DOI: 10.3217/978-3-99161-057-1-199

Increasing the value of bridge SHM data by leveraging network-level open data

Paul R.A. Fidler[1, 0000-0003-4594-4323], Sam Cocking[1, 0000-0003-2782-3464], Farhad Huseynov[1, 0000-0002-5927-2444],
Miguel Bravo Haro[2, 0000-0003-0757-777X], Pedro Ubeda Luengo[2, 0009-0000-4684-3511],
Campbell R. Middleton[1, 0000-0002-9672-0680], Jennifer M. Schooling[3, 0000-0002-4777-0438]
[1]University of Cambridge, Civil Engineering Building, 7a JJ Thomson Avenue, Cambridge CB3 0FA, United Kingdom
[2]City, University of London, Northampton Square, London EC1V 0HB, United Kingdom
[3]Anglia Ruskin University, East Road, Cambridge CB1 1PT, United Kingdom
email: praf1@cam.ac.uk, sc740@cam.ac.uk, fh392@cam.ac.uk, miguel.bravo-haro@city.ac.uk,
pedro.ubeda-luengo@city.ac.uk, crm11@cam.ac.uk, jennifer.schooling@aru.ac.uk

ABSTRACT: Installing and maintaining structural health monitoring (SHM) systems on infrastructure assets can be expensive. These systems may produce large volumes of data that require processing and interpretation before the behaviour of the asset can be understood and assessed. However, in-depth understanding typically also requires knowledge of asset construction details and loading patterns. These data may be produced and stored using disparate systems, databases, and file types, creating additional challenges for data fusion and interoperability.

Additionally, there has been an increasing trend towards public bodies providing access to their data either reactively because of freedom of information requests, or proactively to encourage use by researchers or to allow others to provide innovative products or services using the data in ways not anticipated by those generating and providing them.

This paper presents potential strategies to leverage publicly available data from sources such as Network Rail Open Data Feeds, Rail Data Marketplace, OpenRail Data, OpenStreetMap and others, to contextualise and increase the value of SHM data. Data are considered from four instrumented railway bridges in the U.K., each of varying steel, concrete, and masonry construction. This paper presents scenarios by which these data might be used to gain network-level insights into other structures on the network and discusses the current difficulties in achieving this in practice.

KEY WORDS: Open data, Structural Health Monitoring, Data formats, Digital Twins.

1 INTRODUCTION

It is becoming increasingly common for structural health monitoring systems (SHM) to be deployed to monitor key transport infrastructure such as bridges. Interpretation of data generated by these systems is often challenging and may require additional information. For example, to understand the strains or deflection of a bridge deck it is also necessary to understand the applied loading, including loading from passing vehicles, or from environmental factors such as wind, or temperature. This requirement usually results in additional sensors being specified for the monitoring system.

For road bridges, detecting traffic can be done using lane occupancy sensors, cameras, or weigh-in-motion strips. As road traffic is usually completely unscheduled, there is no information known a priori about the traffic crossing the bridge. Interpreting these data can be challenging and it may be tempting to consider using machine learning (ML) techniques or computer vision to identify the type and position of vehicles. In some cases, such as in Bridge Weigh-In-Motion (B-WIM) systems where the primary purpose of the monitoring system is to weigh the road traffic [1], there may be sufficient sensor coverage to infer vehicle type directly from axle loads. However, in general vehicle identification may not be straightforward.

This need not however be the case for most railways. Railway operations, timetables, and signalling are increasingly digitised. The railway is a known environment, at least as far as those responsible for operating the railway are concerned. These data are already used to provide information to passengers, e.g. through passenger information screens at stations. The data are also available to third-party developers of smartphone apps and websites. Leveraging data from railway timetables and signalling systems as an additional source of information for a bridge structural health monitoring system may mitigate the need for some sensors on the assets themselves.

1.1 Background

In 2015 researchers at the Centre for Smart Infrastructure and Construction (CSIC) at the University of Cambridge installed fibre-Bragg grating (FBG) strain and temperature gauges on two new railway bridges during construction: Bridges IB5 and UB11 in Staffordshire, UK.

Figure 1: Bridge IB5, Staffordshire, U.K.

Bridge IB5, shown in Figure 1, is of steel beam construction with an in situ concrete deck carrying two railway tracks [2],

DOI: 10.3217/978-3-99161-057-1-199

while Bridge UB11, shown in Figure 2, is constructed from nine pre-tensioned concrete beams with an in situ concrete deck.

The initial goal of the project was to evaluate the potential benefits of installing instrumentation during construction and to create 'self-sensing' bridges as technology demonstrators for fibre-optic sensors and for Structural Health Monitoring in general. These were newly constructed bridges and so there was no concern regarding structural integrity, although the fibre-optic sensors were used to investigate creep and shrinkage of the concrete from pre-tensioning, through installation and after commissioning [3].

Figure 2. Bridge UB11, Staffordshire, U.K.

Initially the two Staffordshire bridges did not have permanently installed fibre-optic analysers or data loggers as there was no permanent power supply available on site. However, a power supply was provided in 2021, and permanent monitoring systems were installed, with IB5 upgraded with additional accelerometers, cameras and laser-based axle sensors [4]. These accelerometers and axle sensors were added to augment the existing FBG-based strain instrumentation to create a bridge weigh-in-motion (B-WIM) system [5].

Figure 3. Bridge HDB-19, London, U.K.

Subsequently, the centre also instrumented several other railway structures in the U.K. including Victorian and Edwardian masonry arch bridges, both of which were monitored due to potential concerns with the structures: Bridge HDB-19 in London – a three-span bridge instrumented with FBG strain and temperature sensors, and acoustic emission sensors; and CFM-5 in Yorkshire – a single span bridge instrumented with FBG strain sensors alongside conventional strain and displacement sensors and videogrammetry monitoring [6,7].

Figure 4. Bridge CFM-5, Yorkshire, U.K.

The monitoring systems on the two masonry arch bridges both used solar power. Despite using large deep-cycle batteries and multiple solar panels, these systems do not function 24 hours a day – the systems thus miss the structural response of the bridges for most train crossings.

1.2 Automated Train Identification

To interpret strain, deflection or accelerometer data measured during a crossing of a train over a bridge it is usually necessary to know information about the type of train, including axle loads and spacings. Trains of similar types are likely to produce similar responses, whereas trains of differing types may result in responses that are more difficult to compare.

Various techniques have been used to attempt train identification automatically. Alexakis *et al.* [8] limited their analyses of trains crossing the Marsh Lane viaduct, a masonry arch structure in Leeds, U.K., to only one type of train – the Class 185 three-car diesel multiple unit (DMU). Peak detection was used to identify train bogies and thereby determine which trains consisted of three carriages, while a comparison of readings from two adjacent arches was used to determine the train direction. The purpose of the monitoring was to evaluate whether the condition of the structure was deteriorating over time. Using a single train type allowed a comparison of the structural response from similar loading conditions on different dates. The trains identified as Class 185 trains represented approximately 50% of the traffic crossing the viaduct. Later analyses [9] identified other types of three-car train and Statistical Shape Analysis and a Support Vector Machine (SVM) was used to further classify these trains into Class 185, Class 155/158 and Class 170 respectively. The results were checked by visual observations of passing trains.

Cheng *et al.* [10] used gradient-based decision trees to identify and classify trains crossing Bridge CFM-5. A subset (4,900 out of 7,100) of train crossings identified using FBG strain data from July 2020 to October 2021 was labelled using timetable data obtained 'by scraping publicly available records'. These labels were used to train a model using XGBoost [11] to classify trains based on features in the strain data such as number and spacing of peaks, amplitude and width of peaks etc. This model was then able to classify 930 further train crossing events over an 8-month period in 2023 achieving a classification accuracy of 97%.

To classify trains crossing Bridge IB5 where axles spacing and loading are provided via the B-WIM system, the authors used t-SNE (t-Distributed Stochastic Neighbour Embedding) – a statistical technique to group crossing trains into distinct

DOI: 10.3217/978-3-99161-057-1-199

groups by reducing the dimensionality of input data. Input data include the number of axles, axle spacing, speed, axle loads etc. t-SNE is used to reduce these to just two dimensions which can then be plotted.

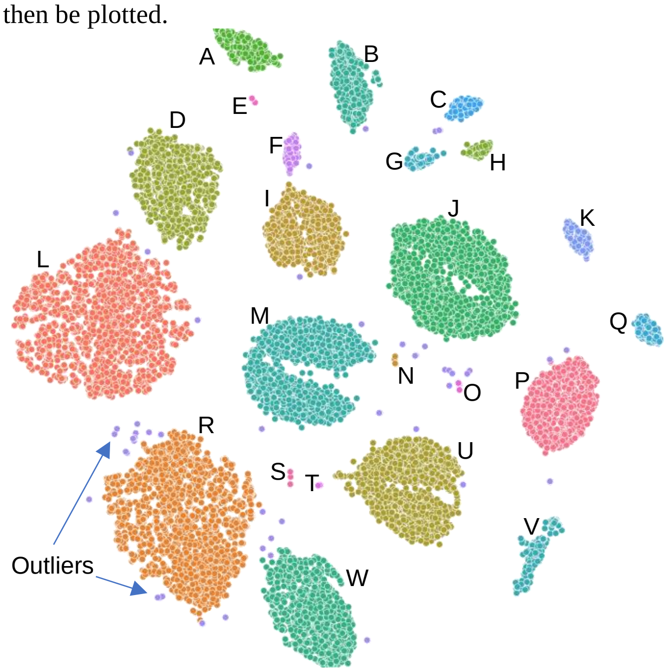

Figure 5. Groups of trains identified by t-SNE and DBSCAN unsupervised train classification

A clustering algorithm such as the Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise (DBSCAN) [13] is then used to distinguish separate groups of trains appearing as clusters within the plot.

Table 1. Labels assigned to groups shown in Figure 5 by manual inspection of video from on-bridge cameras

Label	Description	%
A	Class 390 9 car	2.28
B	Class 390 11 car	3.36
C	2 x Class 221 5 car (10 car total)	1.08
D	Class 350 175 tonne Direction 1	6.64
E	New Measurement Train	0.09
F	Class 220 4 car + Class 221 4 car (8 cat total)	0.95
G	Class 221 4 car + Class 221 5 car (9 car total)	0.86
H	Class 221 5 car + Class 221 4 car (9 car total)	0.72
I	Class 220 4 car + Class 221 5 car (9 car total)	5.68
J	Class 221 5 car	11.27
K	Class 221 4 car	1.20
L	Class 350 165 tonne Direction 1	15.46
M	Class 220 4 car	9.85
N	Tamping Machine	0.18
O	Freight Locomotive	0.21
P	Class 221 5 car + Class 220 4 car (9 car total)	5.79
Q	Class 221 4 car + Class 220 4 car (8 car total)	1.14
R	Class 350 165 tonne Direction 2	14.41
S	Class 350 (axle detectors missed one axle)	0.13
T	2 x Class 350 (8 car total)	0.08
U	2 x Class 220 4 car (8 car total)	8.86
V	Freight Train	1.92
W	Class 350 175 tonne Direction 2	7.00
	Outliers	0.84

Figure 5 shows the plot resulting from 18,800 data points each representing a train. Only northbound trains are included due to limitations of the B-WIM system.

By referencing plots of the axle loads and inspecting video recordings from the cameras positioned on the bridge, labels may be manually assigned to these groups.

Some types of train such as the Class 220 Voyager and Class 221 Super Voyager look similar but may be distinguished using axle weights. Class 221 trains are heavier as they include tilting bogie mechanisms, absent on the Class 220. Similarly, Class 350 Desiro trains appear in four distinct groups on the plot. This is because there are two types within the class with differing total weights. This is likely because early Class 350 trains have dual-voltage capability and are able to use either the third rail system or overhead line equipment, while later Class 350s lack the third-rail pickup [14]. The axle loads are also asymmetric front to back relative to the direction of the train, which leads to two further groups.

Once labelled, the data can then be used as training data for ML techniques such as Random Forest [15] or XGBoost which can then be used to classify future trains.

Other methods that may be used for classifying trains on Bridge IB5 utilise computer vision (CV) and video from the on-site cameras directly. One such CV technique uses a vertical strip of pixels from each frame of a video of a passing train to produce a single 2D image where the x-axis represents time. Examples of such images are shown in Figure 6. A manually labelled set of these images is then used to train a model from a partially pretrained convolutional neural network (CNN) with Keras [16] as the deep learning framework. A dataset of 543 trains was randomly selected, of which 462 were used for image classification, split 85%/15% between training and testing images, respectively.

Figure 6. Examples of (a) Double 4-car Class 220/221, (b) 9-car Class 390, (c) 4-car Class 350, and (d) 5-car Class 221 trains approaching Bridge IB5. The still images are converted from moving video, with time along the x-axis

The results obtained are presented in Table 2 using a confusion matrix. These show classification accuracies of around 90% on average across all train classes. The CNN-based model achieved high classification accuracy despite the relatively small dataset of 543 samples, demonstrating the merit of the approach. However, it also highlights limitations which may not necessarily be solved by increasing the size of the training dataset.

DOI: 10.3217/978-3-99161-057-1-199

Table 2. Confusion matrix of CNN-based train classification algorithm.

Actual \ Predicted: Train Class	350	220/1 4-car	220/1 5-car	2×220/1 4-car	390
350	96%	4%			
220/1 4-car	6%	88%	6%		
220/1 5-car		22%	78%		
2×220/1 4-car				83%	17%
390					100%

In particular, the model performs best when distinguishing visually distinct train types, but struggles with more subtle variations, such as between 4- and 5-car variants of Class 220/221 trains. These misclassifications suggest that incorporating additional features beyond image data, such as train speed or axle spacing, could significantly enhance model performance. Future work will explore hybrid models that combine visual and sensor-based inputs, as well as alternative machine learning techniques such as recurrent neural networks (RNNs) for temporal data or multimodal architectures that can process both image and numerical inputs simultaneously.

Computer vision techniques are however only applicable on SHM systems that incorporate cameras, such as the installation at Bridge IB5. It is also vulnerable to issues caused by poor lighting conditions such as at night or caused by inclement weather conditions leading to water or ice on the lens.

2 PUBLICALLY AVAILABLE RAIL DATA SOURCES

2.1 Network Rail

In the U.K. Network Rail is the organisation responsible for maintaining the track, signalling, most stations and operation of the railway in England, Scotland and Wales. It is not responsible for running train services, which are currently run by passenger and freight train operating companies (TOCs).

Network Rail provides access to some of its operational data including dynamic data on signalling, train movement data, and real-time performance measures, along with static data such as scheduling data and background data needed to interpret these datasets. Accessing the data requires registering an account on the Network Rail Open Data Platform [17] which is free of charge. The dynamic data are streamed via an ActiveMQ message queue connection, which requires a constant connection. Static data is available to download daily or monthly. These datasets have been used by mobile app developers to provide real-time information to passengers, such as the platform from which their train will depart, or whether their train is running late. The available datasets include:

- TD (Train describer): This is a real-time feed of train movements between signalling 'berths'. The signalling ID (or 'headcode') for each train is given, along with a 'to' and 'from' berth number representing a train movement. The signalling ID is only unique within a given signalling region at any one time. Berth numbers are not unique either. Timestamps are to the nearest second.
- TRUST (Train movement): This is another real-time feed of train movements between timing point locations (TIPLOCs), usually stations and junctions. Different message types describe train activation, movement, cancellation. Timestamps are provided to the nearest 30 seconds.
- SCHEDULE and VSTP: These provide details of services that are due to run. The schedule is updated once per day. Each service in the schedule can either be a one-off service, or be valid for a number of days, weeks or months. The VSTP dataset (Very Short-Term Plan) is a real-time feed of additional one-off services for ad-hoc movements not in the main schedule.

Documentation for the feeds is available on a wiki-style website [18] maintained by enthusiasts. Example source code of ActiveMQ clients able to fetch the data feeds is also available in multiple programming languages on GitHub [19].

Historical train movement data are not available from the Network Rail Open Data Feed. This limitation is not generally an issue for app developers but does limit what is available for interpreting past data from monitoring systems.

2.2 Rail Delivery Group

The Rail Delivery Group (formerly the Association of Train Operating Companies) in the U.K. provides additional feeds and APIs collectively known as DARWIN, which offer data for live departure and arrival screens, including estimated arrival and departure times for delayed trains. Also provided is the Historical Service Performance (HSP) API for historic performance data. HSP can be used to query details of past services, such as planned and actual arrival and departure times. However, as the DARWIN and HSP datasets are primarily used for passenger information they do not contain information about freight trains, and only list arrival and departure at stations, not showing the times when trains pass junctions.

2.3 Rail Data Marketplace

The Rail Delivery Group also runs the Rail Data Marketplace. This is a platform on which train operating companies, infrastructure providers, data aggregators, researchers, or rail enthusiasts may release datasets. Datasets include the Network Rail and Rail Delivery Group feeds described above, but also data on train operators' carbon footprints, car park occupancy, train accessibility, fare information, occupancy and loading, complaints etc. Data providers may specify either an open or restricted licence for the data, and have the option of making data available publicly, or only to subscribers. The data may be made available free of charge or require payment.

2.4 OpenRail Data

The OpenRail Data website [24] combines data from the Network Rail SCHEDULE, VSTP and TRUST feeds to provide details of train movements, including cancellations, late/early running arrivals and some details of the type and class of train. Up to three years of historical data may be queried

DOI: 10.3217/978-3-99161-057-1-199

using a variety of web-based forms, with results returned as an HTML page. The source code is available from a GitHub repository, and so it would be possible to run customised instances of this service, modified to produce output in alternative formats such as JSON or XML. However, this would not include historic data.

3 USING OPEN DATA FOR TRAIN IDENTIFICATION

The TD dataset gives timestamps for when trains on the network move between signalling berths. Berths are labelled with four-character identifiers, usually consisting of three or four digits. These do not indicate any human-readable place names, and there does not appear to be any available data linking berths with their geographic coordinates. It is however possible to look at all train movements that occur within a time window either side of a train crossing event as inferred from measurements by a monitoring system. This will result in many candidate berth numbers. Eventually however, after observing enough train crossings it is possible to narrow the berth numbers common to all crossings to find those berths that are likely to be located on either side of the bridge. Figure 7 illustrates one such possible method.

Using data from a typical day:

- Create set U of all timestamped train movements from TD data
- Create a list X of timestamped bridge crossing events inferred from SHM data
- Create initially empty set C of candidate TD berth transitions
- Identify train movements in U that coincide (within a tolerance) of the first timestamped crossing event in X and add these movements (without timestamps) to set C
- For each further crossing event E in X:
    - Identify train movements in U that coincide (within a tolerance) with E and intersect these movements (without timestamps) with set C
- Output candidate TD berth transitions from set C

Figure 7. Possible algorithm to deduce TD berth transitions corresponding to train crossings identified by an SHM system

Alternatively, third-party websites such as OpenTrainTimes [23] provide topological track diagrams showing the real-time position of trains using TD data.

By using both TD and also TRUST data (which does include geographical human-readable locations) the creators of these sites have been able to infer the position of berths relative to stations and junctions. Although these track diagrams do not show bridge locations (except for intersection bridges where one railway track crosses another) by using these track diagrams it is nevertheless possible to narrow the search when attempting to identify TD berths located either side of a bridge.

Table 3 shows the berth transitions that most closely correspond in time to observed train crossings for the four instrumented bridges. For Bridges IB5 and UB11, these could be confirmed using the installed bridge monitoring system, while for Bridge HDB-19 these times were verified using historical data as displayed on the OpenRail Data website [24] and comparing the actual arrival times of trains travelling between Enfield Chase station and Gordon Hill station, the two stations either side of the bridge which are not far apart. There are however points (or switches) north of HDB-19 allowing trains to cross the tracks to access the third platform at Gordon Hill station. Trains heading to or from the third platform have a slightly different berth number transition in the TD feed. For Bridge CFM-5 the crossing times of a few trains were observed and recorded in person during a site visit.

The TD data provides a very good correspondence with bridge crossings for Bridges IB5 and UB11. However, the only useful information provided by the feed besides the timestamp is the signalling ID, known as the headcode, of each train. This provides no train information such as the type of train or possible loading. This must be found either from the TRUST feed or via a third-party website such as OpenRail Data.

Since most of the data from the monitoring systems on these bridges dates from before the authors began logging the TD and TRUST feeds, the possibility of using historical data from OpenRail Data alone was investigated. For Bridges CFM-5 and HDB-19, this turns out to be relatively simple. For Bridge CFM-5 there are no intervening junctions between Church Fenton and Micklefield stations – so any train that reports at both locations consecutively must have crossed the bridge. The only complication is in calculating the most likely time of the crossing as the two stations are approximately 8 km apart, with the bridge located slightly closer to Church Fenton station.

Table 3. TD berth movements that correspond with observed bridge crossings

Crossing	Region	From Berth	To Berth
IB5 northbound	R3	4331	4333
IB5 southbound	R3	4334	4332
UB11 westbound	R3	5611	5615
HDB-19 northbound	Y8	865	869
	Y8	865	X872
HDB-19 southbound	Y8	870	864
	Y8	872	864
CFM-5 eastbound	Y2	709	711
CFM-5 westbound	Y2	714	708

Likewise in the case of Bridge HDB-19, any train reporting at both Enfield Chase and Gordon Hill stations will have crossed the bridge. However, it was noted that in the archived data several services are not shown stopping or passing Enfield Chase station at all. Trains do however all appear to report at Bowes Park station, slight further to the south, so both stations were used when searching for trains crossing the bridge.

For Bridges IB5 and UB11 however, the situation is slightly more complicated. Figure 8 shows the layout of the railway in the Norton Bridge area. Northbound trains travelling between Stafford and Madeley stations either use the West Coast Main Line through Norton Bridge Junction and under Bridge IB5, or they bypass Norton Bridge Junction entirely by crossing Bridge UB11. Northbound trains travelling between Stafford and Stone must cross Bridge IB5, but southbound trains from Stone to Stafford may travel either via Bridge IB5 or via Norton Bridge Junction.

DOI: 10.3217/978-3-99161-057-1-199

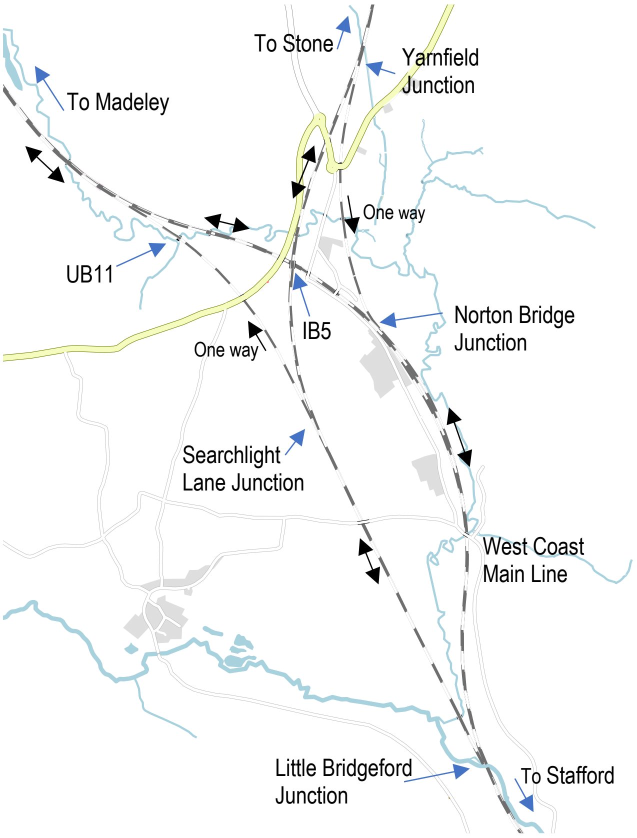

Figure 8. Map showing locations of Bridges IB5 and UB11, railway layout and junctions. (Map data © OpenStreetMap available under Open Database License)

Nevertheless, by inspecting details of several trains crossing the bridges, identifying their headcode ID from TD data, and then looking up details of each train on the OpenRail Data website, sets of rules were developed to identify trains crossing Bridge IB5 and UB11 using only historic data available from the OpenRail Data website.

3.1 Train and route information

Information available from the TRUST and Schedule feeds, or the historical data, includes data such as: The Train UID, the train operator code, signalling id (headcode), power, timing load, speed, catering code, seating class, train status etc. and the planned and actual arrival, departure and passing times at each timing point or station along the route.

Of these the power type, timing load and train status are the most useful for train identification. Power type refers to diesel, diesel multiple unit, electric, electric multiple unit or (for the New Measurement Train) 'HST'. Timing load is an overloaded field. It usually contains a number which if greater than 999 describes the declared load in tonnes. Otherwise, this number may indicate the declared load in tonnes, or the class of train – e.g. 350, 390 etc. There are some exceptions: a value of 506 indicates Class 350 upgraded to run at 110 mph (approx. 175 km/h) while a 'V' indicates a Class 220 or 221 train. The train status field indicates whether the train is a passenger or freight train. Any of the fields may be blank.

3.2 Comparison with Monitoring data

For the railway bridge monitoring systems that are operational 24/7, such as those on Bridges IB5 and UB11, there should be SHM data for each train that crossed the bridges, and there should also be trains identified from TD, TRUST, or historic train movement data corresponding to each event. Where this is not the case, this indicates a possible fault with the monitoring system. For Bridge IB5 it is also possible to assess the coverage of the B-WIM system.

As can be seen from Table 1 the t-SNE+DBSCAN algorithm results in many different groupings for otherwise similar trains or combinations of similar trains. However, the train class information derived from the timing load field in archived data from OpenRail Data does not distinguish between Class 220 and 221 trains and lists these as a single class. Similarly, these data do not distinguish between 9 or 11-car Class 390 trains, or the different types of Class 350.

To allow for easier comparison between the loads classified using t-SNE+DBSCAN and traffic identification based on the archived data, similar load classifications from the t-SNE+DBSCAN results for crossings in July 2023 were added together. The results are shown in Table 4.

The classification of trains using groups found by the t-SNE+DBSCAN algorithm shows generally good agreement with the identifications derived directly from open data. The archived data deduced ten more train crossings than trains processed by the t-SNE+DBSCAN algorithm. This discrepancy is partly accounted for by double-crossing events where two trains cross the bridge in opposite directions within a few seconds of each other. The SHM system treats these double crossings as a single event prior to any processing by the B-WIM system.

Table 4. Comparison of B-WIM t-SNE load classifications and traffic identified from historic open data for trains crossing Bridge IB5 in July 2023

Description	**B-WIM (t-SNE)**		**Archived data**	
Class 220/221	650	50.94%	658	50.04%
Class 350	401	31.43%	421	32.02%
Class 390	154	12.07%	139	10.57%
Other passenger			7	0.53%
Freight	52	4.06%	60	4.56%
Tamping machine	2	0.16%		
NMT	1	0.08%	1	0.08%
Outliers	16	1.25%		
Total	1276	100.00%	1286	100.00%

Both methods are able to distinguish the New Measurement Train (NMT) from other trains. This train is a modified Class 43 High Speed Train, formerly used to carry passengers, but now instrumented with sensors to measure track alignment and gauge, and photograph defects while travelling at line speed. In the t-SNE+DBSCAN data it appears as its own (small) group labelled as 'E' in Figure 5. In the archived train movement data it is the only train with a train status of 'freight' but with a power type of 'HST' and a speed of 125 mph (approx. 200 km/h).

The B-WIM classifier identified two tamping machines. One of these on 7th July 2023 was mis-classified and was actually a 16-axle Class 350 passenger train, as verified by inspecting video of the crossing. However, only 10 axles were detected by the B-WIM system, which may have resulted in the train being incorrectly grouped. The second tamping machine on the

DOI: 10.3217/978-3-99161-057-1-199

14th July as shown in Figure 9 was correctly identified from the B-WIM derived data. However, it was listed in the archived train movement data as a 715-tonne diesel freight train. The accompanying route information for the train showed that it originated at Whitacre Tamper Sidings, which would indicate that it was probably a tamping machine. The declared load of 715 tonnes is significantly larger than the load as shown on a placard on the side of the vehicle (99 tonnes) or the load as measured by the B-WIM system (102 tonnes).

Figure 9. Tamping vehicle heading after crossing Bridge IB5 on 14 July 2023

The archived train movement data are based on SCHEDULE and TRUST data. As previously discussed, these data alone are not able to distinguish between Class 220 and Class 221 trains.

Table 5. Comparison of trains deduced to have crossed Bridge HDB-19 and the FBG strain events recorded by the SHM system.

	TRUST data	FBG events	Proportion recorded
Passenger			
Class 387	35	2	5.7%
Class 700	11	0	0.0%
Class 717	1353	330	24.4%
Class 800/805	11	0	0.0%
Class 802	1	0	0.0%
Other electric	1	0	0.0%
Sub total	**1412**	**332**	**23.5%**
Freight			
400 tonnes	6	6	100.0%
600 tones	17	8	47.1%
715 tonnes	4	0	0.0%
800 tonnes	3	0	0.0%
1200 tonnes	19	2	10.5%
1235 tonnes	12	1	8.3%
1400 tonnes	9	7	77.8%
1600 tonnes	67	17	25.4%
Not declared	17	2	11.8%
Sub total	**154**	**43**	**27.9%**
Other			
NMT	2	1	50%
Total	**1567**	**369**	**23.5%**

However, other datasets available on the Rail Data Marketplace may solve this issue. CrossCountry Trains, which operates the Class 220 and 221 trains, makes the planned train formation of each train available. The data are available as daily CSV files and include serial numbers of the individual trainsets to be used for any given service. This is sufficient to determine the train class. Other train operating companies also make train formation data available.

For monitoring systems such as those used on Bridges HDB-19 and CFM-5 where power is supplied by solar panels and batteries, the monitoring system is likely to miss a significant fraction of train crossing events while the system is powered down. Using the data from TD or TRUST can provide an indication of which trains are missed, and whether the crossings that are recorded are likely to be indicative of the loads that typically cross the bridge.

FBG strain events from the SHM system on HDB-19 were compared with trains crossings from archived TRUST data for the period 1st–15th July 2023. When comparing timestamps between the SHM system and the archived TRUST data, it became apparent that the FBG data did not indicate whether the times had been recorded using UTC or daylights savings time. Since the trains mostly follow a repeating hourly timetable, most train crossings occurring at approximately the same times each hour. When the SHM system was initially installed there was no requirement for the system to synchronize with any external system or data. There was also nothing to prevent clock drift other than an intermittent Internet connection to time.microsoft.com using the NTP client available on Windows. However, after some investigation it appeared that the correlation between crossings and logged FBG events fit better with the archive TRUST data if it was assumed that the FBG timestamps were recorded using daylight savings time and not UTC. This illustrates the importance of looking ahead when specifying and commissioning SHM systems.

Table 5 shows the results of the comparison. It can be seen that approximately three quarters of the trains deduced to have crossed Bridge HDB-19 were not recorded by the FBG strain gauges. As the SHM system is solar powered, it only records data when the solar panels have charged the battery sufficiently for the system to operate, usually from mid-morning to mid-afternoon. Nevertheless, despite missing some classes of train completely, the system was able to record strain data during the crossings of a broadly representative sample of the total population of trains crossing Bridge HDB-19, including freight trains with the heaviest declared loads. However, if all that is required is a ‘standard’ train with which to compare data recorded on the structure from one day to the next to check whether the structural response is changing over time, then the Class 717 train would seem to be a good choice.

4 NETWORK-LEVEL INSIGHTS

The Network Rail data feeds cover the movement of trains throughout England, Scotland and Wales. Once subscribed to the feed, data is available for all train movements on the network, not just in the locations originally of interest. One potential use case of these feeds would therefore be to derive bridge specific traffic models (if not necessarily load models) for every underline bridge, intersection bridge or viaduct on the network. For those bridges located up or down the track from a

DOI: 10.3217/978-3-99161-057-1-199

bridge instrumented with a B-WIM system, it may even be possible to produce a bridge-specific load model, assuming the bridges have sufficient traffic in common. This however would require knowledge of the location of the bridges on the network, and ideally the locations of berth numbers either side of every bridge.

4.1 Bridge Locations

Through a series of Freedom of Information (FoI) requests made by members of the public, Network Rail have released lists of structures on each line of the network. Data from these FoI requests are available in a curated form on the RailwayData website [21]. The bridges listed do not typically have either WGS84 coordinates, or UK Ordnance Survey (OS) grid reference, but instead are described using the number of miles and yards (or sometimes miles and chains) from some datum which is specific to each line.

Recently however, because of work done within Network Rail as part of its Bridge Strike Prevention Strategy, Network Rail has released a list [22] of low bridges at risk of being struck by road vehicles. As the intended use for the data is that they are incorporated into in-vehicle or smartphone-based GPS systems, this list includes bridge headroom data together with WGS84 and OS grid references coordinates in addition to the usual line/miles/yards location. It provides such coordinates for 5792 bridges, but as it only lists vulnerable bridges it is not a complete list of all underline or intersection bridges, or viaducts. The list includes Bridges CFM-5 and HDB-19, both rail-over-road bridges. It does not however list Bridges IB5 or UB11 as the first is an intersection bridge carrying the railway over another rail line, while the second carries the railway over a stream.

Another option is to use data from OpenStreetMap. OpenStreetMap is a collaborative volunteer organisation that is building a database of mapping data that is released under the permissive Open Database License. However, as a volunteer effort, there are no guarantees as to the accuracy of the data, or that they are consistently labelled in a way that makes it possible to query specific features, such as railways and bridges carrying railways.

As an open project, there are services that build upon OpenStreetMap to allow additional functionality such as searching for specific features within the data using simple queries. Overpass Turbo [26] is one such service. It includes a 'wizard' to compose queries from simple prompts.

A prompt such as:

"((bridge=yes or bridge=viaduct) and railway=rail) in England"

will result in a query that can be passed to the Overpass API with the matching features displayed on a map, as shown in Figure 10.

```
// fetch area "England" to search in
area(id:3608484939)->.searchArea;
// gather results
(
nwr["bridge"="yes"]["railway"="rail"](area.searchArea);
nwr["bridge"="viaduct"]["railway"="rail"](area.searchArea);
);
// print results
out geom;
```

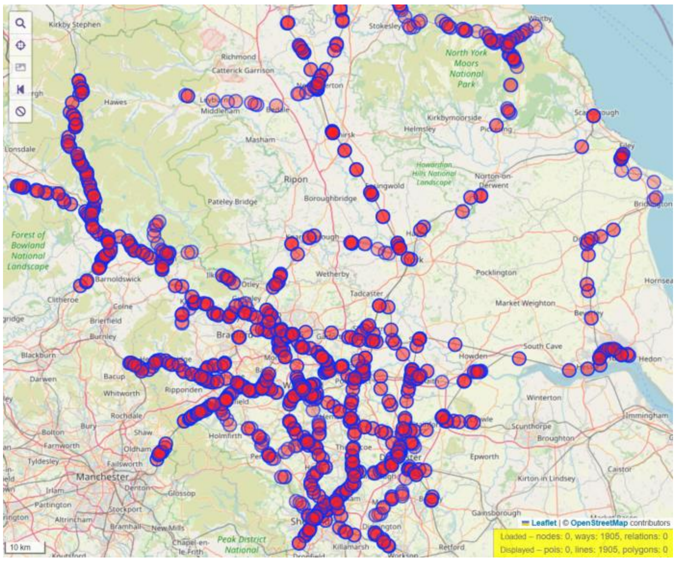

Figure 10. Result of running an Overpass API query to find locations of railway bridges and viaducts. (Map data © OpenStreetMap available under Open Database License)

4.2 Signalling berth locations

There are various ways one could consider obtaining the locations of the signalling berths. One option would be to use the GPS position of trains on the network and compare this with the live TD feed. This would give the approximate location of each berth. These could then be compared with the GPS coordinates of the bridges, where known. These data are not provided in any currently available Network Rail open dataset, but third-party app developers of 'Find My Train'-style apps may have GPS data generated by users of those apps as they travel by train. However, these data may not be open data. SignalBox.io have such a system but require users to sign up for an API key before accessing the data. Other companies such as Raildar, Tracksy.uk, Mistral-data and TrainPositions.com provide similar services under various subscription options.

Train operators in some other European countries do make real-time location data available. Irish Rail (Iarnród Éireann) provide this data via a simple URL, while in Finland train locations may be retrieved using a Real-Time General Transit Feed Specification (GTFS RT) feed or a via a web-based API.

Alternatively, as the location of each bridge is available from data in the FoI requests, albeit in line/mile/yards format, it may be more feasible to calculate TD berth positions in line/miles/yards format too. This could be done by looking at timestamps of TD berth transitions and comparing these with timestamps from TRUST data when trains report at locations with known positions, such as stations. Finding berths either side of a bridge could then be done simply by comparing berth miles and yards locations with bridge miles and yards locations.

5 RECOMMENDATIONS

- When planning to install a monitoring system, look for potentially useful open datasets early, whether related to traffic, weather, or anything else.
- Consider any secondary uses that the planned monitoring system may have that could be enabled with relatively small additions. (E.g. Added axle detectors to enable B-WIM that

DOI: 10.3217/978-3-99161-057-1-199

then provides loading data for other bridges on the network.)

- Treat streaming datasets like sensor data. Data that is used for operational purposes may have little long-term value to the network operator, and these data are often ephemeral. To avoid these data becoming digital waste, log them from day one (or before). Do not assume that somebody else will preserve them.
- Log data first, process later. If datasets are difficult to interpret either because of the sheer volume of information or due to a lack of documentation, log them anyway. They can be processed later once the data are better understood.
- Ensure clocks on any monitoring systems are set accurately. If a permanent connection to the Internet is available this is usually achieved using Network Time Protocol (NTP) to keep clocks synchronised, otherwise, if outdoors, a simple GPS receiver can be used to provide accurate time. When comparing data from multiple sources it is vital that timestamps of data logged by monitoring systems and data from one or more external datasets may be compared. Agree on a time zone.
- Where data from monitoring systems is combined with as-designed and as-built data to form a digital twin, any additional data derived from open datasets should also be incorporated into the digital twin.

6 CONCLUSION

A wealth of data exists that can be used to better understand data generated by bridge monitoring systems on Britain's railways. However, the data are unlikely to be generated in exactly the format needed by a monitoring project. They are instead created (and deleted) according to the needs of the network and train operators.

Nevertheless, once a network-level data source has been identified and its potential benefits and limitations understood, it may have the potential to be used for train identification or bridge-specific load or traffic modelling for multiple monitoring projects. Identifying and logging potential data sources early mitigates issues relating to data retention. Data should be logged as early as possible in the project even if the ability to understand them and use them effectively comes later. Machine Learning and AI remain valuable tools with which to understand data generated from monitoring systems, but sometimes there are simpler ways, requiring less computational resources, to achieve the same goal.

DATA AVAILABILITY

Data supporting this paper are available from the University of Cambridge Repository:
https://doi.org/10.17863/CAM.116750.

ACKNOWLEDGMENTS

The authors would like to thank Network Rail for making train movement and scheduling data available on their Open Data Feeds site, and for their continuing support of the Staffordshire Bridges and HDB-19 projects.

This work was made possible by the Centre for Smart Infrastructure and Construction (CSIC) and the Laing O'Rourke Centre for Construction Engineering Technology within the Department of Engineering at the University of Cambridge. We would like to thank Prof. Changsu Shim (심창수) and the Chung-Ang University Industry Academic Cooperation Foundation, South Korea for funding from the 'Development of Digital Engineering Services for Roadside Structure Design, Manufacturing and Construction' project, and thank Chiho Jeon (전치호) and Jongbin Won (원종빈) for their assistance and support. The support of the EPSRC UKCRIC National Research Facility for Infrastructure Sensing experimental facilities (EP/P013848/1) is also much appreciated.

The authors would also like to thank Peter Hicks for maintaining the Rail OpenData Wiki, Phil Wieland for creating and running the OpenRail Data project and the developers of OpenTrainTimes Live Track diagrams website, and the many contributors to OpenStreetMap project.

For the purpose of open access, the authors have applied a creative commons attribution (CC BY) license to any author accepted manuscript version arising.

REFERENCES

[1] M. Lydon, S.E. Taylor, D. Robinson, A. Mufti and E.J.O. Brien. Recent developments in bridge weigh in motion (B-WIM). Journal of Civil and Structural Health Monitoring (2016) 6:69-81. 2016. DOI: 10.1007/s13349-015-0119-6

[2] L.J. Butler W. Lin, J. Xu, N. Gibbons, M.Z.E.B. Elshafie and C.R. Middleton. Monitoring, modelling and assessment of a self-sensing railway bridge during construction. Journal of Bridge Engineering, ASCE, 23(10). 2018. DOI: 10.1061/(ASCE)BE.1943-5592. 0001288

[3] L.J. Butler, N. Gibbons, P. He, C.R. Middleton and M.Z.E.B. Elshafie. Evaluating the early-age behaviour of full-scale prestressed concrete beams using distributed and discrete fibre optic sensors. Construction and Building Materials. 126 (2016) pp. 894–912. 2016. DOI: 10.1016/ j.conbuildmat.2016.09.086

[4] P.R.A. Fidler, F. Huseynov, M. Bravo-Haro, V. Vilde, J.M. Schooling, C.R. Middleton, Augmenting an existing railway bridge monitoring system with additional sensors to create a bridge weigh-in-motion system and digital twin. Presented at SHMII-11: 11th International Conference on Structural Health Monitoring of Intelligent Infrastructure August 8-12, 2022, Montreal, QC, Canada. 2022.

[5] F. Huseynov, P.R.A. Fidler, M. Bravo-Haro, V. Vilde, J.M. Schooling, C.R. Middleton. Setting up a real-time train load monitoring system in the UK using Bridge Weigh-In Motion technology - A case study. Presented at SHMII-11: 11th International Conference on Structural Health Monitoring of Intelligent Infrastructure August 8-12, 2022, Montreal, QC, Canada. 2022.

[6] S. Cocking, D. Thompson, and M.J. DeJong. Comparative evaluation of monitoring technologies for a historic skewed masonry arch railway bridge. In: Proceedings of ARCH 2019, the 9th International Conference on Arch Bridges, Porto, Portugal, 2019. DOI: 10.1007/978-3-030-29227-0_46

[7] S. Cocking, H. Alexakis, and M.J. DeJong. Distributed dynamic fibre-optic strain monitoring of the behaviour of a skewed masonry arch railway bridge. Journal of Civil Structural Health Monitoring, 11(4), 989 1012, 2021. DOI: 10.1007/s13349-021-00493-w

[8] H. Alexakis, A. Franza, S. Acikgoz, and M. DeJong. Structural health monitoring of a masonry viaduct with Fibre Bragg Grating sensors. In IABSE Symposium, Guimaraes 2019: Towards a Resilient Built Environment Risk and Asset Management - pp. 1560-1567. 2019. DOI: 10.17863/CAM.38693

[9] H. Alexakis, F.DH. Lau & M.J. DeJong. Fibre optic sensing of ageing railway infrastructure enhanced with statistical shape analysis. Journal of Civil and Structural Health Monitoring 11, 49–67 2021. DOI: 10.1007/s13349-020-00437-w

[10] J. Cheng, A. Chen, M.J. DeJong, S. Cocking, Data-derived train classification and monitoring of masonry bridge, Bridge Maintenance, Safety, Management, Digitalization and Sustainability, p.3371-3379, CRC Press, 2024. DOI: 10.1201/9781003483755-399

[11] T. Chen, T. and C. Guestrin.. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International

DOI: 10.3217/978-3-99161-057-1-199

Conference on Knowledge Discovery and Data Mining (pp. 785–794). New York, NY, USA: ACM. 2016. DOI: 10.1145/2939672.2939785

[12] G. Hinton. Visualizing Data using t-SNE. Journal of Machine Learning Research 9. (2008) 2579-2605. 2008.

[13] M. Ester, HP. Kriegel, J. Sander and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231. ISBN 1-57735-004-9. 1996.

[14] British Rail Class 350. Wikimedia Foundation, Inc. Accessed 29-March-2025. https://en.wikipedia.org/wiki/British_Rail_Class_350

[15] T.K. Ho, Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition (Vol. 1, pp. 278–282). IEEE. 1995.

[16] F. Chollet et al. Keras: Deep Learning for Humans. https://keras.io. 2015.

[17] Network Rail Open Data Platform. Network Rail. Accessed 12-March-2025. https://publicdatafeeds.networkrail.co.uk

[18] Rail Open Data Wiki. Peter Hicks et al. Accessed 15-March-2025. https://wiki.openrailda https://github.com/openraildata ta.com

[19] Open Rail Data – GitHub. Peter Hicks et al. Accessed 15-March-2025. https://github.com/openraildata

[20] Rail Data Marketplace. Rail Delivery Group. Limited. Accessed 07-March-2025. https://raildata.org.uk/

[21] RailwayData | Bridges. railwaydata.co.uk. Accessed 15-March-2025. https://railwaydata.co.uk/bridges.

[22] Network Rail Bridge Height Data. Network Rail. Accessed 11-March-2025. https://www.networkrail.co.uk/wp-content/uploads/2024/11/Network-Rail-Bridge-Height-Data-1.xlsx

[23] OpenTrainTimes: Real-time Track Diagrams. OpenTrainTimes Ltd. Accessed 03-March-2025. https://www.opentraintimes.com/maps

[24] OpenRail Real Time Railway Data. Charlwood House Systems. Accessed 07-March-2025. https://www.charlwoodhouse.co.uk/rail/liverail

[25] GitHub - philwieland/openrail: Open Rail Data Processing. Accessed 07-March-2025. https://github.com/philwieland/openrail/

[26] Overpass Turbo. Accessed 16-March-2025. https://overpass-turbo.eu