The Development of an Enhanced Bowtie Railway Safety Assessment Tool using a Big Data Analytics Approach

H J Parkinson*, G Bamford†, B Kandola†

*Digital Rail Ltd, hjparkinson@digitalrail.co.uk, UK and †NTTX Advisory Limited, gary.bamford@nttx.co.uk, UK

Keywords: Big Data Analytics, Railway Safety Management, Accident Analysis.

Abstract

The focus of this paper is on the challenges encountered in safety assessments around the use of multiple data sources in the form of Big Data Analytics. The Grayrigg rail accident is analysed to investigate the type of analytic algorithms that would be needed to extract enhanced safety knowledge out of the available data sources. The current state of play in machine learning is explored and an overview of all the available analytical approaches is given, ways forward are presented and assumptions discussed to determine the attributes and requirements for combinations of data. This was done in order to assess their applicability to the proposed ELBowTie approach. The overall objective is to use this information to develop a prototype machine learning algorithm with a view to training the algorithm to recognise heightened risk situations on the rail infrastructure.

1 Introduction

The concept of Big Data (BD) and associated analytics was not known ten years ago but now it looks like it will be revolutionary for the rail industry. The current thrust in the UK is to move to a ‘digital railway’ using the latest signalling, communication and information systems which will rely on an open data philosophy [1] and the cross industry Defect Recording and Corrective Action (DRACAS) program [2] is also driving this move toward greater sharing of data.

There is a real opportunity to integrate BD within railway safety risk assessment activities [3], to provide timely and relevant information to support intelligent design decisions and ensure risk assessments are fully integrated into the business model for the railway. This paper contains a framework for integrating safety, BD and advanced analytics to provide a real time safety management ‘dashboard’.

Data alone is not the answer though, the added value coming from how you analyse and what you do with these data is the key. In this paper the UK Grayrigg accident [4] is analysed to investigate the type of analytic algorithms that would be needed to extract enhanced safety knowledge out of the available BD sources. The paper then goes on to describe how this information could subsequently be used to provide real time safety management using the ELBowTie methodology [3]. The ELBowTie methodology provides the framework for illustrating the analytical mechanism for introducing BD into safety risk assessments. The thinking behind the approach presented in this paper will help to both enhance risk analysis activities as well as providing an approach for ‘real-time’ or ‘near real-time’ assessment of safety risk profiles.

The Use of the ELBowTie provides a formal means of allocating data to causes, consequences and risk mitigation measures set within an enterprise data analytical framework. ELBowTie is built around the well-known Bowtie risk assessment methodology and has been designed to incorporate BD into the risk assessment activities. The bowtie analysis has been enhanced through linking relevant Enterprise Data Taxonomy (EDT) sources to each of the elements of the bowtie. These data sources include, for example, condition monitoring, social media, and safety management, see Table 1 for the comprehensive list [3].

Once the data sources have been linked to the bowtie elements, resulting in the “Enterprise data taxonomy Linked Bow-Tie” (ELBowTie), they can be used as the basis for tailoring safety algorithms. An example would be evaluation of the, effectiveness of control measures or assessment of precursors to accident causes.

There are 5 main analytical approaches (BASEC) available for BD analysis [5]. These are;

1. Bayesian analytics – probabilistic inference algorithms
2. Analogy analytics – inference algorithms
3. Symbolist analytics – inverse deduction algorithms
4. Evolution analytics – genetic algorithms
5. Connectionist analytics – backpropagation algorithms

The applicability of these analytical methods as they apply to the Grayrigg accident is reviewed and then a demonstration of how each approach could be applied to the safety data is presented. The data used in this analysis includes texts, safety management information and condition monitoring information. The data incorporates real time as well as structured and unstructured data types.

The major focus of this real-time BD analytics is to identify, in a predictive manner, hazards on the railway that could propagate into accidents.
Table 1. Enterprise Data Taxonomy

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Real time</td>
<td>a. Remote monitoring</td>
</tr>
<tr>
<td></td>
<td>b. Traffic flows</td>
</tr>
<tr>
<td></td>
<td>c. Incident data</td>
</tr>
<tr>
<td></td>
<td>d. Close Calls/Alerts/CIRAS</td>
</tr>
<tr>
<td></td>
<td>e. Emergency services communications</td>
</tr>
<tr>
<td></td>
<td>f. CCTV</td>
</tr>
<tr>
<td>2 Asset</td>
<td>a. Maintenance</td>
</tr>
<tr>
<td></td>
<td>b. Risk based inspection and maintenance data</td>
</tr>
<tr>
<td></td>
<td>c. Integrity data, safety, security, environmental</td>
</tr>
<tr>
<td></td>
<td>d. Design Hazard Data (residual risks)</td>
</tr>
<tr>
<td>3 BI - business related</td>
<td>a. Finance</td>
</tr>
<tr>
<td></td>
<td>b. HR related</td>
</tr>
<tr>
<td></td>
<td>c. Quality management</td>
</tr>
<tr>
<td></td>
<td>d. Safety management</td>
</tr>
<tr>
<td></td>
<td>e. Project management/Business Risk Assessment</td>
</tr>
<tr>
<td>4 Operational</td>
<td>a. Complex unstructured data, reports, spreadsheets etc.</td>
</tr>
<tr>
<td></td>
<td>b. Staffing levels, etc.</td>
</tr>
<tr>
<td></td>
<td>c. Manning schedules</td>
</tr>
<tr>
<td></td>
<td>d. Service related, timetables, etc.,</td>
</tr>
<tr>
<td></td>
<td>e. Operational Risk Assessment</td>
</tr>
<tr>
<td>5 Social</td>
<td>a. LinkedIn</td>
</tr>
<tr>
<td></td>
<td>b. Facebook</td>
</tr>
<tr>
<td></td>
<td>c. Other e.g. news items</td>
</tr>
<tr>
<td>6 External</td>
<td>a. Supplier data</td>
</tr>
<tr>
<td></td>
<td>b. Map related data, works location etc.</td>
</tr>
<tr>
<td></td>
<td>c. Environmental trends, weather etc.</td>
</tr>
<tr>
<td>7 Personal</td>
<td>a. Location history</td>
</tr>
<tr>
<td></td>
<td>b. Health related</td>
</tr>
<tr>
<td></td>
<td>c. Education related</td>
</tr>
</tbody>
</table>

2 Overview of the Grayrigg Accident

Grayrigg was the last railway train accident in the UK to result in deaths [6]. One passenger was fatally injured; 28 passengers, the train driver and one other crew member received serious injuries and 58 passengers received minor injuries. The remaining 18 passengers and two crew members were not physically injured in the derailment [4]. The Grayrigg accident caused the death of a Scottish pensioner Mrs Margaret Masson after the train she was travelling on derailed at a set of inadequately maintained points in Cumbria England.

It happened on 23 February 2007 at 20:12 hrs. Mrs Masson was travelling back from London with her daughter and was seated in the leading vehicle of the train. The Virgin train was a Class 390 Pendolino tilting train that was the 17:15 hrs service from London Euston to Glasgow which had 109 people on board including 4 crew and was made up of 9 cars. It derailed at 95mph with 8 cars going down the high embankment into the field below with 5 cars ending up on their side as shown in Figure 1 with the direction of London to the right of the photograph. A paper written to compare different accident causation models using the Grayrigg accident as the example was very useful in providing additional insight into the accident [7].

The immediate cause of the accident was the defective points 2B that were not secured effectively. The arrangement of the crossovers is shown in Figure 2 with London to the left in this picture. These were facing points and the fact that they had moved meant that the distance between the rails effectively decreased causing the wheel flange to climb on the rail and the 1st car and all the following cars to derail. The Formal report on the accident by RAIB provides a rich source of data for use in the ELBowTie analysis. The Grayrigg Time Line is illustrated in Table 1 below.

The investigation report into the Grayrigg accident was published by the RAIB (Railway Accident Investigation Branch) [4]. Subsequent criminal proceeding was taken against the infrastructure maintainer Network Rail (NR) by the UK Crown Prosecution Service. This resulted in NR being found guilty of health and safety violations and ordered to pay compensation to the family of Mrs Masson and £4m in fines. NR made the following apology "The Grayrigg derailment in 2007 resulting in the tragic death of Mrs Masson was a terrible event. Within hours it was clear that the infrastructure was at fault and we accepted responsibility, so it is right that we have been fined. Nothing we can say or do will lessen the pain felt by Mrs Masson's family but we will make the railways safer and strive to prevent such an accident ever happening again."
The problem with the points was caused by successive failures of the three stretcher bars and the lock stretcher bar assembly. The 3rd stretcher bar assembly is shown in Figure 3.

Figure 3. Points Stretcher Bar Arrangement (From RAIB[4])

Three issues were identified that were determined to have caused the hazard:

1. The failure of the joint connecting the 3rd stretcher bar to the right hand switch rail which together with:
2. The excessive residual switch opening causing the left hand switch rail to make contact with the flanges on the passing train wheel causing flange climb.
3. A maintenance intervention of the 18th of February 2007 that did not take place and might have identified the deterioration [7].

Several ‘underlying’ factors associated with the overall safety management systems, organisational arrangements and regulatory structure were identified which were also analyse in the assessment.

3 Accident Assessment Method

The traditional bowtie provides a linkage to Event Tree and Fault Tree Analyses techniques already in common use when conducting risk analysis work. In this respect, the ‘Left-Hand Side’ of a bowtie can be considered to represent a ‘Fault Tree’ and the ‘Right-Hand Side’ an ‘Event Tree’. The bowtie we have developed for Grayrigg is shown in Figure 4 below. In some industries bowties are considered to be an essential component of a Safety Case [8] and the tasks associated with each of the barriers also form critical elements of the SMS.

3.1 Bowtie Diagram Grayrigg Derailment Accident

For the derailment accident scenario, the accident investigation report [4] identified a range of causes and consequences. To illustrate the BD approach, we have selected some of the causes as shown in the bowtie diagram below and detailed in Table 3.

The initiating event is the wheel derailment for which the underlying causes are,

- Loading from trains exceeding design spec (CA1)
- Re-use of threaded fasteners (CA2)
- Increased loads on 2B points (CA3)
- Increased level of flange-back contact (CA4).

Consequences of derailment are identified as,

- Multiple fatalities and injuries (CO1)
- Asset damage (CO2)

The table has been abridged with only one cause arm/barrier and one consequence arm reproduced for illustration, the cause arms not shown are Cause CA2, Cause CA3, Cause CA4 and Consequence arm CO2. For each of the causes and the consequences a number of prevention and mitigation measures (barriers) are in place and are listed and commented upon in Table 3.

4 Data Analytics Assessment

Simplistically and for the purposes of this analysis it is envisaged that ‘live’ information about the condition and performance of a safety ‘barrier’ question can be readily accessed from a BD source. This information is then used as the basis for subsequent analysis to ascertain if the barrier is fit for purpose and can perform its required function.

If, for example, from the BD information it is found that the joint between the permanent way stretcher bar bracket and the switch rail at 2B points (Barrier CA1-B1, see Table 3) was being subjected to forces beyond its design capability, this will be flagged (via the ELBowTie) to the management and/or the maintenance team to take the appropriate action. Similarly, if maintenance was due for a particular component and it was not carried out according to the planned schedule, this would be continuously flagged (on the ELBowTie) until the maintained task was performed satisfactorily.
On the other hand, to minimise fatalities and personal injuries the BD traffic information could be used to warn and stop other trains from approaching the accident scene (CO1-B4, in Table 3).

With the appropriate use of BD and the ELBwTie, accidents potentially could be prevented by ensuring that safety critical aspects are constantly monitored for their required performance and that the consequences are minimised by taking the mitigating actions to prevent a minor incident becoming a major disaster.

Table 3. Description of one Cause-Consequence link of the Bowtie

<table>
<thead>
<tr>
<th>Item #</th>
<th>Description</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause CA1</td>
<td>Loading from trains exceeding design specification.</td>
<td>It was found that the fasteners failed by unwinding. This occurs when the applied load exceeds the clamping force on the joint. The RAIB therefore concluded that, in this case, such loadings were a result of the passage of trains.</td>
</tr>
<tr>
<td>Barrier CO-B1</td>
<td>Vehicle Design</td>
<td>It is recommended that in order to minimise the risk of injury from detachment of seats in the event of an accident, the design standard for seats should include for seats to deform in a ductile manner when overloaded, particularly in the lateral direction. Crash worthiness of vehicles should be considered in design.</td>
</tr>
<tr>
<td>Barrier CO-B2</td>
<td>Infrastructure Containment</td>
<td>None in place at Grayrigg</td>
</tr>
<tr>
<td>Barrier CO-B3</td>
<td>Remote Monitoring</td>
<td>Remote monitoring of rail traffic not carried out at Grayrigg</td>
</tr>
<tr>
<td>Barrier CO-B4</td>
<td>Signalling System to Stop Other Trains</td>
<td>In the event of an accident, remote monitoring to stop other trains from approaching the crash area.</td>
</tr>
</tbody>
</table>

4.1 What is meant by analytics

Before going into a more detailed review of the Grayrigg accident BD opportunities, it is worthwhile defining what is means by ‘analytics’. Analytics is generally a term that is used to denote any type of data analysis. However, it comes in many
forms, it is valuable therefore to illustrate the whole of the available ‘analytics landscape’ to put context around the form of analysis that can be brought to bear in the predictive risk assessment approach described in this paper.

The analytics landscape is populated by various algorithms which are generic mathematical approaches used to problem solve. An excellent review of these approaches is provided in ‘The Master Algorithm” by Pedro Domingos [5] and terminology developed is used in this paper to bring structure to the analytics landscape. In this paper therefore, algorithm types are defined in terms of their ‘BASEC’ elements. The 5 categorises have also been used to describe the 5 “tribes” of machine learning and are shown in Table 4 in no particular order of priority.

Table 4 The Five Machine Learning Tribes

<table>
<thead>
<tr>
<th>#</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Analogy analytics – inference algorithms which are based on extrapolations from similar situations. For example, there are ‘Y’ similar points and crossings to those under study at Grayrigg across the network out of these there are ‘Z’ similar train paths over a subset of points (SSP). Therefore, the points at Grayrigg should exhibit similar characteristics to those contained in SSP.</td>
</tr>
<tr>
<td>B</td>
<td>Bayesian analytics – probabilistic inference algorithms such as decision trees, state machines. Example being asset specific event trees to determine how a particular asset failure mode can propagate into a hazardous situation e.g. points gap to small, leading to wear and tear on the stretcher bar, leading to stretcher bar failure, leading to point failure, leading to derailment.</td>
</tr>
<tr>
<td>C</td>
<td>Connectionist analytics – backpropagation algorithms which look for connections between ideas/data and compare outcomes with those expected from these connections to improve algorithm mechanics. An example being, learning a particular pattern of maintenance activities then looking for variations from this ‘expected’ pattern.</td>
</tr>
<tr>
<td>D</td>
<td>Evolutionary analytics – genetic algorithms which evolve solutions using trends. An example being improving inspection and maintenance activities by understanding the nature of particular asset failure modes and how they change over time.</td>
</tr>
<tr>
<td>E</td>
<td>Symbolist analytics – inverse deduction algorithms which, in general, are aimed at identifying gaps in knowledge and then building a rule set to address the missing information to then help solve problems. An example being, development of a rule set such as: all maintenance tasks are carried out by humans, humans are also error prone. Applying the rule to Jack and Gill who work for a maintenance company, provides an analysis of Jack and Gill as follows; -Jack and Gill work for a rail maintenance company (input data item) -All maintenance tasks are carried out by humans (apply rule) -All humans are error prone (apply rule) -Jack and Gill are error prone (output based on inverse deduction algorithm).</td>
</tr>
</tbody>
</table>

The ‘ideal’ predictive algorithm, or combination of algorithms, operates as a black box that learns and improves its predictive ability over time as more data and experiences are accumulated.

4.2 Mapping Grayrigg data using ELBowTie

Section 3 contains an illustration of how the Grayrigg accident causes, consequences and associated barriers could have been determined using a bowtie risk analysis methodology. The ELBowTie methodology takes the risk analysis a stage further by linking the bowtie elements to the enterprise data sources to provide a ‘live’ predictive view of the hazardous state.

The ELBowTie enhancement to this approach is illustrated in Figure 5 [3]. The top of the diagram (i.e. the main bowtie) details the available EDT for this analysis, this is a bounded set of data items based upon the system under study. This set may be expanded as the safety identification activities get underway and emergent issues from the studies arise.

The data available for the railway that has a potential safety impact has been compiled into a taxonomy as detailed in a previous paper, Parkinson and Bamford [3] and includes geographic, weather, closes call, condition monitoring, social media and other data. The bottom of the diagram contains those data items from the EDT that can be linked (i.e. that could potentially be used in some form of analysis) to the bowtie components such as, threat, controls, outcomes and the recovery measures.

Critical elements to these predictions are the algorithms used to analyse the various data sources. The following discussion illustrates the types of algorithm that would have helped identify the hazardous state associated with the Grayrigg accident.

Table 5 contains an assessment of the enterprise data related to the Grayrigg accident causes, consequences and barriers. Column 1 contains an evaluation of the data sources available for analysis, an indication of the ‘form’ this data takes referenced as;

- Structured data is data that contains clearly defined elements, associated hierarchies and formats e.g. points and their associated components such as stretcher bar bolts and the component specific data such as bolt torque settings.
- Semi-structured data is data that has some structure elements but also contains information that is not easily referenced such as asset maintenance reports held in Microsoft Word.
- Unstructured data is data that is not readily referenced such as contents of pdf files or information contained in photographs.

Column 2 contains an initial assessment of the type of analysis that could be performed on the data sources of column 1. Column 3 then builds on this analysis to demonstrate the advanced analytics that could be brought into effect to provide greater risk insight.

The assessment presented in Table 5 assumes that all the data is in a form that can be accessed to enable computer analysis, for example, in bespoke dB, Excel, Word files etc.
### Table 5. Potential Machine Learning Algorithms ELBowTie Branch CA1 and associated Barriers.

<table>
<thead>
<tr>
<th>Data sources available at the time of the accident</th>
<th>Analysis based on</th>
<th>Examples of potential advanced analytics (BASEC) from Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design specifications/s tards (structured, asset locations, design specs, applicable standards)</td>
<td>Asset locations and applicable standards</td>
<td>B1 – to compare failure trends with standard requirements. A1 – to build correlations between best practice application of standards and geographical locations. S1 – to evaluate applicability of standards to assets e.g. have most recent standards been used. E1 – to monitor asset performance and help improve standards C1 – to connect assets to standards to locations.</td>
</tr>
<tr>
<td>Train movements loadings over points (structured, train timetables and loadings)</td>
<td>Timings, loads specific to locations and assets, correlations with design data</td>
<td>B2 – failure probability determination based upon load and maintenance profiles. A2 – location specific train and asset performance information. S2 – to apply rule sets to timetables (and changes) to predict asset loading and identify issues. E2 – asset performance and degradation profiles correlated to train movements, locations and design data. C2 – connection of track and track assets data to correlate with design data and identify potential for failure.</td>
</tr>
<tr>
<td>Inspection records, times, results (semi-structured records, location, asset specific data items)</td>
<td>Regime, timelines, degradation trending/profile s, key word searching, correlations with design data and train movement data</td>
<td>B3 – analysis of potential failure probabilities and correlation with inspection regime. A3 – investigation and comparison of effectiveness of inspections across the wider network. S3 – rule sets based on key words to analyse inspection records and identify issues. Correlation with historical trends on personnel and asset related information. E3 – to build up a picture of the effectiveness of inspection activities covering both staff and asset performance. Correlated with train and asset design data. C3 – to investigate trends within the inspections regime, staff and asset issues.</td>
</tr>
<tr>
<td>Maintenance records times, results (semi-structured records, asset specific data items)</td>
<td>Regime, timelines, key word searching, correlations with design data and train movements</td>
<td>B4 – analysis of potential failure probabilities and correlation with maintenance regime. A4 – investigation and comparison of effectiveness of maintenance across the wider network. S4 – rule sets based on key words to analyse maintenance records and identify issues. Correlation with historical trends on personnel and asset related information. E4 – to build up a picture of the effectiveness of maintenance activities covering both staff and asset performance. Correlated with train and asset design data. C4 – to investigate trends within the maintenance regime, staff and asset issues.</td>
</tr>
<tr>
<td>Assurance of physical condition records - spot checks (semi-structured, asset specific data items, maintenance personnel)</td>
<td>Regime, key word searching, trend lines, correlations with inspection and maintenance regimes</td>
<td>B5 – analysis of potential failure probabilities and correlation with assurance regime. A5 – investigation and comparison of effectiveness of assurance across the wider network. S5 – rule sets based on key words to analyse assurance records and identify issues. Correlation with historical trends on personnel and asset related information. E5 – to build up a picture of the effectiveness of assurance activities covering both staff and asset performance. Correlated with train and asset design data. C5 – investigate trends within the assurance regime, staff and asset issues.</td>
</tr>
<tr>
<td>Verification records – assessment of compliance with standards (semi-structured, asset specific data items)</td>
<td>Regime, key word searching, trend lines correlations with assurance regime, design data, train data</td>
<td>B6 – analysis of potential failure probabilities and correlation with standards. A6 – investigation and comparison of effectiveness of verification across the wider network. S6 – rule sets based on key words to analyse verification records and identify issues. Correlation with historical trends on personnel and asset related information. E6 – to build up a picture of the effectiveness of verification activities covering both staff and asset performance. Correlated with train and asset design data. C6 – investigate trends within the verification regime, staff and asset issues.</td>
</tr>
</tbody>
</table>

### 4.3 Reasonableness test of the Analytics

This section contains a review of the analytics regime proposed in Section 4.2 that could be applied to rail data. All of the operations presented in section 4.2 can be carried out using standard analysis techniques. There is no need to invoke exotic mathematics, most of the analysis simply requiring trending, correlating and searching type, basic functionality. Although the volume and speed of the data would require modern data processing hardware and software capabilities [9].

The complexity therefore lies not necessarily in the analysis methods but in the supply of data to them and in the interpretation of the outputs. Input data comes in many forms from real time asset monitoring systems through semi-real time train data to non-real time inspection and maintenance, assurance, verification data. Although there are algorithms, as detailed in Section 4.1 that can extract information from the specific sources, correlating results from these calculations on an appropriate timeline to produce robust risk management information is non-trivial. The outline of an approach to automate management of these information streams is presented in Section 4.4. However, the main focus of this paper is on evaluation of effectiveness of the first stage analytical methods detailed in Section 4.2.
The analytics therefore has to cope with an array of issues, such as:

- Setting the sensitivity of the algorithm to flag risks and not produce too many false positives.
- Ensuring appropriate access to the disparate data sources, problems include,
  - Commercially closed data sources
  - Handling of sensitive data
- Keyword searching requires clear definitions to perform effectively. A recent analysis of safety close call data has highlighted the difficulty in targeting keywords to unstructured data sources [10].
- Data quality can also be variable across the data sources reducing its usefulness.

Detailing exactly what aspects of the asset under study are of interest is a critical element of the analytics and the focus of interest may change as the measured performance of the asset changes over time. For example, if an asset’s reliability drops the safety risk profile also varies, which then may require more enhanced monitoring. Understanding what additional monitoring points may be needed is one of the main reasons ELBowTie has been developed.

4.4 Thoughts on a future on alarm flagging for the ELBowTie

One option for combining algorithms uses neural networks. The algorithm learns the relative weight of input data and produces a trigger 0 to 1 to flag risk as depicted in Figure 6. Each data source, for example close calls data or vehicle track geometrical data or social media or any of the other items in the EDT would first be cleansed, normalised and streamed into the appropriate algorithm. The algorithm will have been trained using machine learning methodologies in order to generalise about a given data stream and make inferences whether the data represents a “normal” situation for a period of time on the railway or an abnormal or notable variation.

Each stream would then be fed into, for example a Neural Network that will have been trained using rules for acceptable combinations of inputs. It is likely that the inputs would be weighted based initially on the known safety criticality of the data. Over time this amalgamating Neural Network would be taught to recognise heightened risk based upon the number of potential accidents and occurrence of actual accidents. Over time the algorithm would build up a model of the complex relationship between accidents causes and mitigations.

Figure 5. ELBowTie Diagram

Figure 6. Proposed Structure for Combining Data
that flags up an event would be linked via the ELBowTie to either a cause barrier or a mitigation barrier and this could give real time warnings.

Currently we are researching the use of open source deep learning Neural Networks to predict vehicle derailment quontities. This research is reproducing similar research [11] from several years ago but now with massively enhanced cloud based computing and more effective algorithms which should predict derailment risk more accurately.

Research is also taking place to identify heightened risk based upon close call data with some success [10]. Live social media streaming and word search is already technically feasible as evidenced by the Twitter Sentiment Display [12] which indicate either positive or negative sentiment relating to train operations in real time. It would not be such a stretch to extend this to other social such as GPS data from smart phones. More and more data are becoming available, for example, railway signalling information is in the public domain in the form of train describer data which is already used to show train position in near real time on-line on Rail Radar [13].

It is now planned to address each source of data from the EDT and select a suitable algorithm with the intention of building a complete analysis tool that can then integrate the outputs into a system that can predict the risk in real time and indicate actions to be taken, either by a machine or a person.

5 Conclusions

This paper has used the Grayrigg accident to illustrate how advanced analytical algorithms together with the ELBowTie risk assessment methodology can be used to help reduce safety risk on the railway. The advantages of the approach detailed in this paper are that it provides,

- a visualisation of the data sources that can be considered as part of the safety assessment,
- a means of additional completeness checking of hazard identification workshops through considerations of data sources related to, threats, availability of the controls and recovery measures, and tasks associated with the controls and recovery measures.
- a baseline for building additional analysis, ‘Analytics’, around the Linked data sources with a view to supplementing safety assessments, for example;
  - analysis of individual threat and control based data sources to develop preventative and early warning signals (ELBowTie being a ‘live’ safety management tool),
  - evaluations of correlations between data sources, for example, threats and controls combined analysis to identify vulnerabilities and assessment of control effectiveness levels,
- analysis of pre-cursor indicators through, for example, analysis of near miss information related to the safety hazard outcomes.
- indicators presented in terms of risk, for example;
  - simple enhancements e.g. the use of time and location stamped photos, or,
  - elements that can be changed or improved in a simple manner if they are not working e.g. the quality of reporting on maintenance records.

After the Grayrigg accident a Network Rail spokesman said [14]: “The inspection train runs at speeds of up to 125 mph, or 95 mph on this stretch. There would be no point somebody watching it at that speed as they wouldn't be able to pick up any faults. It has to be run in super-slow motion to spot faults. The train runs for up to 18 hours a day, seven days a week. It would probably take someone most of the month to watch one day's worth of data. It's not what it's there for. It's a backwards reference tool”. What is being proposed here in this paper would have allowed exactly that type of data to be analysed and action taken in a timelier manner.

Furthermore, the data collected would lead to a real time picture of the risks around the network instead of the largely risk based qualitative models currently in operation. As the algorithms are used over longer periods and trained accordingly, the picture they provide of the railway risk gets more and more accurate leading to opportunities for making the railway network even safer and more efficient.

The concept of the Internet of Things (IoT) is being applied in the railway [15] building upon a well-established basis in condition based monitoring. The drive is to have intelligent assets that can gather data from their operation and communicate with other assets and a central control to enable asset management and safety to be optimised. The research detailed in this paper and already in planning should contribute to this trend and provide opportunities for automation and optimisation of many functions in the not too distant future.

References


Accessed 26/2/2016

Accessed 26/2/2016

Accessed 26/2/2016