EDIF: Towards a digital twin for Urban Transport: Final Report & Blueprint

Lee Chapman, Nicole Cowell – University of Birmingham
Philip James, Jennine Jonczyk, Daniel Bell – Newcastle University
James Evans, David Topping, Thomas Bannan, Ettore Murabito, Emma Tsoneva – University of Manchester
Mark Birkin, Dani Arribas-Bel, Dustin Carlino - Alan Turing Institute

Executive Summary

In Autumn 2021, the Department for Transport (DfT) and the three urban observatories of Newcastle, Manchester and Birmingham were successful in obtaining a £300k Economic Data Innovation fund grant. The main objectives of the grant were to promote and improve data sharing across the organisations, establishing best practice were possible and in doing so setting the foundations for urban transport digital twins. The nature of the project meant that other parties were onboarded during the course of the project, namely the Connected Places Catapult and the Alan Turing Institute.

This report summarises the key deliverables of the project, namely the cataloguing of available sensor data, recommendations of metadata protocols and developing a prototype to integrate (harmonise) and visualise disparate data streams from the DfT, the Urban Observatories and other sources. A selection of use cases are also presented for further development which include AI based traffic prediction and air quality modelling and monitoring. A link is also made to the decarbonisation agenda. The report concludes with a blueprint, outlining potential next steps to elevate the work as a core component for the National Digital Twin programme.

1. Background to Project

The last ten years has seen an enormous increase in the number of connected devices collectively known as the *Internet of Things* - IoT. Ubiquitous and cheap data communications from Wi-Fi, 4G/5G and IoT specific protocols such as NBIoT and LoraWAN have supported the growth of connected devices providing real-time information on and from the transport infrastructure such as real-time telemetry from buses and trains, passenger occupancy etc. Additionally, there has been widespread growth in sensors that measure traffic, pedestrians and environmental factors related to transport such as air quality, weather etc. As these have technologies have reduced in cost they are being adopted by local authorities, transport providers and many others to provide operational information in real-time.

The Urban Observatories (UO) are a collective of city/university partnerships that have been researching "*in the wild*" the deployment and management of smart city monitoring systems for the past six years. Funded by BEIS, the £12m project has deployed sensors across six partner cities to gain an understanding of how to build and manage these systems and research questions of sensor calibration, deployment strategies and data management. The UOs have also explored procedural and governance issues related to smart city deployments. In addition to these discovery aspects of the projects, the UOs have been looking at how these new forms of data can support societal change in addressing social and urban challenges such as poor air quality, urban flooding, infrastructure repair and maintenance.

The Alan Turing Institute is the National Institute for Data Science and Artificial Intelligence, and has been funded by UKRI, universities and partners in business and government. It aims to develop and apply world-leading methods to the most important real world challenges today and into the future. The Urban Analytics programme at the Turing is specifically focused on deployments of data science and AI to human systems and mobility patterns.

The UOs share a common ethos around open data and transparency, publishing data freely and in real-time. They share common protocols for data access and have been actively testing data curation technologies, discovery and access mechanisms. However, as more cities and municipalities start to adopt this technology, the danger is that information is siloed within individual systems with poor interoperability and standardisation. This creates serious issues for comparative analysis and integration at regional and national levels. In order to facilitate city, regional and national digital twins (DT), support decarbonisation, improve air quality and enable agile decision making etc. it is imperative that a homogenous approach enabling data integration is pursued whilst acknowledging that operational heterogeneity is inevitable. As such this programme of work set out to achieve a number of goals:

1. Catalogue current deployed sensor technology across the observatories of Newcastle, Manchester and Birmingham
2. Review current and emerging metadata and data access protocols to make recommendations for future IoT systems in the urban transport realm
3. Develop a prototype for integration and visualisation of multiple IoT data sources using metadata and access protocols: a real-time multi-city dashboard of integrated data
4. Develop a range of use-cases to highlight the potential benefits of real-time data for stakeholders and decision makers. The final use cases are:
    a. AI based forecasting and simulation using real-time data
    b. Sensor informed modelling of Low Traffic Neighbourhoods
    c. Deployment of new sensors for monitoring decarbonisation
5. Engage with local and national stakeholders and the user community to understand future needs and requirements, resulting in a shared blueprint

The project is funded through EDIF and includes representatives from the data and analytics team at the Department of Transport, the Universities of Newcastle, Birmingham and Manchester, the Alan Turing Institute and the support of the Connected Places Catapult.

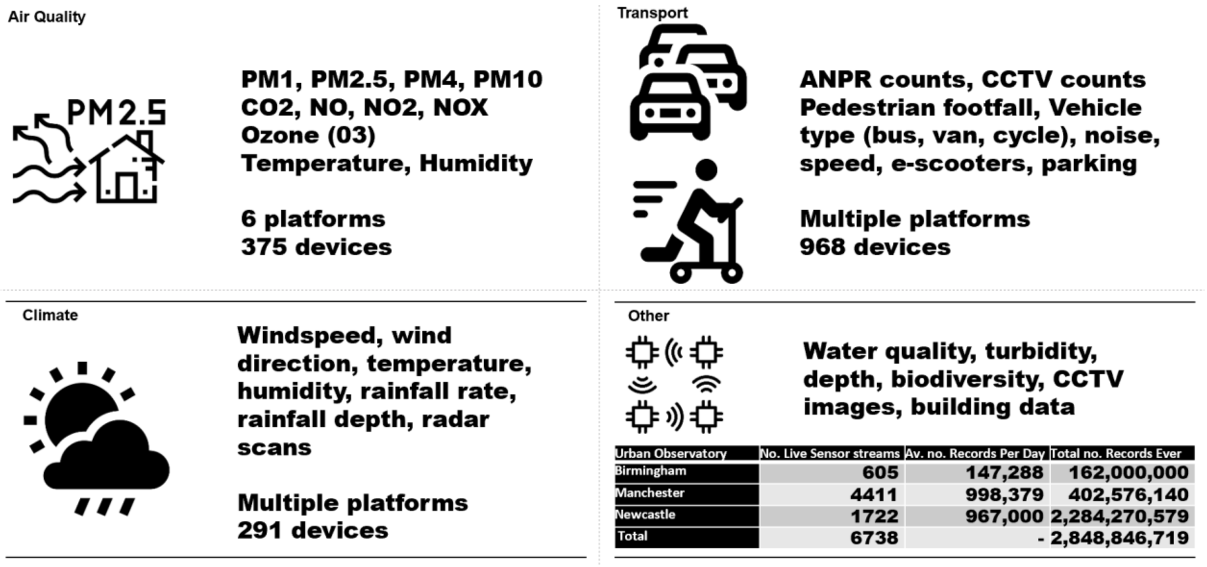
2. Urban Observatories Data Catalogue

2.1 Background

The starting point of the project was to collate the sensor metadata from three UO catalogues – Birmingham, Manchester, and Newcastle. This includes a combined overview of available data (appendix 1), metadata themes per observatory (appendix 2), a discussion of quality metrics, how to access each observatory API endpoints, a discussion on ethics, and questions to consider moving forward. The details of this are available in the separate reports (appendix 1 and 2) but are briefly summarised here.

In total the UO's presently consist of 2146 sensors. However, it must be noted that not all of these are unique hardware devices, but also comprise of software generated detectors. In this regard, the nomenclature is important. An *observation* is a measurement of a single phenomenon in space and time from a single sensor stream. A *sensor stream* is a continuous stream of observations from a sensor device. A *sensor device* is a physical sensor that measures one (but typically more than one) phenomenon. A *sensor platform* is a collection of sensor devices that share similarities (e.g. devices from a particular manufacturer) and access protocols.

The UOs have deployed large numbers of standalone sensing devices, but also operate as a clearing house for 3rd party sensor data that is collected through embedded IoT devices deployed for operational reasons. For instance, data is collated from local authority ANPR (Automatic Number Plate Recognition) devices and from live GPS tracks on buses. These sensing devices are deployed and managed operationally by public and private bodies, but real-time feeds are made available to the UOs. The UOs collate and manage three broad types of sensors: (1) deployed standalone IoT sensor devices e.g. air quality sensing devices (2) embedded IoT devices operated by 3rd parties e.g., ANPR counts and (3) sensing devices that require further processing for data extraction e.g. CCTV. The range of potential sensing devices is enormous, but the majority of urban sensing focuses on mobility or mobility related metrics. Figure 1 below summarises the range of metrics and data volumes.



Urban Observatory	No. Live Sensor streams	Av. no. Records Per Day	Total no. Records Ever
Birmingham	605	147,288	162,000,000
Manchester	4411	998,379	402,576,140
Newcastle	1722	967,000	2,284,270,579
Total	6738	-	2,848,846,719

Figure 1 Summary of current UO data holdings and examples of sensor types

2.2 *Sensor applications*

The applications of UO sensor devices are broad but can be summarised as:

1. Used to *measure trends* especially long- and medium-term trends. Long baselines of data are critical to understanding change against the background noise of a city.
2. Lower cost, and smaller size, enables sensors to be used *to increase geographical coverage* of measurements e.g. to measure air quality in the suburbs or throughout a network.
3. IoT sensors are used to carry out *place-based experiments* to understand the impact of change at high spatial and temporal resolution e.g. to accompany new infrastructure (e.g. a new road) or policy change (Low Traffic Neighbourhoods, Ultra Low Emission Zones, pedestrianisation). Sensors can be moved relatively easily to carry out multiple campaigns.
4. To provide *continuous monitoring across multiple phenomena* to quantify the impact of policy changes and to *capture unintended consequences* (e.g. displacement activity such as rat-running).
5. To provide unprecedented amounts of data to improve and support new *modelling and predictive analytics* such as AI leading to Digital Twins.
6. To provide *operational intelligence* for automated and human-in-the-loop systems such as traffic management, crowd control, automatic signalling.

IoT sensing is both relatively new and currently in a largely unregulated space. This means that care must be taken when analysing data or purchasing equipment. As sensors are deployed "*in the wild*" they suffer mechanical and electrical breakdown, communication issues and degradation. Increasingly data is seen as a saleable commodity and can be locked into proprietary systems. The UOs propose an open, scalable and federated solution to enable data for all and the public good.

It is proposed that sensor catalogue should be updated biannually but should be used in conjunction with metadata also coming from the same sensors (e.g., performance metrics) to ascertain stream health.

3. Metadata

The variety of sensing platforms, devices and observations requires careful thought to capture the necessary information related to an observation. For example NO2 (Nitrogen Dioxide, a critical contaminant from combustion engines) can be measured by a £200 IoT sensor or a fully calibrated DEFRA approved £100,000 sensor. These measurements differ in accuracy and reliability and therefore how they should be used and reported. Metadata must address quality and potential usage applications to ensure that the data is fit for purpose. To this end, metadata has always been of paramount importance in any discipline that involves the production and exchange of quantifiable information, as it represents the means by which such information has to be understood and contextualised.

Metadata has long been a focus of collaboration among the UO's of Manchester, Birmingham and Newcastle, with the goal of exploring options and establish best practices in order to facilitate and standardise the data communication across the cities. This joint effort led to the (partial) adoption of an ontology-based metadata solution and the implementation of APIs that expose and describe the observatories' corpus of data through a shared vocabulary. This project has taken this a step further and allowed the opportunity to further explore viable options, building on the work previously done by the UO's, and ultimately adopting a different approach that would better suit the expectations of DfT in the short, medium and long term. Given the importance of metadata in future work, a report (appendix 3) was produced to indicate a possible way forward in the light of the ongoing collaboration with DfT and other parties.

The summary report concluded that an approach recommended by FutureGov is a pragmatic one for the transport sector, where the design of a metadata standard should start with the short-term goal of meeting the needs of specific target groups and projects. It is suggested that a richer and more expressive metadata standard should be the focus of a long-term effort. This would be achieved through a feedback loop between data publishers and users, which is intended to guide the resource publishers in creating, maintaining and continuously improving their service. However there seems to be no clear roadmap leading to a wide-consensus over a more articulated metadata solution that grows organically from meeting shorter term needs.

For this project, the final adoption of a metadata solution was to be guided by the requirements of (i) adopting an agile approach that would require minimal restructuring to accommodate future innovation in existing functioning systems and (ii) using data models informed by relevant stakeholders' feedback and based on concrete sector-specific use cases. The Open and Agile Smart Cities' (OASC) approach to metadata and data models seemed the most suitable to meet those criteria. In particular, the OASC approach focuses on Minimal Interoperability Mechanisms (MIMs) to achieve interoperability of data, systems, and services based on an inclusive list of baselines and a minimal common ground.

On this basis, the data provided by the UO can be harmonised (in particular, made MIMs compliant) where it can be passed on to a third party to facilitate the sharing and access of live urban data streams on a large scale. This was then demonstrated on the project.

4. Data Integration of real-time IoT

The sheer variety of potential measurements that can be made from real-time transport, infrastructure and environmental monitoring poses a real challenge and cover widely different systems and measurement regimes from real-time GPS bus positions to static road-side air quality monitoring systems. Real-time systems, by their very nature, provide high volumes of data (streams) that present a technical challenge to manage and store and an additional processing challenge to convert raw data into more usable information (e.g. through data aggregation). The skills gap within many end-user organisations who deploy and use IoT sensing devices presents a human and process challenge. Many sensor systems are sold as "data as a service" (DAAS) with the manufacturer providing hardware, deployment, maintenance and tools to view and manipulate data in the cloud. Most end-users lack capacity or skills to carry out integration work in-house. As systems are developed that rely on real-time inputs it is important to understand the health and quality of the data stream (does it have a lag, does it drop out intermittently, how many observations are received every hour etc.).

4.1 Requirements Analysis

To address these challenges a prototype data integration platform should:

1. Be able to access any sensor endpoint and transform or map the data to a common standard data format
2. Provide a flexible metadata system that provides usable domain data but able to accommodate future requirements
3. Be based on open data standards
4. Have low barriers to participation enabling low-tech/low-cost onboarding of data
5. Minimise the impact on any operational systems at end-user locations
6. Provide a single API to access the data for analysis, modelling and visualisation

4.2 Prototype Solution

The timescales of the project meant that a commercial off-the-shelf approach to meeting these requirements was needed. Therefore the prototype makes use of an existing software platform – Urban Data Exchange (UDX). UDX spun out of an EU funded research project to support city sensor data and have previously worked on proof-of-concept with Newcastle UO. The platform provides the means to integrate sensor streams into a standard set of APIs, metadata and data models (Figure 2). More specifically:

1. Data onboarding is relatively straightforward, only requiring a sensor stream endpoint (e.g. API calls to the Urban Observatory data streams).

2. Data is then mapped from its “raw format” to a common data model. In this case, Smart Data Models were utilised (smartdatamodels.org) which provide standardised data models for a large number of metrics. The mapping/harmonisation takes place and requires to be set up for each data stream.
3. The UDX platform provides a searchable discovery engine that is organised through contributors. Data streams can be tagged as public or private.
4. Users can subscribe to sensor streams (if they have permissions) and then data is pulled from the UDX platform using their standard API. One limitation that was discovered is that requests from UDX are limited to 1000 responses at once which only provides a small time window for high temporal resolution sensor data. This is easily overcome by running the request multiple times via a dummy service. Using a 3rd party to provide the data streams avoids the requirements of contributors to maintain or manage complex ICT architecture.
5. Basic sensor metadata is displayed through example data. Metadata on stream health (through a traffic light system) and information on data flow metrics are available from the UDX platform.

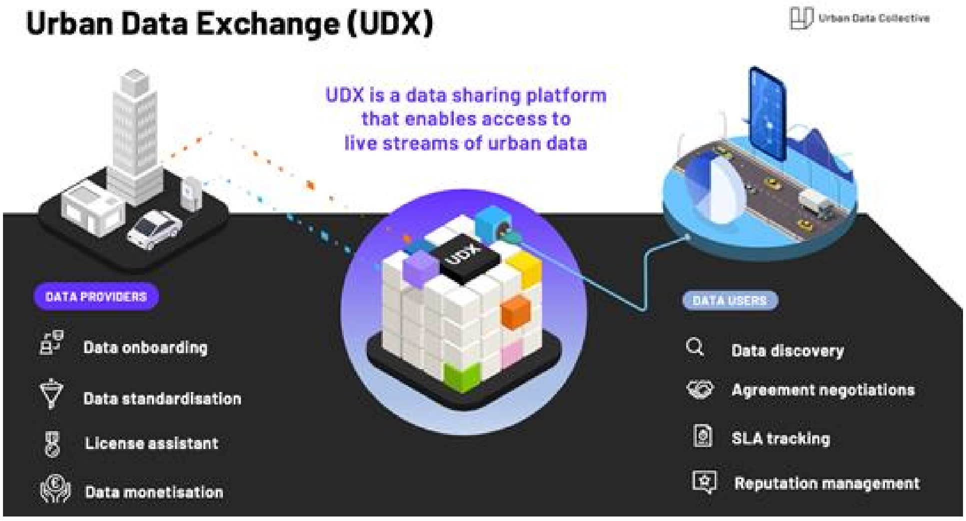


Figure 2: Schematic of the UDX platform

4.3 Integration dashboard

As proof of concept, a subset of data streams were shared with the UDX platform. These included: traffic flow from Manchester and Newcastle, black carbon from Manchester, and PM2.5 from all three observatories. In addition to these, Newcastle temperature, humidity, and noise variables were already supplied to UDX through an existing staging environment.

In terms of metadata and structure, all these streams differed from one another both within and across UO organisations. This highlights the harmonisation challenge as differences exist even in data lakes where communication and standardization are key tenets. It was therefore necessary to decide upon how the payload of these heterogeneous data sources could be restructured into a common format before being republished. Due to its proven track record within smart city and IoT applications, as well as applicability to the task at hand, entities from the Next Generation Service Interfaces Linked Data (NGSI-LD) smart city data model were selected. For PM2.5 and black carbon variables, the ‘Air Quality Observed’ definition was used, where mandatory constraints include ID, type of entity, creation date time, and location attributes. In this case, variable/intensity, sensor altitude, aggregation, date observed, and suspect reading logs were also amended as additional attributes. Likewise, the same mandatory and additional attributes were implemented into the payload structure of traffic flows when using the ‘Traffic Flow Observed’ definition.

Other than writing a description of the stream and selecting a suitable cover image, the final step involved defining an update rate. This was largely defined based on the expected update rate of UO sensor APIs. However, to avoid any data loss, the update rate was set to a slightly longer period than the expected rate in some cases.

4.4 Data Visualisation

By merging UO air quality and traffic flow data sources under the UDX platform, streams could be accessed by data users at a common location and in a common format, thus greatly simplifying the collection process.

Data can then be visualized by visiting a stream on UDX which are available after registering on the UDX site and subscribing to individual streams. However, the project took this a step further by developing a web-based real-time dashboard to demonstrate the relative ease of downstream application development following data integration. After UO data streams had been unified under the UDX platform and in the same formats, this becomes a relatively straightforward task with a range of dashboard software options available.

The dashboard software of choice was Grafana – a multi-platform open-source analytics and interactive visualisation web-application. This was selected over other web-applications, such as Datadog or Dynatrace, for several reasons, namely Grafana's wide range of visualisation options, alerting capabilities, plugins, stylistic choices, and intuitive controls for both developers and viewers. In addition to this, Grafana software is free for use when hosting the site locally. In this project, Grafana was installed and developed locally, but a cloud-based version is available. The infinity plug-in, which provides native support for live JSON inputs, was used to seamlessly connect to the API addresses from the UDX streams.

Within the dashboard, any number of interactive panels can be created and positioned to visualise one or more compatible data sources. Panel types include time series, bar charts, logs, gauges, alert lists, and geomaps. In this case, gauges were used to display Newcastle UO average temperature, average relative humidity, and average noise levels at current time. UO traffic flows and air quality readings were depicted using time-series graphs. In addition to this, interactive heatmap panels were used to display current PM2.5 levels in each city. To display suspect readings contained within the streams, the logs panel was used. Here, the sensor name, coordinates, variable, and value of any suspect reading can be examined. One other panel type used within the dashboard alerts the viewer of general stream health, as well as if any World Health Organisation air quality thresholds have been exceeded.

Figure 3; Grafana Dashboard 22nd March 2022. To view the live dashboard, visit – https://dashboard.view.urbanobservatory.ac.uk/d/6UjxM7Y7z/edif?orgId=1&from=now-7d&to=now&refresh=5m&kiosk.

The final dashboard consists of eleven panels representing real-time air quality, traffic flows, meteorology, suspect reading logs and alerts from three different Urban Observatory cities (Figure 3). Panels can be interacted with in several ways. For example, by clicking on the name of a panel, options will appear to view the data in an enlarged panel, or inspect the raw incoming data in table or JSON formats. Building the Grafana dashboard has demonstrated the ease of downstream application development following on from the data integration stage of the project. As more unified UO streams become available through data sharing platforms such as UDX, the dashboard has the capacity to simply onboard more data sources and present through a variety of panel visualisation options.

Overall, this element of the project has demonstrated how real-time data-streams can be readily harmonized and unified under a smart city model such as UDX. Once harmonised, visualization (or indeed real-time analyses, forecasting and simulation – see use cases) is greatly simplified. Furthermore, it underpins the foundations for scalability, opening the door for data from a broad range of sources beyond the UO's (i.e. other cities).

5. Use Cases

5.1 AI based forecasting and simulation using real-time data

Time-series forecasting methods have often been used to mitigate some of the challenges associated with deploying models that solve, for example, differential equations that represent time dependent processes. This does not remove the usefulness or importance of the latter given their provenance and, often, explainability, but offers an alternative method for utilizing historical data. Additional benefits include the ability to adapt to locally-driven forces and the ability to deploy on edge devices, whether that includes local servers or low power compute systems in the field. In addition, the boom in smart cities has resulted in often substantial investments in infrastructure for capturing a large amount of ancillary data, which in theory can be useful for understanding changes in traffic levels and associated impacts which might be otherwise difficult or impossible to incorporate into a process driven model. Incorporating ancillary data into time-series forecasting could likewise enable a number of stakeholders to develop systems for evaluating a series of interventions in-house before wide scale rollout.

The availability of machine learning and statistical methods, as delivered through common programming languages, has improved significantly over the last decade. This includes widely used package such as Scikit-Learn and Keras to name a few. Statistical methods include simple additive models where non-linear trends are fit with yearly, weekly, and daily periodicity. With regards to machine learning methods, these include Long Short-Term Memory [LSTM] methods, Convolutional Neural Networks (CNN), and hybrid combinations of the two. Despite improved accessibility, a key barrier to implementation of machine learning and/or statistical analytics is a digital skills shortage across the public sector. To take advantage of these methods will require trained specialists with the capacity to help drive digital innovation. One might imagine the creation of 'Data innovation specialist' roles through targeted recruitment from the existing pool of skilled workers in the UK. The recent skills for jobs white paper emphasises importance of digital skills in modern workforce; progression to advanced technical study is a core purpose of the proposed national skills and educational reforms.

Nonetheless, in this use case we provide a small demonstration of what could be achieved. Using hourly data was used from a network of Bluetooth Journey Time Passive Sensors from the Transport for Greater Manchester cloud-hosted data platform, C2 (https://www.drakewell.com/), we evaluate methods that forecast 1-3 hours in advance.

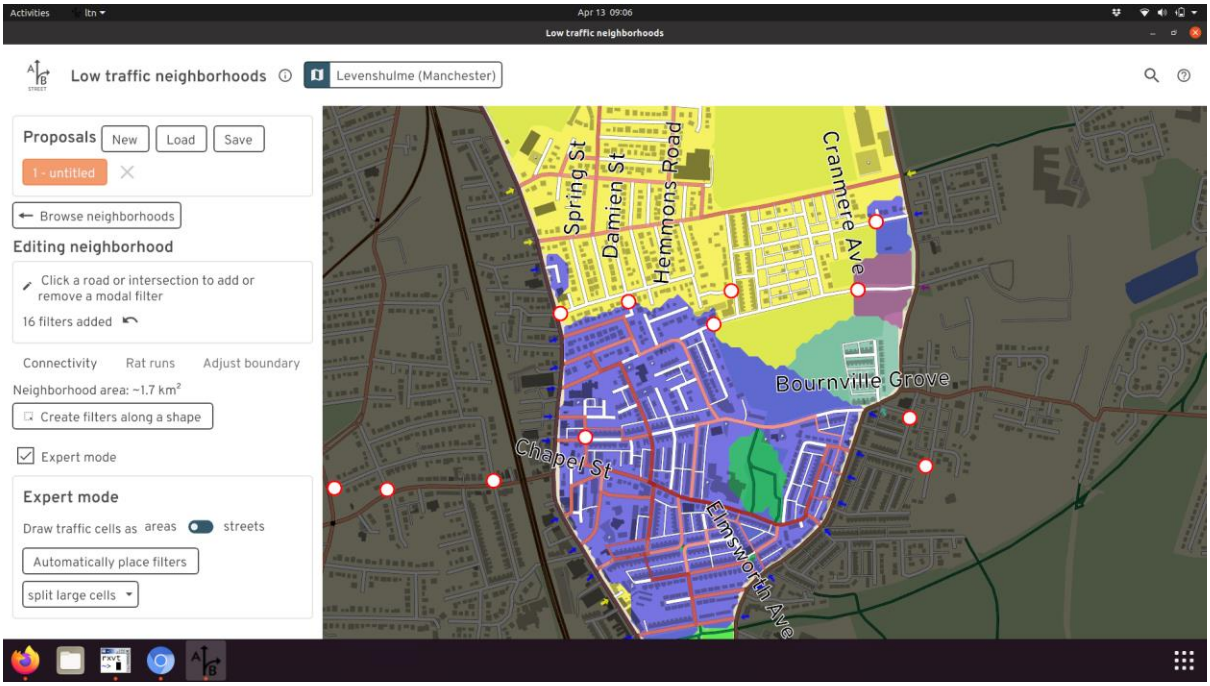
Access to the code, and data, can be requested at this URL: https://colab.research.google.com/drive/1iVHZn6wlg8ZNHWXGm0cq58_21qUNVgJx?usp=sharing

5.2 *Sensor informed modelling of Low Traffic Neighbourhoods*

This use case also demonstrated the potential of simulation using UO data. To tackle rising levels of vehicle traffic on residential streets, city planners detour traffic using modal filters, physical blockers that divert traffic from accessing a street, arranged in a way to produce low-traffic neighbourhoods (LTNs). To assist planners in designing LTNs and performing public consultation, the Alan Turing Institute developed an LTN software tool, which can be run in a web browser at http://ltn.abstreet.org. To predict the overall impact on traffic patterns, the tool uses a travel demand model built from 2011 commute data. This was further developed in this use-case by using traffic volume data collected by the UO in Manchester in order to predict the overall change in traffic volumes due to proposed LTNs. It works as follows:

1. As input, the tool takes a travel demand model, representing a list of origins and destinations of driving trips taken during a typical weekday
2. For each trip, the tool calculates the route a driver would take before any LTNs were installed
3. A second route is calculated, after LTNs
4. The daily volume of vehicle traffic along each road segment and intersection is counted from the routes
5. A visualization shows the relative increase or decrease in volume along the street network

A planner can then look for streets with heavy increased volume and consider additional interventions – like more LTNs in nearby neighbourhoods, or bus lanes and improved crossings for major roads. An example output is shown in Figure 4.


Click a road or intersection to add or remove a modal filter

16 filters added

Neighborhood area: ~1.7 km²

☑ Expert mode

Draw traffic cells as areas ● streets

Figure 4: Model output showing the Levenshulme LTN highlighting overall changes in traffic flows from the demand model.

The results of the use-case indicate that the approach has major potential as a means to understand and enhance the impact of planning interventions. It also points towards priorities for further development. In particular, a mismatch was demonstrated between the near real-time data available from the observatory and the travel demand model (built on 2011 data). This is likely to be a recurring problem in extracting maximum value from the real-time data in high value modelling applications. To this end, there is potential to significantly move forward by exploring other base demand data, like the National Trip End Model.

5.3 Deployment of new sensors for monitoring decarbonisation

Cities account for ~80% of global carbon emissions and have begun to make ambitious commitments, including decarbonisation and net zero targets of multiple sectors through the implementation of large-scale interventions and responsible innovations. Transport is rightly targeted within the net zero ambitions with the introduction of zero emission vehicle mandates, and an increased focus on public and active transport. Unsurprisingly, regional and local methods for supporting net zero often intercept with air quality. Examples include the widespread roll out of Clean Air Zones, School Streets and LTN's which are targeting multiple wins of decarbonisation by encouraging vehicle fleet and behaviour change, improving air quality as well as the health and wellbeing of those living and working in these areas. A measurement and data driven approach with accurate baselines, specifically designed sensor networks and use of novel methodologies is essential to measure the success of such projects. This should include traffic volume and vehicle type information as well as accurate and well understood air quality measurements.

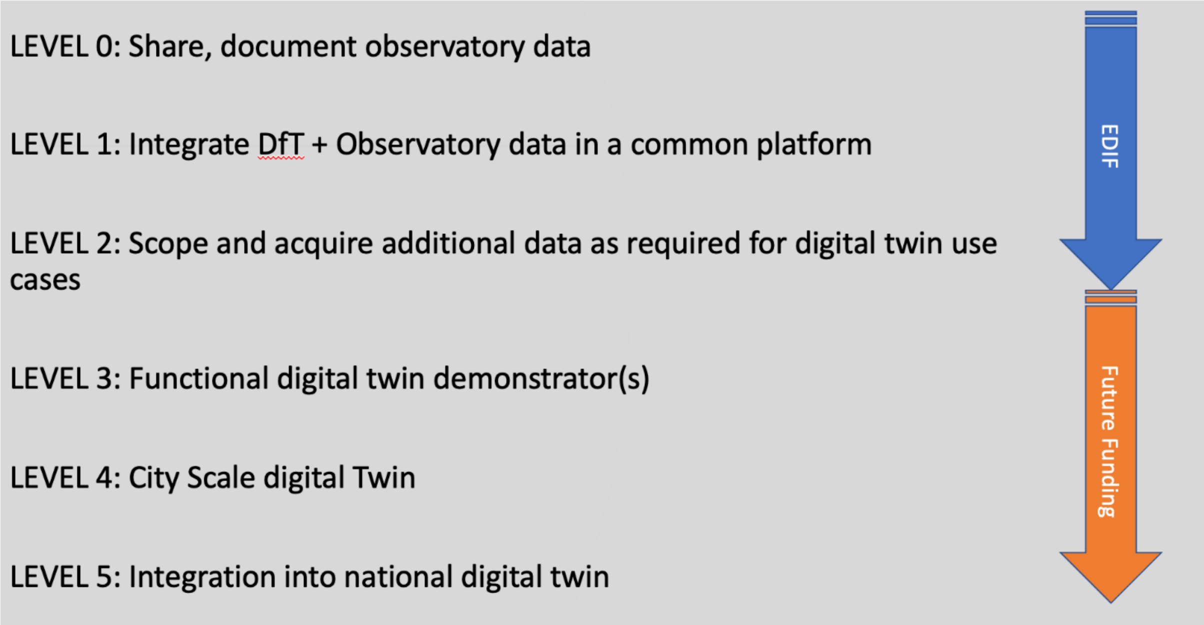
Particulates are a pollutant group of growing interest- historically foci have been led by gaseous emissions such as NO_2 and CO_2 when considering traffic intervention or fleet upgrade, yet as the links between climate change, air quality and public health have become more apparent so has the interest in particulates. An example of this is black carbon, a fraction of $PM_{2.5}$ which can have both major climate and health implications. Black carbon can directly contribute to positive radiative forcing (warming of the atmosphere) as it is able to absorb solar radiation and reduce surface albedo when deposited on ice/snow. Indirectly, it can also interact with clouds to increase warming. Black carbon is also linked to health issues such as respiratory and cardiovascular disease. It is therefore vital to understand its sources and measure its concentration, especially as its reduction could potentially be a faster responding mitigation strategy for climate change over greenhouse gases, given the shorter lifetimes of black carbon. Thus there has been a lot of community interest in black carbon as by the possibility of rapid managed decreases in the rate of climate change could be achieved by controls on the emissions of black carbon, delivering simultaneous beneficial effects on human health, particularly in source regions.

Real-time and continuous monitoring of black carbon are therefore very beneficial to understanding decarbonisation activities and the effectiveness of other air quality traffic interventions. Transport contributes significantly to BC concentrations- particularly diesel engines. Clean Air Zones (CAZ) are either planned or already implemented across all three cities (encouraging uptake of a greener/newer vehicle fleets, especially for diesel vehicles) and other air quality management such as Low Traffic Neighbourhoods (LTNs) and School Streets are altering mobility patterns for citizens. This data then has the potential to reveal insights between mobility changes and air quality (with potential to relate to policy) and subsequently the potential short-long term health and wellbeing response (as already demonstrated in the Levenshulme LTN for NO_x where real-time monitoring has already taken place).

Methods of monitoring black carbon typically rely on light transmission data from an aethalometer. The Manchester Urban Observatory has previously trialled two commercial off-the-shelf sensors for the measurement of black carbon; microAeth MA350 (ambient monitoring) & Ae-51 Aethalometer (personal exposure). Focusing on ambient monitoring to capture traffic interventions, microAeth MA350 were also procured for installations in the other UO's allowing the UO data platform to capture black carbon activity within areas of interest across the 3 major conurbations of Manchester, Birmingham and Newcastle. To date, black carbon monitoring is very limited and it is anticipated that increased observational capacity in this area will bring novel data surrounding policy interventions for both air quality and decarbonisation along with associated public health decisions. Deployment will not only capture the black carbon concentrations in areas of intervention, but also provide regional baseline concentrations ahead of future proposed changes.

6. The Blueprint

Although the project was initially designed to promote greater data sharing between the UO's and DfT, the long term vision of all parties was to use the project as a starting point of a much bigger journey towards producing a digital twin for Urban Transport. The original vision centred around a series of levels which would provide a platform for future spin-off projects:



LEVEL 0: Share, document observatory data

LEVEL 1: Integrate DfT + Observatory data in a common platform

LEVEL 2: Scope and acquire additional data as required for digital twin use cases

LEVEL 3: Functional digital twin demonstrator(s)

LEVEL 4: City Scale digital Twin

LEVEL 5: Integration into national digital twin

The original project funding has ultimately resulted in the level 0 to 2 goals being met in terms of data collection for a series of use cases. Indeed, as demonstrated, the work has gone beyond this and has already started to include analytics / modelling which would underpin future digital twins. This document will now articulate how the next levels could be realised, beginning with a summary of the key partners needed to move forward. Some of these have been onboarded during the project, whereas others respond to the interest expressed by attendees at the three show and tell events organised on the project.

6.2 Key Project Partners moving forward

The Alan Turing Institute: The Alan Turing Institute is the national institute for Data Science and Artificial Intelligence. The institute was onboarded as part of the project and excellent links are now in place with the urban analytics research programme. All three of the UO leads are currently also Turing Fellows.

Centre for Digital Built Britain: Historically based at the University of Cambridge, CDBB is a BEIS funded centre to improve understanding of how 'digital' can lead to better design and operation of the built environment. Of particular relevance to the blueprint is the DT Hub which has been run out of CDBB since March 2020 as part of the National Digital Twin programme. The DT Hub has grown rapidly and now has members from over 1600 organisations and plays a crucial role in the larger connected DT landscape. Although CDBB were not involved in the original project, there is still scope to tap into their vast expertise on future funding bids as part of a bigger consortium. Indeed, members of the original EDIF programme are well networked and have contributed to key publications, such as the Gemini papers.

Connected Places Catapult: The Connected Places Catapult (CPC) has played an integral role in this project (i.e. by ensuring alignment with other CPC projects such as the Active Travel). Moving forward, this link will become further consolidated as the DT Hub transitions from CDBB to the CPC (with effect from the 1st April 2022). This move was motivated to build upon the foundational work of CDBB, so that the hub can play a wider role in projects across the public and private sectors, reaching a broader set of people and sectors. To this end, the CPC, along with DT Hub, will play a pivotal role in taking this blueprint forward.

Regional Transport Bodies: The three UO's have established links to their respective regional transport bodies. Although these were not integrated in the original bid, further engagement was initiated via their attendance at *show and tell* events from the project. It is proposed that a natural progression would be to ensure their direct involvement in follow-up activities related to the blueprint.

Other Stakeholders. The three *show and tell* events on this project were attended by 72 individuals. A summary document is provided in Appendix 4. The future involvement of attendees, and the organisations they represent, in the blueprint is welcomed in an advisory board capacity at the very least.

6.3 Next steps

LEVEL 3 - Short Term:

It is imperative that momentum is maintained in the short term and further funding will be essential to achieve this, even to just maintain the catalogue of accessible live streaming data.

Funding will be sought to expand the consortium to include all partners identified in the previous section. These partners will work together to elevate and mainstream use-cases to a small scale / next stage demonstrator working DT. This will provide the means to stress-test the underpinning work completed on this project (e.g. ontologies, metadata etc), engaging with the broader community to promote data models and to define / adopt lightweight MIM and metadata standards.

An excellent example of a 'level 3 demonstrator' is the Climate Resilience Demonstrator, CReDo led out of CDBB. This was designed to improve resilience by developing a digital twin across energy, water and telecoms networks. Although not strictly a DT, this project actually represented a use-case (i.e. proof of concept) based on urban flooding. However, it was pioneering in the sense that it showed how connected-data and information sharing could be used to improve network performance across sector boundaries. As per the original project, it was also designed to be scalable at a national level. Potential use cases that have been discussed on this project which could be suitable for elevation to this level will include:

1. Decarbonising transport
2. Multimodal network management
3. Mapping best locations for active travel hubs
4. Passive methods to identify the mode and purpose of a journey to supersede periodic census data
5. Emissions data to bus shelters and transport hubs
6. Clustering of people to inform decisions about the location of transport hubs
7. Contextualising the cost of transport modes
8. The impact of congestion charges on behaviour, capturing intended and unintended consequences across traffic / public transport / active travel modes
9. Multimodal transport network management predicting incident analytics, real-time notifications to allow targeted intervention
10. Comparing classic sensor data to 'reused' data from non-classic sensors
11. Out-of-vehicle trip and travel data.
12. Post pandemic travel demand estimation

The Economic Data Innovation Fund has been identified as the most viable funding route to continue this activity and a follow-up bid has subsequently been prepared by the consortium. The project is designed to deliver level 3 of the Blueprint and underpin level 4. It will develop and deliver a subset of the identified use cases to deliver solutions using sensor data for national and local policies on decarbonisation and connectivity/levelling up. This will provide a rich real time dataset for users at national/local level to monitor carbon usage, traffic levels, connectivity, the impact of local and national schemes which at present cannot be easily accessed or used at scale. The proposed project will also build upon the prototype showcased on this project to provide a national sensor network, available on one open platform, integrated into the Digital Twin programme now led by the Connected Places Catapult, and available through the ONS Integrated Data Service.

LEVEL 4: Medium Term

The medium term is about scalability. Fundamentally, there is a need to expand the number of use cases to provide a genuine city-scale DT for Transport. Building on level 3, scaling is also required geographically so that even more cities can replicate the approach, engaging with metadata standards / data platforms and ultimately leading to a national scale digital twin for transport. In this model, use cases become exemplars which regional and national stakeholders can pilot. The consortium will need to lead the consultation on financial and economic models for funding of real-time sensing and associated data curation activities and to engage with regulatory and standards bodies to define common Working Practices. Once the project reaches this level, there is a need to initiate discussions regarding:

1. Who owns the data?
2. Who looks after the data?
3. Who pays for curation?
4. How is the data stored (real time vs archive) - this depends on the nature of the DT
5. Levels of service (open / private / mix)

This would represent a significant body of work. Although the work could be subdivided into two separate projects (i.e. city scale DT and national scale-up), both will require significant funding. At the moment, the only suitable potential source of funding identified for this activity would be the Shared Outcomes Fund.

LEVEL 5: Long Term

Integration of the transport component into the national digital twin where the transport module provides just one of a broader interconnected ecosystem of digital twins. This vision is distant and is the long-established aim of CDBB and the National Digital Twin programme. By evolving the work to level 4, transport will be ready to pull its weight in this space in relation to other sectors. However, it is important to note that foundational cross-sector activities can also be implemented in the short term as has been evidenced by CReDo.

7. Conclusions

The underlying principle of this project remains paramount: transport data that is easier to find, easier to use, easier to share. All parties involved in the project now have improved awareness of the data capabilities of each institution and a greater appreciation of available metadata and data access protocols. The *show and tell* events have ensured that this awareness has been broadened considerably beyond the core team which will be essential as the work progresses beyond the current project.

Arguably the most important component of the work has been the integration and visualisation prototype. This has demonstrated the ease of which how data, regardless of source, can be readily harmonised and made more accessible. This becomes increasingly critical as the concepts begin to scale to include other cities or data sources which could become increasingly disparate, especially under the current situation of only nascent metadata standards.

The use-cases provide an opportunity to continue the work towards a DT and have the potential to combine and expand to provide short-term funding but also underpin funding related to longer term objectives. As the blueprint has illustrated, scale-up to different levels requires considerably more funding then the foundational steps so far achieved. However, this is true across all sectors meaning there remains potential for transport to not just pull its weight, but lead in this area, drawing in other sectors to unlock bigger investment in time.