</td>
<td>Google Maps, OpenTripPlanner, Bing Maps, Transit App for iOS, RouteShout, Tiramisu</td>
</tr>
<tr>
<td>Ridesharing</td>
<td>Applications that assist people in connecting with potential ridesharing matches</td>
<td>Parkio, Avego</td>
</tr>
<tr>
<td>Timetable Creation</td>
<td>Applications that create a printed list of the agency’s schedule in a timetable format</td>
<td>TimeTablePublisher</td>
</tr>
<tr>
<td>Accessibility</td>
<td>Applications that assist transit riders with disabilities in using public transportation</td>
<td>Sendero Group BrailleNote GPS, Travel Assistant Device</td>
</tr>
<tr>
<td>Planning Analysis</td>
<td>Applications that assist transit professionals in assessing the current or planned transit network</td>
<td>OpenTripPlanner Analyst Extension, Graphserver, Transit Boardings Estimation and Simulator Tool, TransCAD 6.0</td>
</tr>
<tr>
<td>Interactive Voice Response (IVR)</td>
<td>Applications that provide transit information over the phone via an automated speech recognition system</td>
<td>BusLine, TransitSpeak, TravelSpeak</td>
</tr>
<tr>
<td>Real-time transit information</td>
<td>Applications that use GTFS data along with a real-time information source to provide estimated arrival information to transit riders</td>
<td>OneBusAway, NextBus, TransLoc, Moovit, next bus arrival signs at bus stops (e.g. in Santiago, Sao Paulo)</td>
</tr>
<tr>
<td>Dedicated SMS applications</td>
<td>Applications designed for feature phones without data capabilities</td>
<td>RouteShout, Transantiago's SMS Bus</td>
</tr>
</tbody>
</table>

Table 1: Technical Innovations Enabled by GTFS Feeds (Mehndiratta & Ochoa, 2014)
2.4 Case Studies

As of June 2013, approximately 1,050 transit operators released official GTFS feeds. Most of the feeds are from operators in the US, Canada, Europe, Australia, New Zealand and Japan, but some are from developing countries. London, UK has been a pioneer in opening up its public data, and is one of the most successful examples of open data use in public transport.

London

Open Data Use in London (https://tfl.gov.uk/info-for/open-data-users/)

- 5,000 registered developers
- Hundreds of apps produced on all platforms
- 30 feeds and APIs including across modes
- Provided in non-proprietary formats
- Easy-to-navigate website (tfl.gov.uk/info-for/open-data-users)
- Societal benefits due to open data use are calculated at up to £58m annually due to customer time saved
  - Based on investment of under £1m annually on Open Data

Why are we committing to open data?

- Public data – As a public body, our data is publically owned
- Reach – Our goal is to ensure any person needing travel information about London can get it wherever and whenever they wish, in any way they wish
- Economic benefit – Open data facilitates the development of technology enterprises, small and medium businesses, generating employment and wealth for London and beyond
- Innovation – By having thousands of developers working on designing and building applications, services and tools with our data and APIs, we are effectively crowdsourcing innovation

Source: https://tfl.gov.uk/info-for/open-data-users/our-open-data
### City characteristics

<table>
<thead>
<tr>
<th>City</th>
<th>Population (thousands)</th>
<th>Metropolitan population (thousands)</th>
<th>Primary transit modes</th>
<th>Regulatory scale</th>
<th>Transit trips per day (millions)</th>
<th>Internet access (national level)</th>
<th>Mobile phone ownership (national level)</th>
<th>Smartphone ownership (national level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico City</td>
<td>8,600</td>
<td>18,000</td>
<td>Microbus, Metro, Bus</td>
<td>National</td>
<td>15</td>
<td>37%</td>
<td>83%</td>
<td>13%</td>
</tr>
<tr>
<td>Santiago</td>
<td>4,600</td>
<td>6,000</td>
<td>Bus, Metro</td>
<td>National</td>
<td>5.25</td>
<td>59%</td>
<td>118%</td>
<td>18%</td>
</tr>
<tr>
<td>São Paulo</td>
<td>11,000</td>
<td>18,800</td>
<td>Bus, Metro</td>
<td>National, local</td>
<td>8.7</td>
<td>46%</td>
<td>124%</td>
<td>28%</td>
</tr>
<tr>
<td>Manila</td>
<td>12,000</td>
<td>26,000</td>
<td>Jeepney, Light rail</td>
<td>National, Local</td>
<td>3.6</td>
<td>32%</td>
<td>99%</td>
<td>14%</td>
</tr>
<tr>
<td>Dhaka</td>
<td>9,200</td>
<td>14,500</td>
<td>Rickshaw, Bus</td>
<td>Local</td>
<td>11.2</td>
<td>5%</td>
<td>56%</td>
<td>&lt;1%</td>
</tr>
</tbody>
</table>

### Transit data outputs

<table>
<thead>
<tr>
<th>Project initiator</th>
<th>NGO assistance</th>
<th>Method</th>
<th>Data collection started</th>
<th>GTFS released</th>
<th>Routes included</th>
<th>Difficulties encountered</th>
</tr>
</thead>
<tbody>
<tr>
<td>City government, World Bank</td>
<td>Yes</td>
<td>Android and iPhone apps, data management portal</td>
<td>2012</td>
<td>2013</td>
<td>475</td>
<td>Fixed stop locations, fixed schedules and headways, vehicle type</td>
</tr>
<tr>
<td>City Government</td>
<td>No</td>
<td>AVL, AFC</td>
<td>2007</td>
<td>2013</td>
<td>376</td>
<td>Group taxis not included due to their flexible operations</td>
</tr>
<tr>
<td>City government</td>
<td>No</td>
<td>AVL, AFC</td>
<td>2008</td>
<td>2012</td>
<td>1,329</td>
<td>None – no flexible transit services</td>
</tr>
<tr>
<td>City government, World Bank</td>
<td>Yes</td>
<td>GPS</td>
<td>2006</td>
<td>2012</td>
<td>906</td>
<td>Fixed stop locations, fixed schedules and headways, vehicle types</td>
</tr>
<tr>
<td>MIT, Urban Launchpad</td>
<td>Yes</td>
<td>Android app</td>
<td>2012</td>
<td>2013</td>
<td>78</td>
<td>Fixed stop locations, fixed schedules and headways, lack of agency websites</td>
</tr>
</tbody>
</table>

#### GTFS Data Collection Experiences Across Five Cities (Eras et al., 2014)

2.5 Open Data in Developing Countries

The ODDC Project, launched in 2012 as a collaboration between the Web Foundation and the Canadian International Development Research Centre (IDRC), aims to provide a platform for researchers to connect and share results and methodologies.
Nairobi, Kenya
Population (city): 3.1 million people
Population (metropolitan region): 6.7 million people
Number of local government authorities: 15

Kenya Open Data Initiative ([https://opendata.go.ke/](https://opendata.go.ke/))
- Government development, demographic, statistical and expenditure data available digitally on a website
- 160 datasets including 2009 census
- **Goal:** provide a “platform for innovation” to produce more efficient outcomes in service delivery, job creation and citizen feedback systems; enable data-driven and better-informed decision making processes; improve transparency and accountability in government operations

Main Challenges:
- many agencies do not have good data to share, or do not want to release existing data
- data often not provided at local scale, only at a regional scale

Organisation: Civic Data Design Lab
- Collaboration between researchers at MIT, UC Berkeley, and the Kenya Institute of Public Policy and Research Analysis (KIPPRÄ)
- **Goal:** Create GIS data and maps for Nairobi that can be shared openly
- Created online wiki space to facilitate download and discussion of data and its challenges
- Partnered with Virtual Kenya

Lessons Learned from Kenya
- Need to build relationships with existing actors
- Need to consider the political landscape in which data is being collected and disseminated and establish trust in order to get organisations to open up data
- Open data can seem threatening to relatively powerful governments and other entities
- Community engagement vital in order to maximise benefits of participatory data sharing

Mexico City, Mexico
Population (city): 8.9 million people
Population (metropolitan area): 20.9 million people
Number of local government authorities: 40+

Transport Overview
- Transport includes both heavily and loosely regulated operators
  - Colectivos (minibuses) make up 50% of motorised trips
- Route and stop location information collected and stored for services on dedicated infrastructure
- Loosely regulated private buses and minibuses have no digital means to collect ridership information
  - 30,000 vehicles in 121 different route associations
  - regulator has little information on routes, drivers or vehicles

World Bank and Open Data
- Trial in spring 2013 to collect basic route data using mobile apps on 10 of the most formalised colectivo corridors
  - Route, travel time, passenger counts
- **Goal:** create a GTFS feed to integrate all transport agencies in Mexico City

Main Challenges:
- GTFS was not designed with flexible services with no fixed stop or schedule information (such as colectivos) in mind
- Need to develop a workaround to estimate headways based on existing knowledge using vehicle counts and speed data
- Data collection is effort- and time-intensive
3. Why Use Open Data in Transport?

Open data in transport has led to technical innovations that allow agencies and providers to communicate with users in new ways, such as through mobile applications and regularly updated transport information. Open data can provide many benefits to a variety of users, including:

<table>
<thead>
<tr>
<th>Public Transport Users</th>
<th>Public Transport Sector</th>
<th>Government and Economy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(better journey experience)</td>
<td>(more efficient operation)</td>
<td>(benefits of a new big market)</td>
</tr>
<tr>
<td>Easy and quick access to transport information through apps to help plan trips and understand system</td>
<td>Extended reach of information to more well-informed passengers</td>
<td>Government as a platform and enabler to disseminate information by using limited resources efficiently</td>
</tr>
<tr>
<td>Visualisation of complex and otherwise incomprehensible data into personalised and relevant travel information</td>
<td>Better informed passengers allows a better use of network capacity</td>
<td>Coordination between different departments and organisations and usage across multiple platforms and applications</td>
</tr>
<tr>
<td>Smarter planning that saves time and money and leaves users better informed and more satisfied</td>
<td>Enhanced image through transparency and openness</td>
<td>New business opportunities for developers to create applications that may have otherwise been cost- or time-prohibitive for the public sector</td>
</tr>
</tbody>
</table>

A move from tightly controlled data and their derived products to publically released data has allowed for applications that benefit users even when governments do not have the financial or technical capacity to provide these innovations directly. Open transport data allows the public to contribute to innovation and the betterment of public services through means that may be cost- or time-prohibitive for the public sector.

Traditional spatial and transport planning tools have often not been successful in developing cities. They are expensive, require high level of training and capacity, and can only be applied towards a single purpose. While they still play a role in large-scale transport planning, there are numerous license-free, expandable and translatable software packages available that can help to build capacity and a culture of data use.

Open data can help transit and planning agencies address questions such as “where can a city densify?” or “which link of transport infrastructure needs to be resilient to natural disasters?” using already existing data. Opening up data can help to achieve economies of scale in supporting wider and more efficient use of transit networks through releasing information to the public.

**Challenges:**
- Need for an institutional environment that prioritises keeping GTFS data up-to-date
- Users unaccustomed to having information and inaccurate/estimated information that may be unreliable could prove challenging

Even though some obstacles to a complete adoption of Open Data in the transport sector still exist, more governments and transport agencies have already embraced the concept and opened up their data. There are various benefits associated with Open Data in the transport sector, for users, public service providers, and the government and the economy.
4. Further Resources

Google Transit and GTFS:
https://developers.google.com/transit/

Open Data Handbook:
http://opendatabook.org/en/

Kenya Open Data:
https://www.opendata.go.ke/

Open Data Essentials from the World Bank:
5. Sources

- World Bank Open Transport Team. "An Overview of Open Transport in East and Southeast Asia." Date?
- Mehndiratta, Shomik and Catalina Ochoa. “GTFS and Transport Open Data: What, why and how to make it work for cities in developing countries”.

Image sources:

- [http://blog.parkio.com/?p=314](http://blog.parkio.com/?p=314)
- [http://mapyc.info/38112/open-trip-planner-portland/](http://mapyc.info/38112/open-trip-planner-portland/)
- [http://windowsphoneapk.com/APK_Bus-Line_Windows-Phone.html](http://windowsphoneapk.com/APK_Bus-Line_Windows-Phone.html)
- [http://www.wmata.com/rider_tools/nextbus/about_nextbus.cfm](http://www.wmata.com/rider_tools/nextbus/about_nextbus.cfm)
- [https://tflnwp.files.wordpress.com/2014/03/apps.png](https://tflnwp.files.wordpress.com/2014/03/apps.png)