Case Study:

TfL Open Data

Type: Website

Organisation(s): Transport for London

Tags: open data, process, metadata, standards

TfL Open Data is a public data portal provided by Transport for London. The portal publishes links to key data sets on London's public transport network for members of the public, as well as a Unified API with live feeds available to any developer who registers. APIs are available for 75% of data sets. Records show 17,000 developers have already signed up. By 2017 a total of 19 apps based on TfL Open Data had over 1 million downloads.

There is a very wide variety of data sets available, ranging from timetable information and facility location to passenger volumes, traffic disruption information, and pollution levels.

Transport for London has a strong vision to make London a 'smart city', exemplified in Project 2030, and is actively looking for new ways to use data to improve quality of life and create economic value. TfL Open Data is a key part of this agenda.

Background

Transport for London (TfL), London's integrated transport authority, began providing open data online in 2007. Since then, TfL has significantly expanded its capacity to cover hundreds of data sets about public transport, as well as roads, cycling, and walking.

TfL has actively encouraged usage of its data by private companies, making it easy to access live feeds. They also provide an open tech forum, where developers can post any questions about how to use their data. TfL has also organised hackathons to generate excitement among the technology sector about the potential of open data.

When registering to use the TfL APIs, users have to agree to terms and conditions, modelled on a modified form of the Open Government Licence. This includes limits on number of calls, and prohibits marketing software products as if they were official TfL projects. TfL also provides Syndication Developer Guidelines, which describe how open data works, including quality standards for data feeds and branding guidelines.

TfL has collaborated throughout the process with a range of external partners. For example, Ordnance Survey input has been important to understand some important 3D data features, such as bridge and tunnel heights. Other collaborations include working with National Car Parks to analyse occupancy levels of London Underground car parks; and with King's College London, who run a network of sensors to track pollution levels.

Important considerations

Commercialisation

The current strategic model has a strong focus on encouraging commercialisation by private sector organisations, driven by strong engagement with the tech community through events such as hackathons.

GREATER MANCHESTER
DOING THINGS DIFFERENTLY

Deloitte has [estimated](#) that providing transport data openly for developers brought significant internal cost reductions and wider economic benefits. Hundreds of journey planner apps are estimated to have added approximately £130 million to London's economy. This has been achieved through high-value job creation in the tech industry and further effects down the supply chain. Travellers on public transport can save time by being able to adjust their routes based on real-time information, and drivers can avoid traffic incidents. This may in turn lead to increased numbers of journeys, providing a further economic boost.

The tools produced by external developers have also reduced TfL's costs by avoiding the need for internal application development. This has reduced the cost of SMS transport tracking services, and brought down contact centre call volumes.

Content and quality

Public data sets are well organised into thematic categories and are generally up to date. However, they cannot be found using a free text search. The format in which data sets are available varies considerably, from PDF to XML, XLS, and CSV, and files are not searchable by data type.

TfL data sets are a mixture of real-time feeds, such as live departure boards, live traffic disruption and the Journey Planner API; and fixed datasets, like timetables and station locations. Among the fixed datasets are transparency-oriented datasets, such as operational performance figures and directors' salaries. Interesting additional datasets published include passenger counts, cycle counters, and cycle hire numbers.

One major additional element is the [Cycling Infrastructure Database](#), which provides a total of 480,000 photographs of cycling infrastructure. Through this database, users can see exactly what cycle lanes, wayfinding signs, and cycle parking facilities can be found on each street.

TfL frequently makes interesting new additions to the data made available, which are publicised in the form of 'data drops'. One important recent data drop has been detailed data on [electric car charging points](#). TfL has worked with electric vehicle charging concessionaires, such as British Gas and Charge Master, as well as the charging station mapping company Zap-Map to curate this dataset. TfL consolidates this aggregate data and makes it available in a more usable format.

However, the website acts as a portal to data, meaning that information is actually stored on a number of different datastores. Some of this data is available on TfL's own data site (e.g. roads.data.tfl.gov.uk), while some is on the [London Datastore](#) or [data.gov.uk](#).

Usage

Although the large ecosystem of mobile applications built for the general public using TfL data presents the most obvious practical application of the data, there are [many others](#). For instance, TfL's [WebCAT tool](#) provides information on London's public transport system that can [help property developers](#) to identify the best locations for retail facilities, offices, and housing. TfL data sets are also [useful for academics](#) researching issues of road safety or cycling.

Also, real-time data releases [helped support commuters](#) during the disruption caused by the 2015 London Underground strikes. This was achieved by reducing stress and aggravation through up-to-date information and suggestions for alternative routes. Open data also helped public transport demand management by government employers during the 2012 Olympic Games.

GREATER MANCHESTER
DOING THINGS DIFFERENTLY

Blockers and challenges

The history of TfL Open Data raises questions about deriving value from open data. Open data provided by TfL has helped to boost London's start-up scene, profiting technology companies that have found ways of using transport data. Whilst this has generated economic benefits through the valuable tools that the private sector has developed, it also suggests that there is an opportunity for TfL to capture some of the value of its own data, or to use the value of its data to leverage other data from private companies.

TfL has faced some minor roadblocks in making data open, such as fees that must be paid for linking [postcodes with addresses](). In 2014, data privacy issues were exposed when one software engineer showed that cycle hire use statistics could be de-anonymised using Customer IDs, leading to the removal of this data set.

User experience

Whilst the TfL Unified API has generated significant enthusiasm and has been widely used, the datasets that are available without registration are of a more intermittent quality. For example, timetable data are not available in one single download, and there is limited metadata available to find out what each file contains before downloading it.

The public-facing segments of the website give the impression that data quality has not always been prioritised. Finding the relevant publicly available data is sometimes more time-consuming than it needs to be, whilst geospatial data are also generally not available in a standardised, navigable map format. However, this merely reflects a strategic emphasis on ties with the developer community. There is an implied assumption that external developers are likely to produce good quality public-facing content for any datasets that are of interest to the public. These quality elements are enforced through the aforementioned Syndication Developer Guidelines.

What can Greater Manchester take from this?

- There is a wide community of developers who are enthusiastic about using open data to design new software.
- A focus on providing as much transport-related data as possible using API and live feeds can generate significant rewards, especially economic returns. However, transport data is, in its fluid and dynamic nature, naturally more amenable to this kind of manipulation than in other areas. Encouraging similar API uptake beyond transport may require careful consideration.
- Targeted collaborations with key external organisations can add significant value to open data, particularly when used to provide greater depths to existing data sets.
- It is easier to encourage the private sector to become a data user of high-quality open data than to engineer a mutual collaboration in which the private sector is also a data provider. However, this may not be a major issue where the tools produced by developers have a significant public benefit.
- It is vital to consider which users to target when designing an open data platform. If the wider public and non-technical users are intended as key data users, it is important to ensure an open data platform is visually accessible and easy to navigate.
- There is an added consideration that data users could be segmented before reaching a dataset location. This would allow open data to be provided in different formats to different groups. For example, providing APIs to developers, while providing visualisations or analysis to more general data users.

- TfL's success highlights feasibility issues faced by comparable projects elsewhere. Data that form the basis of TfL Open Data, including fares and passenger levels, are considered commercially sensitive by private transport companies and so are not shared.
- With a well-established system, and core users, there is the potential for data providers in the public sector to attempt to capture more of the value of data, or to use this value to leverage access to other datasets. However, this is heavily reliant on public data having significant value, a strong and consistent user base and a variety of other pricing and demand-based considerations.

Find out more:

http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf

https://theodi.org/wp-content/uploads/2018/07/Transport-data-in-the-UK-and-France_-A-Series-of-transport-data-case-studies-and-ideas-for-cross-country-collaboration.pdf

https://silo.tips/download/transport-for-london-get-set-go