MOTT MACDONALD

Transport data for artificial intelligence innovation - Appendix

Discovering, prioritising and analysing high-potential data sets

March 2026

Contents

1 Appendix

This is the appendix to the 'Transport Data for Artificial Intelligence (AI) Innovation: Discovering, Prioritising and Analysing High-Potential Data Sets' report, published December 2025. The report presents a structured assessment of existing and emerging data sets in the transport sector. Through a three-stage methodology, it identifies where improved access, quality and coordination of data could unlock high-value AI use cases and drive innovation aligned with strategic policy goals. This appendix contains the detailed outputs from each of the three stages of the report: (i) Discovering and compiling; (ii) Filtering and prioritising; and (iii) Analysis and insight.

A.1 Discovering and compiling

A.1.1 Longlist

The original longlist of 156 data sets was condensed to 123 data sets by consolidating similar entries. The condensed list is displayed below. It is noted that these are the outputs of research and do not represent DfT's views.

Data set/data service name	Theme	Description	Potential use case(s)
Mobile network data - location data analytics	Demographics, movement and behaviours	Location data analytics from mobile devices Specific examples of this data are listed elsewhere in this table	Provides insights into customer behaviour across transport networks allowing a better understanding of how to improve the overall efficiency of the transport system and enhance the passenger experience. Potential for responsive, data driven transport planning and modelling Potential to fuse with demographic and transaction data to understand trip segmentation, for example, by purpose or persona. See also 'Transaction data – banking/transaction data', 'Experian Mosaic 7'
BT active intelligence - mobility data	Demographics, movement and behaviours	Location data analytics from mobile devices	See 'Mobile network data – location data analytics' data set for potential use cases

Data set/data service name	Theme	Description	Potential use case(s)
Apple detailed mobility data (speculative) - mobility data	Demographics, movement and behaviours	Location data analytics from mobile devices	See 'Mobile network data – location data analytics' data set for potential use cases
Google detailed mobility data (speculative) - mobility data	Demographics, movement and behaviours	Location data analytics	See 'Mobile network data – location data analytics' data set for potential use cases
O2 mobile data	Demographics, movement and behaviours	Location data analytics, in at least two data sets: (i) Spatial Insights – mobility data (ii) O2 Motion is mobile-powered data which offers a series of insights (Application Programming Interfaces (APIs)).	See 'Mobile network data – location data analytics' data set for potential use cases
Vodafone analytics – mobility data	Demographics, movement and behaviours	Location data analytics	See 'Mobile network data – location data analytics' data set for potential use cases
BT mobile network data	Demographics, movement and behaviours Transport operations Other	BT's geo mobile network data is anonymised and aggregated using tight GDPR controls, enabling these patterns of movement to be surfaced and analysed. It is composed of various data types including digital identity, location and active intelligence. Includes origin destination (OD), flows, rail first/last-mile insights	Can support understanding of user travel behaviour, loading of public transport, trip purposes and other travel insights
CCTV – in vehicle/at stop	Connected vehicles and sensors	Continuous feeds (live or historic) of video from CCTV cameras in public transport vehicles at stations, stops and platforms. These are traditionally used for live observation for safety and security purposes	This data could support more reliable crowding information at various nodes and on-route in the public transport network, as well as improving bus boarding and alighting counts. This data could also support transport planning and modelling See also 'Bus crowding data'

Data set/data service name	Theme	Description	Potential use case(s)
		but can have tangential benefits. Owned by various operators	
Connected vehicle data including CANBus data	Connected vehicles and sensors	This covers data recorded via CANBus units within connected vehicles across all manufacturers. This could include movement, acceleration, braking, object sensing and location data, alongside vehicle health information and other properties. Such data may also include vehicle occupancy information	Could provide network-wide data to enhance safety, including near miss data, and could provide granular information on the autonomous vehicle fleet, including usage of typical vehicles, which could inform policy Vehicle occupancy information could support dynamic transport (for example, highway network) management and optimisation (for example, through Urban Traffic Control (UTC), or digital twins)
Mastercard tourism insights platform – banking/transaction data	Demographics, movement and behaviours	The tourism insights web platform combines and displays multiple sources of data (for example, behaviour, airline booking and spend) for insights into the travel lifecycle. This allows optimisation of tourism management and marketing for a destination marketing organisation or any tourism related organisation or administration	See ‘Transaction data – banking/transaction data’ use cases
Rail passenger and loading data	Transport operations	The range of data sets on passenger levels are not readily available, as reported during interview, and are only known by train operating companies (TOCs). Passenger counts taken using infrared sensors on train doors, load weight, onboard CCTV and stereoscopic sensors	As TOCs are brought into public ownership, these data sets can be incorporated into data platforms

Data set/data service name	Theme	Description	Potential use case(s)
Ride-sharing platforms	Transport operations	This covers details of trips made: OD, routing information; details of user as per registration with service; payment details (registered payment card, for example); and vehicle type (engine/motor details)	Can support dynamic transport (for example, highway network) management and optimisation (for example, through UTMC or digital twins), more robust transport modelling (consistency of time and resolution of capture of counts across a fully modelled area), monitoring and evaluation, and analytics to support A/B testing
Smart ticketing data	Demographics, movement and behaviours	Statistics on transport usage from smart tickets	Can be used to understand travel patterns by users of smart tickets. A country-wide collection of all smart ticketing of all operators would be considered big data and suitable for longitudinal analysis
UTC and related traffic monitoring highway infrastructure instrumentation	Connected vehicles and sensors	This includes UTC and control data, for example, real-time traffic signal phasing timings and vehicle actuated/active junction control trigger data for bus priority at junctions	Can be used to support transport planning and modelling, as well as highways digital twins, including AI/machine learning (ML) forecasting to support real-time modelling and scenario testing. Fuse data with other sensors, including CCTV, Bring Your Own Device user data (for example, mobile phones) to risk map and potentially dynamically alter the environment to protect more vulnerable users
Transaction data – banking/transaction data	Demographics, movement and behaviours	Debit, credit and prepaid data	Could be used as an indicator of footfall and economic activity, which could help inform transport services. Assuming that temporal attribution is available, it is also an indicator of footfall at different times of day, either in one place or across a larger area. Could be used to examine footfall and spending in stations and, therefore, help inform station design See also 'Experian Mosaic 7', 'BT Active Intelligence'

Data set/data service name	Theme	Description	Potential use case(s)
Wi-Fi connection data – for example TfL data collection	Demographics, movement and behaviours	Connection data that relates IDs (Media Access Control (MAC) MAC address) of mobile devices communicating with wi-fi hotspots, along with timestamp of the communication details to the hotspots themselves. The communication happens before any formal connection between device and hotspot, which means that, theoretically, any connected device can be involved without the need for user interaction	Wi-fi is located extensively in the public and private realms, giving the technical potential to scale this approach. TfL has demonstrated the use of this to understand which routes passengers choose when a change in underground line is required; this can aid with crowd control and planning
CCTV – derived highway information such as traffic flows – TfWM ADEPT Live Labs	Transport operations		See 'CCTV – Highways-based video feeds'
CCTV – highways-based video feeds	Connected vehicles and sensors	Continuous feeds of video from CCTV cameras. CCTV cameras are installed extensively, so these offer potential for wide spatial coverage and source of data. Many different CCTV systems are installed and owned by diverse organisations, with different ownership and access possibilities. Additionally, CCTV installations will have diverse hardware and quality of capture	CCTV is a rich source of data to which Computer Vision (CV) techniques can be applied to extract valuable information. This includes vehicle registration marks (VRM), vehicle classifications, flows and speeds, and potentially 'trajectories'. Such sources can support dynamic transport management and optimisation, more robust transport modelling, monitoring and evaluation, and analytics to support A/B testing CCTV covers pedestrianised areas, as well as areas outside of the highway boundary. This offers considerable potential for addressing a major gap in the walking and cycling modes that are largely invisible in count and flow data

Data set/data service name	Theme	Description	Potential use case(s)
Courier/haulier – last-mile and wider data – telematics providers	Freight	This includes details of vehicle movements covering location (and, therefore, routeing), speed, dwell time, vehicle monitoring information, including idling time, fuel use, tyre pressures and braking information	Supports understanding of risks not realised as collisions (for example, near misses). Potential for fusion with other data (STATS19, OS or Digital Traffic Regulation Orders (DTRO) kerb line mapping, insurance data) for pattern identification around highway design, flows, etc relating to near misses and collisions leading to KSIs and routing of courier and larger freight vehicles (for example, risky interactions with cyclists) See also 'Courier/haulier – last-mile data – Amazon', 'TomTom', 'AppyWay'
Courier/haulier – last-mile data – couriers – for example, Amazon	Freight	Data includes Amazon last-mile routing information, covering route-, stop-, and package-level features for 9184 historical routes performed by Amazon drivers in 2018 in five metropolitan areas in the United States. This data set excludes any personally identifiable information	Could be used to develop ML-based methods to enhance or replace traditional optimisation for routing, using data from Amazon delivery drivers. Live fleet or haulier data could feed into UTMC and/or a digital twin to enable real-time network management that prioritises the movement of goods and people. Additional uses include monitoring courier compliance, identifying opportunities to consolidate last-mile deliveries and supporting enhanced freight modelling See also 'Bus location data (AVL/SIRI)', 'Bus crowding data'
Micromobility (e-scooter, e-bike) – private entrants	Transport operations	Data includes details of trips made by scooters, bikes, e-bikes: distance, routes and speeds; details of user as per registration with service; payment details (registered payment card, for example); pick up and parking locations (docked and undocked); and energy consumption (battery levels)	Detailed movement data to support transport planning and modelling. Particularly rich data set for cycling/micromobility For other modes similarly well covered by real-time data, there is potential to use AI to identify patterns in modal shift preferences Potential to fuse with payment services or other data to infer trip purposes See ‘Sense - smart cycle sensors', 'Transaction data – banking/transaction data', 'Experian Mosaic 7'
National Highways traffic cameras	Connected vehicles and sensors	See: 'CCTV – Highways-based video feeds'	See 'CCTV – Highways-based video feeds'

Data set/data service name	Theme	Description	Potential use case(s)
Rail cab – front-facing CCTV	Transport operations	Video feeds from front of trains	Already exploited for ML-based vegetation detection Additional use cases relate to extending application of ML/CV techniques to detection of issues with ballast condition, items dumped on/near track, trespass and animal incursion onto railway Note: prior art on the railway and National Highways pavement health detection from CCTV surveys
AppyWay	Transport network	Survey to produce high-definition mapping of the kerb line and to support digital Traffic Regulation Orders (TROs)	Support highway digital twin development
Google Earth Engine	Earth observation and environmental	Google Earth Engine combines a multi-petabyte catalogue of satellite imagery and geospatial data sets with planetary-scale analysis capabilities	See 'Meta Data for Good Portal'
Google Maps APIs	Demographics, movement and behaviours Transport operations	A set of tools and functionalities offered by Google that allow developers to integrate mapping features and functionalities into their applications and websites	Mapping of transport of all kinds, particularly vehicle traffic
HERE	Transport network Transport operations	See: 'TomTom' Additionally, LiDAR (Light Detection and Ranging) data captured from the highway network	LiDAR data could support digital twin development and analysis of, for example, urban realm from a usability/walkability/cyclability perspective See also 'TomTom'
INRIX	Transport operations	Includes INRIX IQ and traffic data products, covering location, signal, traffic and trip 'analytics'. Also includes traffic volumes, historic speed and travel time data.	See 'TomTom', 'HERE'

Data set/data service name	Theme	Description	Potential use case(s)
LENNON – rail ticketing	Demographics, movement and behaviours	The LENNON system includes rail ticket sales data for mainline rail operators, pre- and post-allocation data sets and core data sets for official rail passenger usage statistics	See 'Rail Data Marketplace - RDM' use cases
Meta Data for Good Portal	Earth observation and environmental	Source of data from big data/AI fusion covering global hi-res population maps, population during crisis data sets, electrical distribution grid maps, inclusive internet index, commuting zones, places, network coverage maps, Facebook movement during crisis, travel patterns, movement distribution maps, social connections data and canopy height maps	Multiple use cases including forecasting and responding to disasters, as well as automated land use, land cover and land change detection
Microsoft Planetary Computer	Earth observation and environmental	The Planetary Computer combines a multi-petabyte catalogue of global environmental data	See 'Meta Data for Good Portal'
Rail Analytics – mobility data	Demographics, movement and behaviours	Rail Analytics is a suite of services that combines Citi Logik's expertise with mobile network data and Internet of Things data. Products include RailMonitor, StationWatch, StationAnalytics, and DelayAnalytics	Real-time and historical rail data allows forecasting of train occupancy, optimisation of station operations and analysis of delay impacts and passenger responses. Data on station catchments and first-/last-mile travel supports better planning. These high-volume data sets are suitable for ML and AI applications, offering insights to inform and influence rail-related policy
Rail Data Marketplace (RDM)	Transport network Transport operations Demographics, movement and behaviours	RDM brings together rail data sources and related services. Includes rail performance data, fares data, real-time train information, accessibility data, asset data (including bridge strikes, track gradients, OD matrices, safety data) and Network Rail (NWR) wheel impact load detection	Potential use cases include predictive analytics for disruptions, live train time communications, operations support, planning and other passenger experience improvements. Possibility to combine data sets to add value. Attribute-rich and longitudinal data sets are well suited for ML/AI applications. A novel data set example is track gradient data which has been generated

Data set/data service name	Theme	Description	Potential use case(s)
			by LiDAR. This could be used for land modelling to help prevent landslides
TomTom	Transport network Transport operations	Detailed map data covering turn-by-turn navigation, High Definition (HD) and advanced driver-assistance systems (ADAS) mapping to support autonomous vehicles navigating the highway. Data covers gradients, lanes, curvature and speed limits, traffic signs and traffic light (positions) Live and historic traffic data	HD/ADAS mapping could support digital twin of highway HD mapping could support pattern analysis around modal choice and risk mapping (for example, pedestrian and cyclist safety) in fusion with other data sources Traffic data to support modelling, transport planning, digital twin development, risk mapping, active modes modelling and planning See also 'See.Sense – smart cycle sensors'
Transport API	Transport network	Includes routing, stops and timetable data, as well as performance data, fares, etc. Can be seen as a superset of 'Bus Open Data (BODS)', due to multi-modal nature	Supports transport planning, transport modelling and passenger information systems (for example, it can support accessibility/connectivity analyses). Potentially assists with data analysis and insights around wider mobility, in fusion with Mobile Network Data (MND), mobile phone sensor data, etc (for example, it could help disaggregate public transport (PT) from non-PT road-based vehicle movements) See also 'General Transit Feed Specification (GTFS) – public transport timetables and routes', 'Bus Open Data (BODS)'
Data for Road Safety	Transport operations	A scalable solution which any industry partner in the transportation, mobility and traffic data domain and public authorities can join and use to exchange safety related traffic data and information	See 'STATS19 – collisions data'

Data set/data service name	Theme	Description	Potential use case(s)
DfT Street Manager	Transport operations	Service for local highway authorities, utility companies or contractors in England to plan, manage and record street and road works	See use cases for 'National Highways Digital Lab'
DVLA/vehicle registration data	Demographics, movement and behaviours	The DVLA Vehicle Enquiry Service API is a RESTful service that provides vehicle details of a specified vehicle. It uses the vehicle registration number as input to search and provide details of the vehicle. The response data is provided in JSON format	API access available on request. Primary route network (PRN) is accessible from National Highways, police and other Automatic Number Plate Recognition (ANPR) cameras (for example, Transport for West Midlands (TfWM) ADEPT Live Labs), congestion charge zone, clean air zone, (ultra) low emission zone (ULEZ) and car parking (private). There is a wealth of vehicle-specific information which can be gained from PRN (for example, vehicle class, engine type, age and where registered). Could see potential data fusion with mobile network data and other CCTV/CV applications
NIRA Dynamics – vehicle sensors and road surface conditions	Transport operations	Using sensor-fusion vehicle software, fleet data and information from static and dynamic road conditions, NIRA provides services that can optimise electric vehicle (EV) performance and reliability, as well as measure road health. The latter is an automated and objective system for road friction monitoring. Road Health by NIRA utilises road roughness, road friction and road irregularities (such as potholes)	Asset condition mapping and predictions to support maintenance and resilience plans. Potential for a large-scale ML model of impact of vehicle and road conditions on EV efficiency, with a view to optimising energy usage and reducing emissions
OneNetwork planned road disruptions feed	Transport operations	The OneNetwork road disruptions feed provides a view of current and planned disruptions to the road network. These include road works, incidents, public events, road and lane closures, diversion routes and more. Data is captured from over 600 public authorities, utility companies and event organisers	See use cases for 'National Highways Digital Lab'

Data set/data service name	Theme	Description	Potential use case(s)
See.Sense – smart cycle sensors	Connected vehicles and sensors	Data from smart cycle lights. See.Sense lights transform bikes, e-bikes and e-scooters into sensors, generating data insights into the rider's experience. Anonymised data includes information on swerving, braking and speed, dwell times, road roughness, collisions and movement patterns	Relatively unique source of data into general cycling population. This can provide particularly useful insights into risk/danger and, particularly, risks not realised as collisions (for example, near misses). Can be used in fusion with other data sets to support understanding of the causes of risks to cyclists and other road users, target interventions and provide monitoring and evaluation (M&E) as well as analytics for A/B testing.
TfL CCTV	Connected vehicles and sensors	See 'CCTV – Highways-based video feeds', 'CCTV – In vehicle/at stop'	See 'CCTV – Highways-based video feeds', 'CCTV – in vehicle/at stop'
Vivacity Labs traffic monitoring API	Transport operations	Continuous feeds of video from CCTV cameras	See 'CCTV – Highways-based video feeds'
EV charging data	Connected vehicles and sensors Energy generation, Transmission and emissions	Data relevant to EV charging can be sourced from several key stakeholders, including EV charger operators, vehicle manufacturers (original equipment manufacturers), and energy companies offering EV-specific tariffs. This data might include date, time and duration of charging sessions, connection and disconnection times, charger type (domestic, private or public) and charger location. Additional variables might include rate of charge, vehicle model information, start and end battery levels, voltage and current measurements, as well as climate conditions and battery temperature	Understanding charging efficiency across multiple variables to provide system-wide analytics and forecasting of demand to provide enhanced customer experience and predictions of dwell time to reach desired charge level
Bus Open Data (BODS)	Transport network Transport operations	Includes service routing, stop locations and timetabling data for buses, as well as real-time bus locations and crowdedness feeds	Supports transport modelling, transport planning activities and passenger information systems. Potential to support development of a digital twin of highway and wider transport

Data set/data service name	Theme	Description	Potential use case(s)
		See also 'Bus location data (AVL/SIRI)', 'Bus crowding data'	systems, enabling network optimisation to minimise delays and support the movement of people and goods See also use cases in 'Bus location data (AVL/SIRI)', 'Transport API'
Bus location data (AVL/SIRI)	Transport operations	Bus location data (automated vehicle location, or AVL) provides temporal-spatial information about bus locations. SIRI-VM is a European standard for interchange of real-time information, making it appealing to larger players (same tools work in many places)	Supports analysis of bus performance. In fusion with other data sources, for example, traffic/congestion data, mapping data (locations of bus priority measures), intelligent transport system (ITS)/UTC sensors and planned road disruption feeds, this could support a deep understanding of reliability/congestion and safety issues, and real-time operations of the highway network twin See also 'Bus crowding data', 'Courier/haulier – last-mile data – couriers – for example, Amazon'
Copernicus	Earth observation and environmental	The Copernicus Climate Change Service (C3S) provides information about the past, present and future climate in Europe and the rest of the world	See 'Met Office Climate Data Portal' use cases
Rail Delivery Group – Darwin train positioning data	Transport operations	Darwin provides predicted arrival times, platform numbers and shows reasons for cancelled and delayed trains, in real-time	Supports analysis of rail performance. In fusion with other data sources, ML pattern identification and movement prediction could support a deep understanding of rail reliability/congestion and safety issues, and real-time operations of the railway network
Defra environmental data	Earth observation and environmental	A data portal containing all Defra environment-related data sets, such as flood risk maps	Supports improved decision-making strategies that help safeguard natural resources, support food and farming industries, and sustain a thriving rural economy. Can also be combined with transport infrastructure data to understand future risk. See 'Met Office Climate Data Portal' use cases

Data set/data service name	Theme	Description	Potential use case(s)
General Transit Feed Specification (GTFS) – public transport timetables and routes	Transport network Transport operations	Service routing, stop locations and timetabling data for all public transport modes. GTFS-Real-Time (RT) supports real-time positioning data, service and trip updates	Supports transport modelling, transport planning activities and passenger information systems. Potential to support development of a digital twin of highway and wider transport systems, network optimisation to minimise delays and the movement of people and goods. See also use cases in 'Bus location data (AVL/SIRI)', 'Transport API'
OpenStreetMap (OSM)	Other	A detailed open-source map of the UK with features of transport networks and facilities. Known to be good for detailed rail mapping, although OSM generally should be used with caution	Any application requiring a transport network, including transport modelling, journey planning, spatial and network analysis, master planning and novel applications See also 'Ordnance Survey (OS) Data Hub – mapping sources various' use cases
TfL open data (including TfL Unified API)	Transport network Transport operations Earth observation and environmental	Includes a broad set of APIs covering: air quality; tube; bus, coach and river; roads; cycling; walking; Oyster and smartcard; accessibility; network statistics; planning data; and occupancies	Multiple use cases. See related 'General Transit Feed Specification (GTFS) – public transport timetables and routes', 'Transport API', 'National Highways Digital Hub'
Traveline open data	Transport network Transport operations	Traveline offers several open data options including Traveline National Data Set, NextBuses API and National Operator Code database (NOC)	Applications related to public transport operations, including timetabling, journey planning and understanding service patterns
Overture maps	Other	Global, interoperable map data. Seeking to establish a common reference system to allow data from multiple sources to be easily joined with map data	Any application requiring a transport network, including transport modelling, journey planning, spatial and network analysis, masterplanning and novel applications. However, note the *potential* for easier integration of external data source See also 'Ordnance Survey (OS) Data Hub – mapping sources various', 'OpenStreetMap (OSM)' use cases

Data set/data service name	Theme	Description	Potential use case(s)
Datalastic Tracker API and Ship AIS database	Transport operations Freight	Historical and real-time automatic identification system (AIS) data about ships, cargo vessels, fishing boats, cruise liners and more. Simple maritime intelligence for complex data-driven decisions	See 'Vessel Finder'
Elexon Insights Solution	Energy generation Transmission and emissions	Insights real-time information service about wholesale energy data and other data sets also available through National Grid	See ‘National Grid Connected Data Portal’ use cases
Flightradar24 data service	Transport operations	Provides a flight tracking service based on an independent terrestrial Automatic Dependent Surveillance-Broadcast (ADS-B) receiver network. Both live data feeds and historical data files can be configured based on the parameters required	Potential to develop AI/ML tools that predict impacts of delays on airport operations, traffic and public transport systems Potential input into a digital twin
Kpler – marine traffic data services	Transport operations Freight Earth observation and environmental	Live vessel and fleet positions, port calls and expected port arrivals, vessel details based on NEMA data	Transport planning and modelling. Could support ML forecasts of freight movements, modelling of asset deterioration or resilience planning (for example, in the context of shipping delays)
OAG flight data sets	Transport operations	Aviation data sets including global airlines schedules, historical data, connections and passenger bookings, including real-time flight status	Predictive analytics tools for airport operations and multimodal transport planning. Combining OAG’s global flight schedules and real-time status data with ground transport data sets (for example, rail timetables, road traffic feeds) allows forecasting of passenger flows and identification of potential congestion points
Spire – maritime	Transport operations Freight	Spire collects, analyses and delivers detailed vessel, weather and global shipping information in near real-time	See 'Kpler – marine traffic data services' use case

Data set/data service name	Theme	Description	Potential use case(s)
Spire – weather and climate	Transport operations Freight Earth observation and environmental	Global maritime weather, current, forecast and historical weather data, advanced weather forecasts and alerts	Assessment/forecasting of the impact of weather on maritime transport and land-based transport that interrelates with it. Real- or near-time data can be used for preparation ahead of severe weather. Historic data could be used to feed into predictive analyses and to examine the impact of historic severe weather events on maritime-dependent freight (and passenger) transport systems
Strava Metro – active mobility data	Demographics, movement and behaviours	Strava collects large volumes of activity data from users who record rides, runs and walks via phone or GPS devices. The Strava Metro platform processes this data set by removing personal identifiers and aggregating it	Requires application for a partnership. Can be used in infrastructure projects to evaluate, plan, engage and prioritise See 'Strava Labs' use cases
Vessel Finder	Transport operations Freight	Historical and real-time updates of vessel positions data, ship details, port calls (and expected port arrivals), voyage information, arrival times and container tracking	Source of data for digital twin of highway and rail networks, with a particular focus on freight movements. AI/ML use cases could include forecasting impacts of disruption to seaborne freight on road/rail freight to support network resilience
ZapMap EV charging points – EV charging	Energy generation Transmission and emissions| Transport network	Main product is an open map of public charge points in the UK and parts of Europe. Also provides more data and insights into the shape and usage of EV charging infrastructure in addition to the attitudes and behaviours of EV drivers	Historical usage of charge points could be used to inform policy and development by allowing an understanding of in-demand locations, therefore helping to prioritise the installation of future charge points. This could be combined with data from a relevant energy portal to cross reference the capacity of the electricity grid against need for charge points Attitudinal and behavioural data could be used in social research to encourage uptake of EVs
Bus crowding data	Transport operations	Part of SIRI-VM feed. This allows for occupancy data and suitability of a bus for boarding at a departure point	As per 'Bus location data (AVL/SIRI)' but with added benefit of considering the number of people impacted when analysing reliability and who stands to benefit when applying dynamic highway network management techniques through UTMC or digital twin

Data set/data service name	Theme	Description	Potential use case(s)
			See also 'Bus location data (AVL/SIRI)', 'Courier/haulier – last-mile data – couriers – for example, Amazon', 'CCTV – in vehicle/at stop'
Defra air pollution data archives	Earth observation and environmental	Stores monitoring data, descriptive statistics and exceedance statistics	Analysis and mapping of air pollution and particulates can help inform policy and decision-making. Combining additional data sets, such as demographics and deprivation, could help to prioritise mitigation efforts and understand relationships between air pollution and other variables
Met Office Weather DataHub – atmospheric model data	Earth observation and environmental	Weather models	Assessment/forecasting of the impact of weather on transport infrastructure. Real- or near-time data can be used for preparation ahead of severe weather. Historic data could be used to feed into predictive analyses and to examine the impact of historic severe weather events on transport infrastructure
National Highways Digital Lab	Transport operations	Currently includes road and lane closures data service to describe all future and current road and lane closures resulting from incidents, road works or any other reason.	Supports services notifying end users of disruptions. Supports journey planning. Potential data source for a digital twin of the highway network. AI use cases could include ML near-term forecasts of impacts on "local area" traffic, AI/ML enhanced journey planners and optimisers (for example, travelling salesperson problem) See also 'Courier/haulier – last-mile data – couriers – for example, Amazon'

Data set/data service name	Theme	Description	Potential use case(s)
National Highways WebTRIS	Connected vehicles and sensors	WebTRIS provides journey times, speeds and traffic flows on all major roads managed by National Highways in England. The data is collected through a combination of ANPR cameras, in vehicle GPS devices and inductive loops in the road surface. WebTRIS data is backward looking, with data bulk uploaded a month in arrears. However, the underlying sensor data (for example, from Motorway Incident Detection and Automatic Signalling (MIDAS) is captured in real-time and this is considered part of this entry	Can support the development of ML models to forecast the data captured and published through WebTRIS (for example, flows and speeds) for inclusion in real-time and operational models of the highway network. Precedent: proof of concept real-time model of highway network around M42 J6, as part of Innovate UK/Geospatial Commission competition Used for transport planning analysis of changes in work from home patterns as a result of COVID-19
Network Rail – train positioning data	Transport operations	'Movement', ‘Train Running System Total (TRUST) operations processing system’ and ‘Train Describer’ (TD) data feeds. TRUST and TD provide train positioning at the signalling berth level, giving real-time location and speed information	Potential for AI/ML signalling decision support tool to help prioritise minimisation of passenger disruption Input into digital twin of railway
Network Rail – train scheduling data	Transport operations	'Schedule' data feed. Includes daily extracts and updates of train schedules from the Integrated Train Planning System	Would be used in conjunction with real-time location data See ‘Network Rail – train positioning data’ ,‘Rail Delivery Group – Darwin train positioning data’
Waze – mobility data	Demographics, movement and behaviours	Similarly to Google Maps, Waze provides real-time driving directions. It also incorporates user-submitted route details, such as road works and traffic collisions	See ‘Transport data – mobility data’ use cases
Census	Demographics, movement and behaviours	The census is undertaken by the Office for National Statistics every 10 years and provides a snapshot of all the people and households in England and Wales	Census data is often used in projects which require the understanding of population and demographics; transport projects are no exception, as it is important to understand users of services and the demographic make-up of a study

Data set/data service name	Theme	Description	Potential use case(s)
			area. It is attribute-rich, and often high resolution at output area level. It can be combined with other data sets, such as transport services, to understand access to services by demographic
Connected vehicle behaviour	Connected vehicles and sensors	The data set is composed of vehicle journeys undertaken in several UK cities during October 2023. It provides a large sampling of vehicle journeys across a variety of different highly populated areas of the UK. Each record, corresponding to one vehicle journey, includes vectors corresponding to time, location (latitude/longitude) and recorded speed	Use cases might include trying to improve speeding behaviours and promote traffic safety, reduce acceleration and deceleration driving to reduce carbon emission and energy use or optimise road and urban environment to promote transportation efficiency and reduce traffic congestion
Data City business data	Other	Includes a suite of company and sector information, real-time industrial classifications (RTICs), detailed financials, growth metrics, jobs and skills data and more	Similar use cases to ‘Transaction data – banking/transaction data’. This can inform transport decision-making by linking the network to business demand
Experian Mosaic 7	Demographics, movement and behaviours	Mosaic provides rich demographic information across the UK at household level	This, or similar data, could be used to match households to their most likely DfT transport user personas, which would allow an understanding of the spatial distribution of personas and could feed into further analysis and research. Other example applications include where to encourage uptake of active travel based on profiles of interest and site selection for new services, such as air taxis See also 'Transaction data – banking/transaction data’ 'BT Active Intelligence'
Freight Data Hub – MDS transmodal	Freight	Includes freight traffic count data, road freight origin destination (OD), rail freight OD, ports data (those used by traders), port shipping connectivity data, freight vehicle highway	Transport planning and modelling. Depending on the level of disaggregation and temporal resolution, could support ML forecasts of freight movements, modelling of asset

Data set/data service name	Theme	Description	Potential use case(s)
		assignments, freight railway assignments, warehouse space and locations	deterioration or resilience planning (for example, in the context of shipping delays)
High resolution nightlight remotely sensed imagery	Earth observation and environmental	High spatial resolution multispectral night-time light remote sensing imagery is collected by the Jilin 1-03 satellites (JL1-3). JL1-3 enables capture of night-time light images in spatial resolution less than 1m with three spectral bands including 430-512nm (blue), 489-585nm (green) and 580-720nm (red) in 8-bit depth	Facilitates research into night-time activities, urbanisation, light pollution, deprivation, conflict and more. Generally, can be used to monitor change through time and can be combined with additional data sets to generate further insight into safety and risk. It could be analysed alongside road collision data or crime data to determine if there is a relationship between lighting and collisions/crime. It could also be used for asset management of lighting systems
Huq mobile phone data – mobility data	Demographics, movement and behaviours	Huq is a mobile phone app data set. The app collects real-time, anonymised location data from users' smartphones, based on the use of a range of smart phone applications	Real-time insight into human mobility patterns and behaviour at granular spatial resolution and over time. Potential for responsive, data driven transport planning and modelling
National Travel Survey	Demographics, movement and behaviours	Statistics and data about the National Travel Survey, based on a household survey to monitor trends in personal travel	An attribute-rich data set which spans multiple years, enabling longitudinal analysis of trip purpose and mode. Disaggregated data available on special licence which is larger and will yield greater insights. Can be combined with additional data sets for insights into the influence of, for example, deprivation on choice of transport mode and transport purpose
Ordnance Survey (OS) Data Hub – mapping sources various	Other	Open mapping across Great Britain, including base maps and transport infrastructure. OS Features API contains information on buildings, roads, green spaces and more. OS National Geographic Database (NGD) is a single store of all the authoritative data of OS for Great Britain	High quality, high detail base mapping (vector, raster and imagery) and network data sets. Geospatial data is key to understanding the dynamics of transport networks, infrastructure and usage. Provides opportunity for innovation when combined with other data sets. Data is available via download or through API. Note that OS runs the Geovation Hub which supports innovative use of geospatial data

Data set/data service name	Theme	Description	Potential use case(s)
People movement – mobility data	Demographics, movement and behaviours	People Movement Portal provides insights based on mobile network data	Suggested use cases include transport/urban planning, train and station capacity and flow, footfall analysis, engineering projects, park and ride, EV charge point planning, congestion and pollution monitoring, event planning and catchment areas, tourism and spending, out of home advertising, commuter patterns and home working, HGV and trade
TRICS – trip generation system	Demographics, movement and behaviours	TRICS provides trip generation analysis for the UK and Ireland, offering a comprehensive database of traffic and multimodal transport surveys, covering a wide range of development types	See 'Transport Focus Data Hub' use cases
Amazon Web Services (AWS) data marketplace	Transport operations Earth observation and environmental Other	Catalogue of third-party data sets. Categories include retail, location, marketing, financial, resources, healthcare and life sciences, automotive and environmental	Very US-focused but demonstrates the art of the possible. Examples include CargoMetrics Maritime Emissions which includes CargoMetrics' historical and ongoing carbon (CO2) emissions and fuel consumption metrics for dry bulk vessels in aggregate and by vessel type, by month. Maritime is one of the more challenging modes to understand in terms of emissions
GDELT – global monitor of broadcast, print and web news	Other	The GDELT project monitors international broadcast, print and web news in more than 100 languages and identifies key people, locations, organisations, themes, sources, emotions, counts, quotes, images and events	A big data set with many potential applications (for example, an alternative source of information on transport user attitudes, disruptions). Volume of data makes it suitable for data science, AI and ML applications
National charge point registry (NCR) UK – EV charging	Energy generation Transmission and emissions Transport network	The NCR is a database of publicly available charge points for EVs in the UK established in 2011	Primarily utilised by business data users for their products (such as smartphone apps, satellite navigation and route planning) See also use cases in 'ZapMap EV charging points – EV charging'

Data set/data service name	Theme	Description	Potential use case(s)
National Grid Connected Data Portal – energy	Energy generation, Transmission and emissions	A comprehensive source for real-time and historical data on the energy systems provided by National Grid	Real-time information on electricity demand and generation. Extensive archives of historical data to analyse trends and patterns, making it well suited for ML/AI applications. Can be used to inform the development of transport infrastructure, particularly for net zero transition
National Grid ESO Portal – energy	Energy generation, Transmission and emissions	Open data from Great Britain's electricity system operator. Data groups include ancillary services, balancing costs, carbon intensity, connection registers, constraint management, demand, future energy scenarios, generation and more	See ‘National Grid Connected Data Portal– energy’ use cases
Open Charge Map – EV charging	Energy generation Transmission and emissions Transport network	Open Charge Map is an open data registry of EV charging locations. It is developed and operated by volunteers. Points contain an address, equipment details, usage restrictions, network/operator and additional information	See use cases in ‘ZapMap EV charging points – EV charging’
Micromobility (e-scooter, e-bike) – DfT trials	Transport operations	Includes details of trips made by scooters: distance, routes and speeds; characteristics of users (though routing details and user characteristics are not joined); purpose, perceptions of safety and modal shift from post-ride surveys. Not all aspects would be considered “big data”	See ‘Micromobility (e-scooter, e-bike) – private entrants’
Daily domestic transport use by mode	Demographics, movement and behaviours	Daily usage of selected domestic transport by mode for Great Britain as a percentage of pre-Covid baseline. Produced on the second Wednesday of each month	Data is presented by day of year and by mode of transport. Data goes back to 2020 and, therefore, could be interesting for longitudinal analysis and monitoring of change through time. Can be combined with additional GB-wide data sets, such as weather, to understand influences on mode of transport

Data set/data service name	Theme	Description	Potential use case(s)
Google Covid-19 community mobility reports – mobility data	Demographics, movement and behaviours	Google collects location data shared by users of Android smartphones and compares the time and duration of visits to locations to the median values on the same day of the week in the five weeks from 3 Jan 2020	Permits analysis of the impact of COVID-19 on people movement, during and after safety measures were implemented. Also, an example of the type of location data that is held by Google and how it can be used to monitor change through time
Met Office Climate Data Portal	Earth observation and environmental	A portal containing climate data sets for the UK, including precipitation, temperature, sea level, past observations, future projections and socioeconomic data	These data sets could facilitate assessment of the impact of climate change on transport infrastructure against multiple metrics, ultimately contributing to an understanding of the resilience of the transport system. These can be used to prioritise mitigation efforts, particularly using future projection data, by cross-referencing it against the location of sensitive or vulnerable infrastructure
ORR Origin and Destination Matrix	Demographics, movement and behaviours	Office of Road and Rail (ORR) Origin and Destination Matrix (ODM) includes journeys between all pairs of mainline stations in Great Britain in financial year 2022-23	See ‘Rail Data Marketplace – RDM’ use cases
Public transport journeys by type of transport in London	Demographics, movement and behaviours	Number of journeys on the public transport network by TFL reporting period, by type of transport. Data is broken down by bus, underground, DLR, tram, overground and cable car	Allows analysis of public transport journeys by mode. Data has been captured since 2010 meaning it is good for longitudinal analysis but lacks spatial granularity and is not really “big data”. Can be combined with additional London-wide data sets, such as weather, to understand influences on mode of transport
TfWM open data	Transport network Transport operations	Broad set of transport-related data categories including active travel, bus, metro and tram and rail	Multiple: see related ‘General Transit Feed Specification (GTFS) – public transport timetables and routes’, 'Transport API', 'National Highways Digital Hub'
Train timetable data	Transport operations	All national rail passenger train service timetables, along with associated shipping and bus links	Would be used in conjunction with real-time location data See ‘Network Rail – train positioning data', ‘Rail Delivery Group – Darwin train positioning data’

Data set/data service name	Theme	Description	Potential use case(s)
Transport Scotland	Demographics, movement and behaviours Transport operations	Statistics reported on include road casualties, transport statistics, transport and travel, seatbelt and mobile phone usage, air maritime and Scottish Household Survey transport data, including results from the Travel Diary	See use cases for other transport portals and ‘Daily domestic transport use by mode’
Admiralty Marine Data Portal	Transport operations Freight Earth Observation and environmental Other	Data portal provides extensive data on bathymetry, wrecks and obstructions, ships' routeing, astronomical and calendar information, maritime limits and offshore infrastructure	See ‘Vessel Finder’ use cases
Net Zero Market	Energy generation Transmission and emissions	Region-specific, regularly updated data sets	Potential for these data sets to be used by, for example, local authorities to support their net zero efforts. This includes potential locations for renewable heat, EV charge points, renewables and storage, building retrofits and more
Shipping data products – MDS transmodal	Freight	Includes container shipping data, World Cargo Database, Containership Database, Ferry Database, and GB Warehouse Database	See 'Kpler – Marine traffic data services'
UKPN Open Data Portal – energy	Energy generation Transmission and emissions	Covers UK Power Networks (UKPN) data related to its assets across London, the South East and East of England	See ‘National Grid Connected Data Portal’ use cases
Defra survey data download	Earth observation and environmental	Contains LiDAR, bathymetry, airborne imagery, oblique photography, vertical photography, digital surface models (DSMs), digital elevation models (DEMs) and point clouds	Contains a variety of high-resolution data sets which can be used to build models of the landscape (for example, DSMs/DEMs) to detect features (LiDAR point clouds/imagery) and to monitor change over time (imagery)

Data set/data service name	Theme	Description	Potential use case(s)
National Atmospheric Emissions Inventory (NAEI)	Energy generation Transmission and emissions	The NAEI collects and analyses information from a wide range of sources – from national energy statistics to data collected from individual industrial plants – to provide an estimate of annual pollutant emissions	Analysis of changes in pollutants and greenhouse gasses through time. In some cases, the data is categorised into sectors, such as transport. This type of data is important in understanding and tackling climate change
ORR statistics	Demographics, movement and behaviours Transport operations	Large repository of sector-wide statistics, including but not limited to: network passenger and freight usage; safety statistics, for example, Signals Passed at Danger (SPAD) statistics; passenger experience; passenger rail performance; infrastructure and assets; and finance, for example, rail fares	Intelligent understanding of correlations between changes in outcomes (passenger rail performance) and inputs (network condition). Financial information (or more detailed information than currently released) could simplify the passing of money between bodies and divert spend from managing money to delivering outcomes
ORR Origin and Destination Matrix	Demographics, movement and behaviours	In addition to ORR statistics, ODM includes journeys between all pairs of mainline stations in Great Britain	See ‘Rail Data Marketplace – RDM’ use cases
Statistics at DfT	Transport network Transport operations Freight	National statistics releases relating to the transport system across the United Kingdom, including road accidents and safety statistics, road congestion and travel time.	See ‘STATS19 – collisions data’. See also ‘STATS19 – collisions data’, ‘National Travel Survey’, ‘DfT maritime and shipping statistics’.
STATS19 – collisions data	Demographics, movement and behaviours	Contains detailed road safety data about the circumstances of personal injury road collisions in Great Britain from 1979, the types of vehicles involved and the consequential casualties	Analysis of spatial patterns and causes of road collisions through time. Data has been captured over a long period of time which lends it well to ML applications. It also has many attribute fields which allow in-depth statistical analysis. Can be combined with additional data sets, such as road networks and geometries, to generate further insight
TfL active road disruptions	Transport operations	API providing listings and details of active disruptions on the TfL road network. Results	See 'National Highways Digital Lab'

Data set/data service name	Theme	Description	Potential use case(s)
		can be filtered by level of severity and disruption classification	
Tracsis – transport surveys (including CCTV/CV)	Transport operations	Continuous or discrete period feeds of video from CCTV cameras	See 'CCTV – highways-based video feeds'
Mobility metrics for Glasgow City Region – mobility data	Demographics, movement and behaviours	The mobility metrics data set was created using mobile phone application data from Huq and Tamoco data. The data set covers Glasgow City Region with results broken down by Intermediate Zone (417 in the City Region) from 1 July 2019 to 31 December 2021 (2.5 years) with results aggregated quarterly (10 periods)	Insight into human mobility patterns and behaviour at granular spatial resolution and over time. Potential for responsive, data driven transport planning and modelling, particularly if made available real-time
Camera captures and confirmed vehicles seen in the congestion charge zone by month	Transport operations	Data is available on camera captures and confirmed vehicles observed within the congestion charge zone during charging hours, reported monthly. Camera capture data is available from July 2010 and confirmed vehicle data from October 2016	This data set reflects the existence of sensors (cameras) for ANPR capture of vehicle registration marks. These cameras can potentially be a useful source for broader ML activities, including capturing traffic flows (such as per TfWM ADEPT). ANPR could also provide some insight into routeing (at least inferred from timestamped entry and exit points at geolocated camera locations)
Carbon intensity	Energy generation Transmission and emissions	Carbon intensity forecast with a regional breakdown. It uses ML and power system modelling to forecast the carbon intensity and generation mix for each region in Great Britain. Values are the average, maximum and minimum carbon intensity in gCO2/kWh for each day. It includes emissions related to electricity generation only	If historical data is available, analysis in changes by region through time

Data set/data service name	Theme	Description	Potential use case(s)
Civil Aviation Authority	Transport network Transport operations Freight	Data coverage includes details of airport size, aircraft movements, air passenger numbers by type and nationality of operator, freight by type and nationality of operator, air passenger routeing/flows and regular survey of departing passengers	High level aggregate and historic statistics. Could support other applications as part of a data fusion exercise
DfT aviation collection	Transport network Transport operations Freight	Includes airport activities including air traffic at UK airports, punctuality, international passenger movements, mode of transport to the airport and travel purpose. Also includes airline activity including aircraft number of kilometres flown (UK airlines), passengers uplifted, passenger kilometres flown, passenger seat occupancy, cargo uplifted and cargo tonne-kilometre flown, as well as aviation accidents and incidents	High level aggregate and historic statistics. Could support other applications as part of a data fusion exercise
DfT maritime and shipping statistics	Transport operations Freight	Links to 'DfT port and domestic waterborne freight statistics' , 'Sea passenger statistics', 'Shipping fleet statistics'	Aggregated statistics but may support (in data fusion or through validation) similar use case to 'Kpler – marine traffic data services'
DfT port and domestic waterborne freight statistics	Transport operations Freight	Includes freight tonnage traffic by port and year, freight tonnage, international and domestic by direction and year, freight traffic cargo types by year, international freight traffic by route and cargo type	Aggregated statistics but may support (in data fusion or through validation) similar use case to 'Kpler – marine traffic data services'
Strategic noise mapping (NMS)	Earth observation and environmental	Modelled average sound levels from roads and railways in Wales	Can be combined with demographic and environmental data to understand the impact of noise on both the population and wildlife. This can assist with decision-making and prioritisation when considering how to mitigate the impacts of noise

Data set/data service name	Theme	Description	Potential use case(s)
Road traffic bulk downloads	Transport operations	Raw traffic count data and estimated annual average daily flow for points on the road network	Pattern detection to identify relationships between traffic flows and potential contributory factors
Transport Focus Data Hub	Demographics, movement and behaviours	Surveys of transport users, such as bus passengers, rail and strategic roads. Varies by mode but some surveys go back to the 1990s	Behaviour data can be used to inform transport policy for different modes. Some data sets go back many years and could be analysed for long-term shift in attitudes
Use of novel data sources open data set	Demographics, movement and behaviours Transport operations	Deliverable of the FUSION project, commissioned by DfT and being carried out by SYSTRA. It will contain aggregated travel movements by mode over a three-month period, grouped by transport user personas. A closed, more detailed data set will be held by DfT for internal use for a period of three years	This data set supports understanding of the use of transport and how to ensure the smooth operation of transport for all demographics. It has the potential to be considered “big data” if enough participants are surveyed in data creation
Glasgow 3D city models from airborne LiDAR	Earth observation and environmental	Urban Big Data Centre generates 3D city models via the airborne LiDAR point clouds acquired between 2020-2021 on behalf of Glasgow City Council. It is a large-scale 3D city model containing 3D information on terrain, tree canopy and building products in Glasgow derived from high-density airborne LiDAR data	High resolution 3D models can be used for various types of modelling, including transport planning and urban planning. Feature extraction from such models can be used to generate high-definition maps containing attributes, such as street furniture, not typically found in readily available data sets
Sensor-enhanced housing survey data for urban heat investigation	Connected vehicles and sensors	This data set comprises sensor data collected from Smart Citizen Kits (SCK) sensors, combined with survey data collected by the Bureau of Investigative Journalism. The data was collected between July and September 2023, which includes a short heatwave in early September 2023. The small sample size includes 40 SCK sensors	The sensors provide an indication of localised heat stress, air quality and noise disaggregated by deprivation level. This could be used in design of transport schemes that can immediately impact the local environment, including through contributions or mitigations for urban heat, air and noise quality, as well as monitoring and evaluation of existing schemes and infrastructure to gain deeper insights into the wider impacts of transport on local communities

A.1.2 Framework

A framework was developed to gather metadata about each of the data sources, which was then used in onward analysis.

Name	Theme	Description	Example use cases(s)	Ownership	Owner/ controller	Access	Temporal resolution	Temporal extent
Name	Connected vehicles and sensors Demographics, movement and behaviours Earth observation and environmental Energy generation, Transmission and emissions Freight Transport network Transport operations Other	*Description*	List of use cases	Private/police/ local govt/national govt/open	*Owner/ controller*	Open/shared/ closed/not exploited	None/real-time/minute/ hourly/daily/weekly/ monthly/quarterly/ annually/longer	Ongoing/ *Start date*

Update frequency	Scale	Discoverability	Entry type	Spatial extent	Mode	Interoperability	Notes	Source URL
None/real-time/minute/ hourly/daily/weekly/ monthly/quarterly/ annually/longer	PB/TB/ GB/MB	*Recognised data portals, for example:*	Data set/ data source	UK/GB/ E/S/W/ region/CA/ county/LA	Walk/cycle/ micromobility/bus/ DRT/tram/rail/taxi/ride share/car/freight/air/maritime	File/AP/other	Notes	Link

	data.gov.uk, London Data Stor	Data portal

A.2 Filtering and prioritising

A.2.1 Shortlisting criteria

The below criteria were used to perform shortlisting. While not assessed, the effort required to process is another important consideration. Strategic alignment was assessed on a qualitative basis for each of DfT's five priorities.

Scoring criteria	▪ **Strategic alignment:** alignment with the new government's strategic priorities ▪ **AI suitability:** suitability for use in AI application ▪ **Access:** availability and ease of access of the data sources
Scoring	Scores for each criterion are on a 1, 3, 5 scale
Weighting	Total score out of 20 given to each based on the following weightings: • Strategic alignment: 1 • AI suitability: 2 • Access: 1

A.2.1.1 AI suitability

Low (1)	**Medium (3)**	**High (5)**	
Limited potential for AI integration, with minimal benefits from AI applications	Moderate potential for AI integration, with some benefits from AI applications	Highly suitable for AI integration, offering substantial benefits from AI applications	▪ Temporal resolution: frequency at which data is collected ▪ Temporal extent: duration over which data is collected ▪ Scale: size of data sets ▪ Spatial extent: geographic area that the data covers ▪ Update frequency: how frequently data is updated Each subcriteria given a score of 1, 3 or 5, summed, then divided by 25 which is assigned: 5: > 0.7 3: < 0.4 1: 0 - 0.4
Temporal resolution: quarterly or longer Temporal extent: snapshot Scale: MB or less Spatial extent: national/regional/local Update frequency: quarterly or longer	Temporal resolution: weekly or longer Temporal extent: finished before 2024 Scale: GB Spatial extent: GB/UK Update frequency: weekly or longer	Temporal resolution: up to real-time Temporal extent: ongoing Scale: TB, PB Spatial extent: global/continental Update frequency: up to real-time	

A.2.1.2 Access

Low (1)	Medium (3)	High (5)
Data is freely available to access without any restrictions	Data is accessible to specific groups or individuals under certain conditions	Data is restricted and not available for public use, kept within an organisation for internal purposes only or not even exploited internally
Open	Shared, various	Closed, not exploited

A.3 Analysis and insight

A.3.1 Challenge statement

Network management and travel behaviour	**How might we enhance transport conditions for all road users** and meet policy objectives through network management that prioritises movement of people and goods? **How might we make bus services more reliable and responsive to user needs, making it easier for people to rely less on cars, protecting the environment and saving people money?** **How might we help the last-mile delivery of goods to be more efficient for couriers, maintain resilience from multiple suppliers and reduce the number of vehicles and distance travelled, improving environment, congestion, safety and economic outcomes?** **How might data, including closed commercial data, help develop a robust user-centric view of transport and transport users' needs and motivations at much greater levels of detail, contributing to key policy priorities including decarbonisation, vision zero and a more active, healthy and happy population?**
Data access	**How might the public sector make the most of transport-related data, whether sharing their data with others or as consumers of other data themselves**, to enhance people's daily experience of transport, to extend equitable transport access to all communities and to reduce transport's negative environmental and social impacts? **How might data,** including closed commercial data, **support innovation in how (multimodal) transport is planned and provided,** so that it is more responsive to user needs and supports more equitable and sustainable access to education, employment and services?
System outcomes	**How might we make it safer to travel by any mode and at any time of day,** so that equitable access to jobs, services and leisure activities is possible for all however they need or choose to move, including on foot, by bike and by public transport? **How might we better understand the resilience of transport systems** to plan for and equitably mitigate the potential transport-related impacts on communities and economies from external risks including health and climate? **How might we develop a holistic understanding of whole-system carbon emissions across transport** to support better decarbonisation strategies through transport planning and operations?

Infrastructure and systems	**How might transport data integrate with other systems (for example, energy)** to help deliver more efficient and more effective public services and to enable transport decarbonisation? **How might AI and improved data transform operational processes and planning within and related to transport** (for example, local plans, local transport plans, transport assessments and business cases) to accelerate the decision-making process and expedite policy enactment? **How might we transform conditions for active travel** by building a solid understanding of relevant conditions, activities and behaviours at a hyper-local level and, therefore, increase levels of sustainable travel and physical activity, localised economic activity and improve health and wellbeing? **How might we monitor and detect changes in the conditions of transport assets** faster to optimise maintenance and renewal activities and, therefore, minimise impacts on the environment, the population and the economy?

A.3.2 Breadth of challenge statements addressed

	Enhance transport conditions for all road users	Make bus services more reliable and responsive to user needs	Help the last-mile delivery of goods to be more efficient	Make it safer to travel by any mode and at any time of day	Sharing public data and using external data	Integrate with other systems (for example, energy)	Support innovation in multi-modal transport provision	Help develop a robust user-centric view of transport	Transform operational processes and planning in transport	Transform conditions for active travel	Better under-stand the resilience of transport systems	Monitor changing conditions of transport assets	Understand whole-system carbon emissions holistically
#1: Last-mile	✓		✓	✓	✓	✓			✓	✓			✓
#2: Mobile data	✓	✓			✓		✓	✓	✓	✓	✓		
#3: Highway CCTV	✓	✓	✓	✓	✓				✓	✓		✓	✓
#4: UTC	✓	✓	✓	✓	✓	✓	✓						✓
#5 Connected vehicle data	✓	✓		✓	✓		✓		✓	✓		✓	✓

A.3.3 Analysis of high-potential data sets

Each high-potential data set was explored in some detail to understand how it maps to the challenge statements, its potential use cases, key stakeholders for accessing and working with the data sets and barriers to accessing them.

	Last-mile couriers and haulier operational data	**Mobile phone operating system and services providers (speculative)**	**CCTV highway video feeds**	**UTC and connected infrastructure sensor and configuration data**	**Connected vehicle data**
Overview	Up-to real-time information on a range of freight activities and logistics Vehicle locations (parked, in use or in transit including dwell times), vehicle load/capacity information, goods priority information (time-sensitive deliveries), fuel usage, origins and destinations	Operating system providers Data from iOS and Android system services such as GPS, accelerometer sensor data, Apple/Google Pay, Apple/Google Maps and a range of other vendor-provided apps and services that are related through user accounts and devices. Such data would be high volume and might provide highly disaggregated insights into travel patterns and purpose	Fixed position CCTV cameras on the highway network, including National Highways traffic cameras, police and other ANPR cameras Many of these video feeds are suitable for CV data extraction, including object recognition/ detection and ANPR reading, that can identify vehicles and people, classify speed and trajectory and gather other key information about highways	UTC data, including real-time signal phasing timings, loop and other counters, Split Cycle Offset Optimisation Technique (SCOOT) loops, bus priority trigger data at junctions and UTC asset status data, including live information about road closures MAP, which defines road and lane topology near a junction, and signal phasing and timing, give detailed information in a standardised format, on signal phasing	Data that is generated by autonomous and/or connected vehicles across all manufacturers including: vehicle telematics (for example, speed or location); safety-related information, such as unusual braking, road surface conditions and object proximity; and occupancy data. Connected vehicle data richness is increasing exponentially with uptake of autonomous mobility and vehicle sensing. CANBus protocols commonly in use in the sector mean that this information is already widely communicated in standardised formats

Selected challenge statements	#1: Enhance transport for all users through network management #3: Efficient last-mile delivery, reducing vehicle numbers and kms #13: Develop a holistic understanding of whole-system carbon emissions across transport to support decarbonisation strategies	#1: Enhance transport for all users through network management #8: Robust user-centric view of transport and transport users' needs and motivations #10: Enhance active travel with better understanding of local conditions	#1: Enhance transport for all users through network management #4: Make travel safer by any mode and any time of day #7: Support innovation in how multimodal transport is provided and planned	#1: Enhance transport for all users through network management #2: Make bus services more reliable and responsive to user needs #5: Make the most of transport-related data, whether sharing public sector data or as consumers of other data	#1: Enhance transport for all users through network management #4: Make travel safer by any mode and any time of day #12: Monitor and detect changes in the conditions of transport assets
Example use cases	• Enhanced vehicle route optimisation tools using ML • Fusion with other data like STATS19 to better understand traffic collisions • Live haulier data feeding into UTMC systems • Identify last-mile opportunity areas	• Insights into customer behaviour across transport networks and modes • Responsive, data driven transport planning and modelling, including origin-destination matrices or routeing • Fuse with demographic or transaction data to build comprehensive understanding	• Support dynamic network management and optimisation • Support development of digital twins • Increase robustness of traffic counts and transport modelling • Plug gaps in data for walking, wheeling and cycling modes	• ML to support deep understanding of the transport network for forecasting, scenario testing and incident response • Enhanced analysis and risk mapping, including for bus and active modes • Highway network digital twins	• Highly granular vehicle fleet usage data supporting policy-making • Near miss data supporting academic and policy safety research • Support dynamic transport management focused on movement of people

Key stakeholders	• Major freight and haulier companies • Consolidation centre operators • UTMC operators • Highways authorities	• Operating system/services providers • GDPR and legal authorities	• Organisations capturing and processing CCTV, including highways authorities and police • UTMC operators • Highways and regional authorities	• UTC operators • Highways and regional authorities • National Highways	• Vehicle manufacturers • Vehicle data companies such as NIRA Dynamics
Barriers to access	• Fragmentation and complexity of the freight market • Data held by large private organisations • Commercial sensitivity	• Data sensitivity and need for anonymity while still getting sufficient granularity • Commercial sensitivity of data • Cost of procuring data with significant market value • Size of market considerations	• Complex structure of organisations capturing and processing CCTV • Lack of processes to share data • Historic legal agreements with CCTV operators • Knowledge and experience in highways authorities • GDPR responsibilities • Cost	• Technical challenges and a lack of standardisation between different operators • Resource constraints and coordination between organisations • Insufficient vendor incentives	• Huge volume of data being collected • Commercial sensitivity, particularly around safety data given public sensitivity to autonomous vehicle safety