Written evidence submitted by Urban Big Data Centre (FTD0012)

Submitting Team:

Dr J Rafael Verduzco-Torres, Research Associate in Urban Transport Analytics, (Urban Big Data Centre) University of Glasgow

Dr David Philip McArthur, Reader in Transport Studies and Associate Director of the Urban Big Data Centre, (Urban Big Data Centre) University of Glasgow

About UBDC: The Urban Big Data Centre (UBDC) is a research centre and national data service based at the University of Glasgow. We promote the use of big data and innovative research methods to improve social, economic and environmental well-being in cities. UBDC has been jointly funded by the Economic and Social Research Council (ESRC), which is part of UK Research and Innovation (UKRI), and the University of Glasgow since 2014.

Abstract
- Timely, open, and reliable public timetable data not only helps transport users get around more quickly, efficiently and safely but also unlocks opportunities for innovation in research which can support improved planning and delivery of transport services.
- Despite ongoing efforts in the UK, gaps persist, hindering the realisation of the aforementioned benefits. These shortcomings compromise the UK transport system's resilience, limit research capabilities, and position the nation at a comparative disadvantage on the global stage.
- International best practices in the United States and Europe provide examples of how similar issues can be addressed and demonstrate the benefits of adopting an integrated data management vision.
- There is a strong case for a UK-wide integrated public transport open-data programme focusing on data governance, standardised data formats, and encompassing stakeholder engagement. Alongside this, a strategic push towards data standards, making historic data available, and addressing skills development among operators and stakeholders are essential.

Why public transport timetable data?

Sustainable transport options, such as public transport, are likely to be discoverable and perceived as convenient if there is timely and reliable information available. The rising importance of digital navigation has led to an increasing need to collect, organise, and distribute up-to-date public transport information in the United Kingdom (UK). In addition to the direct benefits accruing to passengers, the data can unlock the potential for innovation and improve services through policy analysis and academic research.

The use of structured public transport data in countries like the Netherlands provides a valuable example to guide the planning and delivery of transport infrastructure and land use (Conway, 2021; Conway et al., 2017). The data can also underpin the evaluation of transport projects by facilitating scenario-building (Pereira, 2019). In Great Britain (GB), this data has contributed to research indicating increased employment with improved job accessibility by public transport (Bastiaanssen et al., 2022). It has also been shown that vaccination centres which were well-connected by public transport experienced higher uptake rates during the COVID-19 pandemic (Chen et al., 2023). Private firms have leveraged this data to add value to the economy and drive innovation, as seen with companies like ITO World Ltd, Tracc, TravelTime, TransportAPI, and various journey planning applications.

Effective governance of such data can foster the adoption of international practices and enhance its use in the UK for research and business applications. This can ultimately improve the passenger experience and draw new users. We suggest that these objectives can be met through an integrated public transport open data system.

What are the biggest gaps in available data about transport networks and travel?

Currently, the UK's structured timetable data relies on commercial entities to a considerable extent. Traveline provides data for most transport modes except regional trains; this information comes from the Rail Delivery Group. Translink covers Northern Irish services. Though this data is offered to the public, it is managed and provided via proprietary platforms for registered users. Recently, the Bus Open Data Service (BODS), a Department for Transport (DfT) programme, started offering bus service data across England.

Recent developments include the approval of the 2017 Bus Service Act and the creation of BODS. However, challenges for the use of this data in the future remain. One issue is the lack of genuine open-access data.[1] Other concerns involve fragmented information across transport modes, agencies, service types, and nations. While the National Public Transport Access Nodes (NaPTAN) assists with timetable creation, some stop locations remain unreliable. Often, information is limited to static feeds, or, when real-time data is available, unique identifiers don't match static feeds.

Another concern for the future use is the heterogeneity in data exchange standards. Although efforts to maintain the main UK standard, TransXChange, exist, different sources use diverse specifications. At the same time, the international standard, general transit feed specification (GTFS), is often overlooked in the UK. This oversight causes difficulties in implementing cutting-edge analytical tools and hinders interoperability within the broader context of mobility as a service (MaaS). While user contributions have mitigated some issues (e.g., by creating tools to convert between data types), it creates additional reliance on nonofficial and non-funded solutions. Furthermore, due to a primary focus on users and industry, historical data is not available from leading platforms, limiting analytical applications.

Current conditions reduce the UK transport system's resilience, limit research opportunities, and create a gap between data availability and policy planning capability. This puts the UK at a disadvantage compared to its economic competitors.

What is the emerging best practice internationally, in terms both of developing standards and frameworks for sharing and using transport data, and supporting specific innovations?
"Just GTFS [and similar standard] data practices require significant economic and political investment" (Moran, 2021, p. 282). This understanding has been embraced by countries that have consolidated or are developing a national platform to manage present and future data needs. The US Government, for instance, recently enacted a law allowing the US National Transit Database to require GTFS feeds from transit agencies that receive federal funding.[2] This promotes the use of a single standard format and centralises the information. To aid compliance, the Federal Transit Administration launched the 'National

RTAP's GTFS Builder' platform,[3] offering practical guides, tools, resources, and continuous live support service with training events.

European countries also provide useful examples. Entur in Norway, a state-owned company, operates the national registry for all public transport.[4] It collects, refines, and shares data at a national level, offering open-access, flexible timetable data with real-time information in both NeTex, the European standard, and GTFS. In addition to various services like journey planning apps, ticketing, and open-source software, Entur collaborates with developers, such as Google Cloud,[5] and operators to enhance their services.

France also has a national platform providing static schedules and real-time data, alongside more general mobility data.[6] This is supported by the 2019 French law 2019-1428, mandating static and dynamic data openness in standard European formats, NeTex and SIRI. In many cases, they also use GTFS. Compliance is bolstered with supporting resources, such as personalised support and online tools for operators. The Republic of Ireland recently established a new platform to collect, host, and distribute timetables. They also present an interesting case which moved from the TransXchange format to GTFS as its primary standard.

How can the UK scale up from pilots, pockets of innovation and existing single-mode data sets towards an integrated, comprehensive landscape for transport data?

Best international practices alongside our experience from academic research advocate for a UK-wide public transport open-data platform to manage this complex process effectively. This should aim at integrated information, harmonised specification formats, and stakeholder representation. Such an organisation should not only focus on data distribution but crucially, on data governance. This implies establishing a vision and strategy, dimensioning infrastructure needs, planning in the short- and mid-term, monitoring, hosting and distributing data, reporting, archiving, training, and integrating not just specialists, but also users. There are extremely valuable experiences in the UK that can help to scale up to this step, such as BODS and Traveline.

In terms of sharing standards, it is suggested to evaluate the continuity of TransXchange, transition to more widely supported formats, or adopt a mix of both. Regardless, we see merit in recognizing the benefits of integrating the GTFS standard into the future of public transport data, coexisting with the

official standard. This would allow the use of several open-access resources and tools, which would likely reduce the resources needed for 'in-house' developments. Furthermore, the upcoming system revisions should focus on making additional information available, including fares, real-time feeds, and real-time crowdedness indices.

The future management of this data should consider historical records. This can help, for instance, to characterize trajectories of the accessibility levels by public transport on different communities in the UK. Longitudinal data can bolster the robustness of research conclusions by enabling causal research designs. To date, most researchers and analyst have been missing this opportunity.

How should data availability, and sharing by transport operators, suppliers and other bodies, be encouraged, facilitated and regulated?

To encourage, facilitate, and regulate data availability and sharing by transport operators and other bodies, we envision various approaches. Primarily, identifying key gaps such as accuracy, compliance, or coverage should be systematically addressed, potentially by enhancing funding for academic research. This would identify priority areas.

Convincing stakeholders of the benefits of an integrated system is also crucial. Previous experiences have shown that demonstrating achievements from available data in a clear, concise manner maintains stakeholder interest. This strategy relies on measuring and documenting improved user experiences and the data's value in research and business.

Ensuring data is prepared for future uses involves the ongoing development, consolidation, and updating of skills and knowledge of operators, suppliers, and other bodies. Within the UK, significant expertise exists in academia, government, consultancy, and operator sectors. A leading organisation could harness this experience by identifying and connecting common interests.

Finally, prior cases addressing the goals proposed here demonstrate that compliance and enforcement should be continuously supported with resources such as skills, tools, materials, and dedicated services. In this context, it is also important that regulation prioritises the homogenisation of processes, platforms, and technical standards.

Additional Resources

In 2022 the submitting team published the dataset "Public Transport Accessibility Indicators for Great Britain" which offers open-access information for all 41,729 small areal unit geographies (LSOA/DZ) in Great Britain to key services such as employment, health, and education (Verduzco Torres & McArthur, 2022). https://www.ubdc.ac.uk/data-services/data-catalogue/transport-and-mobility-data/public-transport-accessibility-indicators-data-2022/

Blog by the submitting team using timetable data to study reductions in accessibility at night: https://www.ubdc.ac.uk/news-media/2023/august/unveiling-the-variability-in-public-transport-services-across-great-britain/

Newspaper article based on the work of the submitting team using timetable data to examine the impact of the proposed withdrawal of night bus services in Glasgow: https://www.heraldscotland.com/news/23696053.data-maps-show-impact-losing-glasgows-night-bus-services/

References

Bastiaanssen, J., Johnson, D., & Lucas, K. (2022). Does better job accessibility

help people gain employment? The role of public transport in Great

Britain. *Urban Studies*, *59*(2), 301–322.

https://doi.org/10.1177/00420980211012635

Chen, H., Cao, Y., Feng, L., Zhao, Q., & Verduzco Torres, J. R. (2023).

Understanding the spatial heterogeneity of COVID-19 vaccination uptake

in England. *BMC Public Health*, *23*(1), 895.

https://doi.org/10.1186/s12889-023-15801-w

Conway, M. W. (2021). *If You Zone It, Who Will Come, and How Will They Travel? The Effects of Relaxed Zoning Regulations on Travel Behavior* [PhD]. Arizona State University.

Conway, M. W., Byrd, A., & van der Linden, M. (2017). Evidence-Based Transit and Land Use Sketch Planning Using Interactive Accessibility Methods on Combined Schedule and Headway-Based Networks. *Transportation Research Record: Journal of the Transportation Research Board*, *2653*(1), Article 1. https://doi.org/10.3141/2653-06

Moran, D. (2021). *Extending Open Mobility to Mobility Data Justice: A Localized Case Study of GTFS-oriented (Infra)Activism in Defense of the Mobile-Digital Commons* [PhD, University of Central Florida]. https://stars.library.ucf.edu/etd2020/532

Pereira, R. H. M. (2019). Future accessibility impacts of transport policy scenarios: Equity and sensitivity to travel time thresholds for Bus Rapid Transit expansion in Rio de Janeiro. *Journal of Transport Geography*, *74*(October 2018), Article October 2018. https://doi.org/10.1016/j.jtrangeo.2018.12.005

Verduzco Torres, J. R., & McArthur, D. (2022). *Accessibility Indicators for Great Britain*. https://doi.org/10.5281/zenodo.6759240

August 2023

Endnotes

[1] This refers to access to data without registration or, if necessary, from a reliable and regulated server operating under transparent rules.

[2] H.R.3684 - Infrastructure Investment and Jobs Act.

[3] Available at: https://www.nationalrtap.org/Technology-Tools/GTFS-Builder#Introduction.

[4] Available at: https://developer.entur.org/.

[5] See the case study at: https://cloud.google.com/customers/entur.

[6] Available at: https://transport.data.gouv.fr/.