HOUSES OF PARLIAMENT
PARLIAMENTARY OFFICE OF SCIENCE & TECHNOLOGY

POSTNOTE

Number 472 July 2014

Big and Open Data in Transport

New ways to collect, manage and analyse vast quantities of data present opportunities to provide a smarter and more efficient transport system. This POSTnote examines key factors affecting the growth of big data in transport and current and future applications, focusing on road and public transport. It also explores challenges to opening up, collecting, disseminating and using big data, and to sharing data whilst protecting users' privacy.

Overview

- New technologies and the shift to open data are enabling the use of big data in transport to monitor infrastructure and services, and to understand transport users' needs better.
- Big data is currently being applied in asset maintenance, road traffic management, the planning of public transport services and to inform transport users' decisions.
- Future developments are likely to centre on ways to provide safer, cleaner and more efficient transport and for users to personalise their transport experience.
- In public transport, data collection and dissemination is often fragmented and inconsistent, which may present challenges to exploit fully big data.
- The Law Commission has recommended that the laws that govern how public bodies share data should be reviewed to create a clear legal structure.

Background

The transport sector has always collected and analysed large quantities of data, such as data from timetables, traffic news and air schedules.[1] However, recent developments in the quantity, complexity and availability of data ('big data', see POSTnote 468) collected from and about transport, together with advances in computing technology, are presenting new opportunities to create more efficient and smarter transport systems for people and freight. In addition, 'opening up' data in transport by making it more widely available, and linking it with data from other sectors, is part of the Government's strategy to improve transparency and encourage economic growth.

To maximise these opportunities, the Government is supporting the UK's data infrastructure, most recently with £14 million to make data routinely collected by business and local government accessible for researchers, including for transport research at Leeds and Glasgow Universities. The Government has also established a new Transport Systems Catapult, overseen by the Technology Strategy Board (TSB). Over five years it will receive £46.6 million from TSB and £16.9 million from the Department for Transport (DfT). It has specific objectives to encourage the analysis of big data.[2]

Key Drivers of Big Data in Transport

In the UK, there are a number of transport providers and operators (see Box 1). New technologies and the shift to open data, combined with advances in data linkage, are enabling these providers and operators to monitor infrastructure and services and to understand transport users' needs better.

Technological Developments

The growth of technology in the transport sector has led to an increase in the type and amount of data automatically created and collected about people and freight movements. Examples include smart travel cards (Box 2), sensor equipment embedded in the built environment, location tracking through Global Positioning Systems (GPS) and mobile networks. For example, Transport for London (TfL) collects Oyster card data from 8 million trips per day about when and where a person touches in and out at Underground stations. TfL also uses GPS technology to monitor bus location and advise users on bus arrival times. The Highways Agency (HA) uses road sensors to collect data on traffic flows and GPS data to estimate journey time (Box 3). There are also increasingly large quantities of readily available data on people's activities and behaviour through social media and apps (POSTnote 460).

Box 1. Key Providers and Operators in the UK Transport Sector

Transport operators and providers in the UK transport sector include:

- Infrastructure operators, which own and/or manage assets such as roads and rail. On roads, the Highways Agency (HA) manages the Strategic Road Network (most motorways and major A-roads), and local authorities manage local roads. Network Rail manages the rail network. Transport for London (TfL) manages most major roads as well as the public transport network in London.
- Service operators, which run services using the infrastructure. In London, TfL directly operates Tube services; other public transport services, such as buses, London Overground and the Docklands Light Railway are operated on its behalf by private companies. Outside London, public transport services are mainly operated by private bus, coach and train companies. On the road, equivalent operators are drivers and haulage companies.
- Information providers, which disseminate information to users. In rail, service operators provide data to the Association of Train Operating Companies, which provides information to users via the National Rail Enquiries service. Traveline, a partnership between service operators and local authorities, provides local public transport information. For roads, the HA disseminates traffic information through variable road signs and through websites such as Traffic England. Private companies also provide information through GPS, satellite navigation and maps. Some news and media outlets collate data to offer a range of travel information.

Box 2. Examples of Smart Ticketing in the UK

Transport for London (TfL) operates the largest smart ticketing scheme in England via its 'Oyster card'. This enables TfL to record the location, date and time of when the card is used. TfL links journey data to an individual card for eight weeks, after which the association with the card is deleted and the data are held by TfL for research purposes.[11] TfL uses big data techniques to target appropriate service messages to customers who sign up for information emails. Outside London, examples include the MCard in West Yorkshire, Iff card in Cardiff, the Ridacard in Edinburgh, the Subway Smartcard in Glasgow and iLink across Northern Ireland.

Box 3. Fixed and Mobile Sensors for Monitoring Roads

Road sensors can be used to collect data about traffic flows, speed and travel times. These include:

- **Fixed sensors**, such as loop detectors which are wires embedded in the road that can detect when a vehicle passes a certain point, such as a traffic light. They can also measure the number of vehicles passing a point, providing an indicator of congestion.
- **Mobile sensors**, such as Floating Vehicle Data which use data collected through mobile phones, satellite navigation, or on-board diagnostic systems via Bluetooth or a wireless internet connection. Data are used to determine the general traffic speed on the road or, where the user has allowed access to their data, the time, location, speed and direction of travel of an individual vehicle.[12]

Open Data and the Transparency Agenda

Opening up non-personal data in the public sector is seen as a way to improve public services, increase the accountability of government and generate economic benefit through the re-use of public data to develop new products and services. It is a key part of the Government's transparency agenda, as set out in the Open Data White Paper and the DfT Open Data Strategy, both published in June 2012.[3,4] In October 2013, following the independent Shakespeare review of public sector information, the Government created a National Information Infrastructure (NII).[5,6] The NII is overseen by the Cabinet Office and works with departments and arm's-length bodies to identify data that can be made publicly available. Departments must provide an explanation for not publishing datasets. As well as central government data, the NII contains datasets owned and published by the wider transport industry, such as timetables. In the future, local authority data will also be available through the NII.

The DfT Transport Sector Transparency Board was set up to oversee the opening up of transport data from DfT and other bodies and to engage with the open data community.[7] As of June 2014, DfT and its agencies had published 255 datasets. Of these, 244 are issued under an Open Government licence, which encourages the use and re-use of information, with only a few conditions (Box 4).[8] Datasets available via the central data.gov.uk website are ranked according to how easily they can be re-used.[9] DfT has a further 481 unpublished datasets. Over the course of 2014, Departments are expected to identify and publish further datasets and aim to improve data quality.[10] Advanced computer software can be used to link government datasets with data from different areas, such as health. It can also be linked to publically available data from social media, to gather insight into people's views about transport.

Current Applications of Big Data

Using big data can increase efficiency and reduce costs to infrastructure and service operators, and provide better levels of services to users. The following examples highlight some key applications in asset maintenance, road traffic management, planning of public transport services and to inform transport users' decisions. Other current areas of application, such the use of big data for price comparison and in manufacturing are discussed in POSTnote 469.

Asset Maintenance

Assets such as the road and rail network require maintenance to avoid deterioration and ensure safety. Big data offer new opportunities to identify problems more quickly and to reduce costs. For example, Network Rail uses smartphones and GPS co-ordinates to increase the accuracy of locating track defects to within 5 metres (previously it had been 1 mile). This reduces the time taken by engineers to fix the track. Smartphones can also capture asset data electronically and transfer it to a central system, eliminating the need to process records manually.[13]

On road networks, the HA is using mobile sensors, called floating vehicle data (FVD, see Box 3) to overcome the limitations of fixed sensors, such as the expense of installing and maintaining sensors embedded in the road.[14] As mobile technologies become more widely available at reduced costs they may be used to provide insights into traffic conditions on local road networks, where there is a lack of fixed sensors. In the USA, smartphones have been used to crowd-source information about the state of assets, such as roads.[15] Sensors are also used in aviation, for example by Rolls Royce to collect information from all of its jet engines while they are in service (see POSTnote 469).

Road Traffic Management

Urban traffic management and control techniques such as self-regulating co-ordinated traffic lights, traffic cameras and variable-message signs have been used for decades to reduce traffic delay and congestion. However, big data is providing new insights into traffic patterns. For example, the HA's 'smart motorways' use a range of technologies to monitor driving conditions to allow variable speed limits to be set to improve the performance of the road.[16] Its National Traffic Information Service collects data through FVD in addition to CCTV, traffic cameras and traffic sensors to give a real-time national view of traffic conditions on the Strategic Road Network.[17,18] It also provides real-time traffic data to information service providers, such as media channels, and through the data.gov.uk website. Private companies, such as Inrix and TomTom, use FVD collected through vehicle fleets to gather information on traffic flows and delays. In freight and logistics, big data is being used to optimise delivery, which can reduce carbon emissions and mileage. For example, DHL uses real-time information on traffic condition to calculate optimal travel routes.[19]

Planning of Public Transport Services

Analysis of public transport data can help to understand transport users' journey patterns across the transport network, in terms of where users travel, what mode they choose, how frequently they travel and how reliable their journeys are.[20-23] These insights can be used by transport operators to inform decisions around the planning of services. For example, TfL uses big data tools to analyse Oyster card and bus vehicle location data to infer where users have alighted a bus in London, even though users do not 'touch out'. TfL uses this insight to understand better the end-to-end journeys that customers make for both network planning and for management of its operations.[24]

Analysis of public transport data can also inform transport agencies and operators of how people from different groups (students, children, seniors, regular commuters and infrequent travellers) use the system, thereby providing a deeper understanding of their customers' needs.[25] For example, TfL and academic researchers have used data visualisation to map passengers' travel routes during scheduled changes to services (such as from sporting events), and unscheduled changes (such as temporary station closures).[23] These visualisations can be an additional tool for transport organisations to inform planning and operational decisions about services, and to explain the transport network more visually to passengers.[26]

Informing Transport Users' Decisions

Opening up transport data to software developers can allow users to personalise the information they receive, such as delays on their local roads or public transport route (Box 5). It could also generate economic benefit if reduced travelling times mean that employees spend more time working, or other benefits if it increases time available for non-work activities.[27] For example, opening up TfL data has been valued at £15-58 million per year and has resulted in over 200 travel apps being developed by private companies.[28]

Box 4. The Open Government Licence

The Open Government Licence was designed by the National Archives to allow all public bodies to license the use and re-use of their information under a common open licence, without having to register and apply for a licence. It allows users of licensed data to copy, publish, distribute, adapt, transmit and exploit the data commercially by combining it with other information or including it in a product or application. Users are required to acknowledge the source of the data; not use the data in a way that suggests any official status; ensure that they do not mislead others or misrepresent the data or its source; and ensure that their use does not breach the Data Protection Act 1998 or the Privacy and Electronic Communications Regulations 2003. It was designed to be compatible with internationally recognised licensing models such as Creative Commons.

Future Applications of Big Data

In the near term, developments are likely to centre on ways to provide safer, cleaner and more efficient transport and for users to personalise their experience of transport. For example, in the US, new systems are being developed to enable existing traffic sensors to distinguish bicycles from other vehicles. This aims to provide a better understanding of how cyclists use infrastructure and contribute to the development of safer cycling routes.[29] Big data may also inform decisions around sustainable infrastructure planning, by providing better estimates of future trends in transport demand through enhanced data linkage and modelling.[30]

Vehicles are also likely to be increasingly equipped with the ability to connect to the internet. For example, under proposed EU legislation, all new cars would be fitted with an inbuilt SIM card capable of transmitting and receiving data over the mobile phone network following an accident (POSTnote 322).[31] More broadly, 'connected cars' could open up opportunities for new services, such as remote monitoring and diagnostics, and allow users to send texts and emails via voice commands, and to use apps while driving.[32,33] Further advances in information and communication technologies may also lead to the adoption of autonomous vehicles (see POSTnote 443). In July 2014, the Government announced a review of road regulations and £10 million funding to enable pilots testing autonomous vehicles in the UK from 2015. Transport data are also likely to be further integrated with data from other sectors (for example, energy, utility and health) and an increasing variety and volume of user-generated content. This may enable more efficient and sustainable transport networks as well as more responsive public services in the future.[34-36]

Box 5. Examples of User Apps

Some examples of user transport apps include:

- **Citymapper**, which provides journey planning for multiple modes of transport in major cities by collating open data feeds.[37]
- **Inirx Traffic**, which provides real-time traffic information based on crowd-sourced data from smartphones and vehicle GPS data.[38]
- **Moovit**, which provides real-time public transport journey planning by combining public transport data with information crowd-sourced from users through their smartphone.[39]
- **Waze,** which provides suggested driving routes based on traffic information crowd-sourced from users through their smartphone.[40]

Key Challenges

Opening Up Data

To enable transport service users and app developers to find and re-use data effectively, data need to be archived and made publicly accessible in a useable and standardised format with sufficient information about the data (metadata) to allow others to understand the data (see POSTnote 468). However, overall, systematic retention and archiving of data and metadata in the transport sector is currently patchy, and opening up data might require time and money. There may be a trade-off between opening up data quickly at low cost and making high quality data available at higher costs.[41] Decisions also need to be made about which data have long-term value, and should be retained in accordance with relevant standards, if they are to remain accessible and usable in the future.[4]

Data Collection

Like many sectors, data collected in transport may be incomplete, inaccurate or unreliable at times or in particular places. Investing in new technologies may help to address this in some contexts. For example, not all buses are equipped with the technologies required to provide real-time data, and road traffic data from road sensors can be missing.[42] As such, information disseminated to users may be incomplete. However, depending on the purpose of data collection, the costs of investing in new technologies may not necessarily be perceived as worthwhile. Further, technologies may be unreliable, for example GPS systems are subject to signal loss when coverage is poor. In addition, technology might not be an adequate replacement for traditional methods. For example, DfT considered using GPS devices to automate data collection to reduce the costs of its National Travel Survey. However, it did not go ahead because signal loss from GPS in densely populated areas meant that journey data were not sufficiently accurate.[43]

Data Dissemination in Public Transport

In UK public transport, data collection, analysis and dissemination is often fragmented and inconsistent.[44] This is because there are multiple infrastructure operators, service operators and information service providers that collect data or disseminate information. This can make opening up and sharing data difficult and result in inconsistent information for users.[21] For example, while bus operators make timetable data available to Traveline, bus fare data are generally not shared because of the lack of data in a shareable form and operators' concerns about giving price indications. Some public and private sector organisations have also argued that this leads to a lack of clarity and transparency about access to data.[45] For example, National Rail Enquiries (NRE) received criticism for its licensing restrictions and for how it charged users to access its data.[46] In response to this criticism, in June 2014 NRE changed its policy and now individuals, small commercial users and public bodies do not require a licence and are exempt from charges to access its data.[47]

Using Big Data

Maximising the use of big data requires advanced technologies to link datasets together, and to process and analyse large quantities of data rapidly often in real-time (POSTnote 473). Implementing new IT infrastructure can be technically challenging and expensive, as new systems will need to interface with existing systems to enable datasets to be joined up for analysis. It requires long term investment and planning, which may not be viable for some local authorities with more immediate needs. However, cloud computing services, which provide on demand access to a remote pool of shared computing resources, may offer organisations new ways to exploit big data in the future. For example, cloud services can provide a more efficient, cost-effective model to implement analysis of big data in-house, or enable companies to use on-demand storage space and computing power via public cloud services.[48]

Data Governance

Data collected from transport systems have traditionally been non-personal data, such as vehicle flows. However, privacy concerns have emerged in the context of the growth in the collection of personal data by the private and public sectors. The collection, storage and processing of personal data is regulated by UK and EU data protection laws. However, there are concerns that users are not fully aware of how their data can be used, for example through social media sites and mobile phone apps (POSTnote 460).

In the context of opening up public sector information, the 2013 Shakespeare review recommended that there should be a clear policy on privacy and confidentiality that increased protections for citizens while increasing the availability of public data to external users.[5] For example: appropriate anonymisation (personal identifiers removed) of data for publication; use of 'safe havens' to allow people secure access to personal data where it is in the public interest; and increased penalties for the misuse of data (POSTnote 468). The need to raise awareness of privacy legislation and the rights and responsibilities of a data controller, holder or user was also highlighted. In July 2014, a scoping review by the Law Commission identified widespread misunderstanding and confusion about the statutory framework for data sharing and its relationship with data protection, human rights and the common law.[49] It recommended that a full UK-wide law reform project should be carried out to create a modern and clear legal structure for data sharing between public bodies and with other organisations carrying out public functions.

Endnotes

See overleaf.

POST is an office of both Houses of Parliament, charged with providing independent and balanced analysis of policy issues that have a basis in science and technology. POST is grateful to Stephen Hanley for researching this briefing, to ESRC for funding his parliamentary fellowship, and to all contributors and reviewers. For further information on this subject, please contact the co-author, Dr Abbi Hobbs.

[1] Henry, L. (2013). 'Analytics and Big Data – Rail Public Transportation is a Leader'. Paper presented at *APTA 2013 Rail Conference*, Philadelphia, June 2-5, 2013.
[2] Transport Systems Catapult (2013). 'Five-Year Delivery Plan to March 2018'.
[3] Cabinet Office (2012). 'Open Data White Paper'.
[4] Department for Transport (2012). 'Open Data Strategy'.
[5] Department for Business, Innovation and Skills (2013). 'Shakespeare Review: An Independent Review of Public Sector Information'.
[6] Cabinet Office (2013). 'National Information Infrastructure'.
[7] Department for Transport (2013). 'DfT Transport Sector Transparency Board'.
[8] National Archives (2010). 'Open Government Licence for public sector information'.
[9] Berners-Lee, T. (2009). 5 Star Open Data.
[10] data.gov.uk (2014). Data Published by Department for Transport.
[11] Transport for London (2014) 'Oyster card'.
[12] European Commission (2012). 'ITS Action Plan'.
[13] Network Rail (2013). 'Asset Management Services'.
[14] World Road Association (2006). 'Briefing Note on Floating Car Data'.
[15] City of Boston (2014) 'Street Bump'
[16] Highways Agency (2014) 'Smart Motorways'.
[17] Mouchel (2011). 'National Traffic Information Service'.
[18] Inrix (2012). 'England's Highways Agency Selects INRIX Real-Time Traffic and Travel Time Information for its National Traffic Information Service'.
[19] DHL (2013). 'Big Data in Logistics'.
[20] Batty, M. (2013). 'Big Data, Smart Cities and City Planning'. *Dialogues in Human Geography*, 3(3), pp.274-270.
[21] Detica (2010). 'Unblocking the Transport Network'.
[22]Uniman, D. (2009). '*Service Reliability Measurement Framework using Smart Card Data: Application to the London Underground*'. M.Sc. Thesis. Massachusetts Institute of Technology: USA.
[23] Centre for Advanced Spatial Analysis, UCL (2014). Mechanicity.
[24] Gordon, J. (2012). '*Intermodal Passenger Flows on London's Public Transport Network: Automated Inference of Full Passenger Journeys Using Fare-Transaction and Vehicle-Location Data*'. M.Sc. Thesis. Massachusetts Institute of Technology: USA.
[25] Ortega-Tong, M. (2013). '*Classification of London's Public Transport Users Using Smart Card Data'*. M.Sc. Thesis. Massachusetts Institute of Technology: USA.
[26] Transport for London (2013). 'Olympic Legacy Monitoring: Personal Travel Behaviour during the Games'.
[27] Batley, R., and Laird, J. (2012) '*Theory and Practice of Valuing Travel Time Savings'*. Presentation at the Forge Summer School, University of York.
[28] International Association of Public Transport (2014). 'Opening up to Open Data'.
[29] Gosse, A., and Clarens, A. (2014). 'Estimating Spatially and Temporally Continuous Bicycle Volumes'. Paper presented at *Transportation Research Board 93rd Annual Meeting*, Washington, January 12-16 2014.
[30] International Transport Forum (2013). 'Big Data and Transport'.
[31] European Commission (2014). 'eCall: Time saved = lives saved'.
[32] GSM Association (2013). 'Connected Cars: The Technology Roadmap'.
[33] TNS (2014) 'Time to set the connected car free'.
[34] Thakuriah, P. and Geers, G. D. (2013). '*Transportation and Information: Trends in Technology and Policy*'. Springer, New York.
[35] European Commission Joint Research Centre (2011). 'Mapping Innovation in the European Transport Sector'.
[36] United Nations Economic Commission for Europe (2011). 'Road Map for Promoting ITS – 20 Global Actions 2012-2020'.
[37] https://citymapper.com/
[38] http://www.inrixtraffic.com/
[39] http://www.moovitapp.com/
[40] https://www.waze.com/
[41] UKAuthority (2013). 'Local Digital Today'.
[42] Ding, N., Tan, G., Zhang, W., and Ge, H. (2011). 'Distributed Algorithm for Traffic Data Collection and Data Quality Analysis Based on Wireless Sensor Networks'. *International Journal of Distributed Sensor Networks*, 2011.
[43] Department for Transport (2012). 'Public Consultation on the Future Design of the National Travel Survey'.
[44] Department for Transport (2013). 'Door to Door – A Strategy for Improving Sustainable Transport Integration'.
[45] Office of Rail Regulation (2012). 'Real Time Train Information – A Consultation by ORR on the Findings from its Review'.
[46] Raper, J. (2011). 'Why train departure information is not currently open data'.
[47] National Rail Enquiries (2014) 'New arrangements for accessing live train information for developers'.
[48] Intel IT Centre (2013). 'Big Data in the Cloud: Converging Technologies'.
[49] Law Commission (2014). 'Data Sharing between Public Bodies'.