Asset Analytics
Performance and Safety Management
Series Editors: Ajit Kumar Verma · P. K. Kapur · Uday Kumar

Sarbjeet Singh
Alberto Martinetti
Arnab Majumdar
Leo A. M. van Dongen *Editors*

Transportation Systems

Managing Performance through Advanced Maintenance Engineering

Springer

Chapter 6
Big Data Analytics for Maintaining Transportation Systems

Ravdeep Kour, Adithya Thaduri, Sarbjeet Singh and Alberto Martinetti

Abstract Big Data Analytics (BDA) is becoming a research focus in transportation systems, which can be seen from many projects within the world. By using sensor and Internet of Things (IoT) technology in transportation system, huge amount of data is been generated from different sources. This data can be integrated, analyzed and visualized for efficient and effective decision-making for maintaining transportation systems. The key challenges that exist in managing Big Data are the designing of the systems, which would be able to handle huge amount of data efficiently and effectively and to filter the most significant information from all the collected data. This chapter will draw attention towards the present scenario and future projections of big data in transportation systems. It also presents big data tools and techniques and then presents one brief case study of BDA in each type of transportation system. In this chapter, a broad overview of Big Data definitions, its history, present, and future prospects are briefed. Several tools and technologies especially for transportation are pointed out for maintaining transportation systems. At the end of the chapter, a definitive case studies on each transportation area is demonstrated.

Keywords Big data analytics · Transportation system · Maintenance · Railway · Road · Aviation · Shipping

R. Kour (✉) · A. Thaduri · S. Singh
Division of Operation and Maintenance, Luleå University of Technology, Luleå, Sweden
e-mail: Ravdeep.kour@ltu.se

A. Thaduri
e-mail: Adithya.thaduri@ltu.se

S. Singh
e-mail: Sarbjeet.singh@ltu.se

S. Singh
Mechanical Engineering Department, Government College of Engineering and Technology, Jammu, Jammu, India

A. Martinetti
Maintenance Engineering Group, Design, Production and Management Department, University of Twente, Enschede, The Netherlands
e-mail: a.martinetti@utwente.nl

S. Singh et al. (eds.), *Transportation Systems*, Asset Analytics,
https://doi.org/10.1007/978-981-32-9323-6_6

6.1 Big Data

Various definitions of big data have been suggested to understand its extent. Some of them are given below.

6.1.1 *Definitions*

Gartner research defined big data in terms of three Vs i.e., Volume (growing rate of data), Velocity (speediness of data) and Variety (collection of diverse data types and their sources). Now, the other two more Vs have been attached to it, i.e., **Veracity and Value. Veracity is the** trustworthiness of the data. Value is the most important V of the big data means to extract value out of it. Figure 6.1 illustrates 5 Vs, which remained a challenge for big data handling.

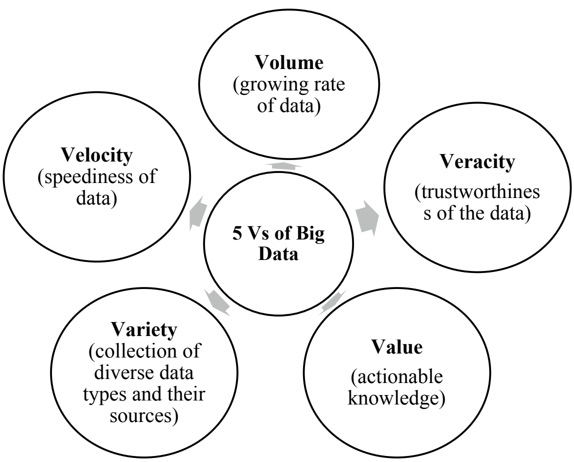
According to NIST (2015), Big Data consists of "extensive datasets primarily in the characteristics of volume, variety, velocity, and/or variability that require a scalable architecture for efficient storage, manipulation, and analysis".

Katal et al. (2013) defined Big Data as huge amount of data, which needs new technologies and architectures so that it becomes possible to extract value from it by analytical process.

Big Data refers to "a dataset whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyika et al. 2011).

Big data means "the broad range of new and massive data types that have appeared over the last decade or so" (Davenport 2014).

Fig. 6.1 Vs of big data as main challenges of treating big data



6.1.2 History

The history of Big Data started in 1663 when John Graunt conducted statistical-analysis experiment to curb the spread of the bubonic plague in Europe. In 1865, Richard Millar Devens, used the term "Business Intelligence" in the Encyclopaedia. Then, in the year 1926, Nikola Tesla predicted that humans would be able to access and analyze huge amount of data using a small pocket device. Afterwards, in 1928, Fritz Pfleumer created a method for storing data magnetically. In addition to this, there is a very interesting estimation made by Fremont Rider in 1944 that the Yale Library in 2040 would have 200 million volumes, which will occupy over 6000 miles of shelves. In 1961, Derek Price published the growth of scientific knowledge by looking at the growth in the number of scientific journals and papers. Afterwards, in the year 1965, US government setup the world's first data center to store 742 million tax returns and 175 million sets of fingerprints on magnetic tape. Then in 1970, EF Codd developed Relational Database Model for storing data. In the year 1991, there is birth of internet and Google launched its Search Engine in 1997. In 1999, there is first use of term Big Data in an Academic paper along with first use of term IoT. Afterwards, in 2005, Hadoop came into existence.

The first use of "cloud computing" in its modern context occurred in 2006. Then, the first study to estimate and forecast the amount of digital data created and replicated each year was conducted in 2007. In the year 2010, Eric Schmidt, executive chairman of Google, told in a conference that as much data is now being created every two days, as was created from the beginning of human civilization to the year 2003. Then, in 2011, the McKinsey report stated that in 2018 the USA alone will face a shortage of 140,000–190,000 data scientist as well as 1.5 million data managers and, then IBM introduced a Twitter hashtag, #IBMbigdata, in the same year. In 2012, the Obama Administration announced a USD 200 million investment to launch the "Big Data Research and Development Plan". Then in 2013, there was World IPv6 launch day and Android Apps has provided more than 650,000 applications, covering nearly all categories. Then in 2014, there was an introduction of SAP HANA and adoption of Cloud ERP. Then, the year 2015 was the Year of IoT and Smart cities. Year 2016 was the year of biggest ever data (90% of the world's data created in the past 2 years alone). In the last, the future of big data according to experts will be an estimated 4300% increase in annual data generation by 2020. The timeline shown in Fig. 6.2 depicts the summarized history and the evolution of Big Data.

6.1.3 Present Scenario

The key challenges that exist for the IT Professionals in managing Big Data are the designing of the systems, which would be able to handle huge amount of data efficiently and effectively and to filter the most significant information from all the collected data. In other words, we can say adding value to the business. Opportunities

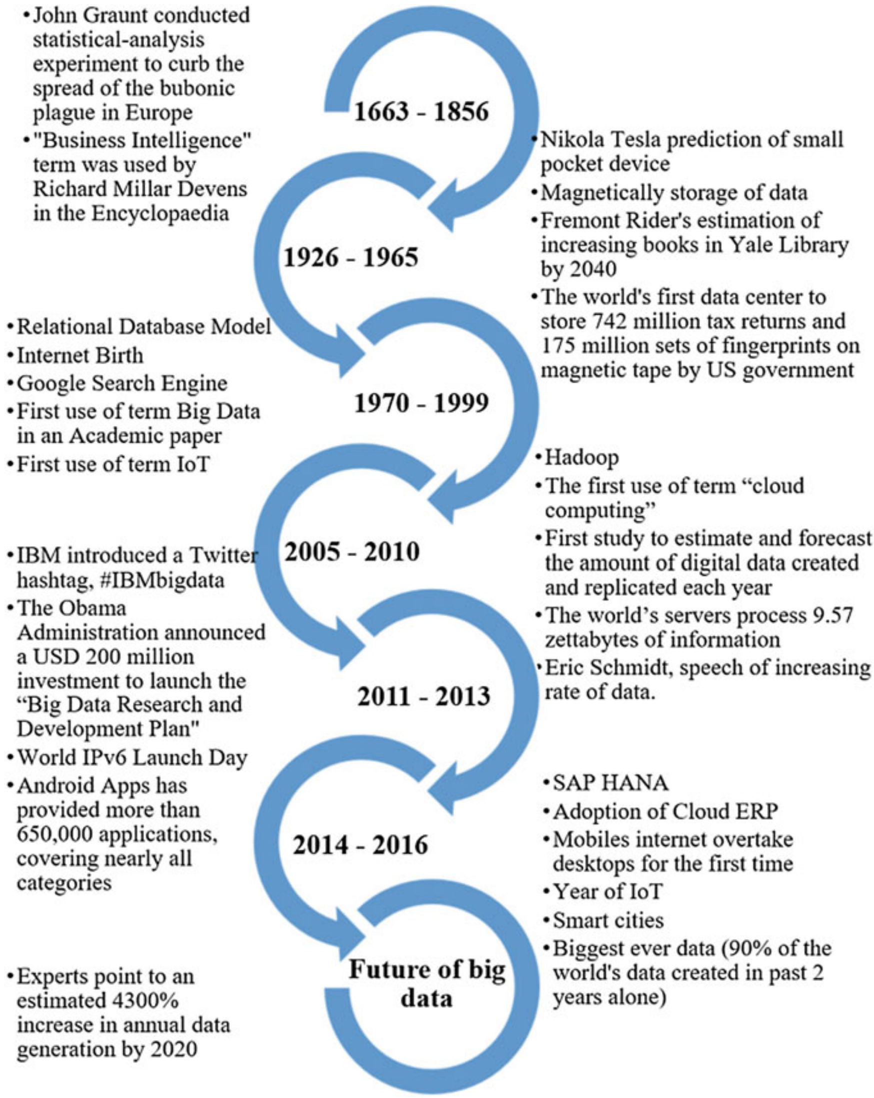


Fig. 6.2 Big data evolution timeline

are always followed by challenges. On the one hand, Big Data brings many attractive opportunities but on the other side, we are also facing many challenges. Big Data challenges include storing and analyzing huge, fast-growing, diverse data stores, then deciding specifically how to handle that data in the best way. Transportation systems like, railways and aviation are very complex systems, so they are organizationally divided into a number of units, such as directorates, sectors, services, etc. To meet their specific needs, each organizational unit uses their own designed applications and databases.

For example, in case of railway transportation (Fig. 6.3), there are variety of diverse databases\systems of rolling stock, track, environmental data and human intervention in Swedish railway infrastructure for different types of data and information (Thaduri et al. 2015). All these data sources are mixture of structure and unstructured data. The most challenging job is to integrate and fuse these diverse data sources to get value out of these. Thus, companies like transportation industry use dedicated tools called BDA to deal with Big Data coming from roads and vehicles sensors, GPS devices, customer's applications and websites. Therefore, large amount of information from different types of sources can be handled by using Big Data techniques e.g. unstructured text, signaling and train data. In addition to this, Parkinson and Bamford (2016) introduced an enhanced ELBowTie methodology to make improvements to railway safety using Big Data analytics.

On the other hand, there are also diverse data sources in aviation transportation. These data sets do not have the standardization or uniformity. Figure 6.4 shows the primary aviation data sets.

A study by Federal Aviation Administration (2013), states that a jet engine generates data equivalent to 20 TB annually. Since most of this data is unstructured, so are not used for any analytics purpose.

In the given present scenarios, it is a very challenging task to utilize different data sources from different stakeholders to form a fused data set and to do analytics on unstructured data. By using existing traditional techniques, it becomes very difficult to perform effective decision-making for maintenance of these assets due to such large size of diverse data. Thus, BDA can solve these problems. Therefore, we can

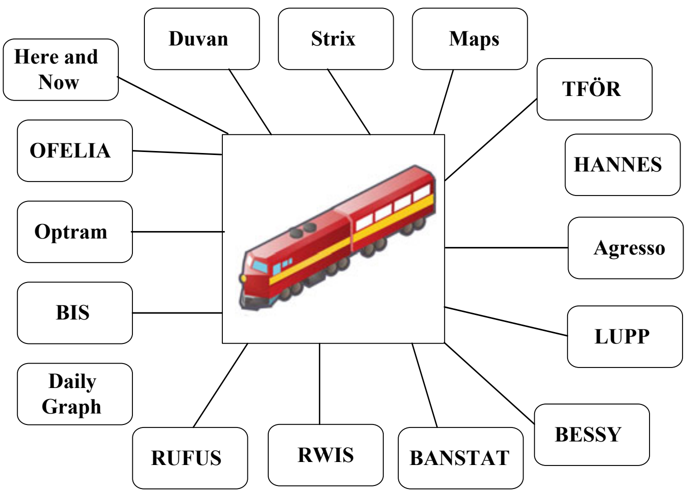


Fig. 6.3 Swedish railway infrastructure data sources (reproduced from Thaduri et al. 2015)

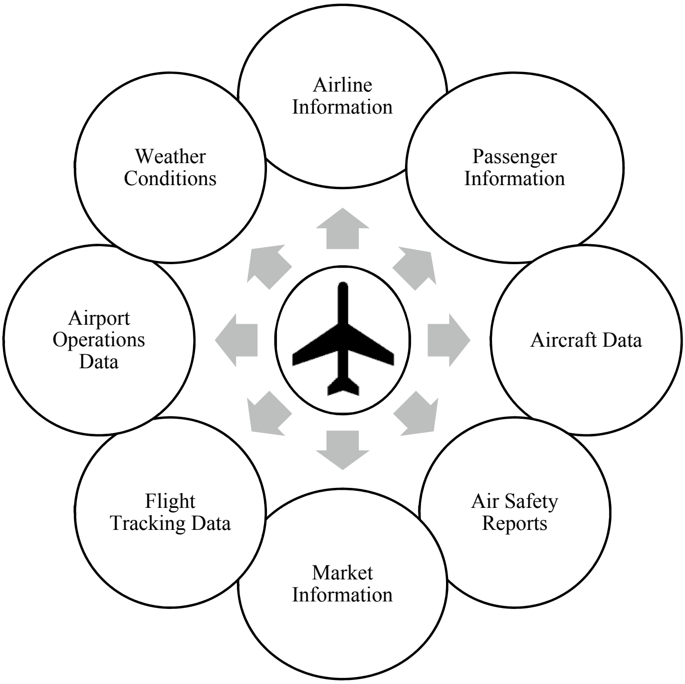


Fig. 6.4 The primary aviation data sets

use BDA to fuse heterogeneous datasets and to predict the fault in the component by analyzing data obtained from the aircraft sensors.

Talking about maritime transportation, marine data is also growing exponentially from last decade, which forms marine big data (Fig. 6.5).

On the one hand, there are very useful values hidden inside this marine big data like a tsunami and red-tide warning, prevention, and forecasting, disaster inversion,

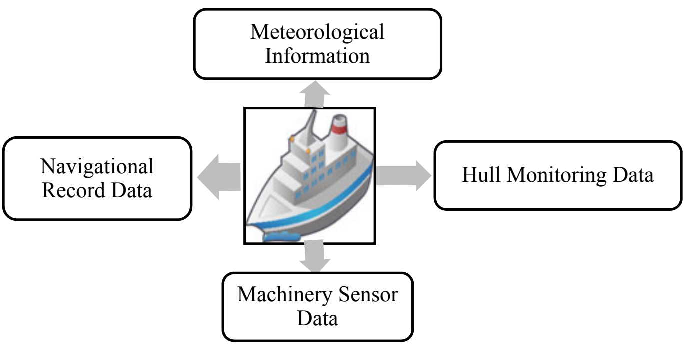


Fig. 6.5 Maritime big data sets

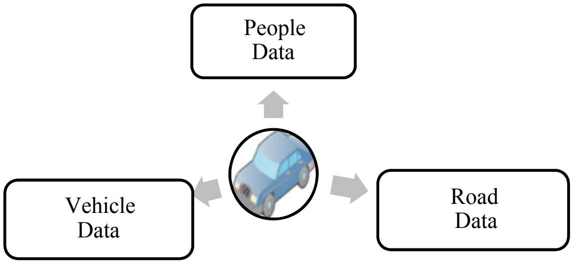


Fig. 6.6 Road big data sets

and visualization modeling after disasters (Huang et al. 2015). On the other hand, marine big data also brings number of challenges like Data Management, Data Transfer, Data Integration, Data Quality, Data Ownership, Data Protection, Adoption and Standard Management, Cybersecurity, Business Model and Human Factor (Zaman et al. 2017). To tackle these issues, researchers are presenting solutions like data management architecture (Huang et al. 2015), the hybrid cloud storage for marine big data (Dongmei and Du Yanling 2014), online ship vetting information system (SVIS), SmartPort Logistics, and ClassNK-NAPA GREEN Solution.

The road transportation data have been exploding over the last few years, and we have truly entered into the era of big data for road transportation. Intelligent Transportation System (ITS) big data in case of road transportation consists of data from three sources—people, vehicle and road (Xiong et al. 2014). Figure 6.6 shows three types of data sets for road transportation.

People data generally define driving, paying and travel behavior. Vehicle data generally define vehicle information, real-time location, operation, and crowdsourcing road conditions. Road data generally define the high-resolution satellite images, aerial photography data, and geometry of road networks and characteristics of road infrastructures. Researchers are using these big data for the traffic flow prediction (Lv et al. 2015) and for designing novel intelligent transportation cloud platform (Xiong et al. 2014).

Henceforth, all the transportation systems are generating huge amount of data from various different data sources and we need to analyze these data to get value out of these data. These data can be used for the prediction of maintenance activities to avoid unnecessary work orders.

6.1.4 Future Projections

The volume of big data is growing day by day and will reach an estimated 4300% increase in annual data generation by 2020 (Fig. 6.7). Technologies like artificial intelligence, prescriptive analytics, and real-time streaming will play a bigger role in Big Data solutions in the future.

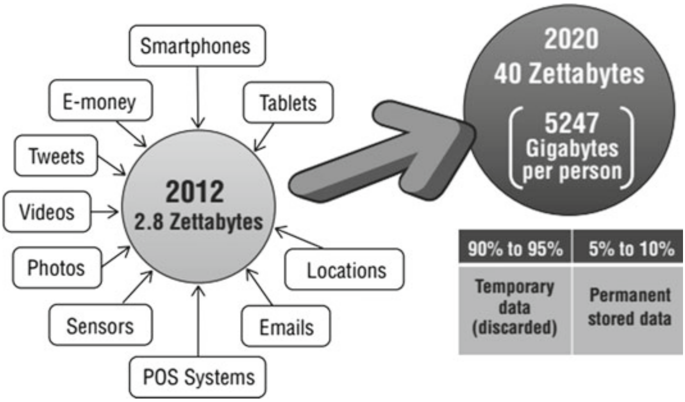


Fig. 6.7 Volume of big data in the future (reproduced from Takikawa 2016)

These BDA techniques can be exploited on the Big Data related to transportation maintenance to achieve early maintenance activities. Figure 6.8 depicts the use of Big data in transportation systems for data visualizations for effective decision-making. Maintenance data related to transportation systems is acquired using IoT technology (RFID, smart sensors, etc.) and send to the cloud via the internet for BDA to trigger early maintenance activities.

With the exploitation of BDA, proper information logistics can be achieved. In addition to this, Big Data Management includes putting the right people, policies, and technologies in place to ensure the integrity, security, quality, and availability of large stores of data.

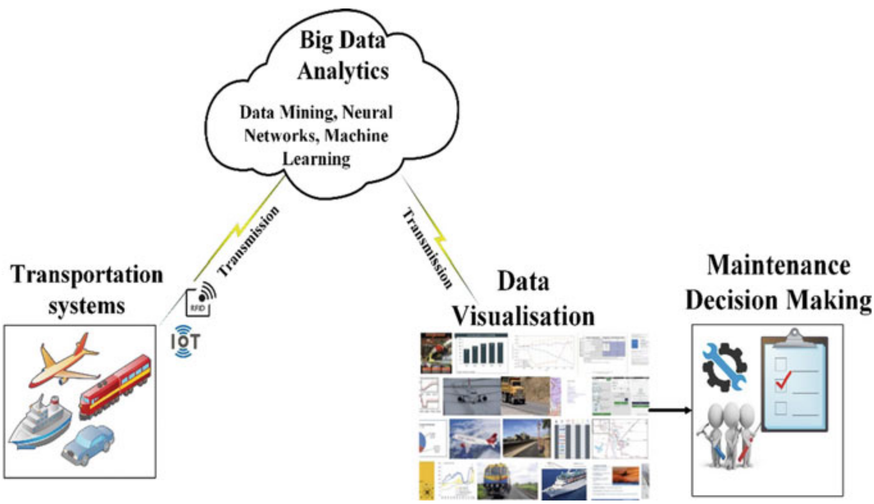


Fig. 6.8 Big data analytics in transportation systems

For the transportation systems, we can use BDA to perform predictive analytics, forecasting and prognosis and the estimation of the remaining useful life of an asset and make decisions based on the analysis of huge amounts of data.

6.2 Big Data Analytics Tools and Techniques for Transportation

Transportation Systems are complex systems and depend on huge and expensive networks of infrastructure and equipment. Maintaining these assets to achieve maximum availability is crucial for increasing revenue and other costs. Unplanned service interruptions can result in lost safety, reliability, maintainability, operational efficiency, capacity as well as improved passenger experience.

We can use Big Data Analytics to predict the fault in the component by analyzing data obtained from the sensors in real-time and reduce unplanned service delays to improve service levels. Some of the big data tools and technologies used to store, clean, integrate, manage, analyze and visualize big data are shown in Fig. 6.9. There were several applications of Big Data in different forms for maintenance in transportation systems. Broadly, they can be categorized into different types; operation or maintenance, types of analytics, application of Big Data analytics, etc. (Ghofrani et al. 2018). First of all, the big data sources of any transportation systems can be of different categorized as in Table 6.1.

The Big Data models and techniques for some applications are summarized in Table 6.2.

6.3 Big Data Analytics for Transportation Systems: Case Studies

6.3.1 Railway

Track Geometry being is one of the most important parameters/indicators to define the track system performance of railway infrastructure. Each country has been collecting huge amounts of data from the measurements cars to acquire all parameters at each specific location that defines the big data. To make the best use of this data, Sharma et al. (2018) developed a data-driven methodology for carrying out effective inspection and maintenance of track geometry for North America railroad transportation. They calculated track quality index (TQI) for each section and isolated defects using random forests to predict probability of occurrence. They also have developed Markov decision process for track geometry maintenance planning.

Different countries use different definitions for measuring TQI of the track. It is a combined indicator of track geometry parameters; longitudinal level, alignment

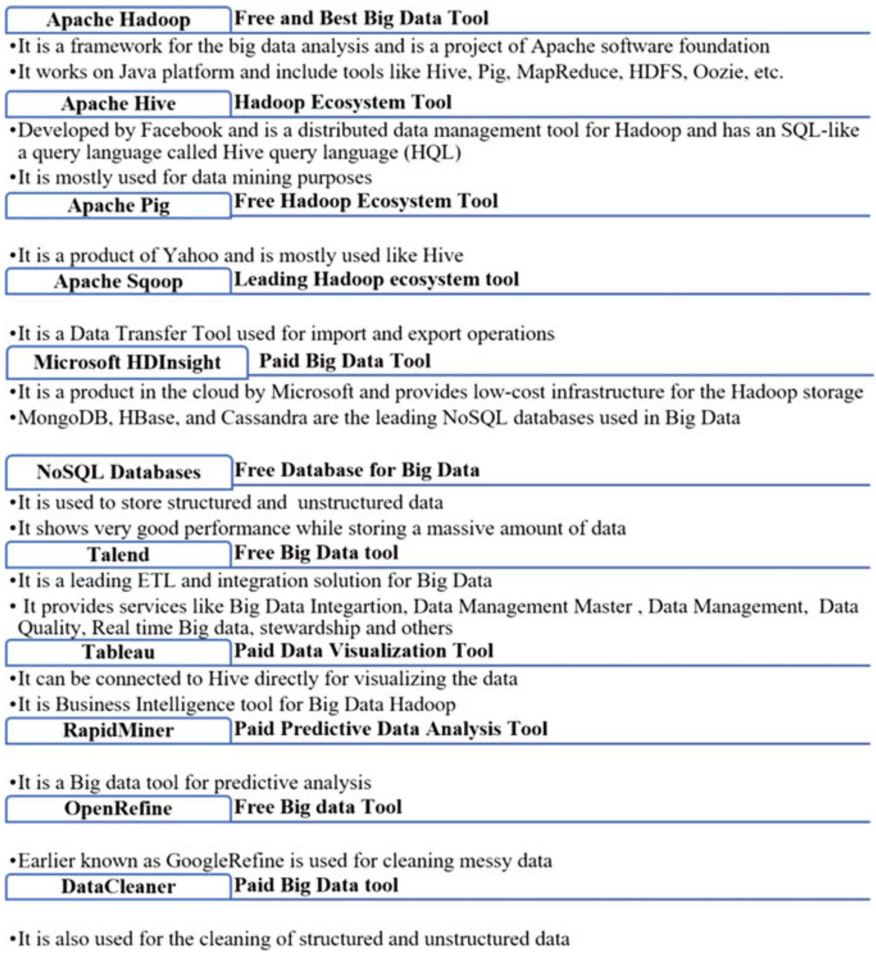


Fig. 6.9 Big data tools and technologies

Table 6.1 Types of big data source for transportation systems

Big data sources	Content types
Vehicle data	Type of vehicle manufacturer, specifications, loading data, etc.
Traffic control data	Time spent in stations, delay in stations, time on path, etc.
Location data	Vehicle positions and stations positions
Timetable data	Start and stop times, train routes
Financial data	Ticket sales, passenger check-in and check-out
Weather data	Temperature, rainfall, floods, etc.

Table 6.2 Application of big data techniques for transportation systems

Technique	Reference	Application
Association	Mirabadi and Sharifian (2010)	Accident analysis in Iranian Railways
	Sammouri et al. (2013)	Floating data sequences for predictive maintenance
	Ghomi et al. (2016)	Severe injury factors at railway crossing accidents
Clustering	Hughes et al. (2015)	Text-based closed call data
	Shao et al. (2016)	Railway accident analysis
	Su et al. (2016)	Railway condition based maintenance
Image processing	Giben et al. (2015)	Railway track images recognition and segmentation
	Jamshidi et al. (2017)	Railway failure risk assessment
Classification	Yin and Zhao (2016)	Fault diagnosis for railways
Artificial neural network (ANN)	Yu et al. (2007)	Fuel cell power conditions of railway vehicles
	Yilboga et al. (2010)	Failure prediction of railway turnouts
Support vector machine	Hu and Liu (2016)	Track geometry degradation modeling
Bayesian network (BN)	Bearfield et al. (2013)	Change of track and safety of railways
Time series models	Stratman et al. (2007)	Structural health monitoring of railroad vehicles

level, cant, twist, and gauge. In the model developed by Sharma et al. (2018), at first they calculated TQI using aggregated parameters and then incorporated isolated defects at the later stage. The flow chart of the study is illustrated in Fig. 6.10.

From the measurement cars, the inspection data is gathered. In the developed methodology, a Markov process models the rail system's deterioration, and the spot geo-defects are predicted by a random forest. The track geometry maintenance mainly follows condition-based maintenance for degradation process by doing necessary preventive and corrective actions appropriately based on the track geometry parameters and indicators. Each activity bares system Reliability, Availability, Maintainability and Safety (RAMS) parameters, and respective maintenance and inspection direct and indirect costs. Though the optimization of these parameters provides specific recommendations for maintenance planning, these alternatives have to be separately assessed by risk parameters. Being the track geometry follows the degradation process and needed optimized for maintenance planning, a Markov decision process is developed. Support vector regression has been used to develop prediction model for localized defects. Three alternatives are considered for their study and they

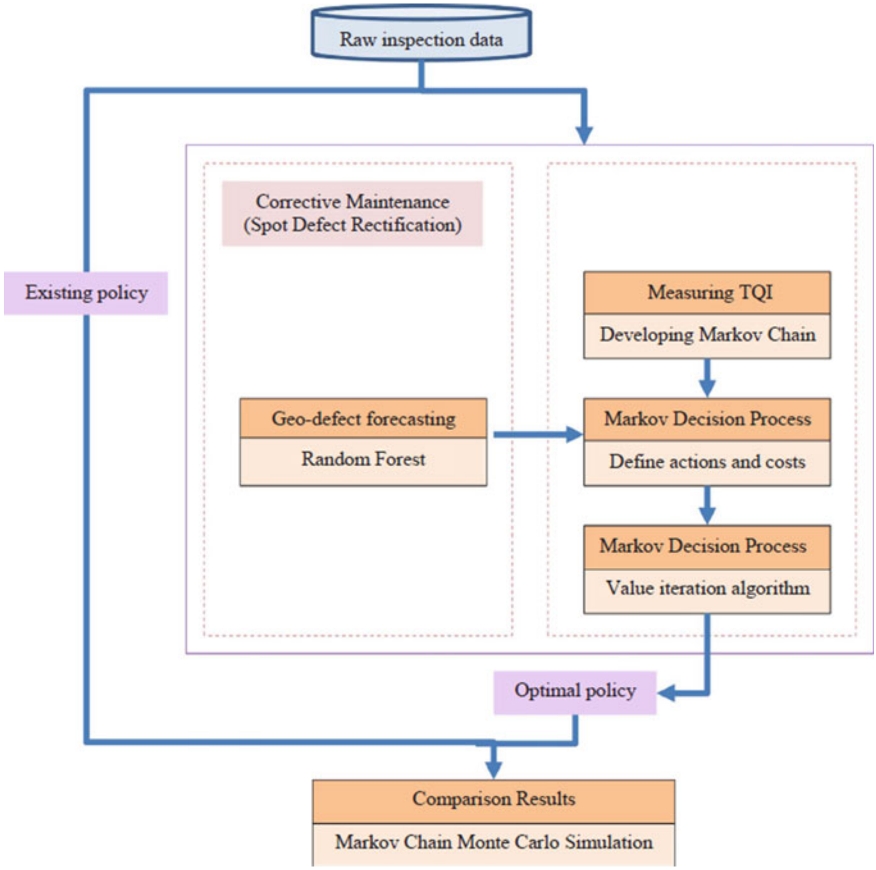


Fig. 6.10 The flow chart for track geometry maintenance planning (Sharma et al. 2018)

have calculated TQI parameters and localized defects and optimized with required performance and maintenance budget. These are shown in Fig. 6.11.

A = 0, TQI is the same or increases over the inspection interval, then no action
A = 1, TQI slight improvement after successive inspections, minor maintenance needed
A = 2, TQI major drop in the inspections, major maintenance needed.

By their simulations, it was concluded that low levels of TQI lead to high risk of track failure due to geo-defects. Depends on the alternative maintenance policies, they obtained optimized inspection plans with reducing cost savings using the application of Big Data.

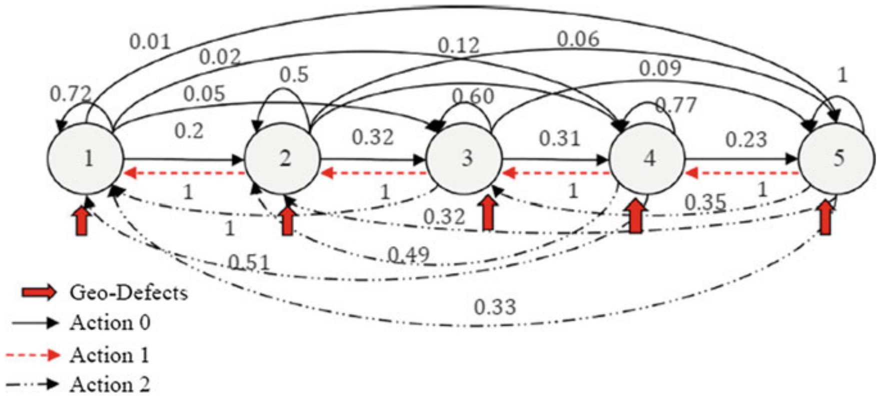


Fig. 6.11 Proposed Markov decision process for 3 actions (Sharma et al. 2018)

6.3.2 *Aviation*

Daily, there are innumerable aircrafts are departing and landing at several airports across the world. They are generating huge amounts of data from different origins of operation and maintenance data that will be useful to optimize their process, increase the efficiency, decreasing the risk, increasing the safety of the passengers and providing the best possible services to the customers. To accommodate a big data analytics of civil aircraft, Li et al. (2017) developed big data platform for civilian aircrafts provides decision-making support for civil aircrafts including maintenance plan, real-time alarm, health management, fuel saving, and airline schedule.

The three main sources of data are aircrafts, airports and maintenance depots. At first, there is a need to develop a cloud-based platform that gathers different kinds of data specific to time and location. The next step is to create an online communication link among these data sources using existing software, base stations, control centers, operational control, and air traffic control services. The third step is to collect relevant parameters from different data base sources such as ACARS (Aircraft Communications Addressing and Reporting System), QAR (Quick Access Recorder), DFDR (Digital Flight Data Recorder), Air Traffic Control (ATC), Airline Operations Control (AOC), Airline Approach Control (AAC) and Maintenance, Renewal and Operations (MRO). These parameters are; flight type, body specifications, flight altitude, Mach speed, elevation, engine monitoring parameters, fuel level, flight controlling parameters, maintenance actions, inspections, spare parts and in addition to passenger tickets and finance.

Li et al. (2017) developed architecture considering IaaS, PaaS and SaaS layers as shown in Fig. 6.12. IaaS consists of several types of devices that provide infrastructure to the platform namely, computing, storage, networking, and other devices. PaaS consists of different modules of data collection, data management, storage, analysis, computing and system services that provide the main core of the architecture. These modules further consist of several advanced submodules that are presently using in the

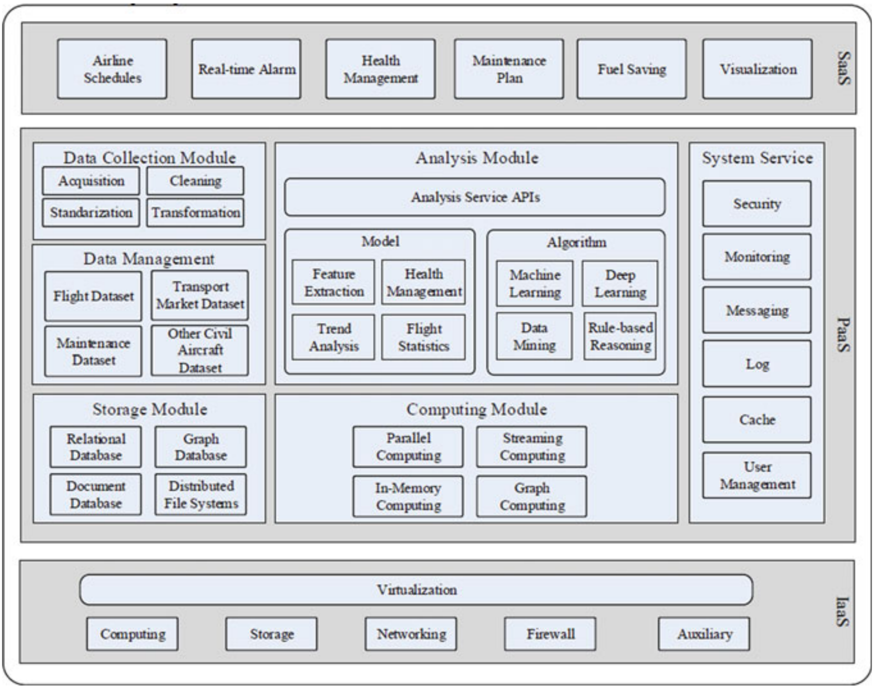


Fig. 6.12 Big data platform architecture for civil aircrafts (Li et al. 2017)

Big Data space. These submodules are relating to one another to achieve the purpose of the operator. Finally, SaaS provides major services for operation and maintenance of aircrafts such as airline schedules, real-time alerts, health management, optimized maintenance plan and visualization of several parameters as stated above. This above-developed platform will be helpful for setting optimized flight arrangement plan, aircraft fault diagnosis and health management, fuel savings and maintenance plan with developed Big data models and technologies with big data analytics. In the next version, they are looking to incorporate more features such as diagnostics, prognostics, deep learning and increase in security and safety.

6.3.3 *Maritime*

One of the Big Data Analytics solutions using predictive analytics to improve maritime safety and efficiency is RIGHTSHIP Qi (RightShip 2018). It is 24/7 available, robust platform to reduce maritime risk and improve efficiency. Figure 6.13 shows that Qi helps to determine the likelihood that a vessel will have an incident in the next 12 months. This platform also helps in showing GreenHouse Gas Emissions (GHG) rating for the world's fleet.

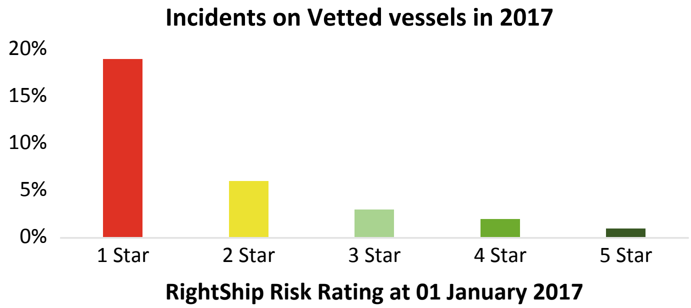


Fig. 6.13 RightShip predictive risk rating as at 1 January 2017 (reproduced from Rightship 2018)

The standard, European energy efficiency scale is used for vessel's GHG Rating. The relative performance of a vessel is rated from A through to G, the most efficient being A, the least efficient being G. There is also more statistics available on the website of RightShip. This platform uses BDA for predictive analytics and real-time assessment to improve maritime safety and efficiency. Qi also has the ability to analyses, integrate and compare big data to identify anomalies and trends in the data. In addition to this, RightShip helps in conducting the following inspections.

- **Dry Bulk Inspection**
  This inspection provides a validation of a vessel's condition and the application of its management system. The feedback after the inspection helps in improving its safety. This feedback includes deficiencies and recommendations along with a list of examples that present an overall impression about the management of the vessel.
- **Tanker Inspections**
  RightShip conducts Ship *Inspection* Report Programme inspections (OCMIF 2018) on oil, chemical and gas tankers throughout the world more quickly, efficiently and with minimal cost. All these inspections are conducted according to the latest OCIMF guidelines and after proper review process are uploaded to SIRE under the BHP Billiton Petroleum Ltd token.
- **Health And Wellness Assessment**
  RightShip has also developed a voluntary Health and Wellness Assessment. This assessment recognizes the impact that living and working standards have on seafarer wellbeing, work performance, safety, and employee retention.

6.3.4 Road

The road transportation data is also exploding over the last few years, and various BDA based methods have been proposed to predict the traffic flow for congestion control and maintenance of the road infrastructure.

A case study related to road management using big data has been done in Korea, which provides a direction for road safety analysis, and traffic information services. Korean Road Management Systems provide maintenance of bridges and road facilities along with the prevention of road slope collapses and traffic congestions (Chong and Sung 2015). Figure 6.14 shows the configuration of Korean Road Management Systems. ROAS (Road Occupation Access System), BMS (Bridge Management System), PMS (Pavement Management System), CSMS (Cut Slope Management System), and TMS (Traffic Management System).

HMS is the Highway Management System that has been established and operated. The types and usability of road data collected through the road management system are given in Table 6.3.

These road management systems provide data sets like maintenance records, crack rates, and grades of pavement status. By utilizing these data sets, it is possible to analyze the grade of pavement cracks to predict risks of road slope collapse and operate a disaster prevention system. In addition to this, it is also possible to analyze traffic congestion by utilizing various road data sets like traffic volume data.

These roads related big data are exploding over the time and, therefore, authors of this study had suggested road management systems and traffic/road information service based on road information (Fig. 6.15).

The expected results of this case study are

- more reliable prediction of road/traffic conditions,
- safe driving,
- enhanced quality of driving conditions,
- save time and
- reduce environmental pollution.

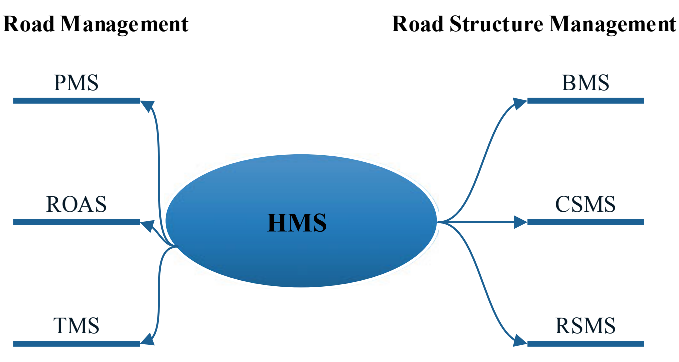


Fig. 6.14 Korean road management systems

Table 6.3 Usability of road data collected through a road management system (reproduced from Chong and Sung 2015)

Road management systems	Road information	Information usability
ROAS	Occupation spot, area, goal, and period	Analysis of risks depending on the occurrence of events
BMS	Bridge type, load, design loading, year of completion, length, width, and structural info.	Analysis of risks depending on the occurrence of events
PMS	Bridge type, load, design loading, year of completion, length, width, and structural info.	Traffic volume control based on the analysis of bridge risks
CSMS	Detailed investigation info., constant measurement info., maintenance status	Annual average volume of traffic, traffic volume of each vehicle type, traffic volume in each road section
TMS	Prediction of risks of road slope collapse	Analysis of traffic congestion in combination of various road data sets

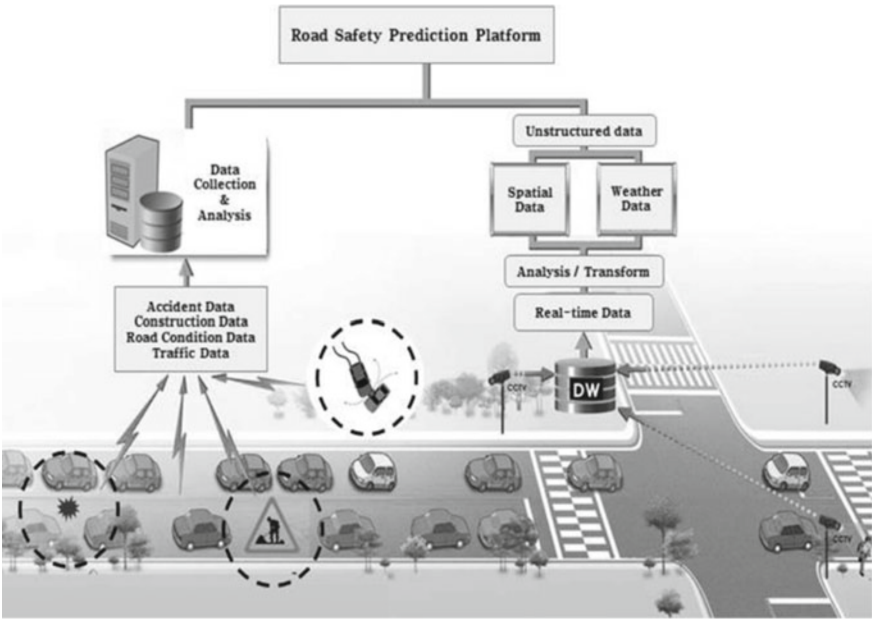


Fig. 6.15 Road safety prediction platform utilizing big data (reproduced from Chong and Sung 2015)

6.4 Conclusion

There are tremendous amount of opportunities that Big Data in transportation can solve some of the issues by effectively implementing in a structure manner. This application should be synchronized in an orderly manner in terms of big data structuring from data acquisition, big data architecture, data analysis, big data algorithms, and big data visualization. The above implementation can be beneficial not only to the maintenance of transportation infrastructure but also for other decisions in other infrastructures too. Based on the literature review and case studies, it is concluded that the application of Big Data architecture and analytics facilitates the usage of advanced data-driven methodologies for classification, prediction, and optimization of transportation systems w.r.t. increase in operational availability w.r.t. effective maintenance policies and strategies. One of the most challenges, still facing the transportation industry is to collect, combine, store, and process the big data so as to serve the industry needs. There are other hindrances such as data security, the involvement of different stakeholders of data owners and investment costs of big data platforms. If these challenges are handled in an effective way in the future, there will be more applications in the transportation industry to increase their performance to serve the societal needs.

References

Bearfield, G., Holloway, A., & Marsh, W. (2013). Change and safety: Decision-making from data. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 227*(6), 704–714.

Chong, K., & Sung, H. (2015, October). Prediction of road safety using road/traffic big data. In *The International Conference on Semantic Web Business and Innovation (SWBI2015)* (p. 23).

Davenport, T. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Harvard Business Review Press.

Dongmei, H., & Du Yanling, H. Q. (2014). Migration algorithm for big marine data in hybrid cloud storage. *Journal of Computer Research and Development, 1*(1), 199–205.

Ghofrani, F., He, Q., Rob Goverde, M. P., & Liu, X. (2018). Recent applications of big data analytics in railway transportation systems: A survey. *Transportation Research Part C, 90*, 226–246.

Ghomi, H., Bagheri, M., Fu, L., & Miranda-Moreno, L. F. (2016). Analysing injury severity factors at highway railway grade crossing accidents involving vulnerable road users: A comparative study. *Traffic Injury Prevention*.

Giben, X., Patel, V. M., & Chellappa, R. (2015, June). Material classification and semantic segmentation of railway track images with deep convolutional neural networks. In *Proceedings—International Conference on Image Processing, ICIP* (pp. 621–625).

Hu, C., & Liu, X. (2016). Modeling track geometry degradation using support vector machine technique. In *2016 Joint Rail Conference* (p. V001T01A011). American Society of Mechanical Engineers.

Huang, D., Zhao, D., Wei, L., Wang, Z., & Du, Y. (2015). Modeling and analysis in marine big data: Advances and challenges. *Mathematical Problems in Engineering*.

Hughes, P., Van Gulijk, C., & Figueres-Esteban, M. (2015, June). Learning from text-based close call data. In *Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015*, 7353 (p. 8).

Jamshidi, A., Faghih-Roohi, S., Hajizadeh, S., Núñez, A., Babuska, R., Dollevoet, R., & De Schutter, B. (2017). A big data analysis approach for rail failure risk assessment. *Risk Analysis, 37*(8).

Katal, A., Wazid, M., & Goudar, R. H. (2013, August). Big data: Issues, challenges, tools and good practices. In *2013 Sixth International Conference on Contemporary Computing (IC3)* (pp. 404–409). IEEE.

Li, S., Yang, Y., Yang, L., Su, H., Zhang, G., & Wang, J. (2017). Civil aircraft big data platform. In *2017 IEEE 11th International Conference on Semantic Computing (ICSC)* (pp. 328–333). IEEE Computer Society.

Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2015). Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems, 16*(2), 865–873.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*.

Mirabadi, A., & Sharifian, S. (2010). Application of association rules in Iranian Railways (RAI) accident data analysis. *Safety Science, 48*(10), 1427–1435.

NIST 1500-1. (2015). *NIST big data interoperability framework: Volume 1, definitions*. Available at https://dx.doi.org/10.6028/NIST.SP.1500-1.

OCMIF. (2018). Available at https://www.ocimf.org/sire/about-sire/.

Parkinson, H. J., & Bamford, G. (2016, April). The potential for using big data analytics to predict safety risks by analyzing rail accidents. In *3rd International Conference on Railway Technology: Research, Development and Maintenance* (pp. 5–8). Cagliari, Sardinia, Italy.

RightShip. (2018). Available at https://site.rightship.com/about-rightship/insights/#how-accurate-is-our-predictive-rating.

Sammouri, W., Come, E., Oukhellou, L., Aknin, P., & Fonlladosa, C.-E. (2013, October). Floating train data systems for preventive maintenance: A data mining approach. In *Proceedings of 2013 International Conference on Industrial Engineering and Systems Management (IESM)* (pp. 1–7).

Shao, F., Li, K., & Xu, X. (2016). Railway accidents analysis based on the improved algorithm of the maximal information coefficient. *Intelligent Data Analysis, 20*(3), 597–613.

Sharma, S., Cui, Y., He, Q., Mohammadi, R., & Li, Z. (2018). Data-driven optimization of railway maintenance for track geometry. *Transportation Research Part C, 90,* 34–58.

Stratman, B., Liu, Y., & Mahadevan, S. (2007). Structural health monitoring of railroad wheels using wheel impact load detectors. *Journal of Failure Analysis and Prevention, 7*(3), 218–225.

Su, Z., Nunez, A., Baldi, S., & De Schutter, B. (2016). Model predictive control for rail condition-based maintenance: A multilevel approach. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)* (Vol. 19, pp. 354–359).

Takikawa, M. (2016). Innovation in railway maintenance utilizing information and communication technology (Smart Maintenance Initiative). *Communication Technology*, 22–35.

Thaduri, A., Galar, D., & Kumar, U. (2015). Railway assets: A potential domain for big data analytics. *Procedia Computer Science, 53,* 457–467.

Xiong, G., Zhu, F., Fan, H., Dong, X., Kang, W., & Teng, T. (2014, October). Novel ITS based on space-air-ground collected big-data. In *2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1509–1514). IEEE.

Years, A. F. F. (2013). *Years 2013–2033* (p. 1). Federal Aviation Administration.

Yilboga, H., Eker, Ö. F., Güçlü, A., & Camci, F. (2010). Failure prediction on railway turnouts using time delay neural networks. In *CIMSA 2010—IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, Proceedings* (pp. 134–137).

Yin, J., & Zhao, W. (2016). Fault diagnosis network design for vehicle on-board equipments of high-speed railway: A deep learning approach. *Engineering Applications of Artificial Intelligence, 56*(October), 250–259.

Yu, X., Starke, M. R., Tolbert, L. M., & Ozpineci, B. (2007). Fuel cell power conditioning for electric power applications: A summary. *IET Electric Power Applications, 1*(5), 643–656.

Zaman, I., Pazouki, K., Norman, R., Younessi, S., & Coleman, S. (2017). Challenges and opportunities of big data analytics for upcoming regulations and future transformation of the shipping industry. *Procedia Engineering, 194,* 537–544.