Paper 66

Proceedings of the Third International Conference on Railway Technology: Research, Development and Maintenance, J. Pombo, (Editor), Civil-Comp Press, Stirlingshire, Scotland.

The Potential for Using Big Data Analytics to Predict Safety Risks by Analysing Rail Accidents

H.J. Parkinson[1] and G. Bamford[2]
[1]Rail System Engineering Limited, Lancaster, United Kingdom
[2]NTTX Advisory Limited, Warrington, United Kingdom

Abstract

We are currently going through an unprecedented era of digital transformation. This change includes everything from the Internet of Things (IoT) through Big Data (BD) to new analytical approaches (Analytics) to analysing business and personal needs. The rail industry in particular is a focus for digital transformation through passenger and infrastructure related initiatives. Digital transformation will fundamentally change the way the industry works particularly in regard to risk assessment and safety. With modern powerful computing and the explosion in data availability, from ever expanding sources, there should be opportunities to use a BD approach to flag up high risk scenarios on the railway before accidents occur. In this paper, an understanding of BD as it applies to railway safety management has been developed in terms of the 5V (Volume, Velocity, Variety, Veracity and Value) model. An enterprise data taxonomy (EDT) has also been suggested as a way of bounding safety data items. Three major accidents have been reviewed and assessed to understand what the high level causation was due to. Each of the causes was then linked, using the EDT, to sources of data that may have highlighted the risk of the hazard together with its potential to propagate into a railway accident. The data sources have been evaluated according to the 5V model and assessed for their respective 'BDness' to provide a score that can then be interpreted as the potential for big data analytics (BDA). The analysis of the accidents has shown that BD could potentially help in mitigating accidents where the causes are systematic and complex in nature. An enhanced ELBowTie methodology has been introduced to provide a mechanism for feeding BD into safety risk assessments. This methodology will also provide a means for linking real time data updates into the ELBowTie thus enabling a risk dashboard to be envisaged. The deliverables from the research presented in this paper therefore lead to a greater understanding of BD and the methods (analytics) that are needed to make improvements to railway safety.

Keywords: big data, analytics, railway safety, risk assessment, accident analysis.

1 Introduction

The world is currently going through an unprecedented era of digital transformation [1]. This change includes everything from the Internet of Things (IoT) through Big Data (BD) to new analytical approaches (Analytics) to analysing business and personal needs. The rail industry in particular is a focus for digital transformation through passenger and infrastructure related initiatives [2]. Digital transformation will fundamentally change the way the industry works. Deep organisational shifts may be required to reskill the workforce or modify business processes to accommodate the new emerging work practices. One area which is already undergoing upheaval is around the convergence of IT and OT (Operational Technology) such as the implementation of the European Rail Traffic Management System (ERTMS) systems where the supporting IT systems (e.g. servers, network connectivity, and desktop systems) are becoming more and more integrated into the operation of the railway.

This paper is particularly timely as it fits well with the UK Digital Railway Program [3] which will facilitate the type of approach outlined here. In the UK there are already BD initiatives for railway risk assessment research [4][5] that are taking into account data available from signalling systems, asset condition monitoring, safety incident reporting and even social media.

With modern powerful computing and the explosion in data availability, from ever expanding sources, there should be opportunities to use a BD approach to flag up high risk scenarios on the railway before accidents occur. This leads to the question that this paper is aimed at answering. "How can this new world of BD be used to help proactively manage safety on the railways and prevent safety incidents?" The management of safety is one of the most important areas of railway operations. However, currently, rail safety is a long way from being a proactive, intelligent digital/data driven management field.

Although there are initiatives starting to address these issues [3], the fractured nature of available data across the industry in spreadsheets, databases, through to bespoke company systems, makes it difficult to implement any coordinated analysis. Even if data sources where available what BD would be needed to reduce major accident risks, indeed, what is meant by BD and Analytics in this context? One poplar idea of BD is to simply collect 'all' available data hoping you will be able to do some clever Analytics on it later? Taking data for data's sake has in some instances led to data overload, with datasets beyond the capacity of ordinary databases leading to poor quality analysis and even no analysis [6]. There is a lot of hype around the topics of BD and 'Analytics'. This paper presents an approach that is grounded in the knowledge of railway operations and systems, which together with a study of the new data initiatives, aims to develop an approach to using them to help in accident prevention.

So the focus of this paper is to:

1. Study the potential scope and context of BD in safety accident risk reduction,

2. Select a number of major accidents to understand what would have been needed to be known to avoid the accidents,
3. Understand what data would be needed to support this analysis (BD candidates)?
4. Evaluate these datasets in terms of the BD requirements to give a measure of the feasibility of this type of Analytics,
5. Review the process followed to understand the types of issues that would face implementation of such an approach,
6. Identify mechanisms for utilising BD to improve safety management.

2 Data Assessment Method

This section contains a summary of how BD could be defined in safety assessment activities together with a brief overview to the safety risk assessment methodology used in the analysis of the selected rail accidents analysed in this paper.

2.1 Big Data Definition

Just what is meant when by the term 'big data'? Is 'big data' something new? Currently, articles and opinions abound, with definitions and views varying [6]. In this paper it is not necessary to get involved too much in these arguments, however, there is a need to state the position on 'data'.

A fairly pragmatic approach is taken in that all data is simply data to be analysed. If this data is not 'big' enough then statistical approaches can be used to 'mine' the data for more information. This approach is a standard engineering approach to data analysis. So is BD just the point at which there is enough volume of data to not worry too much about statistical analysis? That's not entirely correct, as there may be further knowledge to be had by introducing other data sources, making the sample even bigger. Well even that's not exactly correct either, could more frequent data samples be made or data quality could be improved, making the datasets even bigger?

The answer to these questions, from the perspective of this paper, is that all of the above are correct provided they add value to the analysis. There is not much point taking data for data's sake, there needs to be an end rationale in mind for selecting data to be analyse.

From here on the most popular identifiers of a BD approach, the 5 V's [7] are used as a framework for evaluations of data sources:

- Volume (amount),
- Velocity (speed of capture or change),
- Variety (number of sources),
- Veracity (quality), and
- Value (here meaning the safety related value).

Taken together these can be used to rate the level of 'BDness' that would be necessary for the potential to make predictions of the railway safety risk levels. The approach taken in this paper is intended to satisfy all of the 5V BD criteria. The crucial requirement is bounding the available data and then turning it into safety information taking into account the fact that correlations and causations will sometime be counterintuitive.

At a recent BD risk assessment symposium the following quote was made "Instead of constructing a pre-determined structure for risk calculations, such as, for example, fault trees, different sources of safety-relevant information could be brought together to find answers to specific safety questions, and to identify new trends and threats to safety," [4].

It is not intended here to develop a BD system that will actually predict an accident but rather to present an approach that will flag up high risk scenarios based upon the flow of data from the identified sources.

2.2 Enterprise Data Taxonomy

Having specified what is meant by BD, what exactly does that data look like in a rail context? The first step is to define the boundary around the system and related hazards that are being analysed. This boundary could be for example, the train, the infrastructure, the operators, the business, external factors or various combinations of these. In this paper safety data is allocated to one of the areas listed in Table 1. Enterprise Data Taxonomy (EDT).

The EDT listed is aimed at providing a mechanism for systematically classifying data related to safety incidents to facilitate assessment of data sources in a traceable and consistent manner.

For example, condition based monitoring information from sensors, either analogue or digital, that would provide digital information, including vibration (accelerometers), machine vision, heat, displacement, strain, humidity, particle ingress, etc. would be classified as 'Real Time, remote monitoring', which is already an accepted means of classifying this type of data. Other data types are less well defined, for example, data from industry reports, staff morale, organisational culture, but can be equally as important in safety evaluations.

The EDT provides placeholders for these more intangible data sources. The EDT is most likely not complete but is presented here as a structured mechanism for handling BD in rail environments. There are clearly larger data domains such as rail operational data, business specific data and data external to both rail operations and business, which could be used to provide higher level transport linkages. However, in this paper, the focus is specifically on the EDT presented to illustrate how a data taxonomy can be used to help develop BD safety analyses.

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

Table 1: Enterprise Data Taxonomy

2.3 Rail Safety Management

The field of safety management is well established in high-hazard industries such as in the rail transportation, oil and gas and nuclear sectors. Application of standards and enforcement through safety regulations is common place.

The hallmark of modern safety analysis in the railway is that hazards are predicted before they happen to make sure that the appropriate mitigations are in place. The industry has gone from a reactive approach to safety in the early days of rail to this proactive mind-set [8].

A Safety Management System (SMS) normally sits at the heart of these activities providing a structure with which to collect, review, control and demonstrate that safety risk is being managed to an acceptable level, in the UK for example, this is a level that is As Low As Reasonably Practicable (ALARP).

SMS's, however, can become large and unwieldy if not managed correctly. It can sometimes be unclear that their main purpose is as risk control systems for operations and ongoing safety management. Deficiencies (holes) in the SMS caused by either lack of adherence to safety processes or appropriate oversight are major contributors to accidents.

Train accidents, caused by derailment or collision on the UK rail network, that result in passenger deaths, are very rare. If the number of deaths per annum over the last 20 years is looked at, it can be seen that the graph has plateaued recently [9]. To get a further step reduction in safety risk levels the focus needs to be on organisational issues and human elements resulting from the holes in the SMS. These issues are exactly the types of phenomena that should be amenable to a BDA approach.

The question is how can a BD approach be integrated into a SMS? The approach presented in this paper builds upon a recognised technique for visualising and structuring safety risks called 'Bow-Tie' analysis [10]. Bow-Ties are a graphical display of the relationship between:

- Safety hazards (or Initiating Events) – situations that have potential to cause harm, the central theme,
- Hazard Threats and Controls – the situational failure modes that result in a safety hazard together with potential control barriers, the left-hand side of the Bow-Tie,
- Hazard Outcomes and Recovery Measures – the potential outcomes from realising a safety hazard together with mitigation barriers, the right-hand side of the Bow-Tie,

Figure 1 illustrates the generic Bow-Tie diagram. In essence, Bow-Ties are a means for demonstrating that sufficient controls are in place for the effective management of hazards and their associated risks.

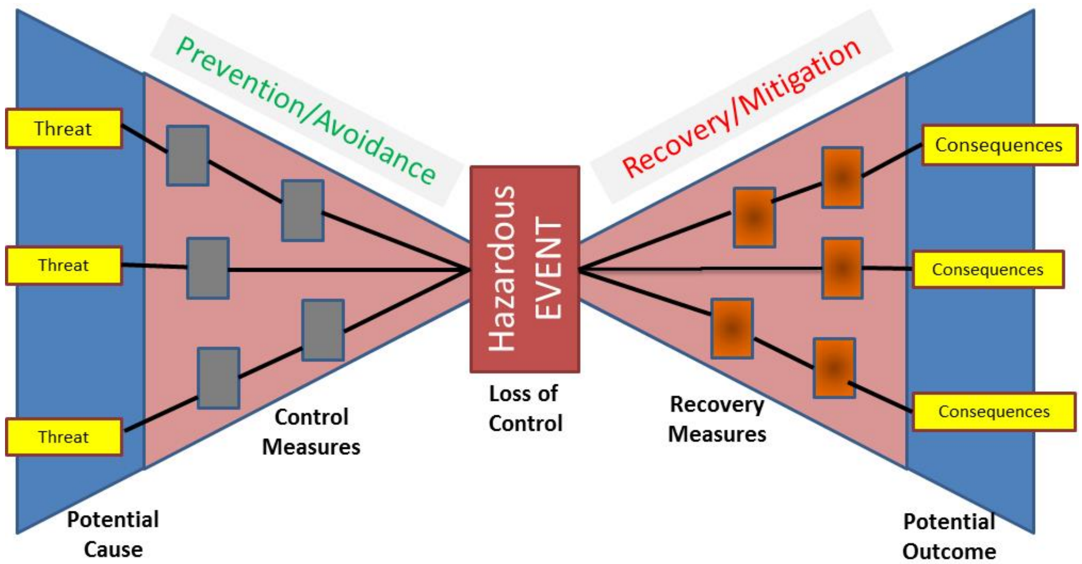

Figure 1: Bow-Tie Diagram

The Bow-Tie approach 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 Bow-Tie can be considered to represent a 'Fault Tree' and the 'Right-Hand Side' an 'Event Tree'. In some industries Bow-Ties are considered to be an essential component of a Safety Case [11] and the tasks associated with each of the barriers also form critical elements of the SMS.

The concept of Bow-Tie analysis is mentioned in the European standard EN50126-2 [12] as reproduced in Figure 2. The EN50126-2 approach depicted in Figure 2 does not explicitly contain specific opportunities for control and recovery measures as shown in Figure 1, which is a limitation. Linking the, causes, consequences and control and recovery actions to data sources is a key aspect of the BD approach presented in this paper.

When data becomes available regarding a particular hazard there may be opportunities to control hazard occurrence and recover using these data sources. To show how this approach can work in the research undertaken in support of this paper, ten accidents have been analysed. However, for the sake of clarity, only the results related to three accidents are provided to demonstrate the mechanics of the approach. The three accidents have been selected to show how the approach can be applied to markedly different hazardous events, as described in Section 3 below.

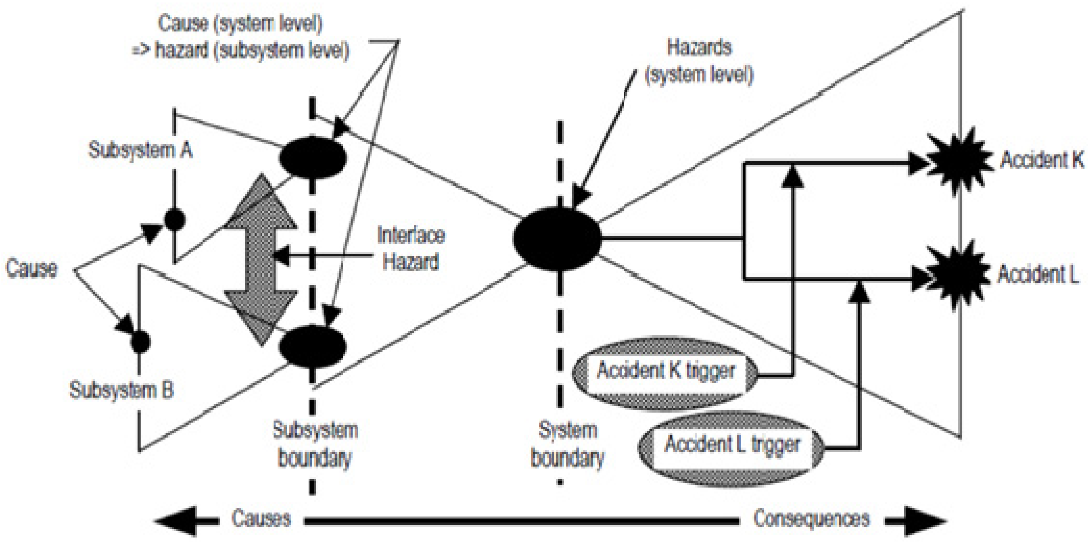

Figure 2: Bow Tie Diagram from CENELEC EN50126-2

3 Data Assessment Method

This section contains a description of the selection and analysis of three major railway accidents to understand the role that BD could have taken in helping enhance safety management related to these events.

3.1 Rail Safety Management

It is known that accidents very rarely have one simple cause and usually have multiple causes such as failure of management, environment and design, to name a few, but what are the main causes? A Study based upon the research of 43 accidents [13] identified 14 high level accident causes as shown in Table 2.

Management Problems	**Operator Errors**
Time and Financial Constraints	Lack of training
Lack of understanding of operation	Carelessness
Engineering Problems	Tiredness or drug impairment
Design d*efects*	Problems from the working culture
Material defects	**Environmental Effects**
Manufacturing process errors	Weather and ambient conditions
Poor quality maintenance	Plant layout interactions
Inadequate analysis of experience	Potentially dangerous equipment or materials.

Table 2: Typical accident Causes

The three accidents used in this paper, two from Spain and one from the UK are outlined in Section 3.2. These accidents have been chosen for review because they are highly complex events with multiple causes. One of the causes in isolation would probably not have resulted in the accident, but it was rather the combination of that cause with others that produced the potential for disaster. A situation like this, therefore, lends itself to analysis using multiple data sources.

The analysis of these accidents takes into account the importance of human factors, management and systems, social media, news, issues and the increasing complexity resulting from, for example, the increased use of software and the disaggregation of railway system management. Failings in the design of processes, standards, rules and procedures, human factors, failings of equipment and degraded modes could also be identified.

The accident analysis has been greatly simplified for the sake of clarity to show the links to potential data sources. In addition, the C-list hazard taxonomy from EN50126-2 [12] has been adopted when describing the hazardous events associated with the accidents. An extract of the C-Hazards is shown in Table 3.

"c-hazard" Reference	Description	Constituents
HP500	Abnormal or Criminal Behaviour	HP0425 Irresponsible behaviour HP0426 Destructive behaviour (all forms) HP0427 Crossing line at station
HP502	Crowding	HP502 Crowding
HP503	Loss of Passenger Compartment Integrity during Movement	HP503
HP504	Passengers in Path of Closing Train Doors	HP504
HP506	Loss of Balance	HP0413 Loss of balance on the ground HP0414 Loss of balance on stairs & escalators HP0415 Loss of balance getting on and off trains HP0416 Loss of balance whilst in a train
HP509	Inappropriate Separation between Running Railways and Passengers	HP509
HP510	Inappropriate Separation between Un-insulated Live Conductors and Passengers	HP0417 Occurrence of DC power arc HP0418 Existence of touch potential HP0419 Inappropriate separation from DC conductor rail HP0420 Structure in contact with live conductor rail HP0421 Inappropriate separation from OHL HP0422 Structure in contact with OHL HP0423 Occurrence of AC power arc HP0424 Inappropriate separation from OHL induced voltage

Table 3: Extract from C-Hazard List in EN50126-2

3.2 Linking Enterprise Data to the Accidents

Three accidents have been appraised by analysing their causes. This appraisal considers:

1. What were the main causes of the hazard?
2. How could the causes be controlled to prevent the hazard?
3. How could the hazard be prevented from leading to an accident?
4. What data would be needed to support this analysis?

Table 4 contains a summary of the results of this appraisal for the three selected accidents. The key issues/causes include both primary and secondary causes and more subtle background issues such as management approaches and organisational culture. The accident causes are then linked to the EDT data sources from Section 2.2.

Accident	Hazards [12]	Key Issues/Causes*	Potential Data Taxonomy
Santiago de Compostela [14], Spain 25 July 2013 Train derailing at high speed on curve	HP602 Loss of Train Guidance (Passenger Trains)	• No high-to-conventional line-speed design control transition (only effective above 124mph; train passed at 121mph); the driver was the sole speed-transition risk control.	1.d. Close calls./ Safety Alert. An alert raised by a driver two weeks before the official opening the line was not acted upon 3.d. Safety Management (Audit Data)
		• Driver distraction: he had been on the telephone to the guard seconds before crash,	1.a Remote monitoring of comms systems
		• Train stability and concerns: top-heavy front diesel generator car seen to topple first and derail set. Articulated mid/rear cars jack-knifed.	2.d. .Train design and lack of stability, could have been known design flaw in hazard log?
		• Poor passenger survivability: 79-deaths; too high for non-head-on derailment/collision. Design for crash worthiness.	2.d Train design. 1.d. Could be a safety Alert/CIRAS
Platja de Castelldefels in Spain [15]. Passengers crossing the line between platforms and being hit by passing train	HP500 Abnormal or Criminal Behaviour **Constituents** HP0425 Irresponsible behaviour HP0427 Crossing line at station HP502 Crowding	• Crowding on platform station. High spirits and subsequent trespassing onto the railway line instead of using the underpass.	5.a/c/ social media Texting the party locations, twitter 1.f Remote Monitoring (Vision System CCTV). Station vision system data. Could detect behavioural norms, crowd density and passengers moving over platform edge. 7.a. Location History GPS. For crowding
		• Station design, members of public claim that it was not clear where the exit was.	2.d Station design 1.d Close Calls 5.a/c/ social media 3.d safety management
		• Time-table non-adherence allowed high speed train to pass through station at same time. Stopping train was 10 minutes late	4.d timetable data
		• Police were present as it was known that there would be a large crowd	1.e. Emergency Services. Police communications
		• It was known either officially or to the public that the crossing of the track as a short cut was quite common.	1.d Close Calls (From Diver or station staff) 7.a Location Data, GPS. Will be used for level crossing safety in USA. Could regular crossing of line be indicated by phone GPS more generally
Hatfield, England [16]. Derailment due to broken rail on curve	HP602 Loss of Train Guidance (Passenger Trains)	• Lack of awareness of risks due to gauge corner cracking, and the failure to develop/communicate the control measures required.	1.d Close Calls Asset management system for track
		• To have proper visual /ultrasonic inspection regime, and to take action.	1.a Remote Monitoring 2.b
		• An inadequate management of the contracts relating to the maintenance and renewal of the track asset.	2.a Monitoring of maintenance contractors, plans, completed, competencies
		• The failure to re-rail as a result of a lack of effective project management, and a failure to take account of the potential risks of delay.	2.a, 2b Asset management system
		• The failure to act on the findings of audit reports relating to a lack of adequate knowledge of the track asset.	2.a 1.d Unstructured corporate data, audits, lessons learned Corporate culture - external data sources potential to be used 3.d

Table 4: Accident and Data Analysis

It is interesting to note the official Spanish accident report for Santiago de Compostela [14] stated that the only contributory issue was the driver speaking to the controller several seconds before crash. The Spanish report for Platja Official Spanish report [15] stated only that the passengers' irresponsible behaviour/ trespassing was the cause.

3.3 Big Data Categorisation of the Accidents

A framework has been devised for evaluating the level of 'BDness' for each of the accidents. The framework uses the enterprise data sources (from the EDT) that have been linked to the causes and gives a rating of the 'BDness' using the BD V's described in Section 2.1.

Tables 5, 6 and 7 contain the results of the analysis. The following simple criteria have been adopted to assess the 'BDness' of the various EDT as follows:

H – Good availability and utility

L – Limited availability and low utility

Santiago data source	Variety	Volume	Velocity	Veracity
	1.d. Close calls./ Safety Alert. An alert raised by a driver two weeks before the official opening the line was not acted upon	L	L	H
	3.d. Safety Management (Audit Data)	L	L	L
	1.a Remote monitoring of comms systems	H	H	H
	2.d. .Train design and lack of stability, could have been known design flaw in hazard log?	L	L	L
Value, BD Quotient, (Number of Hs as percentage of total)			**33 %**	

Table 5: Santiago de Compostela.

Platja data source	Variety	Volume	Velocity	Veracity
	5.a/c/ social media Texting the party locations, twitter	H	H	L
	1.f. Remote Monitoring (Vision System CCTV).	H	H	H
	7.a. Location History GPS. For crowding	H	H	L
	2.d Station design 1.d Close Calls	L	L	L
	3.d safety management	L	L	H
	4.d timetable data	L	H	H
	1.e. Emergency Services. Police communications	H	H	H
	1.d Close Calls (From Driver or station staff)	L	L	H
	7.a Location Data, GPS.	H	H	H
Value, BD Quotient, (Number of Hs as percentage of total)			**63 %**	

Table 6: Platja de Castelldefels.

Hatfield	Variety	Volume	Velocity	Veracity
	1.d Close Calls Asset management system for track	L	L	H
	1.a Remote Monitoring 2.b	H	H	H
	2.a Monitoring of maintenance contractors, plans, completed, competencies	L	L	L
	2.a, 2b Asset management system	L	L	H
	2.a 1.d Unstructured corporate data, audits, lessons learned Corporate culture - external data sources potential to be used 3.d	L	L	L
Value, BD Quotient, (Number of Hs as percentage of total)			**33%**	

Table 7: Hatfield.

The rating of either H or L was based upon expert judgement. The BD quotient from the above tables is intended to give an indication of the total ‘BDness’ of the particular accident scenario and its associated EDT.

BD Quotient	assessment
≥ 50%	High BDA Potential.
< 50%	Limited BDA potential – consider improvements to data sources and understanding.

Table 8: ‘BDness’ assessment.

These scores are based upon an assumption that the railway regime and national law would allow access to this information, which would be a necessity to enable a fully digital railway [3]. It is also assumed that the current or near future potential levels of data availability, computing power and intelligent algorithms were around when the accidents occurred.

4 Analysis of Results and New Safety Risk Model

As stated earlier, the question this paper is aimed at answering is “How can this new world of BD be used to help proactively manage safety on the railways and prevent safety incidents?” In answering this question it has been necessary to define what is meant by ‘big data’, a term that means different things to different people.

The definition used here is somewhat biased as it is based upon the need to focus attention on available industry data to address safety risk issues. There may be even more unforeseen ‘bigger data’ than that presented here that can be used to assess safety risks. However, the data taxonomy, has been designed to cover as much of the data landscape of organisations as is reasonable.

4.1 Analysis of results

The three accidents have been analysed with a view to evaluating if a BD approach would have in any way helped in the mitigation of the events.

The analysis in section 3.3 has assumed that all of the data sources identified both exist and are available for use as BD candidates. Availability of data is a critical issue particularly where there is a perceived safety or security issue related to the data. At present, the rail industry in the UK is working through the implications of opening up data to facilitate its use in a similar way to those envisioned in this paper. Where safety is the focus there will no-doubt be tight controls placed on the access and use of this data. This could therefore limit the effectiveness of any BDA approach.

The tables in Section 3.2 contain a qualitative assessment of the ‘BDness’ of each of the accidents. This has simply been taken to be the percentage of high (H) scoring data items from all of the possible data items (low [L] + H) for the three V’s, Volume, Velocity and Veracity. The remaining 2 V’s, Variety and Value, have been taken to be related to the;

1. input data (Variety of data sources), and,
2. output of the assessment of ‘BDness’ (Value of this assessment).

A level of 50% ‘BDness’ has been set as the point at which a rating for an accident greater than this means that the accident could have benefitted from BDA. The 50% level being chosen simply to represent the point at which there are potentially enough data items available to undertake a more sophisticated analysis.

Using this approach the following issues stand out from the analysis.

1. The Platja de Castelldefels accident is a an example of a ‘systematic failure’ (discussed below) and clearly stands out as having a much higher ‘BDness’ (63%) compared to Santiago de Compostela (33%) and Hatfield (33%). The Platja de Castelldefels accident is therefore firmly in the BDA arena with not only more data sources but also a ‘BDness’ rating over the 50% level. The variety of data sources also exhibit a greater potential for real time data analysis.
2. The Santiago de Compostela and Hatfield accidents are more of ‘random failure’ (discussed below) nature and contain more limited data sources, with the most effective data elements being remote monitoring. Most of the current activities in the UK rail market are focussed on improving remote monitoring of assets which will obviously help in preventing accidents similar in nature to these. However, under the definition of BD laid out in this paper, this approach

would not be considered amenable to BDA. Rather, these are examples of "business as usual" analysis.

What then is the demarcation between business as usual and BDA? Random Failures are probably adequately dealt with by traditional safety approaches which are based upon reliability engineering principles. Condition monitoring using sensors which monitor temperature, vibration, strain, displacement, visual attributes and so on, are all well understood in the railway [18][17].

Systematic failures are the majority contributors to hazards and are generally much more complex and harder to predict. Many are concerned with failures and errors of humans such as operators, designers and management.

One of the reasons there are Quality Management Systems (QMS) is to control systematic error [19]. Because railways have the potential for very high consequence accidents with multiple fatalities there is a need for rigorous Safety Management Systems that figuratively sit on top of the QMS. There are many good examples of Railway SMS in the public domain [20].

The analysis presented in this paper has indicted that a BDA approach can potentially help in the identification of heightened safety risk where the hazard is of a systematic nature. If all of the data associated with the Platja de Castelldefels accident had been available, and a real-time analysis regime based around these data sources has been put in place, then it is possible that the risk of this accident could have been significantly reduced.

4.2 A Proposed Enhancement to Safety Risk Analysis

As a result of the analysis of the three selected accidents a new approach to safety risk modelling is proposed. This new approach is built around the Bow-Tie risk assessment methodology and specifically incorporates BD into the risk assessment activities.

The Bow-Tie methodology is used extensively in the oil and gas market sector and has recently started to be used within the UK rail environment and research into the use of Bow-Ties for human factors in rail is currently being undertaken [10]. The BD approach presented in this paper both enhances the traditional Bow-Tie risk analysis activities and as well as providing a mechanism for 'real-time' or 'near real-time' assessment of safety risk profiles as shown in Figure 3 below which depicts the ELBowTie approach. The Bow-Tie analysis is enhanced through linking relevant enterprise data sources to each of the elements of the Bow-Tie. This in itself is a useful exercise as it can help in identification of additional risk factors over and above those picked up during more traditional hazard identification activities. Once the data sources have been linked to the Bow-Tie elements, resulting in the "Enterprise data taxonomy **L**inked Bow-Tie" (ELBowTie), they can be used in evaluations of, for example, effectiveness of control measures or to look for precursors to accident causes. Use of the ELBowTie provides a formal approach to allocating data to causes, consequences and risk mitigation measures which is set within a big data framework.

The ELBowTie also provides a mechanism for potentially linking to real-time data sources and facilitating development of a real-time risk-dashboard for rail operators. The purpose of this real-time BD analysis is to flag up in a predictive manner hazards on the railway that could propagate into accidents. The setting of the flagging levels and establishing the baseline levels of "normal" operation will no doubt be challenging. If data is available and not acted upon there is also a potential exposure to claims of negligence. Implementation of the dashboard therefore requires careful consideration. The ELBowTie model is illustrated in Figure 3.

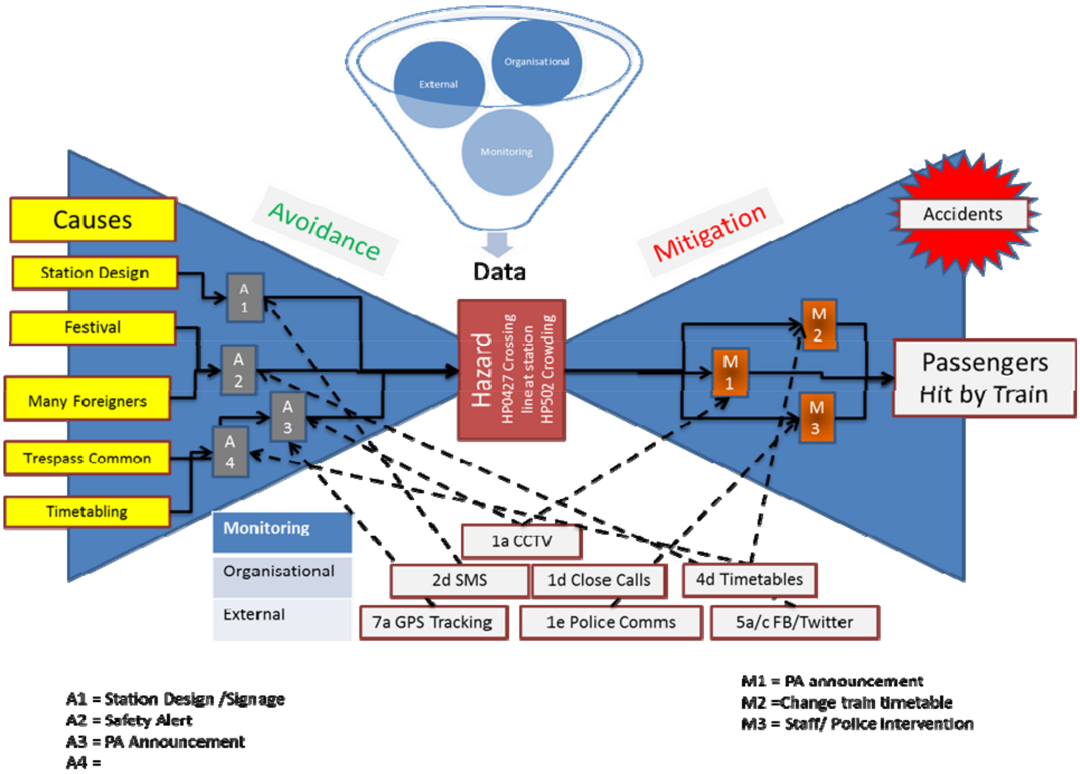

Figure 3: ELBowTie Diagram

The top of the diagram (i.e. the main Bow-Tie) 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 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 Bow-Tie components such as, threat, controls, outcomes and the recovery measures.

The advantages of this approach are that it provides,

- a visualisation of the data sources that can be considered as part of the safety assessment,
- a means of evaluating if these data sources can be defined as BD under the terms of the definition outlined earlier in this paper,

- 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
  - 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 the safety assessment, 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.

5 General Discussion and Conclusions

Big Data was not on the map ten years ago and now it looks like it could be revolutionary for the rail industry. The current thrust in the UK is to move to a ‘digital railway’ using system engineering principles and 3D/4D models. As the design of new railways takes place there will be greater opportunity to integrate Big Data with a Systems Engineering approach providing timely and relevant to support intelligent design decisions. In the future, projects like these could also be initiated with a Big Data Analytics approach to ensure risk assessments are fully integrated into the business model for the railway. This would then form one aspect of a data driven model of the railway and one that would grow and remain relevant throughout the build and into the operation of the line.

This paper has given an insight into how such a Big Data approach can be introduced in the context of the railway safety management. A definition of Big Data as it applies to railway safety management has been developed in terms of the 5V model, and an enterprise data taxonomy has been suggested as a way of bounding safety data items.

Three major accidents have been reviewed and assessed to understand what their high level causation was due to. Each of the causes was then linked, using the enterprise data taxonomy, to sources of data that may have highlighted the risk of the hazard and its potential to propagate into a railway accident. The data sources have been evaluated according to the 5V model and assessed for their respective ‘BDness’ to provide a score that can then be interpreted as the potential for Big Data Analytics.

The analysis of the three accidents has shown that Big Data could be used in mitigating accidents where the causes are systematic and complex in nature. More traditional data analysis methods are seen to be more appropriate for mitigating accidents where the causes are random in nature. However, to fully capitalise on the associated Big Data Analytics will require open access to data sources. These initiatives must be supported by a no-blame culture needed to fully implement the emerging analysis practices.

An ELBowTie methodology has been described to illustrate a mechanism for introducing Big Data into safety risk assessments. This methodology also provides a means for linking real time data updates into the ELBowTie thus enabling a real time risk dashboard to be envisaged. It is intended to build up ELBowTies for the full range of C-Hazards from EN50126-2 [12] and the associated accidents that arose from the hazards to further develop and validate the approach.

It is hoped that the deliverables from the research presented in this paper therefore lead to a greater understanding of Big Data and the methods (analytics) that are needed to make improvements to railway safety management.

References

[1] L.Einav, and J. Levin, (2014) The Data Revolution and Economic Analysis, Stanford University.

[2] Big Data Analytics (S183) RSSB November 2014 Version 2.

[3] Digital Railway Programme, Network Rail http://www.networkrail.co.uk /Digital-Railway-Introduction.pdf

[4] BDRA Seminar Information: http://www.hud.ac.uk/news/2015/may /bigdatatheanswertoincreasedrailsafety.php

[5] P. Hughes, C. Van Gulijk, & M. Figueres-Esteban, 2015. Learning from text-based close call data. In Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015. p. 8.

[6] Seventh Framework Programme for ICT, European Funded BYTE Project: 2014 Deliverable D1.1: Understanding and Mapping Big Data, Final report

[7] N. Attoh-Okine, (2014) Big Data Challenges in Railway Engineering, 2014 IEEE International Conference on Big Data.

[8] G.D. Erdos, , (2004) The Evolution of Rail Safety IRSE Technical Meeting: Adelaide 29 Oct 2004.

[9] RSSB Annual Safety Performance Report 2014/15 http://www.rssb.co.uk /Library/risk-analysis-and-safety-reporting/2015-07-aspr-key-findings-2014-15.pdf

[10] C. Turner, et al..2015, Bowtie diagrams: A user-friendly risk communication tool - Contractor management case study. 5th International Rail Human Factors Conference.

[11] E. S. Yaneira, et al., 2013, Bowtie Diagrams in Downstream Hazard Identification and Risk Assessment, 8th Global Congress on Process Safety Houston, TX, April 1–4, 2012. American Institute of Chemical Engineers.

[12] EN50126-2 (2007) Railway applications -The specification and demonstration of RAMS - Part 2: Guide to the application of EN 50126-1 for safety

[13] M. Neale, , (2007). The Causes of Accidents, IMechE 2007, Published: 16 January 2007.

[14] G. Morse, RSSB Santiago de Compostela derailment. Produced by RSSB July 2014.
[15] Investigación del accidenteⁿº 0030/2010 ocurrido el 23.06.2010; Ministerio De Fomento. En El Apeadero De Platja De Castelldefels (In Spanish).
[16] RSSB Formal Enquiry Summary : Hatfield report and recommendations, 16 November 2004.
[17] R. Ngigi, et al., (2012) Modern techniques for condition monitoring of railway vehicle dynamics. Journal of Physics: Conference Series, 364. 012016. ISSN1742-6596
[18] A. M. Zarembski, , Some Examples of Big Data in Railroad Engineering, 2014 IEEE International Conference on Big Data
[19] CLC/prTR 50506-2 (20007) Railway applications - Communication, signalling and processing systems – Guide for EN 50129 – Part 2: Safety assurance
[20] European Rail Agency : http://www.era.europa.eu/Core-Activities/Safety /Safety-Management-System/Pages/Safety-Management-System.aspx