Report

Getting Smart on Data

Challenges and Opportunities for Transport Authorities from Emerging Data Sources

December 2016

Pedro Abrantes and Clare Linton

Urban Transport Group
Wellington House
40-50 Wellington Street
Leeds – LS1 2DE
0113 251 7445
info@urbantransportgroup.org
Acknowledgements

The authors would like to thank all those who attended our ‘Emerging Data' workshop at the Future Cities Catapult in May 2016:

Chris Mason (Bristol City Council)
Philippe Perret (Citi Logik)
Jarmo Eskelinen (Future Cities Catapult)
Sofia Taborda (Future Cities Catapult)
Teresa Gonzalez Rico (Future Cities Catapult)
David Ladbrook (ITO World)
Steven Sherlock (Merseytravel)
Paul Johnson (Merseytravel)
David Alexander (MyDex)
Alistair Baldwin (NECA)
Neill Davy (Nexus)
Mark Mackenzie (Nexus)
James Howe (Nottingham City Council)
Mark Garlick (Nottingham City Council)
Aran Golden (Nottingham City Council)
David Allatt (formerly at Sheffield City Region)
Dr Andrew Phillips (Sheffield University)
Paul Brindley (Sheffield University)
Martin Bowman (SYPT)
Brenda Worsdale (SYPT)
Fayyaz Qadir (TfGM)
Simon Ho (TfGM)
Ali Inayathusein (TfL)
Gurjit Dulai (TfWM)
Chris Lane (TfWM)
Stuart Lester (TfWM)
Jonathan Bray (UTG)
Tom Ellerton (UTG)
Helen Bowkett (Welsh Government)
Simon Linley (WYCA)
Peter Ratcliff (WYCA)

Thank you also to those who have provided comments at different stages of this project including Damian Bown (Citymapper), Graham Grant (Newcastle City Council), Mark Cartwright (RTIG), Ryan Sweeney (TfL), James Datson (Transport Systems Catapult) and Lee Sucker (TrustHub).
Content

Acknowledgements ........................................................................................................... 1

Executive Summary ........................................................................................................ 3

1. Introduction .................................................................................................................. 4
   Context ......................................................................................................................... 4

2. Data Sharing and Integration ..................................................................................... 8
   Mobility as a Service (MaaS) ...................................................................................... 9

3. Data Ownership and Privacy ..................................................................................... 10

4. Data Quality and Standards ...................................................................................... 12

5. Skills, Capacities and Capabilities ........................................................................... 13

6. Capitalising on New Opportunities .......................................................................... 15

7. Conclusion .................................................................................................................. 17
   Next steps .................................................................................................................... 17
   Wider recommendations .............................................................................................. 17

8. References ................................................................................................................... 19
Executive Summary

Transformative technological changes are shaping the world and unleashing unprecedented volumes of data. Transport is a key generator of emerging data and also stands to be a significant beneficiary of new products, services, insights and ways of working that accompany this. Emerging data will mean transport users will become far more fully informed about their travel choices whilst at the same time transform the ability of transport authorities to plan and manage transport networks and services more efficiently and effectively. The purpose of this report is to outline key challenges and opportunities from the perspective of city region transport authorities, and to articulate a case for action.

Why are we doing this?

Emerging data is a rapidly changing area. It is clear from the actions of some early movers that value can be created by using this data (see case studies throughout the report). However, at present, the transport community does not fully understand the challenges and opportunities that emerging data present. Therefore, this report seeks to identify some of these unknowns and communicate the case for action in this area to our member organisations and beyond.

What is the case for action?

The opportunities that emerging data present could help deliver new products, improvements in existing services, and yield new insights for planning, strategy and operations. At the same time, the growing legislative and regulatory focus on data transparency and openness is likely to require action by transport authorities and other stakeholders.

What does this report do?

This report summarises the insights from a workshop held by the Urban Transport Group and the Future Cities Catapult in May 2016, sets out some initial recommendations and begins to develop the wider case for action. By providing high level perspectives on the opportunities and giving examples and case studies of where different organisations are already beginning to extract value from emerging data, we seek to make the case for a greater number of transport authorities to capitalise further on these opportunities. We also attempt to articulate some key challenges, in particular those which we may be able to overcome more easily through joint working.

Who is the audience?

This report is mainly aimed at city region transport authority officers. It is intended to provide an introduction to the area of emerging data and the value that it can bring to transport. This should hold appeal across organisations, from Chief Executives and Directors to officers handling and analysing data. In addition, it should provide insight for those in between about the potential for capitalising on these opportunities. This report may also be of interest to those beyond the city region transport authorities in understanding the potential opportunities and challenges around emerging data and transport, and help to foster collaborative working across the transport sector.
1. Introduction

1.1. Emerging data, from a range of sources including mobile data and smart cards, offers a wealth of opportunities for the transport sector, such as original insight, ways to improve services and potential new ways of working. But it is not without its challenges.

1.2. This report seeks to identify the challenges and opportunities, from the perspective of city region transport authorities, arising from emerging data. The analysis builds on the insights gathered at a workshop held by the Urban Transport Group (UTG) and the Future Cities Catapult in May 2016. A key feature of the workshop was a series of roundtable discussions, which sought to identify potential actions points for the UTG. The UTG is beginning to develop a wider programme of work in this area and some details of the proposed direction of travel can be found in section seven.

1.3. The rest of this section provides the context in which emerging data sources have developed in recent years and sets the scene. The report is then structured around four key types of challenge that emerged from roundtable discussions:

- Data sharing and integration;
- Data ownership and privacy;
- Data quality and standards; and
- Skills, capacities and capabilities.

1.4. These are followed by a section on how transport authorities and other stakeholders are already capitalising on these opportunities, including how they have overcome some initial challenges. We then conclude by setting out next steps for the UTG and the transport sector as a whole, in order to move this agenda forward.

Context

1.5. Increasingly, data is becoming part of our infrastructure, alongside more traditional physical capital like transport networks, energy generation and water supply (Hapuarachchi, 2016).

1.6. As of 2013, 90% of the world’s data had been generated in the previous two years (Science Daily / SINTEF, 2013).

1.7. Emerging data in transport is multi-dimensional, complex, messy and ill-defined. Big data, open data, commercial data and personal data all play a role, and there may be data sources with significant future value but where potential applications are not yet clear. Hence, we do not attempt to place boundaries on what is included in ‘emerging data’ in this report, or seek to provide an exhaustive list, rather we outline some of the aspects which may be useful and interesting for transport authorities.

1.8. However, ‘big data’ is an interesting starting point, as this term is often equated with emerging data sources and understanding its meaning can also be helpful in getting to grips with the challenges and opportunities created by new technologies. Figure one shows how interest in big data has grown over time, through the number of google searches for the term.
1.9. One prevalent definition of ‘big data’ is information that is of such volume, velocity and variety that conventional analytical approaches are insufficient (Gartner, 2016; Ward and Barker, 2013).

1.10. An alternative, and broader definition, is offered by Mayer-Schonberger and Cukier (2013) who define “Big Data” as “the ability of society to harness information in novel ways to produce useful insights or goods and services of significant value” (p.2). We feel that this definition is more relevant to the way in which transport authorities are likely to approach emerging data sources.

1.11. Another commonly used phrase is ‘open data’, which is defined by the Open Data Institute (ODI, 2016) as “data that anyone can access, use or share”. This is a useful definition for our purposes, though the question of what is meant by ‘anyone’ is an important one for transport authorities to engage with. Another key set of questions is whether ‘open data’ is a good thing and, if so, how this idea can be supported by transport authorities.

1.12. The International Union of Public Transport (UITP), for one, believes that the transport sector should support open data provision, preferably on a cost-free basis, as this can:
   - help with efficiency and innovation in information diffusion;
   - encourage transparency;
   - develop the concept of ‘Smarter Cities’; and
   - stimulate job creation in the technology sector.

1.13. This in turn would help public transport customers, to improve their journey experience, and transport authorities by improving their operations (UITP, 2014).

1.14. Figure two shows UITP’s overview of open data in public transport and also illustrates the complexity of having multiple actors and stakeholders involved in using and generating these emerging data sources for transport.
1.15. Increasingly, public sector organisations, including transport authorities, are required by law to open their data, where there is no specific reason why it needs to remain closed. The 2016 Bus Services Bill is also expected to include regulation to open up private operators’ data about fares, timetables and routes so that developers can produce apps and improve passenger information (Department for Transport, 2016).

1.16. Analysis of big data and the opening up of data held by authorities and operators, should in theory help transport authorities and other stakeholders to improve transport products services and the Transport for London (TfL) case below demonstrates the added value this can bring. However, there are some concrete challenges for this to become the norm which are explored later in the report.

1.17. Mobile data, derived from the location data of individuals’ mobile devices, offers a rich data source, as two-thirds of UK adults now have a smartphone (OFCOM, 2015). This could be a useful resource for transport analysis, modelling and appraisal.
1.18. At our workshop, Citi Logik demonstrated the value of using mobile data in transport in a study for TfL where they examined the destination of customers exiting Holborn Underground station, a previously unanswered question, see figure three.

Transport for London – Innovation in Data

London faces a growing population, with nine new residents every hour. TfL have used innovative approaches to data analysis in order to help them meet this, and other transport related challenges, and improve their customer service.

Using big data from smart cards, tailored travel information is now provided to customers, as well as improved fare and refund offers. Travel behaviour is being influenced to make the best use of the network by providing information about the busiest time to pass through a station (see adjacent image).

In addition, TfL has opened up many data sets and as a result, 5,000 developers have registered to use their open data platform, resulting in hundreds of apps, tools and services.

1.19. This introduction has provided some examples of the potential added value that can be delivered through making use of emerging data source for transport, and further examples are provided throughout the report.
2. **Data Sharing and Integration**

2.1. Knowledge is power but it can be easy to over-state the importance of any single data source on its own. Through data sharing and integration, additional value and insight can be generated; mobile data on individual passengers’ movements, for example, is of less value without information on how the wider transport network is performing.

2.2. In terms of the data that transport authorities themselves hold or generate, there are questions about what to do with that data, with some of the options being to keep it, develop it, sell it, license it or open it, and each of these options pose different opportunities and challenges.

2.3. With these questions of what to do with transport data, comes the question of who should develop applications, tools and services that are driven by this data. TfL’s open data platform has allowed developers to register on their site, with over 5,000 developers at the last count, and has generated many apps as a result (TfL, 2016). Some UK transport authorities are following a similar approach whereas others have preferred to limit external access to their data and develop their own customer facing applications internally.

2.4. South Yorkshire PTE, working with the University of Sheffield, have taken their bus real time information (RTI) data and integrated this with bus network data to generate added value and insight. Figure four shows the process used to integrate the data sources to create a database that can then be queried and used to generate useful outputs and spatial visualisations. The outputs and findings from this can be used to allow better, more informed, decision making and, ultimately, improve services, demonstrating the added value of integration.

![Diagram of data integration process](image.png)

Figure 4 - Integration of bus real time information and network data for South Yorkshire (Brindley et al, 2016)
2.5. Transport API (http://www.transportapi.com/) provides an integrated platform for a range of transport open data, in order that developers can access the data in one location. Figure five shows the range of data types including departures, timetables, planning information, tweet mapping and more (Transport API, 2016).

**Mobility as a Service (MaaS)**

2.6. Data sharing and integration offers opportunities for furthering the ‘Mobility as a Service’, or ‘MaaS’, agenda.

2.7. Mobility as a Service consists of platforms where “users are offered various door-to-door options for their journeys based on a monthly subscription model for all their mobility needs, offering customers the best value for their specific requirements and project partners better visibility for their services” (UITP, 2016).

2.8. Open data is one of the driving factors in the roll out of the MaaS agenda (Hietanen, 2016), as it enables the range of information required to provide a MaaS platform to be brought together. It has been suggested that as MaaS develops, it will offer a viable alternative to owning a car (UTIP, 2016) and this could contribute to a more sustainable transport system.

2.9. It has also been suggested that data shared by MaaS providers could offer opportunities for transport operators to improve their services, by having more detailed information about their customers (Transport Systems Catapult, 2016). However, challenges around access and ownership may still remain, both for MaaS providers and transport authorities. These challenges around ownership and privacy are explored further in the section which follows.
3. Data Ownership and Privacy

3.1. Ambiguity over data ownership and related concerns by individuals over privacy pose a range of challenges for transport authorities, but also creates opportunities.

3.2. Transport data is owned and generated by a range of organisations, from developers to retailers, telecom companies and individuals, as well as transport authorities and operators. This makes for a complex environment in which to ascertain ownership and one which can be difficult for individuals to understand and navigate.

3.3. We are told there are recurring common challenges when transport authorities need to access data sets from other organisations, and there could be an opportunity for collective action in this area. One such example occurs around the use of mobile phone data, which is held by mobile network operators, but could have significant value for transport authorities.

3.4. Legislation is one way in which ownership can be clarified and amended. The Bus Services Bill (Department for Transport, 2016), for example, will require increased openness and transparency on the part of bus operators and the UTG is working with the DfT and other parties to ensure secondary legislation on open data will be fit for purpose. Clearly, this is an area where it makes sense to adopt a joint approach and more of these areas for collaborative working may emerge as the agenda moves forward.

3.5. As the value of data to both public and private organisations becomes increasingly clear, the importance of individuals’ privacy is also likely to rise. In recognition of this, government have for some time begun legislating explicitly on this subject, to create or clarify individuals’ rights and to create obligations on public and private sector organisations to protect individual data (see, for example, the Data Protection Act 1998).

3.6. The new EU General Data Protection Regulation (GDPR) is expected to come into force in 2018, and will still impact the UK after an eventual exit from the EU. This is expected to strengthen data protection regulations, placing accountability obligations on data controllers, requiring them to maintain documentation and conduct data protection impact assessment, and will require a designated Data Protection Officer to hold this accountability (Allen and Overy, 2016). For transport authorities, this will mean adhering to these new regulations which will be challenging as compliance will become more complicated than previous data protection requirements and will need considerable digital architecture.

3.7. Technology can offer some solutions to privacy concerns. For example, data portability offers a novel solution to managing individuals’ data, giving more control back to the user, as the box below illustrates. This offers a potential solution not only to some of the challenges around data ownership and trust but also creates opportunities to support mobility as a service (MaaS) applications, as highlighted earlier.
Data Portability

Data Portability offers the opportunity to put individuals in control of the data they generate and for them to share that data with organisation. Individuals can even enrich their data through providing the ‘why’ of their choices and creating a profile will allow them to share preferences and plans. The presence of secure platform, through which users can choose which data to share, helps to manage issues of trust and security of personal data.

For organisations, this reduces the costs of managing and protecting data, can provide improved insight, offers opportunities for gamification to nudge travel behaviours and can help to improve transport planning and propositions.

Examples of existing data portability platforms in the UK include Mydex (https://mydex.org/) and Trust Hub (https://trust-hub.com/) both of which are seeking to engage in data management in the transport space.

3.8. The UTG recognises that data portability offers an opportunity for transport authorities to protect individuals’ data whilst capitalising on the value of emerging data, and this could be an area for collective working. In addition, we could play a role in promoting greater access to data useful to transport authorities but which is held by other organisations.
4. Data Quality and Standards

4.1. Ensuring that information meets the quality expected by users is critical for maximising the value that can be gained from using emerging data sources. UITP suggests that open data should be "relevant, up-to-date, reliable, easy to use (...) and reusable" (UITP, 2014).

4.2. The quality of open data has been identified as a barrier to both the use and opening of data with regards to the UK Government open data agenda (Martin, 2014).

4.3. The UK Government provides some guidance on the standards required for public sector open data in the UK (Cabinet Office, 2015), see box below. However, this is relatively high level and is of limited value when it comes to specifying passenger transport data where high detail, sustained quality and comprehensive coverage are key to building and maintaining users’ trust in the system. There are also important questions about who should be responsible for monitoring and enforcing data quality in an urban transport context.

Open Standards Principles
The Government’s Open Standards Principles are intended to encourage interoperability within government and beyond, to help support other businesses to engage with the data and embed consistency across organisations (Cabinet Office, 2015).

Data that is truly open is:
- Accessible (ideally via the internet) at no more than the cost of reproduction, without limitations based on user identity or intent
- In a digital, machine-readable format for interoperation with other data
- Free of restriction on use or redistribution in its licensing conditions. (Chief Technology Officer, 2016)

4.4. Data inter-operability presents another important challenge. This refers to the standards, including software architecture and data structures, employed to store and transfer relevant data between different parts of the system and between different organisations.

4.5. The UK has historically been a leader at developing common data standards in the public transport realm, in part due to the fragmented nature of public transport supply, and the need to pool information from across a large number of operators.

4.6. Examples of existing standards include:
- ATCO-CIF standard, a notable early example, which has been subsequently superseded by the TransXchange format, TXC;
- the European SIRI standard for real time information¹;
- Google’s GTFS (General Transit Feed Specification), which has become the de facto standard for public transport network data in the US and many other parts of the world.

4.7. Given the potential for future incompatibilities if different organisations choose to adopt competing standards, questions arise around who should hold responsibility for setting data standards. This is one area where we believe there could be significant benefits from closer collaboration between transport authorities and with other key stakeholders.

¹ Developed by partners in the UK (RTIG), France, Germany and Scandinavia
5. **Skills, Capacities and Capabilities**

5.1. The emergence of new data clearly presents opportunities for the transport sector, as demonstrated in the previous sections. However, some of these opportunities require some new skills, capacities and capabilities.

5.2. Areas where skills gaps may be a challenge for using emerging data include:
- Analysis;
- ICT;
- Legal;
- Governance;
- Commercial; and
- Management.

5.3. Taking the example of analysis, mobile data requires new algorithms and approaches to data management, including machine learning algorithms and cloud-based parallel computing; these approaches are relatively new to the transport sector.

5.4. An additional dimension to the skills challenge for using emerging data sources arises from the high demand for these skills in sectors beyond transport, particularly in the lucrative financial sector.

5.5. On a positive note, the transport sector has historically used and managed large and complex data sets, and through upskilling and collaborative working, the sector can rise to, and capitalise on, the opportunities for using emerging data.

5.6. Working with those outside the traditional transport community can offer solutions to some of the skills shortages within organisations. Innovation contests and ‘hacks’ have been used to raise the profile of data sets amongst the developer community and, in some cases, to address specific challenges or issues, see box below.

**Innovation contests and ‘Hacks’**

Innovation contest and ‘hacks’ offer the opportunity to invite developers to produce applications and services, using open data, through a competition.

In Washington DC, one of the first cities to open its public sector data, an innovation contest called ‘Apps for Democracy’ was held. This generated 47 web or mobile apps in a month, tackling a range of social and service issues (McLaren and Agyeman, 2015).

In the UK, the ODI, and its regional hubs, have held data hacks, to bring together developers and users around specific issues, including flooding and cycling provision (ODI Leeds, 2016).

5.7. UTG is working in the area of skills, recruitment and retention in transport, including by hosting an internal seminar in July 2016 on this subject. UTG has created a web hub to

---

See, for example, Professor Mike Batty’s work in this area [http://www.spatialcomplexity.info/archives/2954](http://www.spatialcomplexity.info/archives/2954)

share knowledge about national initiatives on skills and recruitment and best practice case studies from across transport authorities. This could offer a useful resource to authorities in exploring the skill requirements of managing emerging data and capitalising on the opportunities presented.
6. Capitalising on New Opportunities

6.1. A number of questions arise in moving forward and adapting ways of working in transport including:

- What new ways of working, products etc. are emerging?
- How can we anticipate new trends and opportunities?
- How can we work individually and collectively to capitalise on these opportunities?
- How can we develop new ways of working in transport that have a higher degree of tolerance towards risk and uncertainty?

6.2. This report has begun to address some of these questions, however, an adaptive and flexible approach is required moving forward, as this area is constantly changing as new ideas, data and techniques emerge.

6.3. Our May 2016 workshop provided some examples of specific areas where transport authorities can already begin to capitalise on the opportunities created by emerging data, including:

- Incorporating mobile data into modelling approaches, as suggested for the TfL case in the introduction to this report;
- Visualisation of modelling outputs in order to more effectively communicate results and deliver additional understanding, figure six illustrates approaches to data visualisation used by ITO World;
- Integrating multiple existing data sources to generate new operational value and insight, as demonstrated in the case study of South Yorkshire in section three;
- Drawing on new data sources to complement and improve more traditional information, as TfL have done in integrating mobile app data including Strava into their modelling of cyclist trips (Inayathusein, 2016); and
- Making more of the information held by transport authorities open to third party service and product developers; and work with other stakeholders to ensure they take more open approach to transport data. This can help to support the development of MaaS offers (Transport Systems Catapult, 2016).
6.4. This section has highlighted some of the key areas for capitalising on the potential of emerging data sources for transport and some key questions that need addressing. The final section, which follows, draws some conclusions, provides indications of the future direction of the UTG’s work in this area and provides some recommendations on how the emerging data for transport agenda can be moved forward.
7. Conclusion

7.1. This report has reflected on the landscape for emerging data in transport, building on the insights from a workshop held at the Future Cities Catapult in May 2016.

7.2. Key areas highlighted include the challenges of data ownership and privacy, data sharing and integration, data quality and standards, and the skills, capacities and capabilities required for making the most of emerging data sources. These were drawn together and some quick win areas for capitalising on these new opportunities were suggested.

7.3. We hope that this report serves to inform a broader discussion, both within transport authorities and across industry, on how to make the most of new and emerging technologies. The report will also help to shape the UTG’s ‘Smart Futures’ work programme, which examines the implications of social and technological change for the future of transport. In parallel, we are exploring the challenges around skills, recruitment and retention in transport, both in the context of new technologies and more traditional areas of activity.

7.4. Although we have sought to identify promising areas for greater joint working, we believe that there is a case for continued bilateral engagement between our members and third parties. We hope that the May 2016 workshop, and this report, serve to highlight many of the interesting projects our members and external organisations are undertaking and encourage authorities to build on these where they see potential added value.

Next steps

7.5. The Urban Transport Group is planning to take the following practical steps to advance this agenda through our collective work:

- Setting up a high level strategic group within our professional network, tasked with looking at ‘Smart Futures’ which will take forward our work in this area and promote greater collaborative working between our members. One of the first tasks for this group will be to consider the merits of creating an officer level ‘open data group’ with the specific aim to overcome the challenges identified in this report;
- Development of a ‘Capacities and Capabilities’ work stream to support our members in this changing environment and address some of the challenges for using emerging data in transport, particularly around technical skill requirements;
- Continued development of the UTG’s ‘Insight’ platform, which brings together national datasets and data from our members on an interactive web portal.

Wider recommendations

7.6. This is a rapidly changing agenda and national policy must keep pace. In the Bus Services Bill, government is already making a good start by requiring a more open approach to data by bus operators, in particular around bus fares. However, it is important that the Bill doesn’t stop at general principles and engages with issues of detail around data ownership, integration and standards through secondary legislation and guidance.

7.7. In particular, we call on the government to clarify ownership of bus supply data, including over fares, routes, frequencies, live real time data, historical real time data and other measures of service performance. In doing so, it should acknowledge the value of this
information for transport planning and strategy purposes, as well as for the development of new mobility services.

7.8. The Modern Transport Bill, announced in the 2016 Queen’s speech, offers an opportunity to address the challenges identified in this report in a more comprehensive way.

7.9. On data ownership and privacy, the Bill should focus on emerging data relating to individual travel behaviour, such as smartcard and contactless payment records, as well as mobile, bluetooth and wifi tracking data. The Bill should do two things: re-assert and guarantee individual privacy in line with wider legislation, and consider appropriate legislative measures to ensure that greater social and economic value can be created from this information. The concept of data portability, covered in section three, offers a possible way forward.

7.10. The emergence of new transport data sources creates significant social and economic opportunities, and we expect will continue to do so for some time. It is therefore important that government and the wider industry continue to take these into account as part of their strategic thinking, and act proactively to address specific challenges where appropriate.
8. References


Hietenan, S (2016) The launch of MaaS Global, Presentation to the TravelSpirit Conference, 28/06/2016, Manchester, UK


Inayathusein A, (2016) Using emerging data sources to model cycling, presentation to the UTG Emerging Data Workshop, May 2016

ITO World (2016) Helping the world make better decisions with data, presentation to the UTG Emerging Data Workshop, May 2016


Science Daily and SINTEF (2013) Big Data, for better or worse: 90% of world’s data generated over last two years, https://www.sciencedaily.com/releases/2013/05/130522085217.htm


