INTRODUCTION

Data in public transport (PT) is a complex issue. Coming from an ever-growing variety of sources, different practical and tangible uses for data are emerging: from network planning to predictive maintenance to developing travel planner applications, and everything in-between. Indeed, it is widely commentated that data is the fuel for innovation.

But there are a number of issues: availability, integrity, cost, confidentiality and privacy to name just a few. As public transport companies grapple with these challenges and opportunities, it is necessary for all stakeholders to come together to agree on principles of data sharing, responsibilities and common goals. The objective of this paper is to raise awareness regarding the potential of data, as well as proposing an approach for multi-stakeholder cooperation.

WHAT IS DATA?

It is important to understand the difference between data and information. Put simply:

- Computers need data. Humans need information.
- Data is a building block. Information gives meaning and context.

In essence, data is raw. It has not been shaped, processed or interpreted. Once data has been processed and turned into information, it becomes palatable to human readers and becomes useful for decision-making.

The term ‘Big Data’ is difficult to define but generally refers to gathering large amounts of data from many different sources, to be used for all kinds of purposes. Big Data sets are so large or complex that traditional data processing application software is inadequate to deal with them.
In 2014 UITP published a position entitled ‘The Benefits of Open Data’. The paper acknowledges the increasing pressure for PT companies to provide Open Data, and encourages members to respond positively. It points to experience demonstrating that the provision of Open Data delivers net benefits. It puts forward a preference for Open Data to be provided machine-readable and with no or very limited restrictions in terms of costs, copyrights, patents or other controls. It generally refers to dynamic real-time service information, locations of stations and stops as well as planned schedules, fare products and price structure. It also makes clear that personal information about passengers, anything sensitive or confidential, or data copyrighted by a third party is out of scope.

This position is not without controversy with debate still on-going, particularly concerning ownership and the costs of providing Open Data. Indeed, since the position was published, the sector has seen a rise in platform economics, the value and popularity of information apps and tracking data for all purposes of the business.

In the fast paced digital reality of this era, it is healthy to constantly question these issues as they evolve and progress. UITP’s position still stands but it is clear that there is a need to take a fresh look at data in public transport in general. It is clear that the scope is no longer limited to journey planners. There is a wide variety of uses for data. In this new context, which data exists, how it could be used and by whom, regardless of the Public Transport (PT) company’s policy on Open Data, needs to be (re)examined, together with issues of governance. UITP’s role is to facilitate the debate and propose recommendations going forward.

---

**OVERVIEW OF OPEN DATA IN PUBLIC TRANSPORT**

Once data has been processed, it becomes useful for decision-making

---

Originally, discussions centred on data necessary to create journey planner applications. This is still central to the debate, and as integrated mobility platforms become more and more to the fore, it will only grow in importance. However, there are many different kinds of data and information to be found in the PT domain.

Data coming from the PT domain include:

- Network description: routes, lines, scheduled stop points etc
- Timing information and vehicle scheduling
- Timetable information (static and real-time)
- Travel behaviour data
- Operations monitoring and contro
- Fare management
- Management information and statistics (including service performance indicators)
- Staff management
- Structural assets data (elevators, escalators etc)
- Personal data (names, addresses, credit-card information, phone numbers, email etc)
- Logs of safety and security incidents
- Logs of complaints

There is also data and information coming from other sectors which can be relevant to PT, for example:

- Meteorological data
- Security data (intelligence, crime logs etc)
- Car traffic data
- Travel behaviour data from other mobility providers (bike-sharing, car-sharing etc)
- Mobile phone data
- Wifi data
- Parking data

There are many uses for data. There is much discussion around which data sets are “high value”. In fact, some data-sets are high value for internal purposes (for example operational efficiency), others are only high-value, or at least increase in value, when shared (for example security data).

There are multiple reasons why data is useful in the PT sector. PT is becoming more and more digitised meaning that data and how it is used is becoming crucial. For PT, and therefore urban areas, to thrive in the highly competitive mobility sector in which it now finds itself, it must focus on this opportunity.

Governments are becoming increasingly aware of the added value of Open Data to achieve certain goals, which explains the political pressure towards Open Data policy. For example:

- Stimulating innovation and jobs in the technology sector
- Extending the reach of information beyond that which could be achieved by transport organisations alone
- Ensuring openness, transparency and efficiency in the operation of transport services paid for directly or indirectly by the public

In more concrete terms, here below some key areas are described whereby data can be used to great benefit:

**SMART CITY**

PT data is increasingly considered in the context of the Smart City concept, whereby it is on the one hand one of many data sets required for such a concept to work, and on the other hand part of an ecosystem whereby the PT system can benefit from blending data from external sources, such as meteorological data.
MOBILITY AS A SERVICE (MaaS)
The MaaS concept is an integrated mobility concept whereby different transport services from public and private providers are unified through a platform which creates and manages the optimal trip and for which travellers pay for with a single account. This concept allows the integration of traditional PT services as the backbone of mobility in the urban area, together with complementary modes (bike-sharing, car-sharing) including new trends such as ride-sharing and e-hailing services. The concept is possible thanks to integrated digital platforms. A successful MaaS is dependent of the sharing of data among the contributing modes.

INCREASED OPERATIONAL EFFICIENCY – BETTER SERVICE PLANNING
Increased operational efficiency can be achieved through the smart use of resources. For example, energy consumption on-board vehicles and other installations such as escalators and lifts, eco-driving, building heating and lighting and so on, can all be achieved thanks in part to data collected from the different sources, allowing for smart management of resources.

Better service planning can be achieved through the analysis of travel behaviour data.

PREDICTIVE MAINTENANCE
The maintenance of infrastructure, equipment and rolling stock can be pre-emptive rather than reactive thanks to the availability of an increasing amount of data. Maintenance orders can be generated automatically at the right time, relieving human intervention and avoiding unnecessary interventions or mistakes. It allows to move from interval- and time-based preventive maintenance to condition- and status-based predictive maintenance, intervening ‘just before the failure occurs’. This leads to a higher availability of rolling stock and infrastructure and therefore to better cost-efficiency, ultimately reducing the life-cycle cost.

JOURNEY PLANNERS AND CUSTOMER RELATIONSHIP MANAGEMENT
Integrated, intermodal journey planners are only possible thanks to the availability of both static and real-time data. Additionally, predictive passenger information and predictive marketing services can be developed.

BETTER SAFETY AND SECURITY
The analysis of data relating to safety and security can lead to the reduction of incidents as well as a more efficient management of incidents when they do occur. For example, the identification of hotspots and hot-times for incidents thanks to the analysis of historical data allows the deployment of security staff to the most appropriate points at the right times either as a dissuasive measure or in order to react promptly.
The sharing of security data with law enforcement, in both directions, can help both the operator and the police to achieve their mutual goals of keeping the PT network as free from crime as possible.

**SETTING AND MONITORING KEY PERFORMANCE INDICATORS (KPIs) IN CONCESSIONS**

Data can provide a concrete and tangible basis on which KPIs can be set and then monitored in concessions.

These lists are not exhaustive. A comprehensive work on data should include analysis of the value of the different sets of data from PT in relation to the different group of stakeholders in terms of the costs of collecting the data, and the value it can generate internally and externally. This would be the base for the study of the different options of capturing the value of the data in relation to the different possible governance models.

### CHALLENGES

Enabling the best use of data raises a number of challenges:

**STANDARDS AND FORMATS**

There are a number of existing standards in the PT world. UITP members should think very carefully about the approach they choose. If the PT sector doesn’t come up with solutions and suggestions soon it is expected that other sectors (politics, regulators, ICT and automotive) will dictate data solutions which may or may not be suitable.

It is therefore important to address and discuss the challenge of the interfaces for exchanging data and of the compatibility of data standards.

**DIFFERENT APPROACHES, POLICIES AND REGULATIONS AROUND THE SHARING OF DATA**

Another challenge is related to the different local, national and international regulatory frameworks conditioning the exchange and release of data.

The differences should not prevent cooperation on improving PT data quality to increase its value and sharing it. Neither should it prevent work on the technical challenges mentioned above supporting a greater exchange. This is true for exchange between public transport operators (PTOs) but also for the sharing of data between our sector and others. Here the potential different regulations applying to PTOs and other actors of the mobility sector should also be considered and overcome.

### RELEVANT INITIATIVES AT THE EUROPEAN LEVEL

> General Data Protection Regulation (Regulation (EU) 2016/679) enters into application on 25 May 2018 and envisages a harmonisation of data protection.

> On the 10th of January 2017 the European Commission published a Communication on ‘Building a European Data Economy’, accompanied by a Staff Working Document on the ‘freeflow of data’ and emerging issues of the data economy where it

  - looks at the rules and regulations impeding the free flow of data and presents options to remove unjustified or disproportionate data location restrictions; and
  - outlines legal issues regarding access to and transfer of data, data portability and liability of non-personal, machine-generated digital data.

The Commission’s aim is to define and implement a principle of free movement of data within the European Union as a natural consequence of the free movement of services and the establishment of provisions of the Treaty. The Communication includes elements on the ownership of data and the way data owners will have to distribute their data.
PERSONAL DATA, PRIVACY AND CYBER CRIME

Privacy and protecting personal data are key concerns in the context of data usage, including sharing. It is paramount that data-owners respect applicable laws in terms of personal data, which usually forbids that information, for example travel behaviour, be traceable back to the individual, as well as protecting credit card information and so on. It is also an issue in terms of cyber-security, whereby personal data is a target for certain cyber-crime. As mentioned in the above section on European Union regulation, privacy concerns are legitimate and important, and measures must be in place to protect such data.

STAKEHOLDERS, GOVERNANCE AND COOPERATION

Data ownership is a decisive issue and central to the debate regarding data-sharing. Often, the PTOs are considered data-owners because they are the ones processing it. But is it theirs to sell? Should personal data not be considered property of the individual? Or should society claim ownership, in the sense that its value ought to be returned by means of either making it available for services to be developed from it, or sold with the revenue returned to improving the PT service?

THE OWNERSHIP OF DATA AND RESPONSIBILITY FOR THE EXACTITUDE OF DATA

Furthermore, ensuring the exactitude of data is both a technical issue and organisational issue. From the point of view of the traveller regarding, for example real-time journey planners, in case of a problem, who is responsible, the service-provider (using the data provided by the data-owner) or the data-owner? Who does the customer complain to?

COSTS

Maintaining data, ensuring and improving its quality, and sharing is said to be costly. Sharing data and making it available has a cost, and some operators are facing multiple repeated requests. Depending on the regulatory framework but also on services and place, this cost can vary. The question of the responsibility for bearing this cost, and/or for sharing it needs to be addressed.

The quality of data must meet certain requirements for it to be processed. Bringing transport data to the required level of quality can represent a very significant burden. Guidelines, best practices and initiatives enabling the sharing of costs for instance between operators and other actors could be considered.

Data licencing is also an important issue to consider. When data is made open by an operator, it is important that the long-term implications of the licencing of the data be taken into account.
How does today’s connected PT traveller get his/her information?
A Channels provided directly by PTO or Public Transport Authority (PTA) to the customer, mainly through apps and internet travel planners
B Channels provided by the private sector, also mainly apps and internet travel planners e.g. Google, Apple etc
C Channels provided by (semi-) public parties. Public parties would be, for instance, local or national governments.

GOVERNMENTS
Governments are responsible for the safety and security of citizens, including related to PT network. There is a constant change in the relationship between government and society due to technological innovation and societal changes, changing threat scenarios and so on. For example, the German federal government has just passed a bill to allow more video surveillance, effectively shifting the balance between privacy and security a step closer towards the security end.

In terms of data, governments have a role in ensuring legitimate legal frameworks are in place to protect personal data and privacy, to ensure fair competition, to create an environment which delivers economic growth and so on. National governments are the first in line to set the framework for the use of data, data sharing, Open Data policy and so on. Indeed, whilst differences of opinion reign in terms of who can do what with which data, the legislature should offer an unambiguous framework that provides clarity to both data providers and data users in terms of producing, storing and reusing data, in the context described above.

PUBLIC TRANSPORT OPERATORS
PTOs are the main collectors of operational data in the PT sector. Their role is thus crucial in terms of data-sharing. As has been discussed above, there is potential for PTOs to use data to improve their own services and processes. However, whether it is making the most of common business intelligence or getting into Big Data, it is a relatively new area for PTOs and many lack the knowledge, skills and resources to tap its full potential.
A PTO can have its own policy regarding Open Data, but is often bound by the legal framework in which it operates.

PUBLIC TRANSPORT AUTHORITIES
The role of the public transport authority (PTA) is terms of data is complex but essential. PTAs, compared to private operators or other private sector stakeholders, have a responsibility to make sure that PT remains the backbone of urban mobility. In this sense, it has an essential role in making sure data-sharing policy is going to create the right environment for this to be possible. An Open Data policy which leads to loss of revenue and ridership of PT due to the creation of competitive services is clearly a risk that must be addressed. This means that data policy must be considered in the context of the entire eco-system of PT to make sure that the final aim is supported: an integrated mobility system with PT as its backbone to ensure a thriving urban area.

TRAVELLER
Thanks to extensive digitalisation of the PT experience on the side of the traveller, riders generate a lot of data. There is no clear idea about who owns this data and who should have access to it, particularly personal data. Currently by default, the PTO is generally considered the ‘owner’ of the data it sits on, and has a responsibility to protect personal data. Recent regulatory initiatives at the European Union level point to the traveller as the owner of his/her own data, however how this would be managed practically remains to be seen.

The traveller tends to find and use the most convenient services available, and services made possible by Open Data and the data-centred platform economy are increasingly popular. This puts pressure on the PT sector to get the data-sharing issue right.

PRIVATE SECTOR
It is no secret that developers are hungry for PT data. From local start-ups to tech giants such as Google and Apple, mobility is the new attractive sector to be a part of. It is in part thanks to popular new services developed by the private sector that PT has an opportunity today to redefine its role and reputation in urban mobility. Some cities choose to work with the private sector to this end, indeed they would rather that the private sector intervenes where the PTO or PTA in question does not have the required knowledge, skills, resourcefulness or imagination to create added-value services. In some places, however, a win-win collaboration with the private sector has so far proved elusive.

Besides developers, the traditional PT supply and service industry also has a role. Even for traditional products and services, suppliers are hungry for data which will help them improve their offerings.

OTHER STAKEHOLDERS
In the financial sector analysts foresee an increase in electronic and digital payments with no human interference. Financial transactions which mainly consist of transferring data will be increasingly from machine to machine, mostly via mobile devices with no distinction between the types of payment. This creates a lot of (sensitive) user data. Globally, banks are experimenting with block-
chain and other forms of decentralised data storage. The attention to digital payment also confronts banks with new challenges. Data safety, security, accessibility and the possibility of integration are essential to the success of the use of digital payments.

Now digital payments are becoming more accessible and generally accepted, security of these data transfers become increasingly important. It is a main condition to fully exploit the benefits of these digital payments. PTOs indicate that security of digital payment is essential to its long-term success. Responsibility for the use of electronic payment systems is also shifting to other parties such as PTOs and passengers.

**RECOMMENDATIONS**

- **UITP should provide a framework for discussing the governance of data**
  - For all actors in the debate
  - But also allowing each group of stakeholders to work with their peers if they wish to do so
  - With the aim of achieving robust agreements on the governance of data

- **One of the main goals of this exercise should be to facilitate and improve the framework for the exchange of data:**
  - to provide the traveller with coherent and coordinated information
  - to enable complementary mobility services, irrespectively of the entity providing them
  - to explore other areas in which data-sharing between different stakeholders can bring benefits

- **For this purpose it should attempt to identify the possible roles of PTOs and PTAs in particular.**

  It should also assess the technical barriers limiting this exchange, whether related to the need for stronger interfaces or for data standards

- **Under the leadership of TEC, identify the opportunities to capture the value of data (shared or otherwise) for various type of actors and depending on governance models**

This is an official Action Points of UITP, the International Association of Public Transport. UITP has over 1,400 member companies in 96 countries throughout the world and represents the interests of key players in this sector. Its membership includes transport authorities, operators, both private and public, in all modes of collective passenger transport, and the industry. UITP addresses the economic, technical, organisation and management aspects of passenger transport, as well as the development of policy for mobility and public transport worldwide.

This Action Points document was prepared by the Information Technology and Innovation Commission and the Information Technology and Services Industry Committee.