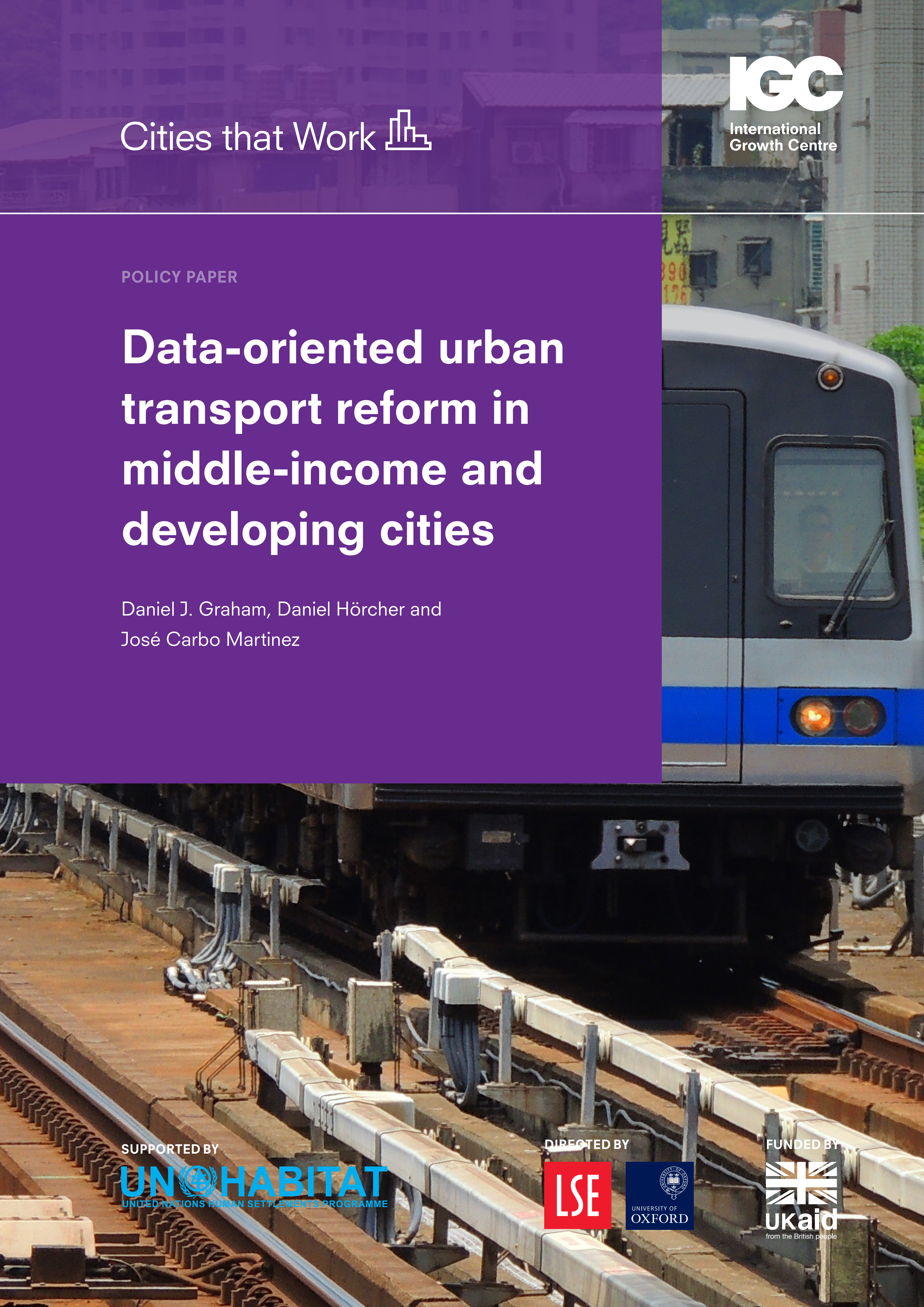
Cities that Work

IGC International Growth Centre

POLICY PAPER

Data-oriented urban transport reform in middle-income and developing cities

Daniel J. Graham, Daniel Hörcher and José Carbo Martinez



SUPPORTED BY UN-HABITAT UNITED NATIONS HUMAN SETTLEMENTS PROGRAMME

DIRECTED BY LSE UNIVERSITY OF OXFORD

FUNDED BY UKaid from the British people

CONTENTS

LIST OF CASE STUDIES

Introduction

Automated data sources are increasingly being used in almost all fields of contemporary transport science. Empirical observations and quantitative findings derived from data can support decisions of various actors of the urban society, including policy makers, road and public transport operators, private service providers and travellers. The cost of data collection is rapidly decreasing, as most of the new data sources are necessary side products of automation and digitisation of the transport industry. Also, new technologies and telecommunications generate tremendous amount of information on transport users' daily habits and travel preferences. Thus, planning and policy making can be facilitated with efficient data utilisation on both the demand and supply side of the mobility market. Despite the inevitable availability of automated data sources, making good use of them requires an investment of additional knowledge and analytical efforts. As the amount of data grows significantly faster than our learning and processing capabilities, much of the information remains hidden in untouched databases, or even get lost after a while due to the increasing cost of data storage.

The cost of data collection is rapidly decreasing, as most of the new data sources are necessary side products of automation and digitisation of the transport industry

Is the relevance of data use different in developing and middle-income cities compared to developed ones? With the acknowledgement of some exceptions, including the ones reviewed in the case studies of this report, it is fair to say that data utilisation lags behind in developing and middle-income cities due to the limited amount of investment and possibly the lack of application of data analytics techniques. However, the late adaptation of large scale data sources in transport management allows the possibility of leap frogging certain development stages, and investing in state-of-the-art data infrastructure, thus amplifying the expected returns. This paper intends to identify areas where the use of transport data could have a significant impact on local economies and society as a whole, and draws attention to the critical prerequisites of efficient use of the planned data infrastructure.

The paper highlights the key success factors determining the efficiency of data management interventions, and the benefits that the successful implementation of such policies could deliver for local communities and their surrounding economies. Key success factors include (1) a healthy balance of investments in data technology and human knowledge; (2) equal focus on data applications in both demand estimation, customer experience and the technological side of transport operations; (3) adequate attention to earlier pitfalls in the implementation of data related policies with a specific focus on privacy; (4) assurances of seamless flow of existing data within institutions and between them, including data sharing between various modes of urban transport; and (5) the efficient recycling of the footprint automatically generated by digitalised systems. The paper argues that data provides opportunities to ameliorate the disadvantages of missing investments in heavy transport infrastructure, and highlights examples where transport data can be a catalyst for positive change in future middle-income and developing cities.

Data provides opportunities to ameliorate the disadvantages of missing investments in heavy transport infrastructure, where transport data can be a catalyst for positive change

Review of transport data sources

Transport data can be categorised into two main groups based on the source of information and the method of collection.

— **Observed data:** Information is often generated with the direct intention of collecting data, when actual effort is made to observe and record transport related activities. A classic example of this type is survey based data procurement with human involvement, sometimes with manual data processing. Digitalisation and the availability of advanced information technology can improve the efficiency of targeted data collection by replacing human efforts with automated communication through online platforms or other means of information exchange.

— **Automated data:** Data can be gathered from the transport process itself, i.e. the production and use of mobility services, provided that such processes are digitalised to some extent. If the movements of people, goods and vehicles leave digital footprints, then these data items can be aggregated and processed in a systematic manner. The great benefit of this kind of data is that no additional effort is required to observe behaviour or ask travellers about preferences, because the source of such data is a digitalised process that is in operation regardless of whether the data is actually utilised later on or not.

The great benefit of this kind of data is that no additional effort is required to observe behaviour or ask travellers about preferences

In contemporary transport science both types of data are extensively used, although one may observe a gradual shift towards the second source, due to the increase of digitalised transport processes, and consequently, the falling cost of automated data collection. Therefore, for this paper we focus on the availability and applicability of automated data sources in the developing cities context.

Demand side data

One of the main goals of quantitative transport analyses is to understand human behaviour, and the resulting mechanisms underpinning transport demand. To achieve this goal one may ask people about their habits explicitly. However, with this method, there may be a bias in some of their responses for a number of reasons, including personal interests or inaccurate judgements. Automated data sources allow for the replacement of stated travel preferences with observed actual behaviour, including one-off travel decisions as well as regular mobility habits. This leads to significant improvements in terms of the quantity of data and the size of sampled users, and the accuracy or quality of observations, given that human intervention is not needed for data collection.

Automated data sources allow for the replacement of stated travel preferences with observed actual behaviour, including one-off travel decisions as well as regular mobility habits

The most frequently used sources of demand-side data are the digital signs of personal electronic devices, including mobile phones, smart phones, other consumer electronics devices with online communication, and personal vehicles equipped with GPS tracking. The method of using such data sources is straightforward: if people carry some of these devices with them when travelling, then the digital signs generated at various geographical locations highlight the most relevant attributes of trips, such as the time and duration of travel and the route and mode chosen. In developing and middle-income countries the penetration rate of smartphones has exploded rapidly in recent years, out-pacing the stages of Internet use with immobile desktop personal computers. This creates tremendous opportunities for using smartphone location data for transport analytics, as the number of active smartphone users may even exceed the usual rate in developed countries. However, local privacy laws and the fragmented nature of the structure of many mobile phone markets can make it difficult access data that are comprehensive enough to serve as a representative source of information. With a high number of operators present on the market, lengthy negotiations have to be undertaken to reach this critical amount of data.

Another typical source of automated demand data is sourced via the collection of public transport fare payment records, when digital technologies are involved in payments or ticket validation. The availability of this data source normally depends on the tariff system within a city: if fares are dependent on distance, for example, then it is likely that the main payment medium is a smart card for public transport. In cities where long-term travel passes are also available, paper tickets may still be used by the majority of passengers on shorter term journeys. Operators normally introduce smart cards due to the ease of use they provide for passengers, but the data retrieved from smart card systems is also highly valuable if the operator has the knowledge and technical background to utilise it for planning and operations[1]. In a middle-income city context, we see that tariff systems are normally based on usage dependent fares. Consequently, there are more public transport networks where smart cards are available (especially in Asia and South America) than in Europe or North America. Our experience suggests, however, that very few developing and middle-income cities, for example, Santiago[2], Chile, have in-depth experience of smart card data analysis.

The data retrieved from smart card systems is highly valuable if the operator has the knowledge and technical background to utilise it for planning and operations

Route planners and other online applications may also serve as a rich source of travel demand data, as route planning requests are normally proportional to the level of demand realised[3]. This source is highly informative because not only the attributes of the *realised* trip are revealed, but also the alternatives that the user considered before commencing the trip. With so many details of the travel decision disclosed, transport operators and planners can infer the main driving forces behind travel behaviour. Present trends pointing towards the on-demand provision of personalised bundles of transport services, often called as *Mobility as a Service*, suggest that quick data transfer could make transport

The on-demand Mobility as a Service suggests that quick data transfer could make transport supply highly responsive not only to variations in realised demand, but also to instantaneous changes in people's intention to travel

1 Pelletier, M. P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557-568.
2 Munizaga, M. A., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, 24, 9-18.
3 Llorca, C., Molloy, J., Ji, J., & Moeckel, R. (2018). Estimation of a long-distance travel demand model using trip surveys, location-based Big Data, and trip planning services. *Transportation Research Record: Journal of Transport Research Board.*

supply highly responsive not only to variations in *realised* demand, but also to instantaneous changes in people's *intention* to travel.

This concept of Mobility as a Service can be an opportunity for developing and middle-income cities to integrate a wide range of currently informal services under the umbrella of a virtual online platform, provided that information channels linking users and suppliers become available in the future. Currently, the market of informal services is highly fragmented, which makes it difficult for users to combine various modes of transport provided by multiple agents and recognise the most suitable services available[4]. Inefficiency exists on the supply side as well, as informal service providers have no information about the spatial and temporal distribution of competing suppliers, which may lead to low fleet utilisation in the long run. Mobility as a Service can coordinate supply and offer users a personalised mix of transport services. In a fragmented market of informal providers such coordination is even more impactful than in developed cities having well-organised public and shared transport networks already.

Currently, the market of informal services is highly fragmented, which makes it difficult for users to combine various modes of transport provided by multiple agents and recognise the most suitable services available

There are a number of additional devices installed in vehicles used for transport services that are capable of recording demand related information, including CCTV cameras, among others. The footage recorded inside vehicles, for example, can be evaluated with picture recognition software, thus allowing the operator of public transport service to better understand ridership patterns[5]. Some urban rail operators use train weighing systems for effectively the same purpose, while others install dedicated sensors to doors that count the number of boarding and alighting passengers at each stop. Most of the above mentioned technologies have been available for a while. However, industry-wide experience suggests that very few operators process demand data collected with isolated solutions in a systematic manner. Data fusion, in other words matching multiple sources of data, can exponentially increase the usefulness of demand data, as a comprehensive picture of ridership patterns renders causal inference, forecasting and other statistical analyses significantly more robust.

Mobility as a Service can coordinate supply and offer users a personalised mix of transport services. In a fragmented market of informal providers such coordination is very impactful

Social media is an increasingly popular source of customer satisfaction related metrics, as it has become the main platform where passengers express their impressions and frustrations regarding transport and other services. To derive reliable conclusions from large quantities of such data, the use of artificial intelligence and text recognition are definitely needed. Indeed, the primal precondition of using feedback from social media for transport planning is that a significant share of the local population of the city should use social media on a regular basis. Statistical inference from this source may be biased if just a limited part of the population uses such online applications, which is often the case in developing and middle-income cities.

4 Guillen, M. D., Ishida, H., & Okamoto, N. (2013). Is the use of informal public transport modes in developing countries habitual? An empirical study in Davao City, Philippines. *Transport Policy*, 26, 31-42.

5 Zhou, L., & Wu, G. (2018, April). An overload behavior detection system for engineering transport vehicles based on deep learning. In *AIP Conference Proceedings*, Vol. 1955, No. 1.

So far we focused on the observation of short-run demand data that gains increased relevance when fast information technology computation power allows for real-time interventions based on data collected just moments before supply decisions are made. For strategic and tactical planning, however, long-term trends of demand are more meaningful. Such trends can be inferred from short-run data, provided that computation time does not prevent the use of large-scale disaggregate data. Additionally, automated data sources can be enriched by traditional aggregate statistics from various governmental databases.

Automated data sources can be enriched by traditional aggregate statistics from various governmental databases

Supply side data

Supply-side transport data includes metrics related to vehicle movements, technological performance, service availability, timetable reliability and pricing when relevant. Most of these data sources are generated by sensors in vehicles, by the infrastructure, or traffic/train control systems. Supply-side data serve as a primary source of feedback for operators and transport agencies on the quality of the service they provide and the efficiency of engineering solutions behind it .

Supply-side data serve as a primary source of feedback for operators and transport agencies on the quality of the service they provide and the efficiency of engineering solutions behind it

Live tracking of rail or road vehicles is of primary importance among supply-side data sources. This data source was unlocked by the spread of consumer GPS devices and other positioning systems installed on public transport vehicles, and wireless data transfer techniques that enable the GPS unit to communicate with a central data base. Information provision and regulation, particularly in case of informal transport services, such as motorcycle taxis, privately operated minibuses or ride sourcing[6], real-time tracking and communication with customers can lead to significant efficiency gains. For these informal services, matching with customers and exploiting system-level scale economies due to geographical coverage is a key success factor, as the competitiveness of Uber (US), Didi (China) and other ride sourcing applications clearly demonstrated[7] in many cities globally.

Tracking vehicles is beneficial in timetabled transport modes as well, especially in managing disruptions and unexpected shocks to demand. In this case, vehicle movement data allows the operator to implement real-time interventions to reschedule or re-dispatch services. The same data can be used to inform passengers, reduce uncertainty, and offer alternative travel options during disruptions. Thus, the resilience of the entire system improves, and the flow of urban mobility can remain stable.

Tracking vehicles is beneficial in timetabled transport modes as well, especially in managing disruptions and unexpected shocks to demand

In-vehicle data collection can be used for maintenance and monitoring the technological state of vehicles on the entire fleet level. Such data is generated in large quantities by on-board equipment. Similarly, the digital output of signalling and track-side devices is often utilised for the benefit of higher technical performance.

6 Leng, B., Du, H., Wang, J., Li, L., & Xiong, Z. (2016). Analysis of taxi drivers' behaviors within a battle between two taxi apps. *IEEE Transactions on Intelligent Transportation Systems*, 17(1), 296-300.
7 Zha, L., Yin, Y., & Yang, H. (2016). Economic analysis of ride-sourcing markets. *Transportation Research Part C: Emerging Technologies*, 71, 249-266.

Major fields of application

Data can be used for all types of applied transport analysis, planning and operations. In this sense, the availability of new, large-scale data sources can benefit a wide range of potential fields of applications. In this paper we highlight some of the examples of the areas where benefits could be greatest in a middle-income and developing city context.

Real-time traffic management, transport operations

The summary of data sources in the previous section revealed that the main benefit of transport data is that it captures the movements of people and vehicles in urban networks. This is a pre-requisite for any kind of intervention aiming at regulating interactions between users and suppliers in the network. In road transport this includes the management of traffic flows[8] through signals and the allocation of road space. The equivalent set of interventions in rail modes include headway regulation, timetabling, and real-time dispatching policies when unexpected events are observed in the network[9]. In middle-income and developing country contexts, traffic management can be seriously limited by the absence of information on vehicle and passengers flows, as the traditional infrastructure of traffic counting loop detectors and traffic surveillance cameras are not established[10]. With the adequate use of immediately available automated data sources, the need for such infrastructure would be eliminated.

With the adequate use of immediately available automated data sources, the need for such infrastructure would be eliminated

The management of shared modes is even more relevant in middle-income and developing cities, as public transport in developing cities is usually much more granular than in more developed cities[11] with a dense network of high capacity rail infrastructure: the average vehicle size is much smaller, and trip patterns are closer to a point-to-point system with direct minibus services than the hierarchical network of feeder and trunk routes. Such fragmented public transport systems operated by a large number of private owners were considered much less efficient than the hierarchical structure. However, the

The use of automated data could reduce the efficiency gap with better information provision, more coordination between suppliers, smoother traffic flows, lower prices due to competition combined with higher rate of capacity utilisation

8 Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2015). Traffic flow prediction with big data: A deep learning approach. *IEEE Trans. Intelligent Transportation Systems*, 16(2), 865-873.

9 Van Oort, N., & Cats, O. (2015). Improving public transport decision making, planning and operations by using big data: Cases from Sweden and the Netherlands. In *18th IEEE International Conference on Intelligent Transportation Systems.*

10 Cracknell, J. A. (2000). Experience in urban traffic management and demand management in developing countries. *Final Report, World Bank Urban Transport Strategy Review.*

11 Gwilliam, K. (2003). Urban transport in developing countries. *Transport Reviews*, 23(2), 197-216.

use of automated data could reduce the efficiency gap with better information provision, more coordination between suppliers, smoother traffic flows, lower prices due to competition combined with higher rate of capacity utilisation.

Traffic modelling and demand estimation

For planning and policy making purposes, the core challenge is not the real-time monitoring of travel flows, but the accurate estimation of future demand for transport. The theory of travel demand modelling has advanced considerably in recent decades in terms of the methodological toolbox and practice-ready software for quick implementation in any geographical areas of application. However, the availability of reliable input data is still a major prerequisite of reliable forecasting. In developed countries, demand modelling is supported with a series of costly household surveys performed at regular intervals. For many middle-income and developing cities this may be too administratively cumbersome and costly to become reality in the immediate future, due to the high cost of manual data collection as well as the unavailability of traffic and demographic data in high spatial resolution. It is more likely that the fusion of automated data sources[12] will replace household surveys, and even outperform traditional traffic models in terms of forecasting efficiency.

It is more likely that the fusion of automated data sources will replace household surveys, and even outperform traditional traffic models in terms of forecasting efficiency

Travel demand modelling is used to forecast the direct impact of transport interventions on traffic and passenger flows in urban areas. Demand modelling is highly relevant for policy making in the transport sector, as decisions are almost arbitrary without reliable estimates of future flows or a well-functioning traffic model calibrated for the city of interest, using local data. Thus, it is not an overstatement that sharing data to improve local traffic models is in the interest of all parties and stakeholders owning transport data.

Sharing data to improve local traffic models is in the interest of all parties and stakeholders owning transport data

Bus rapid transit system, Santiago de Cali, Colombia. Photograph by David Alejandro Rendón.

12 El Faouzi, N. E., Leung, H., & Kurian, A. (2011). Data fusion in intelligent transportation systems: Progress and challenges–A survey. *Information Fusion*, 12(1), 4-10.

CASE STUDY 1: DHAKA, BANGLADESH: DEVELOPMENT OF ORIGIN–DESTINATION MATRICES USING MOBILE PHONE CALL DATA[13]

In the city of Dhaka, Bangladesh, although transport modelling was carried out, it had very low reliability. This was due to the lack of demographic data on people's residential and working locations in sufficiently high spatial resolution, and the absence of fixed roadside loop detectors or other infrastructure normally used for traffic counting. However, the penetration rate of mobile phones reached 90% by 2012, which created an opportunity for researchers to estimate the spatial pattern of transport demand from this newly available automated data source. The combination of mobile phone Call Detail Records (CDR) collected over a month, together with video data processed with picture recognition allowed researchers to provide a much more accurate picture of travel demand. Both data sources provided a cheap alternative to the other data sources usually required: CDR information is stored by mobile phone companies for billing purposes, and the estimation method worked successfully with only 13 temporary cameras installed at key locations of the road network, such as bridges and other major bottlenecks.

The supplier of the CDR data was the mobile phone operator with the biggest market share in Bangladesh. The dataset contained calls from 6.9 million users in total, which is more than 65% of the population of 10.7 million of the central part of Dhaka city. They made 971.33 million calls over the month under investigation, which implies around 140 calls per person over the same period. This call frequency, combined with the timing of calls, is sufficient to infer the geographical area where each anonymised user lives and works, thus providing a baseline knowledge of commuting patterns.

The researchers assigned the 1360 mobile phone towers to 29 major traffic zones in the city, which gave 812 origin-destination pairs. Their estimation algorithm recovered the daily activity pattern of users, including the route chosen, such that the resulting flows in the modelled network fit observed traffic flows on real-world links where the cameras were installed. This had been translated into an optimisation problem, in which mobile phone data was used as a constraint on where selected people had to be in certain points in time, while traffic camera footage served as a sample of real traffic flows in the network. Data on missing trips had been recovered by the algorithm in the way that best reproduced traffic flows in the sample. Four additional traffic counting experiments had been performed to validate the algorithm, which gave promising results overall.

13 Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. (2014). Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40, 63-74.

This pilot project proves that the need for heavy infrastructure and detailed administrative data collection can be replaced with immediately available automated data (mobile phone data in this case) to deriver reliable estimates of transport demand in a busy and highly congested city of 10 million inhabitants. The incremental costs of such experiments can be further reduced if traffic counts are also performed with CCTV footage or other sources of visual data collection devices, already installed for non-transport purposes.

Traffic in Dhaka, Bangladesh. Photograph by joiseyshowaa/flickr/.

Infrastructure investment appraisal

In the case of large-scale infrastructure projects, the assessment of competing alternatives goes beyond demand modelling: the comprehensive economic appraisal of costs and benefits can also be supported by new data sources. Several developed countries, including the UK, have made efforts to improve the acceptance of cost-benefit analysis (CBA) as a decision making tool superior to purely political arguments[14]. In middle-income and developing countries, however, this process is lagging behind, partly due to the absence of reliable data available for economic analysis[15]. Investments in this area can therefore have a crucial impact on the politics behind large-scale transport infrastructure projects as well as other high-level investments such as rolling stock acquisitions.

14 Vickerman, R. (2017). Beyond cost-benefit analysis: the search for a comprehensive evaluation of transport investment. *Research in Transportation Economics*, 63, 5-12.
15 Talvitie, A. (2000). Evaluation of road projects and programs in developing countries. *Transport Policy*, 7(1), 61-72.

The monetary valuation of travel attributes, such as values of travel time[16] or crowding inconvenience[17], is an area where behavioural models built on automated data can significantly improve the measurement of the benefits of investments, with a simultaneous reduction in measurement costs compared to survey-based methods. Furthermore, with more reliable estimates of demand elasticities, it can simultaneously improve the assessment of the benefits and external costs of *induced demand* as well. This is the demand that results as a direct consequence of transport improvements and is often neglected as part of political considerations. On the supply side of transport provision, the quantitative (i.e. statistical) analysis of engineering data can help bring cost estimates closer to reality.

Besides allowing for better measurements of the elements of traditional CBA, big data further opens up the possibility for the measurement of wider economic benefits (WEB's) of transport investment. These are the impacts that are normally not considered in traditional appraisal[18] and are mostly external benefits of improved mobility that lead to (1) higher productivity due to the compactness of labour markets, (2) social inclusion through equitable access to public utilities and jobs, and (3) more diverse societies as a result of intensifying trade[19]. In middle-income and developing countries, however, these WEB's can be even more important in magnitude relative to the traditional benefits of transport, normally dominated by travel time savings. For example, access to education, health care and other basic social services can make much bigger difference in terms of external benefits than reducing travel times for practically immobile groups of society.

External benefits of improved mobility can lead to higher productivity, social inclusion and more diverse societies

Market regulation

A better understanding of the transport system and the behaviour of economic agents through big data analytics is helpful for the efficient regulation of formal and informal transport services. In middle-income and developing cities where organised public transport services may have a relatively small market share compared to informal providers, including taxis and minibuses, the need for price and quality regulation is of utmost importance. If information on customer experience can be shared between users, service providers and the regulator, then the chances of the prevalence of abusive competitive behaviour decreases. Transport is a *credence good,* meaning that customers often have to make decisions on consumption before realising the quality of the service they choose. Thus, abusive behaviour cannot be prevented, unless

In cities where organised public transport services may have a relatively small market share compared to informal providers, including taxis and minibuses, the need for price and quality regulation is of utmost importance

16 Brownstone, D., & Small, K. A. (2005). Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A: Policy and Practice*, 39(4), 279-293.

17 Hörcher, D., Graham, D. J., & Anderson, R. J. (2017). Crowding cost estimation with large scale smart card and vehicle location data. *Transportation Research Part B: Methodological*, 95, 105-125.

18 Gibbons, S., & Graham, D. J. (2018). Quantifying wider economic impacts of agglomeration for transport appraisal: Existing evidence and future directions. *Discussion paper, Centre for Economic Performance, LSE*.

19 Grisé, E., & El-Geneidy, A. (2017). Evaluating the relationship between socially (dis) advantaged neighbourhoods and customer satisfaction of bus service in London, UK. *Journal of Transport Geography*, 58, 166-175.

the post consumption evaluation of customer experience is forwarded to other customers or a supervisory body. Large-scale data collection, coupled with subsequent analysis and communication to the consumer, can help overcome this challenge[20], and thus may be used to improve the trust of users in formal and informal service providers.

Market regulation is also a fundamental precondition of the economic efficiency of the urban transport sector , in which the spatial dimension of service provision gives ground for the establishment of local monopoly power. That is, suppliers can raise their fares if competing services are not available in their close geographical proximity, even if multiple competitors operate on the level of the entire city or region. With automated data collection, fares and other transport tariffs can be monitored in real-time. Moreover, the regulation of prices can also be turned from a static approach to a dynamic, time and location dependent, solution. For example, just as ride sharing applications regulate the fares of their drivers and make changes to these commensurate to demand, e.g. surge-pricing, within the matter of seconds.

Market regulation is also a fundamental precondition of the economic efficiency of the urban transport sector

The presence of consumption externalities in the transport sector, first and foremost time lost in road congestion, is another market failure that requires regulatory intervention at a disaggregate level. Time loss is an external cost of driving imposed by one driver on fellow road users. Economic theory suggests that the external cost of driving should appear in the price of infrastructure usage, which implies a congestion-dependent road toll. The practical application of congestion-dependent pricing was prevented by technology until recent years. However, the availability of mobile data transfer devices (e.g. smart phones) equipped with GPS systems, for example, could make dynamic road pricing a feasible alternative of other means of car taxation that do not react to the external cost of driving in congestion. Advocates within the transport economics community argue that pricing delivers a significantly more efficient allocation of road space, a resource desperately needed in busy developing cities, than queuing and burning millions of unproductive hours of its inhabitants . Transport data is once again a critical element of this line of policy proposals.

The DiDi app allows taxis, private cars and designated drivers in China to be hired via smartphone. Photograph by iphonedigital/flickr.

20 Abenoza, R. F., Cats, O., & Susilo, Y. O. (2017). Travel satisfaction with public transport: Determinants, user classes, regional disparities and their evolution. *Transportation Research Part A: Policy and Practice*, 95, 64-84.

Potential benefits identified

Investment in data use has several potential benefits for transport service operators, the actual travellers, and society overall. In this paper we highlight the key benefits from data use to be considered in the decision-making process, to help identify the interventions which will likely yield the best return on investment. The basic principles outlined here can also be utilised later on in quantitatively more accurate cost-benefit analyses.

Efficient individual travel decisions

The transport system within a city offers a wide variety of travel options for daily commuting as well as irregular business trips. This is especially the case in middle income countries where the hierarchical structure of transport networks is far less developed than in developed countries. In other words, developing cities offer more travel options including informal and irregular ones. Reliable information available for commuters can therefore significantly improve the efficiency of individual travel decisions and consequently the transport system as a whole. Additionally, real-time travel information can reduce uncertainty in large transport networks facing challenges in service reliability. Such interventions create value for society through reductions in the perceived user costs of urban mobility, in a cost efficient way compared to large infrastructure projects. Reducing commuting times can have an effect in labour supply patterns. The positive impact of reducing commuting times in female labour supply[21], in sustaining dense concentrations of economic activity[22], and in allowing urban areas to specialize in sectors with emphasis on human interaction[23] is well documented in the transport economics literature.

Reliable information available for commuters can therefore significantly improve the efficiency of individual travel decisions and consequently the transport system as a whole

Performance measurement and benchmarking

The quality of urban transport services is at the heart of many local political debates, as it affects the everyday life of almost all members of urban communities. The measurement and benchmarking of the efficiency and overall quality of service provision is therefore crucial to achieve an efficient allocation of resources within the transport sector. This can further help shape the direction of public debates with objective and comparable performance metrics. New trends in transport benchmarking emphasise the importance of customer-

21 Black, D. A., Kolesnikova, N., & Taylor, L. J. (2014). Why do so few women work in New York (and so many in Minneapolis)? Labor supply of married women across US cities. **Journal of Urban Economics,** 79, 59-71.

22 Heblich, S., Redding, S. J., & Sturm, D. M. (2018). The Making of the Modern Metropolis: Evidence from London. *Technical Note No. 25047*. National Bureau of Economic Research.

23 Michaels, G., Rauch, F., & Redding, S. J. (2013). Task Specialization in US Cities from 1880 to 2000. *Journal of the European Economic Association*.

focused performance metrics, e.g. the evaluation of service quality through the actual experience of travellers. Transport data applications enable the recovery of passenger experience in large quantities, which is a key prerequisite of effective policy decisions.

CASE STUDY 2: SANTIAGO, CHILE: COMMERCIAL BUS SPEED DIAGNOSIS BASED ON GPS-MONITORED DATA

The urban bus network of Santiago, the capital of Chile, underwent a major overhaul in 2007, leading to an impactful reform of the city's entire transport system. The package of interventions labelled as Transantiago included the complete restructuring of the regulatory framework as well as the network of bus lines, the acquisition of new buses, and the entire fleet was equipped with smart card readers and on-board GPS devices that report the vehicle's position every 30 seconds to a central database. This created conditions for detailed monitoring of bus speeds throughout the network, one of the most relevant determinants of service quality from a passenger's point of view. However, the local transport agency also faced a major challenge in utilising the new sources of data: the volume of new information made it difficult to deduct general conclusions about the performance of the network as a whole in various time periods, or even the state of individual bus lines sometimes served by multiple dozens of vehicles. The root cause of the problem was partly the absence of representative travel time metrics for a network as a whole, and the amount of data that made processing lengthy.

A joint research project with the Transantiago Authority and researchers from University of Chile[24], published in 2011, proposed a new methodology to derive bus speed metrics capable of transforming disaggregate bus location data into service performance measures at various levels of aggregation. This methodology required processing more than 44 million GPS pulses collected from 6178 individual buses serving over 700 different routes, within the period of just a single week. To make this experiment even more challenging, GPS positions had to be assigned actual streets and line segments. This is not a straightforward exercise given that the precision of GPS tracking allowed for deviations of several tens of metres, which could make the position relative to the two sides of a junction ambiguous, for example. On the level of bus lines, they computed the representative speed as a virtual metric of service quality, by interpolating observations from all buses travelling along the line in a given moment in time. By analysing the daily pattern of this representative metric, Transatiago Authority can now monitor the fluctuations of service quality between bus lines, as well as the daily, weekly or seasonal variations of service quality for any given line. At a higher level of aggregation, the same metric can be

24 Cortés, C. E., Gibson, J., Gschwender, A., Munizaga, M., & Zúñiga, M. (2011). Commercial bus speed diagnosis based on GPS-monitored data. *Transportation Research Part C: Emerging Technologies*, 19(4), 695-707.

extended to entire districts and regions within the city, or even the bus network as a whole.

The new data processing methodology is currently being applied, together with many other applications, to identify traffic management interventions with the highest potential for improvement in service quality. For example, several bus lines received temporary bus lanes in periods at critical bottlenecks of the network, where the mean bus speed dropped significantly relative to other segments of the line. Earlier on, such management policies were implemented in an intuitive manner, based on bus drivers' feedback followed by manual measurements on-site. With the new data processing method, suitable network sections for bus lines can be identified in a systematic way, considering all candidates in the network, all equipped with a comparable metric of relative service performance. This can significantly increase the efficiency of low-cost interventions such as implementation of bus lines.

Scale economies of shared transport modes

Transport technology is characterised by scale economies, i.e. the more people that use the same vehicle, the lower the average social cost of travel becomes[25]. However, exploiting scale economies requires that passengers with similar travel patterns are matched efficiently. Public transport is a traditional example of shared mobility, and recent spread of car sharing and ride sourcing follows the same economic logic, albeit on a smaller scale.

Developing and middle-income cities can reach high efficiency in the transport sector without costly investments in new infrastructure

In public transport, matching is ensured by large-scale infrastructures, such as stations. Advanced data exchange between travellers could ensure high performance matching of individual travel needs in various transport modes, thus exploiting scale economies without the need for costly infrastructure development. In other words, modern cities built on a data infrastructure no longer need physical *meeting points* to exploit economies of sharing vehicles: meeting can be ensured by fast and efficient data exchange. As a consequence, developing and middle-income cities can reach high efficiency in the transport sector without costly investments in new infrastructure . Efficiency in this context have a number of wider economic and social dimensions: it includes low average social cost per passenger as well as other benefits derived from accessibility to social services and jobs.

25 A seminal contribution to this recognition in the transport economics literature: Mohring, H. (1972). Optimization and scale economies in urban bus transportation. *The American Economic Review*, 62(4), 591-604.

Key success factors

This paper aims to identify the key prerequisites of achieving of a healthy balance of investment costs and social benefits in case of data related interventions. Success factors include the combined development of technological tools and human capacities, several aspects of institution requirements related to the flow of information between various decision makers, and privacy related considerations. It has to be emphasised here that the first success factor of making use of data is even being able to access it. Therefore, we would like to draw attention to ensuring that the regulatory and organisational environment in a middle-income and developing city enables access to transport data. Our aim is to direct attention to complementary policies that have to accompany any investment in data-based technologies.

Knowledge and human capacities

Data based methods are often considered as part of an isolated branch within the transport planning profession, especially when it comes to the application of large datasets that cannot be handled with conventional software. In fact, the fundamental problems of transport planning have not changed with the advance of big data[26]. The problems themselves have remained the same, and what new data sources can provide is simply an opportunity to improve the effectiveness of planning and operations. Thus, it is crucial to develop human capacities and quantitative skills within the transport planning profession in parallel with investments into data technology , to make sure that the information captured via expensive datasets can be turned into actual benefits for users and operators. In the short run this may be an ambitious to form a precondition for new investments in data infrastructure. Therefore, at the initial stages of improvements in data infrastructure, communication and cooperation between those who are familiar with the local peculiarities of transport operations, and those who bring new technologies into an organisation, should be ensured right from the first preparatory steps of such projects.

It is crucial to develop human capacities and quantitative skills within the transport planning profession in parallel with investments into data technology

Balance of demand and supply side interventions

As noted earlier, state-of-the-art data applications can be grouped into two main categories, depending on the end user. Customer experience is an area where new data sources have gained significant ground in the form of advanced data provision, including route planners, online timetables, incident notification, new sales channels and service rating, among others. Monitoring of the transport system and the optimisation of the supply of mobility is another key field where information can contribute to significant efficiency gains.

26 Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016). Urban planning and building smart cities based on the internet of things using big data analytics. *Computer Networks*, 101, 63-80.

From a technological point of view, however, most of these applications can be fed from the same data sources. Therefore, a crucial precondition of the efficiency of data-oriented interventions is that the demand and supply sides of the mobility sphere should have balanced access to new data sources. In practice this often implies that data should be made available for various departments of the same transport organisation, or even be shared between multiple agencies within a city . Without these channels for shared information, it may often be the case that the impact of improved service provision (supply side) on customer experience (demand side) remains hidden, and therefore the operator will not be able to monitor of the effectiveness of interventions.

Data should be made available for various departments of the same transport organisation, or even be shared between multiple agencies within a city

Privacy and the essential trust in data-based applications

The culture of data use, i.e. everyday habits in the use of technology, varies widely between different developing and middle-income countries. The lack of trust on the user's side can become a solid barrier hindering the leverage effect of data related interventions. New interventions in developing countries have the potential to leap-frog the privacy concerns that may otherwise jeopardise the value of personal information in urban transport, by leveraging lessons learnt from recent efforts in developed countries to improve the safety of everyday data use. This requires that before the implementation of new data-based policies, experience from other cities should be reviewed and benchmarked in details. Cities can learn from the failures of their peers , in case of a wide range of policies ranging from protection against cyber-attacks to the handling of fare payment data.

Before the implementation of new data-based policies, experience from other cities should be reviewed and benchmarked in details. Cities can learn from the failures of their peers

Transport data should not become mode or institution specific

The majority of the information content of transport data has no special features that prevent its application in other transport modes. Nevertheless, institutional barriers and legal requirements may, for example, limit the use of travel data to a single mode, or even only one operator within that mode. Making transport data interoperable across modes is one of the key success factors of an impactful intervention. Interoperability is a challenging task within the same transport mode in a liberalised market, e.g. when it comes to exchange of data between rail operators. Regulating data exchange in this case is a complex exercise that may require the use of economic incentives, rules of access to the market, and adjustment of the conditions of subsidisation to data sharing policies.

One way of rendering data interoperable is to actually make it publicly available for anyone. Open data policies have proven to be successful in many cities of the developed world

One way of rendering data interoperable is to actually make it publicly available for anyone. Open data policies have proven to be successful in many cities of the developed world. The success of these policies depends on the understanding of the trade-off between risk and openness, the motivations of government organisations, and the development of a culture in which open data is part of daily life[27].

Reusing (recycling) transport data for multiple purposes

Using transport data efficiently does not necessarily require investments in data collection, as data is constantly generated in various parts of the technological process of travelling. For example, the original purpose of smart card systems in public transport was to improve customer experience, but as a side product, operators can extract useful information from the digital footprint of e-ticket use. Data owners should have the ability to find all areas within transport planning and operational processes where new data sources can be re-used successfully . At some point this success factor meets the requirement of interoperability: the beneficiary of reusing available data sources can often be other service providers, or other institutions within the same industry.

The benefits are predicated on having an effective institutional and regulatory framework in place to be able to access and merge various types of data, as well as the in depth knowledge of data analytics

PLUS Traffic Monitoring Center (TMC) provided by Teras Teknologi. Photograph by Suryasuharman.

27 Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), 17-29.

Conclusions

This paper has argued that emerging automated data sources have the potential to ameliorate some of the disadvantages middle-income and developing cities face, by not having all the major transport infrastructure investments in place. The benefits, however, are predicated on having an effective institutional and regulatory framework in place to be able to access and merge various types of data, as well as the in depth knowledge of data analytics. The high penetration rate of smartphones in middle-income and developing countries is a good example of a phenomenon that leads to outstanding data availability with little investment into costly hardware. The example of Dhaka, Bangladesh, shows that real efficiency gains can be achieved in the transport network if already available data sources are utilised to better understand travel habits and weak elements of transport supply.

The paper has identified a number of potential benefits of data oriented transport reforms. Such benefits may emerge on the travellers' as well as operators' side. Information provision for the end user can improve the way people plan and undertake their journeys by (1) considering multiple travel alternatives; (2) preparing for incidents and service disruptions on the go; and (3) adjusting everyday travel habits to fluctuations in transport supply. On the operator's side, data availability helps understand the determinants of the quality of transport products, and the impact of operational interventions on the end user's behaviour and perceived benefits. In middle-income cities, can make the complicated mechanics of the transport sector more transparent, more reliable and more resource efficient.

To realise these benefits, this paper argues that key success factors include (1) a healthy balance of investments in data technology and human knowledge; (2) equal focus on data applications in both demand estimation, customer experience and the technological side of transport operations; (3) adequate attention to earlier pitfalls in the implementation of data related policies with a specific focus on privacy; (4) assurances of seamless flow of existing data within institutions and between them, including data sharing between various modes of urban transport; and (5) the efficient recycling of the footprint automatically generated by digitalised systems.

We believe that the notion of leap-frogging is especially relevant in the context of urban transport in middle-income and developing citie s: what had been perceived as chaotic or underdeveloped in the past has the opportunity to make large efficiency gains, so long as the available data sources are used effectively. The boundaries between private and public transport, from the user's point of view, can be more easily removed in middle-income and developing cities due to the absence of heavy public transport infrastructure built on the concept of a rigid hierarchy of feeder-trunk structures in most developed cities. Informal transport services such as jitneys, minibuses, taxis and rickshaws, can be transformed into a highly efficient network of direct point-to-point routes

The notion of leap-frogging is especially relevant in the context of urban transport in middle-income and developing cities

with no transfers, if traffic management and the matching between users and vehicles is improved with information. These are just two examples where transport data can be a catalyst for positive change in future middle-income and developing cities.

Smart card machine for a Moscow bus. Photograph by Stolbovsky.

Cover photo: TBC

Cities that Work

International Growth Centre