











# Platforms to help manage the transport data explosion

The transport sector has seen a recent clamour for new data platforms to make it easy for us to share transport data. **Teresa Jolley** and **Nic Cary** set out to investigate what types of services are available

Table 1: Data types and sources

	Data source categories	Types of data	Uses	Sources	Access: Open / Managed (Mgd) / Mix	Machine Readable (M/R) / Offline (O/L) / Mix
	Geographic and administrative and accounting definitions, ID's and boundaries	Shapefiles, Way-ID's, OS codes, CIPFA codes, place and road names	Location accuracy, jurisdiction / responsibility, connecting between datasets	ONS; DHCLG; GOV.UK register; Ordnance Survey; OpenStreetMap; National Gazetteer, GeoPlace and OS MasterMap	Mix	M/R
	Physical highway network: geometry, inventory and condition	Roads, footways, cycleways, bridges, streetlights, cycle stands, cameras, sensors, bus stops, lines and signs, charging points, parking locations	Routing/navigation, asset management, value and budgeting of works, ensuring the network meets demands of all users	Condition and some assets: Highways Authorities; Geometry and some assets: OpenStreetMap; GeoPlace National Street Gazetteer and OS MasterMap	Mix	Mix
	Controlling use of, and restrictions on the network, temporary and permanent	Traffic Regulation Orders (Traffic Management Orders in London) e.g. HGV route restrictions, parking restrictions, road closures, bus lanes, 20mph zones, speed limits	Journey planning, satnavs, ITS applications and CAVs, traffic modelling and planning	Highways authorities	Mgd*	O/L*
	Managing disruptions and access to use of the network Accidents, incidents, impact of major events	Street works Road works Diversions Gritting routes	Helps planning, routing, congestion management, air quality, saves time, makes better use of limited resources  Saves lives, reduces trauma, makes better use of healthcare resources	Highways authorities Separately held Mostly presented through Roadworks.org: free map view	Mgd*	M/R
	Major works: spoil + containment locations + routes for infrastructure e.g. Crossrail, HS2, large housing developments	Construction design and management plans	Better management and coordination of capacity, incidents, works - network resilience	Construction works data	Mgd*	O/L*
	Counting, measuring and monitoring traffic speed, density and flow (vehicles, cyclists, pedestrians, assisted users)	UTMC, ANPR, CCTV, MAC address harvesting, bluetooth / WiFi sniffing, movement monitoring, count surveys	Journey planning Reducing congestion and emissions, improving air quality Better management and coordination	Local authorities; mobile network operators; CAVs; companies with: satnav, ride hailing, app, vehicle tracking data; crowd-sourced data; freight/logistics companies	Mix	M/R*
	Parking, fuelling and charging – off-street + on-street for cars, HGVs, bikeshare docking, seating	Parking spaces in real time, prices, locations, total capacity, vehicle type restrictions  Cycle hire/docking locations, capacity Freight/HGV bays and parks Charging and refuelling facility locations for different fuel types	Helps planning, routing, congestion management, air quality, saves time, better use of resources, limits congestion caused by searching for amenities	Local authority equipment, private car park operators, CCTV, ANPR	Mix	M/R*
	Air quality	CO2, PM10, PM2.5, 'NOx' (NO, NO2) O3, CO, SO2, NH3, Pb	Public health	DEFRA and local authority equipment, weather stations, ANPR	Open*	M/R*
	Road safety data, including accident hotspots	Killed and Seriously Injured, 'Stats 19', Crashmap	Saves lives, reduces trauma, makes better use of healthcare resources, reduces congestion	DfT	Open	M/R
	Planning, purchasing, routing, licensing and controlling public transport and licensed providers operating on the network	<b>Rail, bus and coach data:</b> fares, timetables, routes and stops, real time movements and locations, <b>Licensed</b> taxi and coach drivers and operators / other providers	Journey planning, modal shift Public safety Transport planning and modelling	DfT NaPTAN stops, Traveline timetables and routes Fares: rail companies and some bus companies  Stops, and bus routes, timetables: open Rail fares: open Bus fares: closed	Mix	M/R*

\* Mostly

“When I hear someone say they have a data platform I reach for my shotgun,” Professor Jonathan Raper declared earlier this year. Raper was in a Q&A session with Graham Hanson, the DfT’s Intelligent Transport Systems (ITS) policy lead, at the Catch! showcase event at the Future Cities Catapult in Clerkenwell Green.

Raper reflects mounting concern amongst transport data experts that people have a misplaced understanding of what data platforms should provide to qualify as useful in meeting transport’s data-sharing challenges. Currently only a few local transport authorities use such platforms. So why should they invest?

At many conferences about digital transport, speakers often tell us that we have entered a fourth Industrial Revolution. This time, they tell us, the revolution is data-driven. However, many go on to point out that the transport sector has lagged behind many in making better use of data to provide better services for customers, as well as those who plan, operate and maintain transport on behalf of these customers.

## Why is data sharing so important?

“A very important job cities and their transportation partners have to do is gather the data they generate within one unified data-sharing platform,” says Jean-Marc Lazard, CEO at OpenDataSoft. “Doing so helps them understand how the data can be reused by the community and what the community needs in terms of quality and relevance to reuse the data. Before Waze pushed their data from their community-based app onto Lille Métropole’s data platform, both stakeholders didn’t have a precise idea of what usages would come out of such a collaboration. Having Waze data as well as their own traffic sensor data gathered in the same hub actually enabled Lille to build a comprehensive and cross-referenced information base. They now rely on this information base to make more responsive decisions when traffic issues occur.”

Many types of data are useful to local authorities in running local transport. Not all of these are created by them. For example, they depend on knowing what happens on Highways England’s Strategic Road Network (SRN) to respond to incidents and congestion that cause traffic to spill onto their local road networks but, often, they are not set up to receive and act on these in real-time.

By sharing transport data, beyond the immediate needs of the data creator, data can yield many additional benefits. Whilst data needs to be shared safely and responsibly, most transport data contains no personal or sensitive information. In March 2017, the Transport Systems Catapult published a briefing paper *The Case for Government Investment to Incentivise Data Sharing in the UK Intelligent Mobility Sector* with the Open Data Institute and Deloitte. This quantified the national economic and social benefits of transport data sharing. In his foreword to the paper, Professor Sir Nigel Shadbolt said that “not sharing and making more transport data open could cost the UK £15bn by 2025”. But data sharing also brings many internal and local benefits to local transport authorities. Data can often provide these

authorities with essential evidence in a business case for change, as Martin Tugwell, programme director for England’s Economic Heartland, emphasised in his talk at the Chartered Institution of Highways and Transportation’s annual conference last year.

Taking one example of how local authorities can benefit from easy access to data, the DfT’s Local Highways Maintenance Incentive Fund rewards councils who demonstrate, by means of the data they have gathered, that they are delivering value for money in carrying out cost-effective improvements. In another, data sharing provides better visibility of high-risk assets that often require emergency and unplanned fixes – manhole gullies, for instance.

“Data takes the emotion out of the conversation,” said Peter Parfitt, Buckinghamshire County Council’s Improvement Manager, at a workshop on asset management data. Sharing data, both within local authorities and across local authority boundaries, can give them a much better view of their transport operations and asset maintenance priorities. Amongst many other benefits, it can help local authorities measure modal shift initiatives, reduce congestion, improve air quality and increase parking efficiency.

For these reasons, although the world of data is an arcane one, if we are going to deal successfully with some of the transport capacity challenges that face us, it is clear that we are going to have to get more fluent with the tools and techniques for exploiting data.

## Modern data platforms and APIs

Ash Wheeler, director of data platform and data marketplace provider oneTRANSPORT, says: “Modern data platforms that exploit a combination of the ubiquitous availability of the internet, cloud computing and big data tools and techniques for harnessing and sharing real-time data are rendering the building of traditional ‘relational databases’ to store transport data largely redundant.”

Provided data are reasonably clean – that is to say error free and shorn of duplicates – and, better still conform to some type of standard, raw or source data can and often should be published as locally to its source as possible, via something called an Application Programming Interface (API). Think of this simply as the machine equivalent of a human face, the main interface through which we engage with each other’s minds.

## Data aggregation

This so-called ‘API’ announces its existence to other machines. If this interface is connected to the internet and efforts are made to ensure that it is as near constantly available as possible, this API can be discovered by internet crawling tools and any number of different data aggregation techniques can be deployed to harvest data from it. Think of these data aggregation services simply as a means of putting lots of disparate sources of data together so that machines can analyse these and visualise them and turn them into information that is useful to us when we make transport decisions – for instance when we use a journey planning app, or when network operators take decisions about routing vehicles in a traffic control centre.

The API plus aggregation approach to data removes the need for a single, central source of data, and a consequent dependency on a single point of failure, and enables us to deploy multiple solutions to harvest and aggregate data. There is simply too much data to ever hope that transport data could or should be concentrated in one place. The Internet of Things (IoT) is one area where this is especially true. IoT devices can report on anything from traffic flow, to air quality, to the strain on a road bridge or the state of a drainage gully. Such devices hold out the promise of revolutionising our approach to operating and maintaining transport networks efficiently. However, by the next decade, they are predicted to become far too numerous, with estimates into the billions, for it to be practical to store the data they create and transmit these to a traditional relational database. This is where Big Data tools and inexpensive cloud storage have a key role to play in handling the massive amount of data that these devices generate.

## DfT’s approach to data platforms

During 2017 and the early months of 2018 the DfT let three ‘Discovery’ projects to investigate what should be done with bus data, streetworks and roadworks data, and local transport data respectively. We await DfT’s views on whether these categories of transport data should be concentrated and centralised or whether the Department will adopt a decentralised data publishing model. On 20 March a transport expert tweeted a recommendation from Deloitte’s recent work with the Open Data Institute on buses open data which favoured the latter approach, stating, under Recommendation 2, that: “Data should be published as close to the source as possible to realise its potential.” By contrast, another part of the same Department seems to favour concentrating data on something that sounds remarkably like... a single platform. On 28 November 2017, Highways Authorities and Utilities Committee (HAUC) published a report on its website stating that the DfT has appointed Kainos to produce ‘service to hold real data’ on streetworks and roadworks, subject to further testing.

## Overview of data platforms and services

So far, we have found a small range of national platforms, which are shown in the table overleaf, that can exert some claim to be national. These offer a variety of transport data cataloguing, aggregation and mapping services.

The table is a first iteration of what is out there and will feature at **Digital Transport Exchange**. In future editions we will look at the regional counterparts to these platforms, where there has been strong growth recently – for example Data Mill North – and other types platform focused on providing modelling and analytics tools. The authors are planning to transform this into an online resource and invite further input and comment from practitioners and platform providers. If you have a platform, please get in touch: @teresacjolley @geektwo geek.

The **Digital Transport Exchange** at **Birmingham’s iCentrum** on **11th / 12th July** will have sessions on data platforms.

DIGITAL  
TRANSPORT  
EXCHANGE

MAKING TRANSPORT WORK





11.07.18  
WORKSHOP DAY  
& NETWORKING EVENING  
Creating a digital transport  
ecosystem in a city region

11-12.07.18 | iCentrum, Birmingham

12.07.18  
CONFERENCE, EXPO  
& NETWORKING  
Meeting our future  
transport challenges

www.digitaltransport.uk

### Table 2: Data platform providers

Table 2: Data platform providers								
Platform	Self-service data store for 3rd party data providers	Access to individual data sources published by others	Access to unified, clean and accurate data feeds as a service	Trust in consistency, geographic interoperability and timeliness of data provided by others	Can process, publish and licence data for others	Self-service analytics tools	Can advertise, set charges / criteria for use of data	
 <b>BaseMap</b> http://www.basemap.co.uk/ tools to work with data such as DfT's Trafficmaster national data set of traffic flow, density and speed	✗	✓	✓	High	✗	✓	✗	
 <b>Data.gov.uk</b> Recently revamped searchable catalogue of a large number of public datasets	✗ii	✓	✗	Low	✓	✗	✗	
 <b>Highways England WebTRIS</b> http://webtris.highwaysengland.co.uk/ Speed at junctions at 15-minute intervals across the Strategic Road Network taken from MIDAS loops and other sensors	✗	✗	✓	High	✗	✗	✗	
 <b>ITO World</b> http://www.itoworld.com/ Aggregator of static and real time multi-modal transport data for journey planning apps, transport operators, authorities and consultancies in the UK and key metropolitan areas around the world	✓	✓iii	✓	High	✓	✗	✗	
 <b>oneTRANSPORT</b> https://onetransport.io Data aggregation platform for open sharing of real-time sensor data including traffic movements (speed and flow), parking and city IoT device data	✓	✓iv	✓v	N/Avi	✓	✗	✓	
 <b>OpenDataSoft</b> http://opendatasoft.com/ Data publishing and sharing platform that makes data explorable and reusable via APIs	✓	✓	✓	N/Avii	✓	✓	✓	
 <b>Ordnance Survey</b> http://www.ordnancesurvey.co.uk and related sites e.g. OS MasterMap and GeoPlace	✗	✓	✓	High	✗	✗	✗	
 <b>roadworks.org</b> https://www.elgin.org.uk/ Data aggregation platform providing data on current and planned roadworks and streetworks, diversions, closures, winter gritting routes, live congestion, incidents and accidents, information from traffic cameras and signs. Professional view combines with real-time traffic view	✓	✓	✓	High	✓	✓	✗	
 <b>Transport Scotland – Traffic Scotland</b> https://trafficscotland.org/developarea/ Traffic status and other road data in Scotland in DATEX II format	✗	✗	✓	High	✗	✗	✗	
 <b>Traveline National Dataset</b> http://www.traveline.info/ Data aggregation platform providing bus, coach, tram, ferry, light rail and rail data	✗	✓	✓	High	✓	✗	✗	
 <b>TransportAPI</b> http://www.transportapi.com Data aggregation platform with many types of transport data including buses and trains. Full-service with data sets that can drive apps. Provides live departures and arrivals, timetables, journey planning, tweet mapping, performance indicators and fares for the whole of UK public transport	✓	✓	✓	High	✓	✗	✗	
 <b>Transport Systems Catapult Intelligent Mobility Data Hub</b> https://ts.catapult.org.uk/innovation-centre/imdh/ Managed data catalogue of transport data sets for research use by planners, modellers and academics	✓	✗viii	✗	Low	✗	✗	✗	
 <b>Urban Big Data Centre</b> http://ubdc.ac.uk/data-services/data-catalogue/transport-data/ Aggregates anonymised data from Strava users in Scotland and the North East of England. Strava is a social network for athletes, recording data from its users who upload cycle rides and running activity via smartphone or GPS device	✓	✗ix	✗	Low	✗	✗x	✗	
 <b>Urban Transport Group Data Hub</b> http://www.urbantransportgroup.org/insight The Urban Transport Group brings together and promotes the interests of Britain's largest urban areas on transport	✗	✓	✓	Med-High	✗	✓	✗	
 <b>Zipabout Transport Graphxi</b> http://www.transportgraph.uk/ Data aggregation and communications platform which combines static, operational and behavioural data in real time to predict and manage transport network flow. Enables disruption dissipation, door to door wayfinding, personalised two-way customer service and smart city planning	✓	✓	✓	High	✓	✓	✗	
<div>i Taken from the platform's website or preferably hi/print res</div> <div>ii Inclusion and metadata governed by public sector organisation responsible</div> <div>iii ITO provide the pipework for others to do this</div>				<div>iv With user account</div> <div>v As independent data feeds from each publisher</div> <div>vi Dependent on publisher</div> <div>vii High for provided data, N/A for data uploaded by users</div>		<div>viii Managed access on request</div> <div>ix Managed access on request</div> <div>x No tools, but research expertise</div> <div>xi Designed for local authorities</div>		