Procedia Computer Science
Volume 53, 2015, Pages 457–467
2015 INNS Conference on Big Data

Railway assets: A potential domain for big data analytics

Adithya Thaduri[1], Diego Galar[2], and Uday Kumar[3]*

[1] Lulea University of Technology, Lulea, Sweden
adithya.thaduri@ltu.se
[2] diego.galar@ltu.se
[3] uday.kumar@ltu.se

Abstract
Two concepts currently at the leading edge of todays information technology revolution are Analytics and Big Data. The public transportation industry has been at the forefront in utilizing and implementing Analytics and Big Data, from ridership forecasting to transit operations Rail transit systems have been especially involved with these IT concepts, and tend to be especially amenable to the advantages of Analytics and Big Data because they are generally closed systems that involve sophisticated processing of large volumes of data. The more that public transportation professionals and decision makers understand the role of Analytics and Big Data in their industry in perspective, the more effectively they will be able to utilize its promise. This paper gives an overview of Big Data technologies in context of transportation with specific to Railways. This paper also gives an insight on how the existing data modules from the transport authority combines Big Data and how can be incorporated in providing maintenance decision making.

Keywords: Big Data, Railways, Maintenance, Transportation

1 Introduction

Recent technological advances have sparked what amounts to a revolution in the application of Big Data cognitive and informational tools. There is enormous amount of data that has been generated by the systems using sensors by data acquisition techniques. These type of data varies primarily by 5 Vs; volume, velocity, variety, value and veracity. The types of data can be of different characteristics structured, semi-structured and unstructured. Due to the advancements in sensor technology, data handling and storage and capabilities of faster computing algorithms, the usage of Big Data is quite predominant in several sectors. This huge amount of data that are collecting from the various sources has invoked several possibilities and insights for the future technology that will embrace a new wave of application areas.

Big data is defined by as data that includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable

*Masterminded EasyChair and created versions 3.0–3.4 of the class style

Selection and peer-review under responsibility of the Scientific Programme Committee of INNS-BigData2015

doi:10.1016/j.procs.2015.07.323

elapsed time [1]. It is quite difficult to define quantitatively the Big Data by numbers because the definition comes to the question, How big is big data? We are talking the data in terms of 100 Terra bytes, 10 GB/s from 100 sources and this figures will change because of increasing installation of Internet of Things (IoT), Wireless Sensor Networks (WSN), cloud based services and condition monitoring of heavy machines. Hence there is a need of Data mining capabilities to handle this massive amount of data for efficient storage and accessible information. What is the use of this massive data? That is where the importance of Big Data Analytics is emerged. IBM [2] defined the big data analytics as the use of advanced analytic techniques against very large, diverse data sets that include different types such as structured/unstructured and streaming/batch, and different sizes from terabytes to zettabytes. By analyzing the Big Data, the data analysts, researchers, technicians, operators and business users can make better understanding of the system and provide faster and better decisions that were inaccessible before. There are tools that are still applying in the research arena that can also be utilized for big data analytics are machine learning, natural language processing, predictive analytics, data mining, statistics and strategic management methods.

Big Data has been applying in several areas like computing, telecommunications, web and mobile services, manufacturing, process industries, transportation, scientific simulations, etc. The most prominent applications of Big Data as per McKinsey Global Institute analysis are healthcare, retail, transportation, manufacturing, public services and personal data [3]. It is also worth to mention that also, from the more operational point of view, the latest trends in data visualization focus on two different lines:Augmented Reality [4] and Infographics tools [5].

2 Transport Industry

In virtually any city, one is likely to see the results of analytics in the operation of transit buses and trains that are essential for maintaining the mobility of the metro area. Furthermore, it is useful for both the public and the industry to realize how significantly public transportation has been a leading pioneer in the rich and extensive historic development of these tools. There are some published studies over usage of Big data in area of transportation. It applied for Istanbuls automated fare collection system and pricing for BRT-Bus Rapid Transit line planning with visualization metrics to obtain better recommendations for consumers [6]. Using Markov-chain approach, a multi-modal transport network in London was developed with better information clusters for efficiency of transport [7]. City Intelligent Energy Network used statistical data including city economy, construction, population, and different energy parameters to develop the comprehensive model for low-carbon emissions [8].

Using Big data, several key indicators for transport was demonstrated to measure performance of cities [9] and for sustainability [10]. Big data processing and data mining has been used to develop architecture for traffic cloud data mining and optimization of strategies (TCD-MOS) and related data processing and network optimization methods [11]. There are other case studies such as big data analytics for safety management [12], an assistant decision-supporting method for urban transportation planning using GPS (Global Positioning System) data [13], crowdsourcing for intelligent transport system[14], crowd sourcing geo-social network [15], intelligent transport system for predicting drivers behavior [16], realtime monitoring from traffic for TomTom [17] and in Netherlands [18]. Several opportunities in public transport and processing techniques and their challengers [19] were described in transport domain [20].

The are endless possibilities of Big Data analytics in the transportation and one of the main areas is to look into maintenance aspects for maximum customer experience. It can identify behaviour of bottlenecks, maximum loads, variation in traffic, unplanned delay timings, inspection timings and accidents that will impact customer's comfort, business's losses and asset's reputation. The importance of maintenance in this context to utilize Big Data analytics to improve the asset's performance such as scheduling the windows when there is less traffic, do maintenance tasks at bottlenecks, rescheduling the assets with respect amount of traffic by passengers and enhance overall efficiency that can lead to reduce in operating costs.

3 Railways and Linear Assets

The European specific policies stated that the railway infrastructure managers should focus are more and more on reducing operational costs at the same time increase performances in financial assets and safety[21]. The main component of the system is a commercial CMMS (Computerized Maintenance Management System) that implements LAM (linear asset management) that allows the dynamic segmentation of the line, permitting to specify attributes and features along the railway lines. This system also allows the definition of Linear relationships indicating intersections, parallel or grade crossing of linear assets [22].

There have been studies on the application of big data in area of Railways. It is stated that although Big data is at nascent stage, for railways, both quality and quantity are going to increase further time and several challenges of data analytics are discussed [23]. EIT (Enabling Innovation Team) are discussing on future prospects in applying of Big Data in railways [24]. Even the top business leaders of Hitachi were also brainstormed on effective utilization of big data in railways [25]. One of the applications of maintenance on railways was carried out by Dutch railways on Axle Box Acceleration (ABA) measurements with 1 terabyte of track degradation data for performing adaptive and self-learning mechanisms [26]. Big data was applied in Utrecht, Netherlands to handle the traffic and explain the usage of mobile phones, smart cards and computers to predict the traffic and improve operations accordingly [18]. The maintenance of railways was pointed out on application by using big data by Markov state classification [27]. The metaheuristics can be seen as sophisticated and intuitive methods which mimic natural phenomena and explore the solution within a feasible region in order to achieve specific goals and applied in railway engineering [28]. The implementation of a railway asset monitoring system based upon semantic data models that offer greater capabilities for data integration, extensibility, and compatibility over traditional approaches was carried out for railway asset management [29]. The use of support vector machine (SVM) technique that effectively utilizes large-scale data and provides valuable tools for operational sustainability as described for alarm prediction in railways [30]. The typical coomunication protol for European railways is shown in Figure 2. There are several data sources generated from the railways such as timings, speed, number of passengers, etc., from different trains dynamically and that constitute big data.

The future of railways with Big data and the Internet of Things will allow transportation modes to communicate with each other and with the wider environment, paving the way for truly integrated and inter-modal transport solutions by Arup Report [31]. Rail travel has to become a safer, cheaper and more efficient means of transport, as well as a source of revenue generation. This is why Big Data solutions have been designed to enhance business and travel experience, surpass the existing silos of systems and processes, drive innovation and build performance [32]. The prospects of using big data for railway management on handling large quantities of data was discussed [33]. For integrated maintenance analysis and perspective of innovation in railway sectors, there is a need of complex and centralized big data management to cope up with technological and engineering aspects [34]. Big data technologies for railway freight marketing decision by data acquisition, data preprocessing, storage and management using Hadoop was

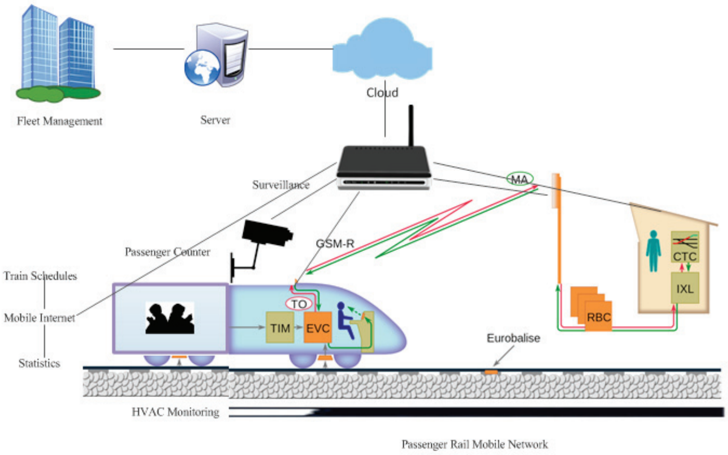

Figure 1: Communication in typical European railway transport management system

utilized [35]. The characteristics of increasing data volume related to equipment management of high-speed railway to summarize large datasets by using modern management methods to create a more complex situation of data management were demonstrated with emphasis on value and vital [36].

4 Swedish Railways

The Swedish Rail Administrator is one example of the current situation in Europe where Infrastructure Managers deal with data from hundreds of sources that are disparate in nature and have different granularity. The data can be segregated as structured (timings, speed, traffic), semi-structured (images and videos) and unstructured (maintenance records). This disparate sources can be handled by Big Data to reduce the barriers of data management.

This scenario is partially solved by subsystems like BIS where connectivity and data exchange are secured. However the amount of stored data is increasing exponentially and context engines that link data sets in sensical ways increases just linearly. Moreover the systems often store the wrong data from malfunctioning sensors or due to wrong tags like time stamps, GPS, etc., causing terrible consequences for the data quality and the further decisions taken based on them. For this purpose, data cleaning to remove or isolate non-sensical data is needed and anomaly detection algorithms working within the context engines will partially solve this issue. Therefore new tools and methodologies must be brought to the railway arena in order to establish meaningful associations and remove meaningless outliers within the gathered data for further exploitation and support in the future decisions.

4.1 Data management

For a linear asset such as a railway, much information needs to be captured and analyzed to assess the overall condition, maintenance, capital spending, and inspection of the railroad tracks. Examples of information that can be collected include track availability, use of track time, track condition, performance history, and the work performed. Measurements of the condition of the track typically include continuous and spot measurements from automatic inspection vehicles,

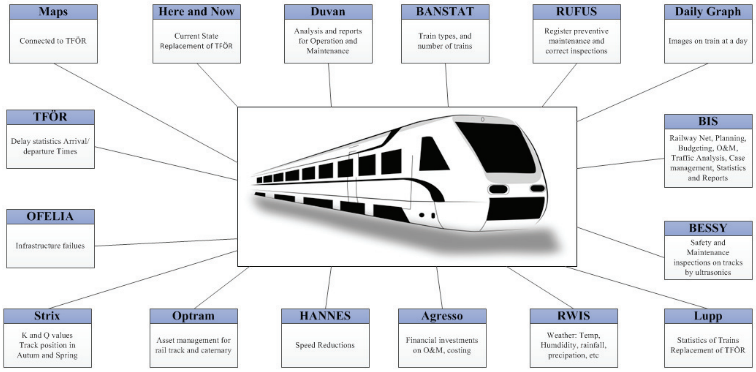

Figure 2: Big data management of Railways by Trafikverket

visual inspections from daily walking inspections, and records of in-services failures. Examples of conditions measured by automatic inspection vehicles include geometry car measurements (deviation from design curves, geometry exceptions to railroad standards, vehicle ride quality exceptions), rail measurements, gage restraint measurements, track deflection and stiffness measurements, clearance measurements, and substructure measurements.

Even with an accurate map of the corridor, rail, ties, and other corridor assets do not have any physical characteristics that lead to easy identification. Furthermore, problem areas for targeted maintenance often do not obey discrete physical boundaries such as beginning and end of a rail section. Transport Administration continuous collects large amounts of data about the railroad and rail traffic. This information is divided into a variety of databases/systems and it is not easy to get an overall picture of what information is available. There are a number of explanations for this problem. The meaning of the past becomes clear when one asks questions to users of a system. It is seldom able to present a comprehensive list of the type of information that you can get out of a particular database. Due to feasibility of data collection in the Swedish Railways where enormous amount of disparate data is collecting continuously by several sources from rolling stock, track, environmental data and human intervention, there are few challenges to be considered. One of them is the decision making for maintenance of these assets with this complex disparate data and finding the best solution is quite ambiguous. These challenges can be attained by using Big Data.

4.2 Traffiverket data modules

There are several systems in Swedish railway infrastructure for different types of data and information.

- TFÖR: TFÖR is the system that has been previously used to collect delay statistics of trains operated under service led by Banverket. Data is gathered on planned departure and arrival times per train number and per station, actual departure and arrival times for these stations and cause reporting the delays that were five minutes or greater. TFÖR decommissioned in 2010 and is replaced by "Lupp" and "Here and Now".

- BIS: The track information system BIS describes the whole railway net, where everything from the curves, the signals, catenary and switches is reported. The information is reported in the BIS is a snapshot, which means that the BIS cannot see how the infrastructure changed over time.

- Strix: Strix is a measurement carriage used to produce K - values and Q - values. Two dimensions that measure track position, which in turn is important for passengers' comfort. K - values and Q - values can be obtained in a very detailed level, but in this project we have only data on K - values and Q - values of a spring and autumn measurement per track portion.

- Optram is a web-based system for analysis of measurement data. Here measurements from Strix and other food are coordinated. Also some measurements regarding the catenary is compiled here. The system is used by managers and entrepreneurs to analyze asset information from track and catenary. The analysis is used to plan for an optimized maintenance.

- AGRESSO: Agresso contains information on how much money is spent on different sections of the track, accompanied by a chart of accounts. Here you can see how much is spent on operations, maintenance, reinvestments and new investments. Data are in current prices. The operation cost is difficult to distribute to track portion.

- BANSTAT: Banstat is a database (where the information is subject to confidentiality) from which data is retrieved using SQL queries. The information that can be obtained from the database are: number of trains of different types, number of GT (Gross Ton) for different train types and thus can also a number of train-km and number of gross tonne km be calculated. Banstat contain large amounts of data, but to get the details of the above, require a manual process that often occurs with in Excel.

- OFELIA: In Ofelia records disruptions in infrastructure plants, as symptoms of the bug reports. Then Ofelia infrastructure failure and problems are recorded and track contractors details of actual cause and what action is performed, the time of appearance and action and close the failure report.

- BESSY: BESSY is a system for safety and maintenance inspection of the Transport Administration (former Banverket) permanent facilities. There are also performed automated ultrasonic inspections to detect cracks and defects in the rails, switches and welds before they cause broken rails. The results of these ultrasonic inspections are stored in Bessy.

- Rufus (New): Rufus is a new system (2010), where operation and maintenance contractors should: 1) register the preventive maintenance measures, and 2) re-submit corrected inspection complaints arising in the inspection system Bessy. The goal of Rufus is to simplify the reporting of what action is performed and in conjunction with this change, it is thought that BESSY will focus on the implementation of the surveys.

- Daily Graph: Daily Graph contains graphic schedules with information on how the trains is moving in time and space. There are no digital numbers but only images. Daily graph is useful because it shows only the trains that run on a particular day, making it much easier to count the train than in a standard graphical timetable where all days of the week and variants for different time periods are included in a single image.

- Maps: In Maps, information was gathered on train services. Information on delays was brought from Tfor. Maps has now been replaced by Lupp.

- HANNES: HANNES gather information on speed reductions.
- RWIS: SRA system for weather information. RWIS includes point values from the measuring stations. The type of information collected is: air temperature, road surface temperature, dew point temperature, relative humidity, precipitation type, rainfall and wind speed.
- Lupp (New, replaces Train Plan, Opera, tto): Lupe is an internal system that also targeted at those who need to produce statistics for the trains. Lupp is from January 1, 2010 replacing TFÖR/SAMST and maps as a tool to produce statistics. These systems are phased out starting in april.
- Here and Now (New, replaces TFÖR): Here Now is aimed at companies who need to see the current state of the trains. The company can also do some reporting current train services. In the system, one can see how the situation is for a particular train, what the situation is at a certain place, and when delays occur, one can see any reason given for the problem. Here Now replaces as of November 29, 2009 TFRas a tool for managing the current situation.
- Duvan (New): Duvab is an analysis tool for studying the operation and maintenance. In the analysis, the system can perform searches based on final reports. The searches are based on information derived from systems Bessy, BIS, Ofelia and MAPS.

TRV (Trafikverket) has over the years have collected huge amounts of data, and it should be possible to achieve a lot with these data. The information is not readily available and it often requires a lot of database processing to set up the data sets so that the analysis can be conducted. There is a need of new services like forecasting and nowcasting techniques for the railway assets to increase capacity, punctuality and other performance indicators. These techniques are used along with the machine learning algorithms analysis the big data to capture the snapshots of requirement of maintenance activities at the earliest stages so that the asset will be safe and robust. The big data from all the sources from the trafikverket is shown in Figure 2.

5 The conceptual proposition: Discussion

Common taxonomies and ontologies to connect systems, description of asset hierarchies and events, tags like GPS or time stamps, harmonization in the reporting of maintenance and operation performed of the assets to be exploitable afterwards, integration of disparate data sources like video pictures sounds or information coming from people experience etc. Since railways is an infrastructure with linear and non-linear assets, there is need of defining the taxonomies and ontologies of several factors. There is need of defining of disparata data with ontologies and combined to form a common place for Big Data. Use of cloud based approach is best suited for this approach to analyze this data to perform maintenance actions. The goals of this application is to acquire big data for disparate sources, provide insights on big data analytics and join the taxonomies with suitable recommendation for maintenance activities from the railways standards data described in above section. Each of the data sources are maintained in cloud and utilized within process of data collection, data analytics and decision making. The decision making can be carried out as shown in Figure 3. There are three types of methods that can be performed from the railway asset; physics based, symbolic based and data-driven based methods. The hybrid modelling of these methods provide accurate diagnosis

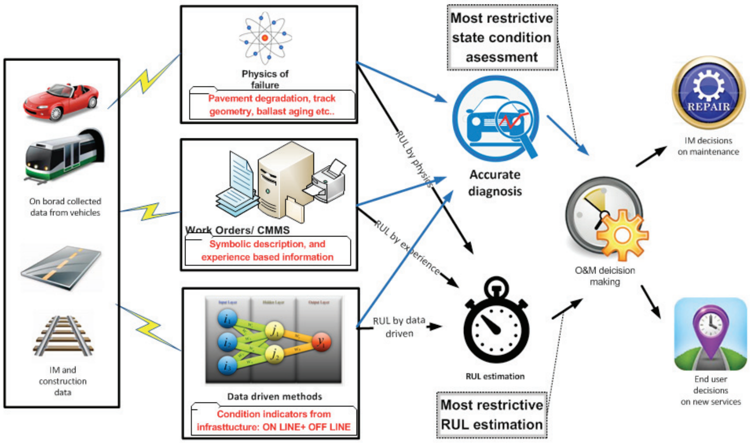

Figure 3: OM decision making for Maintenance using Big Data

of condition of asset and to determine remaining useful life (RUL). These two characteristics helps for decision making of operation and maintenance for end users.

6 Conclusion

Massive business, susceptible to incorporate the Big Data concept have not done yet, either for the lack of specific tools or the excessive cost to involve all the required stakeholders. One of these sectors is the maintenance of infrastructure assets and especially those related with railway transportation. This traditional sector presents many unique features from the Big Data point of view that is worth to cope with. The coming points justify that infrastructure assets maintenance and specially the case of railway systems comprised by railway network and vehicles is the case of choice to focus Big Data fundamental research due to the following challenges:

- Amount of data: hundreds of TB/day of different sources for the European railway system as a whole,
- Heterogeneous sources of information: such as dynamic conditions of the vehicles, geographic information, weather characteristics, results from maintenance tasks, etc.
- Real time requirements: for data analysis combined with years of records, scalability requirements, the problem cannot be easily divided in smaller problems since, the wearing of an infrastructure of hundreds of kilometers is dependent on the geographical layout, weather conditions, complete fleet actual characteristics, maintenance requirements, etc.
- Algorithms to predict wearing: current algorithms are limited to lab conditions but not applicable to real life conditions, algorithms based on Big Data are application specific and very incipient,
- Huge business opportunities: for companies in the SW analytics, maintenance and infrastructure operators estimated over 1billion per year in Europe.

Seems big data has huge potential in railway due to need of analytics and amount of information processed. However the introduction of information like physical models and experience based makes a more challenging scenario when it comes to a point of predicting the behavior of the assets in operation and maintenance with detailed emphasis on performance of assets.

References

[1] C. Snijders, U. Matzat, U.-D. Reips, Big data: Big gaps of knowledge in the field of internet science, International Journal of Internet Science 7 (1) (2012) 1–5.

[2] IBM, What is big data analytics?, http://www-01.ibm.com/software/data/infosphere/hadoop/what-is-big-data-analytics.html, (Visted on 2015-06-09).

[3] McKinsey-Global-Institute, Big data: The next frontier for innovation, competition, and productivity.

[4] R. T. Azuma, et al., A survey of augmented reality, Presence 6 (4) (1997) 355–385.

[5] H. Chang, Book review: Data-driven healthcare & analytics in a big data world, Healthcare informatics research 21 (1) (2015) 61–62.

[6] I. Gokasar, K. Simsek, Using big data for analysis and improvement of public transportation systems in istanbul.

[7] M. Faizrahnemoon, A. Schlote, L. Maggi, E. Crisostomi, S. Robert, A big-data model for multi-modal public transportation with application to macroscopic control and optimisation, International Journal of Control (2015) 1–28.

[8] M. Wang, J. Wang, F. Tian, City intelligent energy and transportation network policybased on the big data analysis, Procedia Computer Science 32 (2014) 85–92.

[9] R. Li, A. Kido, S. Wang, Evaluation index development for intelligent transportation system in smart community based on big data, Advances in Mechanical Engineering 7 (2) (2015) 541651.

[10] C. D. Cottrill, S. Derrible, Leveraging big data for the development of transport sustainability indicators, Journal of Urban Technology 22 (1) (2015) 45–64.

[11] J. Fiosina, M. Fiosins, J. Müller, Big data processing and mining for the future ict-based smart transportation management system.

[12] C. E. Otero, M. Rossi, A. Peter, R. Haber, Determining human-perceived level of safety in transportation systems using big data analytics, in: Proceedings on the International Conference on Internet Computing (ICOMP), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2014, p. 1.

[13] X. Xu, W. Dou, An assistant decision-supporting method for urban transportation planning over big traffic data, in: Human Centered Computing, Springer, 2015, pp. 251–264.

[14] K. F. Yan, Using crowdsourcing to establish the big data of the intelligent transportation system, in: Advanced Materials Research, Vol. 791, Trans Tech Publ, 2013, pp. 2118–2121.

[15] A. Z. Mohamed, A review on crowd sourcing geo-social related big data approaches as solution to transportation problem, in: Applied Mechanics and Materials, Vol. 663, Trans Tech Publ, 2014, pp. 622–626.

[16] H. J. Yu, Z. G. Wang, X. Y. Liu, D. Hu, A big data application in intelligent transport systems, in: Applied Mechanics and Materials, Vol. 734, Trans Tech Publ, 2015, pp. 365–368.

[17] E. De Romph, Using big data in transport modelling, Data & Modelling Magazine,(13) 2013.

[18] N. Van Oort, Big data opportunities in public transport: Enhancing public transport by itcs, in: IT-TRANS, 18-20 February 2014, Karlsruhe, Germany, 2014.

[19] A. A. Chandio, N. Tziritas, C.-Z. Xu, Big-data processing techniques and their challenges in transport domain, ZTE Communications 1 (2015) 010.

[20] S. Buckley, D. Lightman, Ready or not, big data is coming to a city (transportation agency) near you, in: Transportation Research Board 94th Annual Meeting, no. 15-5156, 2015.

[21] C. Trifon, Benefits of linear asset management applications, `http://www.mermecgroup.com/news-e-events/media-coverage/627/1/benefits-of-linear-asset-management-applications-.php`, (Visited on 06/09/2015).

[22] F. Cores, N. Caceres, F. Benitez, S. Escriba, N. Jimenez-Redondo, A logical framework and integrated architecture for the rail maintenance automation, in: European Transport Conference 2013, 2013.

[23] N. Attoh-Okine, Big data challenges in railway engineering, in: Big Data (Big Data), 2014 IEEE International Conference on, IEEE, 2014, pp. 7–9.

[24] EIT, Route mapping capability for gb and international rail markets, `http://www.futurerailway.org/innovation/Documents/EIT-Railwaycapabilitymapping-PhaseB-Finalreport.pdf`, (Visited on 06/09/2015).

[25] S. Iwata, J. Domme, Information and telecommunication systems business information meeting: effectively utilizing big data, `http://www.hitachi.com/IR-e/library/presentation/111207/111207.pdf`, (Visited on 06/09/2015).

[26] A. Núñez, J. Hendriks, Z. Li, B. De Schutter, R. Dollevoet, Facilitating maintenance decisions on the dutch railways using big data: The aba case study, in: Big Data (Big Data), 2014 IEEE International Conference on, IEEE, 2014, pp. 48–53.

[27] B. Y. Song, Y. Zhong, R. K. Liu, F. T. Wang, Railway maintenance analysis based on big data and condition classification, Advanced Materials Research 919 (2014) 1134–1138.

[28] S. G. Nunez, N. Attoh-Okine, Metaheuristics in big data: An approach to railway engineering, in: Big Data (Big Data), 2014 IEEE International Conference on, IEEE, 2014, pp. 42–47.

[29] J. Tutcher, Ontology-driven data integration for railway asset monitoring applications, in: Big Data (Big Data), 2014 IEEE International Conference on, IEEE, 2014, pp. 85–95.

[30] H. Li, D. Parikh, Q. He, B. Qian, Z. Li, D. Fang, A. Hampapur, Improving rail network velocity: A machine learning approach to predictive maintenance, Transportation Research Part C: Emerging Technologies 45 (2014) 17–26.

[31] ARUP, Future of rail 2050, http://www.arup.com/Homepage_Future_of_Rail.aspx, (Visited on 06/09/2015).

[32] Altran, Tomorrows trains moving towards smarter rail networks, http://www.altran.com/fileadmin/medias/0.commons/images/Industries/SmartInsight_Railway_Connectivity.pdf, (Visited on 06/09/2015).

[33] M. TANAKA, Prospective study on the potential of big data, Quarterly Report of RTRI 56 (1) (2015) 5–9.

[34] R. Nappi, Integrated maintenance: analysis and perspective of innovation in railway sector, arXiv:1404.7560.

[35] X. Zhang, D. Gong, Application of big data technology in marketing decisions for railway freight, in: 2014 International Conference of Logistics Engineering and Management (ICLEM), 2014.

[36] Y. Q. Shao, R. K. Liu, F. T. Wang, M. D. Chen, Research on big data management for high-speed railway equipment, in: Applied Mechanics and Materials, Vol. 462, Trans Tech Publ, 2014, pp. 405–409.