Safety, Accessibility and Sustainability Panel

TRANSPORT FOR LONDON
EVERY JOURNEY MATTERS

Date: **13 October 2015**

Item: **Leveraging Our Data for Maximum Impact**

This paper will be considered in public

1   Summary

1.1   To provide the panel with an initial overview of how we are using our data to deliver better policy, operational, efficiency and customer service outcomes and how this is being integrated into our existing and future plans.

2   Recommendation

2.1   **The Panel is asked to note the approach and action set out in this paper**.

3   Introduction

3.1   There are a number of ways to think about this, but there are three broad areas which capture the place of data in our business strategy:

   (a)   'Big data': our own analysis of one or more large data sets, owned by us or someone else, to gain insight and thus support better resource allocation and outcomes. Examples of Big Data include that derived from ticketing and physical assets like trains or buses;

   (b)   'Open data': making our own proprietary data sets freely and openly available in machine readable form to third parties to develop products or analysis to benefit stakeholders and the wider digital economy. Examples include travel information to help customers save time and collision data to reduce the number of people killed or injured on the roads. We also consume open data from other organisations (e.g. on health and housing) to put alongside our own; and

   (c)   'Data driven decision-making and innovation': what actually happens when Big and/or Open Data are successfully exploited. Examples include prioritising spend such as targeting accessibility improvements to benefit the greatest number of people, automated refunds when things have gone wrong and travel information products.

4   Big Data – Background

4.1   Data has always been extensively used to inform decisions about the development of transport in London and the products and services we deliver for our customers. Data is recognised as a highly valuable asset class in its own right.

4.2   With London's rapid population growth and pressure to deliver more with less resource, the effective analysis and application of data to make better decisions and to become even more efficient and productive has never been more important.

4.3   Millions of public transport and road journeys are made every day offering vast quantities of information. For example, there are nearly 20 million daily 'taps' captured through our ticketing system alone alongside huge quantities of data from sensors on vehicles and signalling. This is combined with insight from customer surveys and social media to increase its practical application to: plan services more efficiently; prevent asset failures and thus improve reliability; provide customers with more personalised travel information; and improve customer service.

4.4   Investment in modern analytical tools and skilled people is allowing us to push further and faster in discovering more from the data we and others hold. Data from different areas of TfL are shared across teams of analysts.

4.5   An essential element of the challenge is retaining a clear focus on the questions being posed of our data to meet the needs of modern London. Only with a clear definition of the improvements we must make can the oceans of information available deliver practical benefit at proportionate cost.

5   Current Exploitation of Big Data

5.1   There are numerous examples of where data has driven change in transport and we are going further almost every day.

5.2   For many years we have used rail station entry and exit data for network planning from touches in and out with Oyster and then contactless payment cards. We have now gone further and infer where customers are leaving a bus (where no exit touch is required) through a Big Data tool combining bus location and ticketing to create origin and destination pairs. As a result, we now have a comprehensive picture of travel patterns across rail and bus networks to improve network and interchange planning and review the impacts of closures or diversions. This information was previously only available as a static snapshot from costly manual surveys.

5.3   Customer service is also being improved. Through our analysis of journey patterns, we try to avoid customers incurring a maximum fare by automatically completing journeys if customers have been unable to touch out – for example, due to a major sporting or entertainment event where the gates have been left open for safety reasons. We also provide automatic refunds in the event of major disruption so that customers themselves are not put to the trouble of having to 'apply' for a refund.

5.4   We also use data to improve road safety. Analysis of trends from death and serious injury on the roads has enabled us to identify the major contributory factors and to target preventative action. We have also recently released the London Collisions map carrying extensive information on the location of accidents so that our partners can also help reduce the numbers of people killed or seriously injured.

6 Plans for Big Data

6.1 We are constantly looking for new ways to unlock the power of data to improve public transport and roads and increase operational efficiency. In particular, we see further scope to harness existing and future roads data sets.

6.2 We are also working to identify complementary data sets from other organisations to combine with our own. Road safety is a priority where hospital emergency data could be used to better understand the pattern and nature of road traffic injuries and help target action even more effectively.

6.3 The Surface Intelligent Transport System programme, which will upgrade our legacy traffic control systems, will provide us with vastly improved operational data. For the first time we will be able to bring together our roads data sets in one place. This data, gathered for example from on street sensors, cameras and counters, combined with third party data sets and our extensive traffic models, will improve our operational management of the road network and enable us to provide more information to customers and road users.

6.4 We will continue to build a deeper understanding of our customers across the transport network – both public transport and private services. We can gather information from all our customer 'touch points', including our vehicles, website and customer information campaigns as well as ticketing. This end-to-end view will improve our already strong communication with customers, offering even further personalised information.

6.5 We are also integrating our ticketing back office to provide better connections and pricing incentives between our services. The latest example of this is moving towards integrating cycle hire payments into the Oyster/contactless back office.

6.6 On London Underground, services are becoming more reliable year-on-year and big data is helping us analyse the performance of our tracks, signals and trains. We will use it to develop efficient predictive and preventative maintenance regimes that will help us prevent delays by spotting and stopping major faults before they actually happen.

7 Open Data – Background and Current Use

7.1 As reliance on data increases, it is vital that we release information in a usable format to enable the analytical and other skills of third parties.

7.2 Our published policy is that our data should be made freely and openly available unless there are compelling reasons not to do so, such as prohibitive cost. This approach has: stimulated innovation in information provision faster than we could ever have delivered ourselves; extended the reach of information to harness public adoption of mobile devices; saved money because we have not had to develop and maintain products ourselves; and has made a positive contribution to perceptions of our openness and transparency.

7.3 To date, this has largely concentrated on public transport travel information where we lead the industry globally in real-time service status such as arrivals and departures. With National Rail Enquiries now also making their data freely and

openly available, we are seeing products which combine our respective data to deliver more integrated information and journey planning products.

7.4   The Open Data section of the TfL website has more than 5,000 developers registered to receive the data sets which power over 360 apps and other products in the form which best suits our customers. This has created a community of developers with whom we engage to explain our approach and plug gaps in the market.

7.5   An independent study, commissioned by the Department for Business, Innovation and Skills in 2013, reviewed the value of public sector information for the UK as a whole. As part of this, Deloitte estimated that our open data approach saves passengers up to £58m per annum in time against our £1m investment in making the feeds reliably available. Further work is underway to quantify the additional 'reach' of information through third party products beyond the information we provide directly ourselves.

7.6   As mentioned, last month we launched the Capital's first interactive digital collision map. This is part of a continued drive to reduce the number of casualties on London's roads. The map allows easy search for collisions anywhere within London with information about when, where and how severe incidents were back to 2005. On the first day, the map was accessed by over 20,000 people. An Application Program Interface (API) will be made available shortly to make it even easier for third parties to manipulate the data.

7.7   We are also running a six month trial with twenty three boroughs of the Surface Playbook. This is a geographical information system that provides access to all the projects and programmes planned for the London road network, including significant London borough projects and those led by utilities and third parties.

7.8   The information will be used by the boroughs to help inform their decision-making when planning future schemes and programmes on their roads. Once the trial is complete in January 2016, the feedback we receive will be used to develop a solution for the wider external publication, helping us to extend the benefits to as many people as possible.

8   Current Organisation

8.1   Our approach to data is overseen by the TfL-wide Customer Group, chaired by the Managing Director, Customer Experience, Marketing and Communications, and driven by our overall business priorities.

8.2   The TfL Leadership Team, chaired by the Commissioner, sets the organisation-wide data policy. Groups of analysts work together to tackle specific issues and pool knowledge. It is by definition neither possible nor desirable to 'centralise' this work because that could unduly slow things down in a world where agility is key. But we are working on greater visibility, co-ordination and prioritisation via the Customer Group and are supporting our teams to share outputs and learning.

8.3   A sharp focus on Privacy and Data Protection guides our work. Our use of Privacy Impact Assessments and the adoption of 'privacy by design' is key for ensuring that our big data analytics are compliant with privacy and data protection

requirements. We also explain how we use the data we collect including via www.tfl.gov.uk/privacy.

8.4 We are evaluating the extent to which the future use of mobile device connection data (captured by both TfL and telecoms service providers across our public transport and road networks) could be used to make further improvements to transport while continuing to protect personal data.

9   Conclusion and Next Steps

9.1 Our Big Data and Open Data approach is directed towards delivering a better city for London's growing population. We will continue to use agile prototyping and proof-of-concept to demonstrate the value of big data and to produce useful products quickly.

9.2 We are currently capturing an inventory of all data sources to ensure analytical teams have visibility of the assets available to them and that we have a clear plan for making even more of them openly available to third parties. This will also help us to identify gaps and find ways of addressing them. This inventory and what it means for the future will be brought to a future meeting of the Panel.

List of appendices to this report:

Appendix 1: Science & Technology Select Committee: Big Data Inquiry Submission

List of Background Papers:

None

Contact Officer: Vernon Everitt, Managing Director, Customer Experience, Marketing and Communications
Number: 020 3054 7167
Email: VernonEveritt@tfl.gov.uk

House of Commons Science & Technology Committee

The Big Data Dilemma

Written Evidence from the Greater London Authority and Transport for London

1.   Introduction

1.1   The GLA and TfL welcome the opportunity to contribute evidence to the Science & Technology Committee's inquiry into the opportunities and risks of big data. This submission is based on these core arguments:

1.1.1   Large cities like London are huge data ecosystems in their own right. A 'city data' focus is therefore necessary, and can deliver additional benefit to a national approach.

1.1.2   Across a range of city challenges and opportunities (e.g. to better co-ordinate infrastructure planning and the impending Internet of Things) there are significant opportunities for big data to be more thoroughly exploited to deliver social, economic, and environmental value, and service innovation. These opportunities for data analysis must be considered in the context of data protection and data privacy. However, with appropriate safeguards in place, Big Data tools and analysis are compatible with a city's responsibility in this area.

1.1.3   The role for city government in driving value through the data ecosystem needs to be more sharply defined. City government needs to set out a) how big data can be deployed in pursuit of desired outcomes and b) to actively address a range of underpinning issues - legal, regulatory, organisational and technical – so the city data ecosystem functions at an optimal level. This is a significant advance on today's open data publishing model and one which requires orchestration through a City Data Strategy and strong political support at city level around issues like data sharing.

1.1.4   While skills are present and growing, city government is still not fully equipped to deal with the Internet of Things and big data. City administrations need to be more intelligent customers of big data services and to develop their own data science skills.

2. London the Data City

2.1   London is an entity in which a vast array of data is created, published and exploited (e.g. TfL holds data on 11 billion journeys per year). Data drives a detailed understanding of population, assets, the economy, indeed almost every other aspect of city life. Economic activity is detailed through workforce statistics; mobility patterns through automatically collected transport data; housing patterns through occupancy and sales data.

2.2   Data and sophisticated modelling techniques have long been at the heart of city planning. Projects such as WITAN (see para 6.5 below) use cutting-edge tools and analytical capabilities to push data discovery further and faster.

2.3   (Big) data should be treated as a valuable asset class in its own right. As the realisation of its power to deliver better policy, operational efficiency and customer outcomes it deserves more attention at city level.

3. Open Data is no longer enough

3.1   A very large proportion of the data made open in cities is generated by the public sector. The internationally acclaimed[i] London Datastore contains over 600 datasets, but few among these can be described as big data. Transport data feeds aside, it is a very high quality but largely static data catalogue.

3.2   The public sector's strategy of opening up proprietary data in machine readable form so that third parties can develop products or analysis to benefit stakeholders and the wider digital economy, has been an undoubted success. The Open Data section of TfL's website has more than 5,000 developers registered to receive data sets.

3.3   The approach has stimulated innovation, especially in the field of transport information provision - around 360 apps have been created - and extended the reach of information to harness public adoption of mobile devices.

3.4   An independent review of the value of public sector information for the UK, commissioned by the Department for Business, Innovation and Skills in 2013, specifically highlighted TfL's effort. In the accompanying case study, Deloitte estimated that this open data approach saves passengers up to £58 million per annum in time against a small investment of making the feeds reliably available.

3.5   However, without further incentives to encourage more consistent, higher quality and higher volume 'big' data from a wider set of suppliers, cities like London will fail to capitalise on the potential of big data to tackle the complex questions with which cities have historically grappled and deliver potentially transformative innovation.

[i] Open Data Publisher Award 2015; Open Data Institute Annual Awards; http://theodi.org/news/odi-2015-award-winners

4.   Big Data in City Policy and Practice

4.1   Increasingly, city policy recognises the power of big data. The London 2050 Infrastructure Plan recognises the potential of tools such as Building Information Modelling (BIM) to define assets and generate high quality big data to be used in the planning and future operation of places. In development schemes like Old Oak Common, we are considering how a range of smart themes, including big data, can be incorporated into planning processes.

4.2   Big data is being actively considered in the London Energy Plan (to improve electricity demand and supply calculations) and in medical research to explore how through a Digital Health Institute big data exploitation, with careful regard to patient privacy, can enable commercialisation of innovation in digital health.

4.3   TfL is already drawing on big data to transform transport services. Twelve million daily public transport trips make Oyster and contactless payment cards a significant source of big data. Nineteen million daily 'taps' from these systems allow travel patterns to be studied, bringing a depth to our understanding of customer profile and behaviour. The additional 18 million car, cycling, and walking trips provide a phenomenal 30 million daily journeys on the TfL network that are fit for big data analysis.

4.4   London's richly detailed travel data feeds into transport planning models to predict the impact of development in our City. TfL has long used station entry and exit data for network planning. We can now also infer where people are leaving a bus through a big data tool (ODX) that combines bus location and ticketing data to create origin and destination pairs. This creates a comprehensive picture of travel patterns which network planning teams can use to minimise the impacts of closures or diversions.

4.5   Data captured through ticketing, vehicle sensors and signalling is combined with insight from customer surveys and social media, to increase its practical application. This information is used to plan services, monitor assets and provide customers information on new services or work affecting their journey.

5. **The Role of Government - Data Supply and Harmonisation are Vital for a Big Data Ecosystem in Cities**

5.1   The realisation of the big data opportunity is contingent on supply. Undoubtedly, more can be done to unearth and enable the (secure) sharing of data.

5.2   This is best delivered through a 'city data' approach, in which city government sets out in relation to city challenges and innovation opportunities the big data it wishes to see exploited. It is also the job of city government through policy, regulation and political pressure to ensure that other bodies presiding over large value generating tracts of data (e.g. utilities, developers, banks and housing associations) share it to allow for its exploitation.

5.3   There remains a serious need for more extensive data sharing across the public sector (e.g. between health services and city government). Much opportunity lies in melding together data to develop an even better understanding of how London operates. For example:

    5.3.1   TfL is identifying complementary datasets to drive more targeted action (e.g. road safety statistics and hospital emergency data can be merged to better understand the pattern and nature of road traffic injuries).

    5.3.2   The GLA and Mayor's Office for Policing and Crime 'Information Sharing to Tackle Violence' project aims to improve public safety and reduce violent crime by data sharing and analysis between health services, police and community safety partners.

5.4   A London-specific issue is that of data sharing across Borough Councils. Here, legal and resource issues persist, but stronger collaboration can deliver harmonised and, with regard to the internet of things, interoperable data. There is a clear value case for this in London, where policy issues frequently cross administrative boundaries.

5.5   The GLA's Schools Atlas Service[I] is a powerful example of how harmonised data - in this case used under license from DfE - can be used to address a strategic issue like school places demand.

5.6   The further potential of harmonised, interoperable and shared big data to drive reform and efficiency across a range of city services and market-related issues means that an examination of potential use cases should be a priority. Areas we see as prone to data gaps and/or which are candidates for big data approaches include:

    5.6.1   energy markets - more granular and near-real time smart meter data can be used in aggregate to better manage the peaks and troughs of energy

---

[I] http://www.london.gov.uk/webmaps/lsa/

supply and potentially, with proper data protection principles in place, at a household level to deliver better energy deals.

5.6.2 infrastructure - through the London Infrastructure Map[I], London now has its fullest understanding of the development pipeline ever. In terms of resilience though, we can undoubtedly learn more about the interdependencies between different pieces of infrastructure and how these are affected by a variety of events.

5.6.3 waste management - urban waste analytics projects[II] can draw together operational and other data sources (e.g. meteorological, socio-economic, housing type) and use predictive analytics techniques to set out future waste generation rates across communities. This could potentially increase recycling or lead to new, efficient service models that cross administrative boundaries.

5.6.4 city skills markets - currently a devolution 'ask', a better supply of data from government and companies like Linked-in can provide a more holistic view of skills demand, to be used to develop a market for the provision of more effective skills, better tailored to the needs of the city economy.

5.6.5 connected citizens - powerful personal devices consume and transmit large amounts of data. With permission, cities can use this big data input to provide better services (e.g. active management of health conditions) and to manage local environments.

5.7 It is the role of city government to provide the enabling platform which will allow for the open, and where necessary, secure sharing of big data flows. This next-stage Datastore - in effect an urban operating and data innovation platform – will, supported by a City Data Strategy, be a mainstay for further economic growth and social value.

5.8 Our view is that in pursuit of a broader, assured supply of relevant data, all data need not be open. Indeed, for some areas of interest to cities, there are compelling reasons why data must be shared securely and without open dissemination. Not least is the prevention of misuse of sensitive data, where there is a risk of re-identifying individuals whose identities have been masked, or where even non-personal data sharing might result in risk of harm.   The ODI Open Data Spectrum[III] is a good reference point here.

[I] http://www.london.gov.uk/webmaps/2050_Public/
[II] Urban Waste Analytics; Cazap, Ianiuk, Lui; Centre for Urban and Scientific Progress, NYU; 2015
[III] http://theodi.org/data-spectrum

6. Big Data Skills in City Government, Leadership and Partnerships

6.1   Compared to domestic and international cities, London has a sophisticated approach to big data, with successful practices existing in parts of city government. Some of the large cohort of analysts are equipping themselves with the skills - coding, solving problems in new programming languages, hacking data, high standard visualisations - of this relatively new profession, and migrating towards it. There is, however, a potential shortage of data science skills in waiting.

6.2   Dramatically increased data volumes and velocities require more data scientists with the combination of analytical and data science skills to make best use of big data across a wide range of sources (e.g. to deploy the large-scale distributed data pipelines which enable powerful analysis). There is also a need to build capacity around experimenting with emerging (open source) technologies.

6.3   Leadership is needed within city government to embed these new skills in policy making. More explicitly, the role of this leadership is to ensure that data science exploration is creative, but also generates practical outputs for highly relevant policy areas. This must be supported by educating senior leadership to create a big data culture, meaningful to both technical and non-technical audiences. Management skills are also required to ensure that IT policy can accommodate future technology developments and optimise, rather than blunt, data science skills.

6.4   This leadership must also drive external partnerships to build capacity. We advocate stronger links with the Cabinet Office's Data Science Team (and entry into their Data Science Accelerator Programme). The GLA and TfL already enjoy big-data-oriented academic partnerships. While the Mayor's Smart London Board is an excellent conduit for these projects, we will forge even stronger links with academic research departments. The relationship between the Centre for Urban and Scientific Progress at New York University is an excellent example of how City policy can drive, and in turn be underpinned by, complex big data analytics.

6.5   An area we would like to see grow, the GLA is already active in partnerships with small businesses, expert in big data analytics. With Mastadon C, we were successful in winning Innovate UK grant to build a city modelling platform (WITAN), drawing together economic, labour market, population, transport and housing data in one interrelated model to aid city planning. While small businesses have the opportunity to commercialise the product for resale, city government benefits from skills transmission.

7.   A City Data Strategy

7.1   The GLA is producing a City Data Strategy to underscore the importance of big data and to prepare the city data ecosystem to take full advantage of it. Making necessary links to broader city policy aims, this strategy will communicate city government's role predominantly in data publishing, needs identification and governance. It will also recognise the needs of the data value chain and those of related technology fields.

7.2   At the core of our proposals will be a common reference architecture. This will be vital in marshalling city data (both open and proprietary) produced by a multitude of systems, devices and applications, the logistical distribution of which varies according to suppliers, their sectors and distribution channels, and the policies and regulations to which they are subjected.

7.3   Other non-technological aspects to the strategy include business models (e.g. monetisation and experiences in city data economies like Copenhagen), value networks, standards to ensure data quality—and how to indicate the confidence level of data quality from source systems—treatment of costs, and data sharing between agencies and their broader use in accelerating smart technology in cities[I].

7.4   The implementation of data sharing models will require legal consideration: of licensing arrangements, access to third-party-input data, Intellectual Property Rights and data ownership of outputs and algorithms. The strategy can provide helpful guidance for approaching these complex negotiations between agencies, and also for highlighting contractual considerations of IPR and any personal data processing if analytic services are outsourced.

7.5   This strategy will also address data privacy and data protection to ensure legal compliance, and to preserve the trust between the city and those whose data is being analysed. Data Subject rights apply for big data—people have the right to be told what personal data is being processed about them, the purpose, and to whom it may be disclosed. This is particularly a consideration for public bodies, where customers do not generally have a choice of services. Therefore, Big Data applications should adhere to best practice for data security and data privacy. To achieve that, the strategy will discuss the approach of embedding privacy by design and use of Privacy Impact Assessments (PIAs) for big data databases and applications.

7.6   Many smart solutions offer greater personalisation, but in order to gain insight, disaggregate-level data must be shared and combined. Currently, there is a deficit of public trust and a reluctance to share personal data[II]. To address these concerns, the GLA has contributed to the Digital Economy Catapult's

---

[I] PAS 182 *Smart city concept model. Guide to establishing a model for data*, currently under development by the British Standards Institute

[II] 'Trust in Personal Data'; Digital Economy Catapult; http://www.digitalcatapultcentre.org.uk/pdtreview/

work to build a personal data sharing 'Trust Framework', designed to create guidelines which build consumer confidence. This is an area where guidance from the Information Commissioner will also be of paramount importance, and the GLA and TfL are working closely with them[I].

7.7   The strategy will also consider the Legislative framework, in particular the impact of the new EU General Data Protection Regulation, which is likely to have a significant impact on how big data algorithms are applied, and how personal data is considered vis-à-vis anonymised and psuedonymised data. The Freedom of Information Act and the Re-Use of Public Sector Information Regulations that apply to public bodies also must also be considered.

8.   Recommendations

8.1   We believe greater determinism from city government is needed to extract maximum value from data ecosystems which correspond to cities as social and economic entities. Lessons learned - cultural, legal, technical, but also those about meeting city innovation and growth challenges - from implementing a City Data Strategy in London should be transferred into other settings.

8.2   City data ecosystems will only achieve maximum effectiveness if a wider range of big data is shared. As city data is defined as much by the private as the public sector, and because some data may be shared and monetised, we suggest that under strong political leadership, cities publish on an annual basis the case for what data is to be used (and the conditions of use) to help deliver against policy commitments.

8.3   Further, we request that the committee supports the importance of 'free to access and use' digital national infrastructure data to generate maximum value (e.g. compare how free Ordnance survey mapping data has contributed to economic growth versus limited use of Royal Mail PAF data that is not freely available).

8.4   In pursuit of greater harmonisation and interoperability of city data, Government needs to commission an organisation like the Open Data Institute to investigate the costs, benefits and potential use cases for city-wide, harmonised data in complex service areas (e.g. in adult social care). Organisations like London Councils should be involved to build on early stages data sharing exercises.

8.5   This same exercise could also reasonably examine how central government exercises, like the Competition and Market Authority's inquiry into energy markets, would benefit from big data approaches.

[I] The recent Information Commissioner's Office guidance: "Big Data and Data Protection" sets out considerations for applying big data where there personal and/or sensitive data is involved.

8.6   A group of City authorities should be convened to establish opportunities for big data analysis and associated disciplines like machine learning to achieve impact. A keynote project - formed around 'testbed' partnerships with central government, academia and industry - could run over the course of a year, with a defined outcome based on an agreed city policy priority (e.g. how to use big data flows to deliver better energy management across city assets). Upon completion of the exercise, the analysis and methodology will be shared across interested cities. At the very least it should have dedicated support from the Cabinet Office's Data Science team.

8.7   In support of this, we suggest a 'Big Data' version of the successful Release of Data Fund[I].

8.8   To build a supply of data scientists, city government will actively engage with the growing number of relevant courses (e.g MSc in data science or smart cities) and deliver big data in government modules within them. These should involve work placements actively addressing city issues.

8.9   Next-stage city data or urban operating platforms, able to handle big data volume, speed and breadth will be more costly propositions than their catalogue predecessors. So they are developed with city needs in mind, a City Data Platform Technical Working Group should be established by the Cabinet Office. This should link to other exercises (e.g. the European Commission's Innovation Partnership's work on urban operating platforms). To encourage wide and ongoing market engagement, the technical group should produce a set of agreed open standards to complement those being developed for data, along with an assessment of the cost to government of implementing them.

8.10   The potential to 'build once and share many times' across government should be exploited through a Government Data Innovation Exchange. The Innovate UK model is helpful in allowing small businesses to commercialise innovation and products. Consideration should also be given to how city government can generate maximum value from its own products and promote their wider adoption (here the London Schools Atlas and GLA/MOPAC's Neighbourhood Confidence and Crime comparator are good examples suitable for export).

9.   Annex

9.1   We include as an annex a document that outlines TfL's approach to Big Data for further details.

[I] https://www.gov.uk/government/publications/breakthrough-fund-and-release-of-data-fund

ANNEX

September 2015

TRANSPORT FOR LONDON'S APPROACH TO BIG DATA: LEVERAGING BIG DATA FOR MAXIMUM IMPACT 9 SEPTEMBER 2015

1 INTRODUCTION AND SUMMARY

1.1 Transport for London (TfL) operates a huge transport and customer service business; but we need to consider ourselves also as a huge data and information business.

1.2 Data is a highly valuable asset class in its own right. It has the potential to deliver better policy, operational, efficiency and customer service outcomes. It is core to enabling us to: do more with less while responding to sharp population growth; meet rising customer expectations; and placing ourselves at the centre of the national economy and Government drive for greater productivity.

1.3 There are a number of ways to think about this, but there are three broad areas which capture the place of data in our business strategy:

- 'Big data': our own analysis of one or more large data sets, owned by us or someone else, to gain insight and thus support better resource allocation and outcomes. Examples of Big Data include that derived from ticketing and physical assets like trains or buses.

- 'Open data': making our own proprietary data sets freely and openly available in machine readable form to third parties to develop products or analysis to benefit stakeholders and the wider digital economy. Examples include travel information to help customers save time and collision data to reduce the number of people killed or injured on the roads. We also consume open data from other organisations (e.g. on health and housing) to put alongside our own.

- 'Data driven decision-making and innovation': what actually happens when Big and/or Open Data are successfully exploited. Examples include prioritising spend such as targeting accessibility improvements to benefit the greatest number of people, automated refunds when things have gone wrong and travel information products.

1.4 None of this is new to us - data has always been at the heart of transport planning. But modern tools and developing analytical capabilities are allowing us to push further and faster in discovering more from the data we

and others hold. Our customers are also looking to us to set the pace in these areas.

1.5   In order to do so, we will continue to ask the following questions as we develop our business strategy:

- What are the data sets available to us and what more can we do to exploit them to solve the challenges we face, including through private sector suppliers?

- What questions should we and others be asking of our data?

- How can we best act quickly and in an agile manner, to undertake proof of concept explorations to evaluate potential for future big data tools and analysis?

1.6   And our use of big data tools and analysis and our approach to open data is guided by the following principles:

- Transparency –communication to our customers and stakeholders on how we collect and use data, particularly important where personal data is concerned.

- Sharp focus on privacy and protection in our data collection and in our use of big data tools.

2   BIG DATA – DETAIL

2.1   With London's growing population and pressure on finite resources, the effective analysis of data to help make better decisions and become more productive has never been more important. We need to be better at understanding how our network operates, how much it costs and how customers are using it.

2.2   Millions of trips are made offering up vast quantities of data. This is captured through ticketing, sensors on vehicles, signalling and mobile devices. This can be combined with insight, such as customer surveys and social media, to increase its usefulness and practical application. We use this information already to plan services, monitor assets and provide customers with information. For people registered with us, we can personalise information tailored to their regular public transport journeys - eg when new services are introduced or work is taking place at their local station.

2.3   This also feeds into planning models which predict the impact of future development in London. For many years we have used station entry and exit data for network planning. We can now also infer where people are leaving a bus through a Big Data tool called ODX that combines bus location and ticketing data to create origin and destination pairs. So we now have a comprehensive picture of travel patterns which our network planning teams can use to improve efficiency. This information was previously only available as a snapshot gathered from costly manual surveys.

2.4   Bringing this together allows us to respond quickly to a rapidly changing city, improve network and interchange planning and review the impacts of

closures or diversions. We use data analysis to understand the impact of planned and unplanned 'events' to develop information and advice for customers. This allows us to relieve the pressure of events and recurring peak-time congestion at stations and on the roads.

2.5   To enable better use of other data sets, we have research partnerships with several of academic institutions. They work with us to identify and analyse various data sources to enhance our understanding of travel patterns.

3   PLANS FOR BIG DATA

3.1   We are constantly looking for ways to unlock the power of data to improve the transport network (public transport and roads) and organisational efficiency. In particular, there seems further scope to harness existing and future roads data sets.

3.2   We are also working to identify complementary data sets from other organisations to combine with our own. Road safety is a priority where hospital emergency data could be used to better understand the pattern and nature of road traffic injuries an help target action more effectively.

3.3   We are considering emergent technologies and Intelligent Transport Systems to provide better tools to manage the road network. New data sources from in-road sensors will provide more intelligence to help improve journey reliability and provide better information for all road users.

3.4   We are interrogating our existing customer data to better understand who are customers are and how they use our services. This deeper understanding helps alleviate customer 'pain points'. For example we can now fill in journeys of customers who accidentally miss a tap when travelling through the public transport network, automatically refunding their maximum journey charge. The understanding of customer types help us in planning the retail offering at stations.

3.5   Our aim is to build a deep understanding of our customers across the transport network—both public transport and private services. We can gather information from all our customer 'touch points', including our website and customer information campaigns as well as ticketing and customer service. This end-to-end view will improve our already strong communication with customers, offering even further personalised information.

3.6   Our vast customer data will also help us better plan and tailor our future services to be even more customer-focused. Because we can understand travel patterns in a way we could only dream of years ago, this will also unlock further efficiencies in our operational services.

4   OPEN DATA

4.1   As reliance on data increases, it is vital that we release information in a usable format to harness the skills of third parties.

4.2   Our published policy is that data should be made freely and openly available unless there are compelling reasons not to do so, such as prohibitive cost or where the data is personal.

4.3   This approach has: stimulated innovation in information provision faster than we could ever have delivered ourselves; extended the reach of information to harness public adoption of mobile devices; saved money because we have not had to develop and maintain products ourselves; and has made a very positive contribution to perceptions of our openness and transparency.

4.4   To date, this has largely concentrated on public transport travel information where we lead the industry globally in real-time service status such as arrivals and departures.

4.5   The Open Data section of TfL's website has more than 5,000 developers registered to receive the data sets which power over 360 apps and other products in the form which best suits our customers. This has created a community of developers with whom we engage to explain our approach and plug gaps in the market.

4.6   An independent study, commissioned by the Department for Business, Innovation and Skills in 2013, reviewed the value of public sector information for the UK as a whole, and specifically highlighted TfL's effort. In the accompanying case study on our work, Deloitte estimated that our open data approach saves passengers up to £58 million per annum in time against our small investment in making the feeds reliably available. Further work is underway to quantify the additional 'reach' of information through third party products beyond the information we provide directly ourselves.

5   PLANS FOR OPEN DATA

5.1   We now need to examine areas beyond travel information in which open data provision could drive better outcomes.

5.2   Road safety is the most promising area identified so far. Data is already available in machine readable format and a road safety API will shortly be made available enabling analysis of the data underlying our London Roads Collision map. We are also working with the Department for Transport and Police to enhance the timeliness and quality of casualty and collision data to better inform our near-term road safety. A separate exercise is underway to look at other roads and transport data that ought to be made openly available, including to enable more 'personalised' information products. This might include GPS data from those willing to provide it to map their entire door-to-door journey. This potentially opens a window into travel behaviour that has never previously been available.

6 KEY CONSIDERATIONS FOR GAINING VALUE OUT OF BIG AND OPEN DATA

6.1 As noted above, a sharp focus on Privacy and Data Protection guides our work in Big Data. Our use of Privacy Impact Assessments (PIAs) and the adoption of 'privacy by design' is key for ensuring that our big data analytics are compliant with privacy and data protection requirements.

6.2 We also explain to our customers about the data that we collect, its use, and its retentions. We notify customers at the point of collection and also on our website: www.tfl.gov.uk/privacy.

6.3 This is crucial for building and retaining trust amongst the individuals whose data is going to be analysed. This is especially important for TfL as a public service provider. Our customers are reliant upon our services and products in order to travel around London—they do not have the option of choice that one would have in other sectors.

6.4 Our Big Data approach is business-led, not technology-solution-led. There has been a considerable interest in the opportunities of Big and Open Data, and London is enthusiastic about the prospect for gaining better outcomes for cities from our approach. However, we would caution cities who are considering a Big Data approach to focus the following areas:

- A strategic business-led approach and not a technology driven approach.

- Using prototyping and proof-of-concept work to demonstrate the value of big data approaches before investing in significant expense.

6.5 TfL has adopted the approach of starting from key business questions to drive our investment into big data analytics. Before investing in expensive systems, we have demonstrated the value of the analysis to leadership teams. This approach ensures that big data analyses are focussed on city outcomes rather than being simply an exploration of IT solutions. To that end, we encourage cities who are embracing big data, do so through a strategic business-led/policy-led approach and not a technology-driven approach.

7 CONCLUSION

7.1 New capabilities and a focus on the right questions potentially unlocks faster progress in using data and insight to solve the challenges of a rapidly growing global city.

7.2 It is convenient to think about leveraging this within a framework of Big Data, Open Data and data driven decision-making and innovation and action is being taken under each heading.

7.3 As part of our business processes and strategy, we need to be constantly asking:

- What are the data sets available to us and what more can we do to exploit them to solve the challenges we face?

- What exam questions should we and others be asking of our data?

- How can we best act quickly and in an agile manner, to undertake proof of concept explorations to evaluate potential for future big data tools and analysis?

7.4 These questions, alongside our focus on data protection and data privacy, form our approach for Big Data analysis to solve urban mobility challenges.

September 2015