"Mining the Datasphere: Big Data, Technologies, and Transportation"

DISASTER MANAGEMENT

Version 1 (January 2017)

Project Steering Group

Dr Ken Michael AO (Chairperson), SBEnrc

Professor Keith Hampson, CEO SBEnrc

Professor Peter Newman, Curtin University

Dr Charlie Hargroves, Curtin University

Professor Bela Stantic, Griffith University

Kamal Weeratunga, MRWA

Jannatun Haque, NSW RMS

Kim Thomas, Aurecon

University Research Team

Program Leader:

Professor Peter Newman, Curtin University

Project Leader:

Dr. Karlson 'Charlie' Hargroves, Curtin University

Researchers:

Associate Professor Bela Stantic, Griffith University,

Daena Ho and Daniel Conley, supported by Andrew Hojem, Oliver Pyke, Josh Wood, Rohan Aird, Khiem Nguyen, Georgia Grant, and Harry Carpenter, Curtin University.

CONTENTS

EXECUTIVE SUMMARY

Interruptions to transportation through natural disasters or on road collisions or obstructions is a key issue facing transport planners and managers with many now asking if there are any promising technologies offering new solutions? With the rapidly growing availability of data and the ability to analyse large data sets this report seeks to answer the question "*What value can 'Big Data' provide to assist with disaster and emergency responses?*". There is great interest and hype around 'Big Data' and this report provides a summary of an assessment of its value to assist in informing responses to natural disasters and emergencies informed by world's best practice. In basic terms, 'Big Data' refers to a very large and rapidly expanding pool of data that is collected from multiple sources and platforms often at high speed and in real time. In recent times the term has become a popular topic for discussions and research on smart urban transportation and mobility. This report seeks to investigate the hype around the term and identify some tangible value for road agencies in Australia that has yet to be captured. It is important to distinguish between 'small data' and 'Big Data', with the difference being that the tools used to assess small data are not sufficient to assess big data and new approaches are needed.

The volume and variety of Big Data give it enormous potential to aid in disaster management through the integration of multiple emergency-sensing platforms and response systems. This enables cohesive and well-executed mitigation techniques, which enable first responders to:

I. Identify at-risk areas to inform affected communities well in advance of the disaster,
II. Identify areas which need the most urgent relief immediately after the disaster and coordinate crisis responders and logistics during the crisis,
III. Provide guidance to other members of the public to avoid hazards, including the provision of a safe transport pathway away from the danger, and
IV. Develop of a database of emergencies, allowing future evaluation of the effectiveness of the emergency management response and identify optimal strategies for responding to future disasters.

While many countries are working towards using multiple data streams in disaster response, emergency identification is usually separate from emergency response, as illustrated in Section 4.1.1. Rio de Janeiro is an early leader in terms of integrating disaster sensing and management technologies (4.1.2). Researchers in Korea have also started using data from multiple social media platforms to identify disaster hotspots, and aim to use their platform to predict hazards based on analysis of social data as well as structured disaster information (4.1.3). In addition, Transport for London provides insights into what an integrated Big Data system for both congestion management and incident response looks like (4.1.4). NSW's State Emergency Service is already taking steps towards integrated response systems by sharing crisis data and information across local and state governments (4.2.1), but their approach has yet to be specifically tailored to the field of transport resilience.

Essential to the harnessing of Big Data is the use of analytics technology which sifts through information and identifies the key data. This report identifies three of the main platforms in Big Data analysis and management: Hadoop, an open source data collection and analysis platform; Spark, a Hadoop

extension with improved processing performance; and SAP HANA, a licensed, customisable software proven effective in collecting and analysing Big Data for disaster management.

The huge amounts of data exploding across a huge array of platforms have become increasingly relevant in providing high-resolution information to aid in emergency management efforts and make transport systems more resilient. The exciting implications of Big Data are only beginning to be realised, especially in Australia, which predominantly draws upon on real-time, for-purpose emergency detection data, effectively 'small data'. This report presents various case studies to provide city planners and transport organisations with examples of how multiple data streams can be harnessed to increase the resilience of transport systems in Australia.

In conclusion, there is huge scope for the development of truly integrated disaster management systems which will act to improve transport resilience in the face of emergencies and unexpected events. This is especially the case in Australia, which has yet to explore and develop the potential of growing sources of data to inform disaster management. With data exploding across a dizzying array of platforms, and the promise of high-resolution information which can accelerate disaster recovery, many cities across the world are beginning the shift to multi-platform data in order to respond effectively and efficiently to disasters and unexpected events. Ultimately, harnessing multi-platform data to create a more robust and resilient transport and communications network will allow communities to be better prepared for emergencies, respond rapidly during emergencies and recover swiftly afterwards.

1 INTRODUCTION

1.1 What is 'Big Data'?

There are multiple definitions of Big Data. Most commonly, the term is used to broadly characterise data sets so large they cannot be stored and analysed by traditional data storage and processing methods. The research firm McKinsey describes 'Big Data' as a large pool of structured and unstructured data which can be analysed, aggregated, and communicated.[1] A large volume of data is now available for a growing number of sources; however, this is only one dimension of its complexity. The velocity at which data is received and the variety of information available adds to the challenge of creating value. Further, data is now produced in multiple formats, languages and software configurations depending on where the data is sourced. These formats include relational, textual, multimedia and document mark-up languages such as XML.

It is these three characteristics (referred to as the three V's – Volume, Velocity and Variety) that distinguish 'Big Data' from other forms of data. The emergence of such large and complex data sets has primarily been the result of a decrease in the cost of sensory and observational technologies in conjunction with mass digitisation of systems and processes around the globe. Combined with large-scale sensor networks and computer simulations, there is now a vast amount of information stored in a worldwide network of distributed archives. Such changes have allowed for a proliferation of platforms that have enabled the transformation of the analogue world into one made of vast realms of digital information.[2]

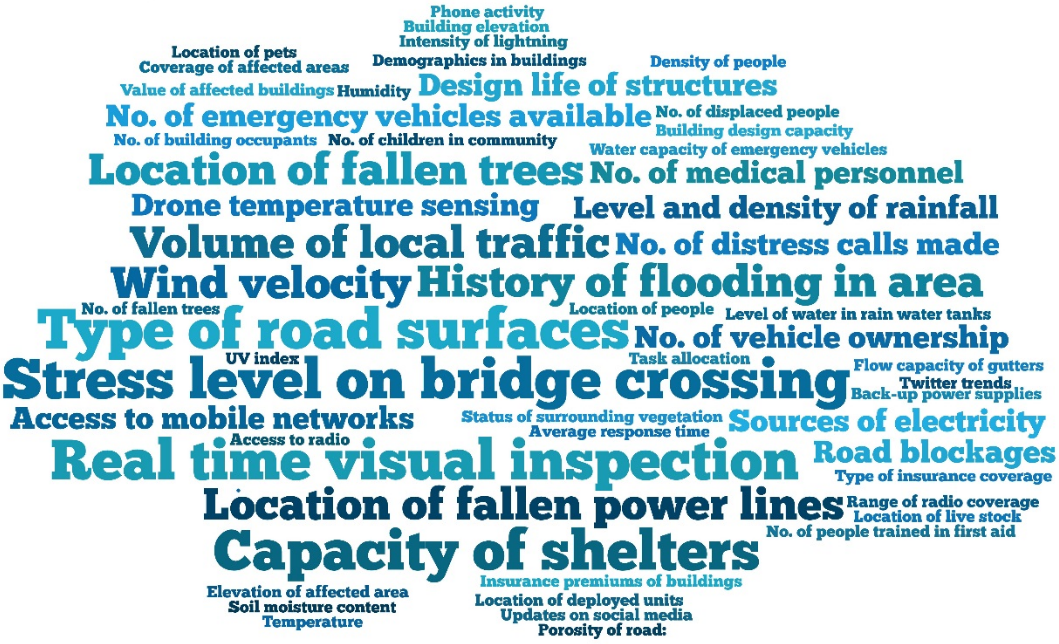
Not only can data be used to observe direct phenomenon (such as streamlining traffic signal timings) but the interrogation of data streams to identify commonalities in response to perturbations provides a unique potential to identify linkages between previously seemingly unrelated data. Because of the extremely large volumes of information produced, 'Big Data' requires analysis in order to produce meaningful results. The term 'Big Analytics' is used to describe the processing of multiple massive data sets to extract useful algorithms and information. If well-harnessed, Big Data can help to improve the resilience of transport systems, allowing a swift return to efficient operation of transport corridors following disasters and unexpected events.

When considering data related to emergencies and disaster management such as traffic counts, average velocity, temperature conditions, length of quest etc. would be classified as 'Small Data' however as the word cloud below presents there are literally hundreds of data sources that stand to inform disaster management efforts.

---

[1] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H. 'Big data: the next frontier for innovation, competition, and productivity', *McKinsey Global Inst.*, 2011.

[2] Hassanien, A.E., Azar, A.T., Snasael, V., Kacprzyk, J., Abawajy, J.H. (ed) 2015, *Big Data in Complex Systems: Challenges and Opportunities*, Springer International Publishing, Switzerland. (ISBN: 978-3-319-11056-1)



1.2 Data Superhighways: Big Data in Transport Systems

1.2.1 What value can 'Big Data' create?

The link between 'Big Data' and transportation is not a new phenomenon; traffic systems have long produced streams of observational and sensory information both directly and indirectly. In 2014, a study by the Australian Government Bureau of Infrastructure, Transport, and Regional Economics (BITRE) identified a number of available data collection technologies and concluded: '*Recent and emerging technologies offer significant opportunities for collecting more information, more cost effectively, about personal travel activity and road use, that can better inform day-to-day network management, long-term infrastructure planning and road user travel choices*'.[3]

Smart transport systems using Big Data can achieve a higher level of efficiency, which leads to cost savings, reduced energy demand, better delivery of services, improving life quality and reducing environmental impacts. In this report, transportation refers to systems of mobility including vehicles, roads, railways, subways, buses, taxis, bicycles, ferries and share-rides. Each transportation mode plays an essential role in mobility of a city and if properly harnessed can move people and products to their destination safely and efficiently at a reasonable cost.

In addition to congestion management and improving the flow of mass transit options, data can be used to identify emergencies (either from natural disasters or crashes and obstructions on roadways) informing swift and effective deployment and communication of alternative transport routes to commuters.

[3]BITRE 2014, 'New traffic data sources – An overview', *Australian Department of Infrastructure and Regional Development,* Canberra.

Further, the integration of multiple emergency-sensing platforms with response systems enables cohesive and well-executed disaster mitigation techniques, that include:[4]

- The early identification of at-risk areas,
- The identification of the areas which need the most urgent relief during events,
- The coordination of crisis responders and logistics during the event,
- The provision of advice to the public to avoid hazards,
- The provision of a safe transport pathways away from the event,
- The development of a database of emergencies, allowing emergency management authorities to both evaluate the effectiveness of the emergency management response, and identify optimal strategies for responding to future events.

For-purpose seismic data and social media platforms are currently being used to identify disaster hotspots in the event of earthquakes (See details in Section 4.1.1), but have yet to be extensively used in conjunction with each other. Early movers in this area are the Rio de Janeiro Operations Centre that was significantly upgraded to preventing deaths during annual flooding events and serves as a prime example of the potential rewards of integrating disaster identification and management techniques (Section 4.1.2). Researchers in Korea have also begun using Big Data from multiple social media platforms to identify disaster hotspots, and aim to use their platform to predict hazards based on analysis of social data as well as structured disaster information (Section 4.1.3). IBM and Integritie has collaborated to create software which prioritises post disaster insurance claims based on assessment of multiple data streams across multiple platforms, such as social media and email, in order to aid in recovery efforts.

Transport for London (TfL) provides insight into what an integrated Big Data management system for both congestion management and incident response would look like. Using data from multiple sources across the bus network, trains, roads, taxis, ferries and cycle paths, TfL produces a real-time view of traffic conditions and sets variable speed limits across its transport network accordingly. Citizens of London also receive real-time information covering various subjects such as weather, air pollution, delays in public transport, availability of public bikes, and real time camera feeds.[5] During an unexpected event, real-time and historical data analysis can be used to quantify the affected passengers and transport systems set up to meet their predicted travel needs. Although still at what would be called a 'Small Data' TfL's use of data has allowed better management of higher numbers of passengers: operational performance across 2011-2016 shows improvement in the quality of passenger journeys, including a reduction in wait times and traffic flow, increased customer satisfaction, and an increased number of kilometres serviced by public transport.[6] (See Section 4.1.4)

---

[4] Emmanouil, D., Nikolaos, D. 2015, *Big Data analytics in prevention, preparedness, response and recovery in crisis and disaster management*, WSEAS Press, The 18th International Conference on Circuits, Systems, Communications and Computers, Recent Advances in Computer Engineering, Series 32, p.476-482

[5] Kitchin, R. 2014, 'The real-time city? big data and smart urbanism', *GeoJournal*, 79(1):1–14. ISSN 1572-9893. DOI: 10.1007/s10708-013-9516-8

[6] Transport for London 2015, 'Annual Report and Statement of Accounts', *Greater London.*

In Australia, the NSW State Emergency Service shares information across local and state government in real-time in order to enhance emergency response (Section 4.2.1), but has yet to apply this technique specifically to transport resilience. In WA and VIC, two advanced data collection and analysis systems are also in development to detect and predict bushfires (4.2.2). Both systems have a large amount of untapped potential in terms of integrating extensive emergency response and mitigation techniques. In the field of transport systems, the traffic management platform STREAMS and QLD's Managed Motorways initiative have produced flow on benefits in accident reduction, while QLD's Emergency Vehicle Priority program adjusts traffic controls in order to improve the response time of emergency vehicles (Section 4.2.3 and 4.2.4). Both systems are smaller picture in terms of their scope and application, focusing mainly on improving existing systems; neither addresses the wider scope of transport resilience planning, which includes pre-emptive collision prevention and the strengthening of vital transport corridors (**Error! Reference source not found.**).

1.2.2 Collecting Big Data

The rapid rise in the capacity of data storage options (both in-house and remotely) along with the increase in computational ability means that there is great potential to harness additional value from data that can inform the working of a modern city, especially its transportation systems. Data from transport systems is highly varied and comes in three broad categories:

i. *Highly structured datasets* that originate from technology implemented to address well-defined problems (e.g. data from automatic toll road payment transponders for the use in processing toll road payments or data from intersection sensors on traffic flows and time of day usage of the road network) which can more aptly be defined as 'Small Data',

ii. *Unstructured datasets* that are produced from any interaction between road users and digital infrastructure. Given the explosion of mobile phones, personal computers, sensors, cameras, and devices, there is huge (yet largely untapped) potential to harness these data streams to inform congestion and disaster response, which is now moving into the realm of 'Big Data', and

iii. *Data from seemingly unrelated sources* that stand to provide insights into the behaviour and functioning of the transport system, such as the price of parking at particular public carparks, the level of fines for illegal parking, the amount of people walking more than 1 kilometre to public transport, real time weather conditions, the location of downed trees and infrastructure that is interfering with roadways etc., which now makes the overall data set unmanageable using small data techniques.

Currently access to data is not a concern as there are a multitude of data sources available which produce a wealth of information (however it may be the case that additional data sets that are currently not available may be more valuable to congestion management than those that are currently openly available). The challenge is to harness the data by processing and interpreting it both at the higher levels of trends and scenarios and at the lower levels related to the day to day management of transportation infrastructure. High data volumes mean that it takes time to process and advanced computing

technologies are required to improve response times.[7] Currently, data is used to inform trip times and route selection; however, Big Data can be used to inform predictive analyses and the development of advanced user information platforms.

This analysis requires programs and technologies that extract value from what our research team refer to as the 'Datasphere', which contains data that may seemingly be disconnected from transportation but when assessed shows correlations that would otherwise be hidden. It is the combined 'Big Analytics' processing of all available data streams within which the true potential value of Big Data exists.[8] Effectively harnessing such data can provide significant benefits due to the development of temporal, spatial, and historical correlations between key factors during emergencies.[9]

1.2.3 Data Analysis: 'Big Analytics'

Because of the volume and complexity of the data produced, there are inherent difficulties in data analytics and challenges exist in the analysis and harnessing of this information. In particular, the different data formats and languages in which data is stored may lead to difficulties in processing using data mining algorithms.[10] However, the potential rewards are impressive. The availability of 'Big Data' provides insight into actual passenger and road use behaviour, as opposed to reported behaviours and preferences which may not present the whole picture.[11]

1.2.4 Data Visualisation and Communication

Communicating value from 'Big Data' to road users, planners and operators is crucial for the improvement of transport networks. To curtail road congestion before the congestion becomes severe, Big Analytics algorithms must be able to communicate with traffic lights and other traffic control systems when real-time congestion pre-cursors match historical Big Data information on severe congestion events.[12] Strategic control of traffic lights and congestion information for emergency vehicles can greatly shorten the response time when travelling to an incident.

[7] Mullich, J. 2013, 'Drivers avoid traffic jams with Big Data and Analytics', *Bloomberg L.P.,* New York.
[8] International Transport Forum 2015, 'Big Data and Transport', *International Transport Forum.*
[9] International Transport Forum 2015, 'Big Data and Transport', op. cit.
[10] Hassanien, A.E., Azar, A.T., Snasael, V., Kacprzyk, J., Abawajy, J.H. (ed) 2015, op. cit. ISBN: 978-3-319-11056-1
[11] van Oort, N. & Cats, O. 2015, 'Improving public transport decision making, planning and operations by using Big Data: Cases from Sweden and the Netherlands', *IEEE 18th International Conference on Intelligent Transportation Systems*. DOI 10.1109/ITSC.2015.1
[12] Sawyers, P. 2015, 'How Microsoft's using big data to predict traffic jams up to an hour in advance', Venturebest, April 3.

2 HARNESSING DATA TO BETTER MANAGE EMERGENCIES

2.1 Role of Big Data in Emergency and Disaster Management

As dynamic networks, transport systems are subject to unexpected events while operating, leading to disruptions in services and severe congestion. A 2014 *Transport Resilience Review* by the UK transport authorities defines transport resilience as ensuring that systems are able to withstand the impacts of emergencies, operate in the face of these events and recover promptly from its effects.[13] However, before looking at strategies that can minimise the duration and impact of unexpected events, two types of unexpected events must first be distinguished:

1. *Man-made events such as road accidents and emergency repairs which cause traffic congestion due to lanes or entire roadways being blocked*: One of the most appealing potential applications of Big Data is to predict the location of future traffic incidents so that traffic management systems can act to circumvent congestion and loss of life. However, road accidents are caused by traffic dynamics and human responses that are near impossible to predict. Even a perfectly safe vehicle under perfect geometric and environmental conditions may still crash due to sudden changes in road dynamics and disruptions in the normal flow of traffic. Current models assume vehicle speed to be equal to the speed limit;[14] however, in order to better improve collision prediction models, real-time data can be used. Next, the program must act to prevent the predicted collision, a technically challenging task due to the volumes of information received, the complexity of projection algorithms and the limitations of processing speed.[15]

2. *Natural disasters or extreme weather events such as bushfires, storms, and cyclones, which can cause severe congestion due to inaccessible roads and congestion from people leaving the area*: This second cause should be of key interest to transport authorities, as the Fifth Assessment Report of the Intergovernmental Panel on Climate Change predicts that climate change will increase the incidence of natural disasters in the future.[16] Experts have recognised the power of Big Data to manage and respond to natural disasters and Japan and the US recently launched a joint research program to examine and develop methods to use data to reduce pressures from natural disasters which kill thousands each year and cost the global economy billions. Multiple platforms and software have been used to track areas most affected by disasters and aid in sending relief, such as Rio de Janeiro's crisis management centre which was created after the 2010 floods and mudslides (Section 4.1.2).

In general, the level of resilience should be proportional to the intensity of use of that transport corridor, as well as the availability of alternative routes. Critical highways, bridges and railways must be identified and prioritised for strengthening in order to protect 'points of failure' in transport networks. Big Data emerges in the integration of information from weather forecasters, transport operators and infrastructure planners: during emergencies, each group understands the 'trigger points' that lead to failures in transport systems. Forecasters can disseminate detailed natural disaster or severe weather predictions (including information on the likely location and severity), with sufficient warning time for

---

[13] OGL 2014, *Transport resilience review: a review of the resilience of the transport network to extreme weather events*, Department of Transport, UK, London.
[14] Hossain, M. 2012, 'A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways', *Accident Analysis & Prevention,* vol. 45. (http://dx.doi.org/10.1016/j.aap.2011.08.004)
[15] Hossain, M. 2012, 'A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways', op. cit.
[16] IPCC, 2014: *Climate Change 2014: Synthesis Report*. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland

transport operators to take precautionary measures. With information about the wider network at their fingertips, transport operators can make well informed and quick decisions on route selection and can clearly communicate the data they receive to first responders and road users. Finally, transport planners can harness multiple large databases from both forecasters and operators to build future contingency plans that mitigate disruptions to transport networks, perhaps by providing alternative transport corridors to supplement critical networks and are less affected by emergencies.[17]

In the event of a natural disaster or other emergencies, access to accurate real-time situational analysis is critical. To provide relief where it is needed most, emergency responders need an easily accessible snapshot of the situation based on rapid aggregation of information from a range of sources. This is where Big Data comes in. Big Data analytics platforms have the potential to both predict disaster areas and provide high-resolution, real-time information so that emergency 'hotspots' can be identified. The potential to harness Big Data in emergency management is hence two pronged:

1. *Predictive Emergency Management*: These strategies are based on real-time data from a variety of transport, seismic and weather related sources, combined with predictive algorithms based on blending this data with historical data sets in order to predict the severity and likely location of emergencies.
2. *Effective, cohesive and prompt deployment of emergency response systems*: Sensors can be used to produce real time high-resolution information to inform the deployment of emergency services and route the public to safety. This can include the use of tilt sensors on trees, poles and other infrastructure that is likely to interfere with traffic flow in the case of a disaster to provide real time appreciation of the potential routes that are blocked. As the previous project in the 'Future of Roads' theme outlined (Project 1.35), drones can also be used to assess the damage and identify quickly accessible routes.

So can using Big Data really predict incidents? The short answer is 'yes, but not perfectly, and not right now'. The long answer is that while high-resolution transport information collected from millions of data sources contains the necessary information to predict travel patterns and weather conditions to identify problem areas, the challenge is to process this information to extract the useful information and correlations to be compared to historical data in real time before and during disasters.[18] In addition, no prediction is perfect and there is always a margin of error. Many Big Analytics traffic prediction systems are still in development, and disasters or emergencies can be difficult to predict.

2.2 Big Data in Disaster Response

There are four main phases of disaster management: prevention, preparedness, response and recovery and two major sources of Big Data: dedicated sensor networks (e.g. earthquake detection sensors, tilt sensors, pavement integrity sensors etc…) and multi-purpose sensor networks (e.g. social media, smartphones, technology enabled vehicles). The ever increasing developments in advanced technology

---

[17] OGL 2014, *Transport resilience review: a review of the resilience of the transport network to extreme weather events*, op. cit.
[18] Jie Xu et al. 2015, 'Mining the Situation: Spatiotemporal Traffic Prediction with Big Data', *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4.

provide new innovative solutions to improve disaster management. People are empowered by the combination of computer systems and networks that include sensors, smartphones, cyber-physical systems and drones providing critical information to those in need during disasters. The systems need to be reliable and responsive as when disaster strikes, it is crucial that emergency personnel and decision makers have access to real-time information in order to assess situation and respond appropriately. As author of the book 'Digital Humanitarians' Patrick Meier reflects, in times of emergency '*access to information is equally important as access to food and water*'.[19]

Big Data is able to help identify and notify emergency response personnel of areas that need the most urgent attention. Real-time monitoring provides guidance to disaster victims in taking the best route away from the dangerous zone in order to prevent congestion or mistakenly move to another dangerous zone. With the study of the collection of data from previous incidents, it is easier to identify the most effective response strategies and development of infrastructures for effectively responding to future disasters. Crowd-sourcing allows for the potential of real-time communications on displaced locations and aids in the tracking of people's location and behaviour during a disaster however this information can be difficult to verify.

Big Data can also be utilised not only in disaster response or resilience but in recovery activities such as safety confirmation, logistics, volunteer coordination and provision of relief supply. Big Data provides a great global opportunity for disaster control however there are challenges in analysing and processing the wide ranging variety of data sources and veracity of data content.[20] Overcoming these challenges will unlock the benefits of applying Big Data to disaster management, and Section 3 investigates cities across the world and here in Australia where data has been used to provide improved and targeted disaster response.

2.3 Privacy Concerns

The exponential growth of mobility-related data will trigger significant changes in the transport industry accompanied by rising concerns relating to the adequacy of regulations ensuring privacy.[21] Even data that is said to be 'anonymous' may still be able to be linked to specific individual sources if cross-referenced with other sources of related data. Not only do traffic management centres have to tackle this issue, they also have to decide on whether the data is reliable enough. Much of this data right now has to be verified with other data sources such as sensors and camera footage or still-shots. In addition, companies may need to migrate to non-relational (NoSQL) databases to accommodate and process large unstructured data sets. These NoSQL databases usually use external security enforcing mechanisms; hence to reduce security breaches, companies have to use additional security software, reviewing security policies for the 'middleware' between the operating system and the NoSQL

[19] Meier, P. 2013, 'How to create resilience through Big Data', *iRevolution.*

[20] Pu, C. and Kitsuregawa, M. 2013, 'Big Data and Disaster Management: A report from the JST/NSF Joint Workshop', *Georgia Institute of Technology,* Technical Report No. GIT-CERCS-13-09.

[21] International Transport Forum 2015, 'Big Data and Transport', op. cit.

database, while also toughening the NoSQL database itself to match its counterpart relational databases.[22]

The multi-tiered nature of Big Data means that transaction logs are stored in multi-tiered media. In smaller datasets, IT managers can manually move data between tiers, giving them a measure of control; however, as the dataset grows exponentially, auto-tiering is likely to become increasingly necessary for big data storage management. As auto-tiering does not keep track of where the data is stored, unauthorised access to data stores is less easy to detect and data breaches may occur. Thus, new mechanisms must be developed to prevent data theft and maintain the 24/7 availability.[23] In Australia, the Privacy Act regulates and protects personal information, including the Australian Privacy Principles (APPs) which define the standards, rights and obligations in relation to handling and assessing personal information. Big Data changes how key privacy principles—which include data collection, minimisation of data retention and use limitation—are applied. However, as the APPs are technologically neutral, corporations and other organisations can adapt their Big Data handling policies to protect personal information while also retaining the maximum use from the information derived from Big Data analysis.[24]

According to the Privacy Act, organisations must take reasonable steps to implement practices, procedures and systems that protect personal information. These organisations must also be able to deal with privacy related complaints from individuals. A systematic risk management approach must be used to identify reasonable steps according to the size of the entity, its resources and the complexity of its operations. Organisations dealing directly with Big Data must take more rigorous and detailed privacy protection procedures than an entity handling the results of Big Data analytics. [25] One possible technique to ensure privacy is to use de-identification to remove the personal identifiers such as addresses and date of birth, as well as any other unique individual characteristics. This means that the Privacy Act no longer applies.[26] However, this technique is not foolproof and if de-identified datasets are matched to other datasets or other information, it is possible that individuals can be re-identified.[27]

---

[22] CSA 2012, *Top ten Big Data Security and Privacy Challenges*, Cloud Security Alliance, Rolling Meadows, Illinois.
[23] CSA 2012, *Top ten Big Data Security and Privacy Challenges*, op. cit.
[24] OAIC 2015, *Consultation draft: Guide to Big Data and the Australian Privacy Principles*, Office of the Australian Information Commissioner, Canberra.
[25] OAIC 2015, *Consultation draft: Guide to Big Data and the Australian Privacy Principles*, op. cit.
[26] OAIC 2015, *Consultation draft: Guide to Big Data and the Australian Privacy Principles*, op. cit.
[27] OAIC 2016, *Privacy business resource 4: De-identification of data and information*, Office of the Australian Information Commissioner, Canberra.

3 PLATFORMS AND TECHNOLOGIES FOR BIG DATA COLLECTION AND ANALYSIS

3.1 Emerging Digital Platforms for Big Data

3.1.1 Overview

There are a number of promising emerging software platforms that can potentially be applied to further harness the potential of 'Big Data'. Some of these platforms both collect and analyse data, while others require the input of raw data collected by the user. For emergency and disaster management applications, the input datasets come from a range of sources including GPS data from mobile phones, taxis and buses, as well as information taken from traffic light sensors, fixed sensors, and other datasets. The data analysis platforms then analyse and interpret the information to respond to user queries. While previous platforms have mainly focused on small data mining, there are now a robust set of digital platforms that are leaders in Big Data processing when users have numerous flexible requirements. The use of cloud based or in house processing platforms each have their own advantages. Cloud based services provide a less expensive startup cost, however the user is generally charged for the amount of data they wish to process. On the other hand, in house systems incur larger initial costing and maintenance fees, but offer an unlimited amount of data processing.

Hadoop distributes data collections across multiple nodes within a cluster of servers, meaning custom hardware does not need to be bought or maintained. Spark is a data-processing tool which operates on the data distributed using Hadoop. Although Hadoop can process the data, Spark does so at a significantly faster rate. Both programs are open-source, free and can be used in conjunction to increase processing speed; however, modifications have to be made to the program in order to customise it to a user-specific application. Because Hadoop and Spark are open source, many companies build their own data analytics software based on these frameworks. These companies can be hired to develop a more specialised system and establish a support network. Many of these companies are also start-ups, which means that while complex, high level services may be obtained from such companies at a lower price (compared to more experienced competitors), there is a level of risk that the company may collapse or fail to perform to the expected standard.

Alternatively, SAP HANA provides an all in one platform which has been proven to be effective in handling the data required to analyse traffic congestion. Being run through cloud based or in house servers makes SAP HANA versatile and the most cost effective SAP HANA setup can be chosen based on how the system must perform. However, this platform is not open source and licensing must be purchased or rented. This is not necessarily a bad thing, as SAP HANA provides large amounts of support, especially with hardware, and can help set up a Big Data system quicker than using an open source platform.

Table 1 compares the three technologies and tools often used to analyse Big Data.

Table 1: Comparison of the three Big Analytics platforms

<table>
<tr><th>Hadoop with MapReduce</th><th>Hadoop with Spark</th><th>SAP HANA</th></tr>
<tr><td>Open-source</td><td>Open-source</td><td>Closed-source</td></tr>
<tr><td>Disk memory, which uses batch processing where data is stored and then processed at specific intervals</td><td colspan="2">In memory, allowing real-time continuous processing of incoming data</td></tr>
<tr><td>No support</td><td>No support</td><td>Hardware support</td></tr>
<tr><td colspan="2">Machine learning capability has to be specifically programmed into the platform</td><td>In-built machine learning capability for better predictive ability</td></tr>
</table>

3.1.2 Hadoop

Hadoop is a framework platform that can be used for the processing of large data sets across clusters of computers. It is designed to be highly scalable, meaning it can be scaled from a single machine to many thousands, with each offering local computation and storage. Rather than rely on hardware, the library itself is designed to detect and handle failures at the application layer, making it a highly robust and efficient tool. Scalability also gives it a large range of uses and users can modify it for specific needs. Hadoop has multiple systems that can be used in conjunction with each other to enhance specific properties. These include:

- *Hadoop Common*: The common utilities that support the other Hadoop modules.
- *Hadoop Distributed File System (HDFS)*: A distributed file system that provides high-throughput access to application data.
- *Hadoop YARN*: A framework for job scheduling and cluster resource management.
- *Hadoop MapReduce*: A YARN-based system for parallel processing of large data sets.[28]

Figure 1 shows how the applications can be used for both single use and multi-Purpose Platforms.

[28] Apache Hadoop 2016, 'What Is Apache Hadoop', The Apache Software Foundation.

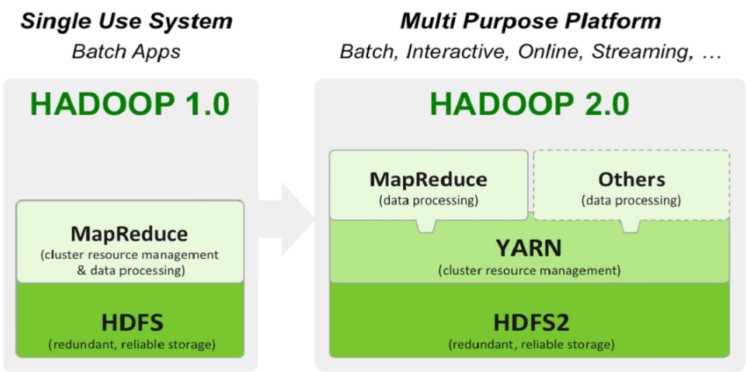


Figure 1: Data Analytics News 2014[29]

Hangzhou Trustway Technology Co. Ltd. has significantly improved its transportation management capability using Apache Hadoop on Intel® Xeon® processors (specifically using MapReduce, Hive and optimised HDFS). Hadoop's software is connected with traffic monitoring equipment such as the city's checkpoints, video monitoring, traffic flow detection, signal systems, and devices to provide the city with a big data storage system with high throughput and fault tolerance. This allows for a fast and efficient dynamic monitoring system that enables vehicle track analysis and searching, fake plate number analysis, vehicle control and traffic violation data storage.

> *Testing has shown the system is capable of carrying out collision analysis on 2.4 billion plates in only ten seconds.*[30]

From members of the Hadoop framework, different software can be produced for different needs, as highlighted in the following section with Apache Spark. Spark was implemented through Hadoop YARN, HDFS and MapReduce and is discussed in the following section.[31]

3.1.3 Spark

Apache Spark is an open source framework designed to perform real-time processing. As a whole, Spark cannot distribute files throughout the system, however it is used in conjunction with Hadoop as an alternative to MapReduce when data sets are in the processing stage. Both Spark and MapReduce rely on resilient distributed data (RDD) sets, which are representative of the incoming data for which the analysis takes place. RDDs have the capability to recompute their own data if a systematic failure were to occur due to their long lineage.[32]

The development of Spark is to be implemented alongside Hadoop due to the various limitations of MapReduce, which essentially processes the data as a batch, and stores the processed data on a disk. Conversely, Spark allows for analysis to be undertaken on a single cluster of both analytical and

[29] Bigdata 2014,'Hadoop 2.0 and YARN Architecture', BDAN.
[30] Intel 2013, 'Improving traffic management with Big Data analytics', Intel Xeon.
[31] Hortonworks 2015, 'Apache Spark'.
[32] Anoop Daware 2015, 'Apache Spark vs. MapReduce', MapR.

operational data using real-time processing. Additionally, the storage of the data in memory creates a significantly quicker run-time.[33]

Both programs differ in the way they process the data, however since MapReduce was designed solely to perform batch processing, its performance can plateau. As a result, Spark has been known to process the same dataset 100 times faster than Hadoop's MapReduce using memory, or 10 times faster using its disk.[34] Figure 2 shows the structure of the Spark platform.

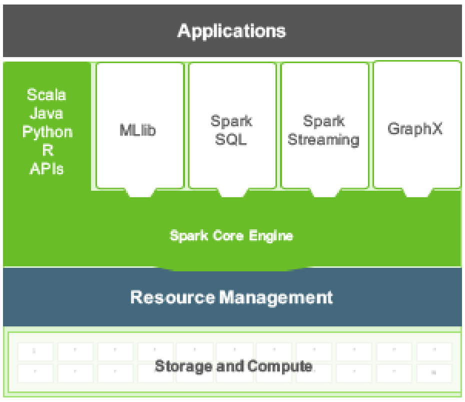


Figure 2: Structure of the Spark platform[35]

A case study undertaken in India illustrates the power of the Spark platform. Using gathered sensor values containing vehicle speed, count of the vehicle and the time taken for the vehicle to pass by the sensor area, a significant amount of data points is converted into a comma separated value (CSV) file. The CSV file is processed in Spark in order to predict the severity of traffic congestion, and the study found that …

> *…where the existing system took a significant amount of time, the implementation of Spark predicted traffic conditions in half a second.*[36]

3.1.4 SAP HANA

SAP HANA is a single in-memory platform which combines application services, high-speed analytics and data acquisition tools to deliver an all in one package. SAP HANA can be integrated as either an in house system or through a cloud based database service such as SAP HANA Cloud Platform, HP Helion, Microsoft Azure and many others. The platform lends itself perfectly to be applied as a Big Data analytics tool and is being used around the world for this purpose. SAP HANA can handle multiple data inputs and provide predictive analytics as well as spatial and graphical data processing.[37] The platform also excels at delivering deeper insight from Big Data and the Internet of Things due to its strong machine learning capabilities. Due to the software's ability to perform analytics in-memory on a single

[33] Vardhan 2015, 'Apache Spark vs Hadoop MapReduce', Edureka.
[34] Xin 2016, '$1.44 per terabyte: setting a new world record with Apache Spark', databricks.
[35] Hortonworks 2015, 'Apache Spark'
[36] Prathilothamai, M. Lakshmi, A.M. and Viswanthan, D. 2016, 'Cost Effective Road Traffic Prediction Model using Apache Spark', *Indian Journal of Science and Technology*, vol. 9, no. 17. DOI: 10.17485/ijst/2016/v9i17/87334
[37] SAP HANA n.d., What is SAP HANA?

data copy, outputs from the raw data can be obtained in real-time allowing for instantaneous reactions in dynamic systems,[38] such as traffic management.

SAP HANA is being used around the world for many different purposes such as business analytics, bank processing and more recently traffic management. Both Japan and China have used SAP HANA to reduce congestion on city roads. The Nomura Research Institute (NRI) conducted an experiment using GPS data from 18,000 taxis located in Tokyo using Intel®Xeon® processor E7's to run SAP HANA.[39] Using the platform, they managed to process approximately 360-million items of data in a single second, as discussed in Section **Error! Reference source not found.**. In Nanjing, China, SAP HANA has been used in a similar way to provide real time traffic congestion analyses. Nanjing also used Intel®Xeon® processor E7 as their main processor and were able to process 100 million points of data from both floating car data (GPS) and fixed device data (eg. traffic sensors) each day.[40] The software then gives a graphical "temperature" rating to the roads in the city which is used by city management and accessed by over 800,000 members of the public.[41] Figure 3 is an example of the graphical output SAP HANA is able to provide.

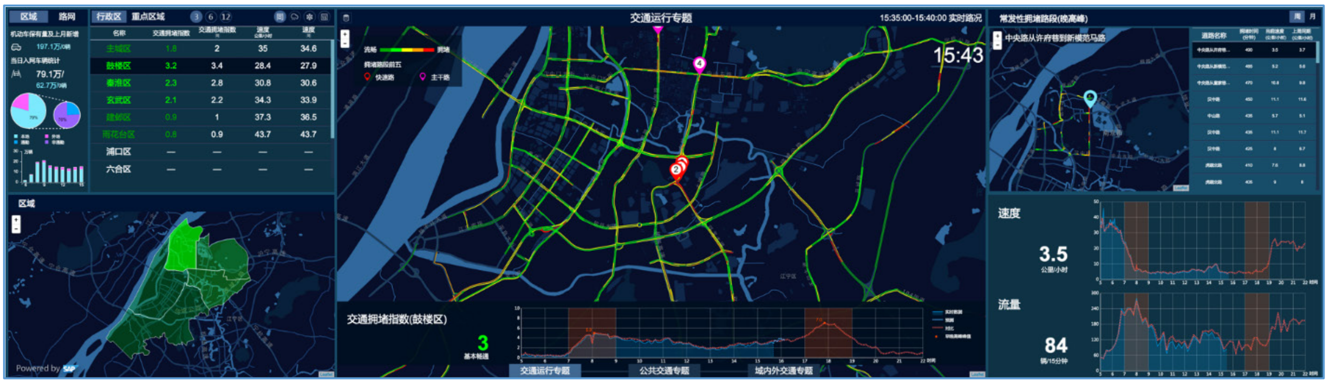


Figure 3: Output from SAP HANA in a traffic management application[42]

The data used by SAP HANA to produce the analysis of traffic congestion above is taken from a variety of sources, such as GPS. The New South Wales Special Emergency Services (NSWSES) have proposed a system that could be processed using SAP HANA, and allows for sharing of real-time information in order to enhance the response of emergency services.[43]

[38] SAP HANA n.d., What is SAP HANA?, op. cit.
[39] Intel 2011, Not limited to ERP applications alone, the Intel® Xeon® processor E7 family also provides new business opportunities via in-memory technology, Intel.
[40] Chen, 2016, Smart Traffic: an IOT solution, SAP HANA Innovation Award
[41] Chen, 2016, Smart Traffic: an IOT solution, op. cit.
[42] Chen, 2016, Smart Traffic: an IOT solution, op. cit.
[43] Fosso Wamba, 2012, 'Big data' as a strategic enabler of superior emergency service management: Lessons from The New South Wales State Emergency Service', University of Wollongon

4 BIG DATA CASE STUDIES AND APPLICATIONS

Harnessing Big Data and associated technologies across the transportation system has huge potential to aid in disaster response. The challenge is to find effective, efficient ways to use Big Data to inform disaster response policy and decision making on both planning and active real-time levels. Multiple different techniques, methods and programs have been implemented internationally and in Australia to apply Big Data to disaster response. This section provides snapshots which cover the following four components of emergency management:

1. Identification of at-risk areas to inform affected communities well in advance of the disaster.
2. Identification of areas which need the most urgent relief immediately after the disaster and the coordination of crisis responders and logistics during the crisis.
3. Provision of guidance to other members of the public to avoid hazards, including the provision of a safe transport pathway away from the danger.
4. Development of a database of emergencies, allowing future evaluation of the effectiveness of the emergency management response and identify optimal strategies for responding to future disasters.[44]

4.1 International Examples

4.1.1 Earthquake Prediction and Response

Multi-source data from past earthquake patterns, real-time satellite images and abnormalities in the atmosphere due to the release of energy or gas are processed to provide an estimate of when an earthquake will occur. These estimates are accurate to within one to thirty days and allow people in the area to be alerted. In February 2015, it was estimated that an earthquake of 6.5 on the Richter scale would hit Sumatra in the coming weeks, and less than ten days later the island was hit by an earthquake measuring 6.4.

Information collected from a variety of sources can be used to relieve complications and assist the recovery process for those effected by earthquakes. Following the Haiti earthquake of January 2010, social media data was used to produce maps which displayed the location, extent of damage and time of occurrence over the country in order to assist those in most need. Information posted to Facebook or Twitter from a variety of people assisted in the production of the maps. As many parts of Haiti are disconnected from the internet, many individuals struggled to post things online and a lot of the information on the worst affected areas had to be sent by volunteers and the government. The database of data produced was used to create an aerial 'Google Maps style' image of the most disaster stricken areas, where the response was most imperative. With the information provided by multiple social media sources, as well as from first responders and aid workers, additional relief workers were

[44] Emmanouil, D., Nikolaos, D. 2015, *Big Data analytics in prevention, preparedness, response and recovery in crisis and disaster management*, WSEAS Press, The 18th International Conference on Circuits, Systems, Communications and Computers, Recent Advances in Computer Engineering, Series 32, p.476-482

able to travel to the disaster zone. This approach can be extrapolated to more developed countries which are able to obtain higher resolution results due to the greater detail in incoming data.

Crucially, this earthquake mitigation case study highlights the existing issues with disaster mitigation. A large majority of emergency sensing and emergency response platforms exist as separate components which do not automatically communicate with each other. The next case study highlights the benefits of integrating both platforms: Rio de Janeiro makes use of integrated emergency management technologies, processing information from a large variety of structured and unstructured sources to provide effective, centralised emergency response.

4.1.2 Rio de Janeiro Operations Centre, Rio de Janeiro

In response to a series of devastating floods and mudslides in April 2010, the Brazilian city of Rio de Janeiro overhauled its city operations. In collaboration with the financial service company IMB, the city launched the Rio de Janeiro Operations Centre (ROC) in late 2010, with the primary aim of preventing deaths during annual flooding events. In preparation for the 2014 World Cup and 2016 Summer Olympics, the operations centre was expanded to cover all emergency response situations.

The ROC collates information from city departments and agencies and analyses the diverse data sets with the aim of better resource allocation and coordination during crisis events. The centre receives over 900 video streams which are processed, aggregated and displayed on a 'Smart-Wall'. With 80 square meters of screen, the Smart-Wall consists of a digital map of the city with access to over 120 layers of data, as shown in **Figure 4**.[45]

Figure 4: The Smart-Wall in the ROC[46]

Big Data is gathered from multiple sources, including 30 government departments and agencies, as well as 560 city-wide municipal and 350 private sector owned cameras. The aggregated data is visualised in real-time on an 80 square meter tiled screen comprised of 120 information layers. The centre employs 400 people and operates at all times. The centre uses data feeds on various items such as weather,

[45] Security Middle East 2015, 'Smart Cities: a smarter operations centre for Rio', *SME*.
[46] Security Middle East 2015, 'Smart Cities: a smarter operations centre for Rio', op.cit.

energy, water, police, gas, electricity, rubbish collection, traffic and medical services.[47] Data layers are updated in real time through the use of fixed sensors, video cameras and GPS devices.[48]

By integrating multiple city emergency agencies into one command centre, the ROC can take various stages of crisis management into account. It can predict, mitigate, prepare and ensure rapid response to emergencies. The ROC is the first city-wide system which acts to mitigate disasters on multiple levels: through complex data analysis algorithms and fusion software, the centre is able to identify and extract patterns and trends as well as the ability to predict where incidents are likely to occur.[49] Historical and real-time Big Data from various sources can be used to predict the locations in which an incident will happen. For example, the system can predict heavy rain 48 hours in advance. ROC focuses on monitoring and responding to weather-related emergencies, but also makes sure daily operations of the transport system are functioning smoothly and continue to function during emergencies. Since the ROC began operating, emergency response time in the city has decreased by 30 per cent.[50]

4.1.3 Smart Big Board, Korea

Smart Big Board is a real-time monitoring system using social data for disaster management developed by Choi and Bae.[51] They observe that during Hurricane Sandy and a flood incident in Gangnam, Seoul, the most affected areas corresponded to 'hotspots' for social media activity. Concluding that geographical data from social media outlets such as Twitter can be used as part of an integrated data analysis system, they subsequently developed the Smart Big Board platform, which identifies areas where emergency services are most needed. During emergency situations, the platform uses social media platforms to generate information on crisis zones by data crawling through social media, analysing disaster-related information and displaying disaster situations and trends on a map in real time.

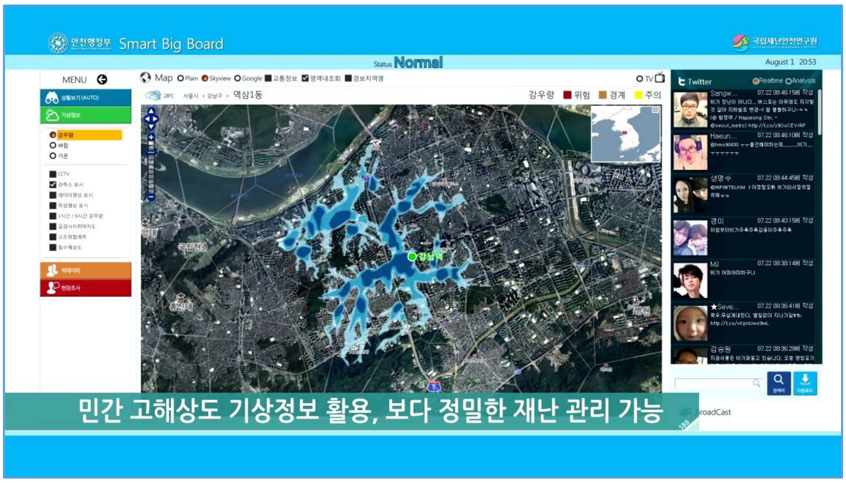


Figure 5: The "Smart Big Board" mapping social media data flows during a disaster

---

[47] Copenhagen Cleantech Cluster 2012, *Danish smart cities: sustainable living in an urban world*, Copenhagen, vol. 2, no. 41.
[48] International Transport Forum 2015, 'Big Data and Transport', op. cit.
[49] International Transport Forum 2015, 'Big Data and Transport', op. cit.
[50] Berst, J. 2013, 'Why Rio's citywide control centre has become world famous', *Smart Cities Council.*
[51] Choi, S., Bae, B. 2015, 'The Real-Time Monitoring System of Social Big Data for Disaster Management', *Computer Science and its Applications*, Springer Berlin Heidelberg, pp 809-815.

The Smart Big Board has been test-operated in Korea since 2013 to monitor cases of contagious disease, heavy snows, and dust storms. The designers plan to improve system performance and usability by developing a disaster sign sensing system which analyses and monitors social media data constantly and activates the full Smart Big Board system when an impending emergency is detected. The end goal is for Smart Big Board to be used to predict hazards based on analysis of social Big Data as well as structured, for-purpose disaster information.

4.1.4 An Integrated International Example: Transport for London

Big Data collection and analysis is integrated extensively into the Transport for London (TfL) system, which handles bus network, trains, roads, taxis, ferries and cycle paths. These connected networks provide huge amount of information through smartcards, ticketing systems, vehicle sensors, traffic signals, social media, GPS location and mobile networks. TfL collects data for both traffic management and emergency responses.[52] Currently, data collected from TfL services is shared through third-party application developers which provide tailored solutions. A number of Microsoft and Oracle platforms are in use for running the TfL systems. Transport planners also plan to integrate open source solutions so the system can cope with future data demand, including increasing the capacity for real-time analytics, integrating a wider range of data sources and improved plan services.[53]

Using data from its extensive transport network, the citizens of London receive real-time information covering various subjects such as weather, air pollution, delays in public transport, availability of public bikes, level of the river, twitter trends in the city, and traffic camera feeds. These sites produce visualisations of complex data, increasing ease of interpretation and analysis for non-experts.[54] The use of data in real-time and future (planned) traffic management has led to direct economic benefits, with increased transport efficiencies. TfL's data was valued at £15-58 million per year and triggered 20 travel apps developed by private companies. Public transport usage has also increased to record highs, with the number of customers on public transport during the 2012 Olympics becoming a daily occurrence.

Emergency repairs on Putney Bridge, which is crossed by 870,000 people every day, led to an unexpected transport bottleneck in surrounding areas. Using Big Data analysis from historical and real-time databases, TfL addressed this disruption by quantifying the affected passengers. Since half of the journeys started or ended very close to the bridge, TfL concluded that this group of passengers would be fine and they needed to manage the other half. A transport interchange was set up and bus services on alternate routes were increased. The affected passengers were also informed by personalised messages. As such, the UK government has strategized the use of Big Data on a broad scale, maximising the opportunities of using such datasets. The fundamental focus behind the Government's strategy is 'opening up' data in transport by making it more widely available across various sectors to improve transparency and encourage economic growth. The Government established the Transport Systems Catapult,[55] overseen by the Technology Strategy Board (TSB), and is set to receive £46.6 million from

[52] Marr, B. 2015, 'How big data and the internet of things improve public transport in London', *Forbes,* May 27.
[53] Cohen, B. 2012, *The top 10 smart cities on the planet*.
[54] Kitchin, R. 2014, 'The real-time city? big data and smart urbanism', *GeoJournal*, 79(1):1–14. ISSN 1572-9893. DOI: 10.1007/s10708-013-9516-8
[55] Yianni, S., Zanelli, P. 2016, *Technology Strategy 2016 for Intelligent Mobility*, *Catapult Transport Systems*, London.

TSB and £16.9 million from the Department for Transport (DFT), with the specific objective to encourage analysis of Big Data.[56]

4.1.5 Transportation Decision-Making (TransDec)

TransDec[57] has been developed by the Integrated Media Systems Centre in the University of Southern California. It includes real-time large-scale traffic sensor data collection, efficient real-time and historical spatio-temporal data processing and data analysis in order to provide a visualisation of transport networks. The system has a three-tier architecture: data tier, presentation tier and query interface. On the data tier, TransDec employs various real-time traffic data such as NAVTEQ and RIITS (Regional Integration of Intelligent Transportation Systems). The RIITS dataset is made of historical and minute-to-minute real-time data provided by various organisations such as Caltrans D7, Metro, LADOT, and CHP, which includes CCTV snapshots, events, bus locations and arterial congestion. TransDec also uses data mined from the US transportation network map and points of interest. Next, the presentation tier and query interface provides a web-based map to the user, enabling spatio-temporal queries which can be used to identify unexpected delays and emergencies on the network and allows faster and better coordinated responses. The TransDec intelligent transportation system provides monitoring, decision-making and management services. However, it is not exempt from the processing challenges faced by many Big Data and Big Analytics platforms as outlined in Section 5.1.

4.2 Australian Examples

4.2.1 New South Wales State Emergency Service (NSWSES)

The New South Wales government aims to increase the use of open data, as '*sharing data allows government agencies to focus on delivering core public services. It encourages innovative solutions to our citizen's problems*'.[58] Formed in 1955 after a series of floods, NSWSES' 24-hour service develops and dispenses information on severe weather warnings. Collaborating with state agencies such as the Bureau of Meteorology, the firm can maintain a bidirectional direct link between its website and the Bureau of Meteorology website during major operations, offering real-time access to accurate weather or emergency information. NSWSES shares human and capital resources with other states during major disasters.[59]

Using multi-platform, multi-source data, NSWSES enables real-time access and sharing of information across local and national government agencies for improved decision making to enhance emergency service response, allowing the movement of critical assets across the state to deliver emergency services where they are required. This also allows planners to make strategic decisions about where to invest in the future, including transport infrastructure that develops new capabilities to reduce system

[56] The Parliamentary Office of Science & Technology 2014, *Big and Open Data in Transport*, UK Government, London.
[57] Integrated Media Systems Center 2016, *Transdec: Big Data for Transportation*, University of Southern California, Columbia.
[58]NSW Government 2016, 'Premier's Innovation Initiative – Open Data', *New South Wales Government*, Sydney.
[59] Wamba, S.F., Edwards A. and Sharma, R. 2012, 'Big Data as a strategic enabler of superior emergency service management: Lessons from the New South Wales State Emergency Service', *University of Wollongong Faculty of Business*, Wollongong.

vulnerability.[60] NSWSES processes both structured and unstructured data from a large variety of sources and formats, providing information to support emergency operation on the field. Privacy issues are handled by NSWSES 'Big Data' governance, which handles trust, accountability and transparency in relation to the collection, storage and sharing of data. While NSWSES is focused on emergency services deployment, their platform and structure can be tailored to suit specific, transport-based requirements.[61]

4.2.2 Bushfire Management, WA and VIC

In Western Australia, more than 5,000 bushfires occurred within the twelve months after July 2013. There is currently a detection system in place in the state called the Aurora Bushfire Detection System, which uses data to predict at-risk areas for bushfires and relays this information back to the public via the government's website.[62] The Aurora system calculates the direction and intensity of the bushfire using several data sets: threat analysis data, time since last burn, forecast weather, drought factors and fire hotspots. The information processed by the system enables Aurora to determine whether there is likely to be any impact on any surroundings whilst also enabling a message to be sent to any residents that may need to evacuate.

On the other side of the country, Victoria is currently in the process of implementing the Resilient Information Systems for Emergency Response (RISER), which investigates new weather monitoring networks, cloud-sourced information and the use of drones as a large collaboration of research to attempt to respond to bushfires more efficiently.[63] RISER's new weather monitoring networks collect a detailed set of data outlining the temperature, radiation, wind speed and direction, soil moisture and humidity to predict not only the likelihood of any bushfires occurring but also the destructiveness. It is also stated that the process will collect information from social media platforms such as Facebook and Twitter.[64]

Both of the techniques outlined above follow very similar principles: the use of multiple data sources to form a high-resolution database which provides projections of the likelihood of a disaster. As highlighted in Section 3, it is the integration of detailed, extensive emergency sensing and response databases which will facilitate effective emergency response networks, aided by resilient transport systems. The current systems in place in Australia use detailed emergency sensing networks, but have yet to go deeper beyond the surface in terms of emergency mitigation techniques.

4.2.3 STREAMS, Australia (except NT)

Brisbane based company Transmax has developed the software platform STREAMS which uses real-time transport data and integrates CCTV, variable message signs and vehicle detectors to produce a

---

[60] Wamba, S.F., Edwards A. and Sharma, R. 2012, 'Big Data as a strategic enabler of superior emergency service management: Lessons from the New South Wales State Emergency Service', *University of Wollongong Faculty of Business*, Wollongong.
[61] Wamba, S.F., Edwards A. and Sharma, R. 2012, 'Big Data as a strategic enabler of superior emergency service management: Lessons from the New South Wales State Emergency Service', op. cit.
[62] SNWA, 2015, 'In focus: big data keeping track of bushfires', ScienceNetwork Western Australia
[63] Kealy, 2016, 'RISER provides new data streams to assist bushfire decision-making', The University of Melbourne
[64] Park, 2012, 'The cutting edge of bushfire surveillance will use drones and social media updates to feed an incredibly sophisticated data engine, making it possible to predict and monitor the unpredictable', SBS

map-based, browser-style intelligent transport system (ITS) interface for transport network management. Despite being an example of small data, the combination of this integrated data set allows STREAMS to build a Geographic Information System (GIS) that models transport network infrastructure in real-time under a single software interface, with the ability to send control requests and manage ITS devices under the existing infrastructure. Currently operating in all Australian states (Northern Territory excluded), managing 1,955 intersections with 48,500+ STREAMS connected devices,[65] the system delivers efficiency and performance across the entire road network and provides the necessary integration of information to make large scale increased performance possible.

While STREAMS is primarily designed to reduce traffic congestion, VicRoads reports that the installation of this system also delivered a 30% reduction in motorway accidents,[66] an indication that better traffic management also reduces unexpected man-made events. The STREAMS system also delivered economic benefits of $94,000 per day, travel time savings of 42% during peak periods, and an 11% reduction in greenhouse gas emissions.[67] Transmax has recently partnered with Parsons Brinckerhoff to introduce the emerging technologies of Big Data in transport management to the Colorado Department of Transport to assist the U.S. in reliability, capacity and safety of its freeways.[68]

4.2.4 Traffic Management and Emergency Response, Queensland

'Managed Motorways' is an initiative of Queensland's Department of Transport and Main Roads that draws upon smart technology to manage the South East Queensland Road network, reducing congestion and improving safety.[69] Again despite being an application of small data, the program manages the flow of traffic through utilisation of variable message, speed limit and lane signs. Merging traffic is effectively controlled using ramp signalling in order to minimise the interruption of flow. Electronic message signs display real-time travel time information to drivers, while roadside data systems such as sensors and CCTV monitor the conditions and enable rapid response to incidents. The Department aims to use the system to reduce stop-start travel, increase the ability to predict journey times and increase the capacity of the roadway, with the flow on effect of reducing traffic accidents and carbon emissions.[70]

More directly aligned with traffic resilience is the Queensland government and Transmax's Emergency Vehicle Priority (EVP) solution, which allows emergency vehicles to travel more quickly and safely. The EVP system uses the location of the emergency vehicle and the flow of surrounding traffic to estimate the vehicle's time of arrival at intersections. Using this information, EVP pre-emptively changes traffic lights to green before arrival to ensure the emergency vehicle can continue through without having to slow down. This system has been found to reduce the travel time for emergency vehicles by 10-18%.[71] Not only is there reduced impact to the surrounding traffic, drivers of emergency vehicles no longer

[65] Transmax 2016, 'STREAMS', *Transmax Pty Ltd*, Perth.
[66] Victorian Auditor-General's Office 2010, 'Using ICT to Improve Traffic Management', *Victorian Government*, Melbourne.
[67] Victorian Auditor-General's Office 2010, 'Using ICT to Improve Traffic Management', op. cit.
[68] Mena Report 2016, *Australia: Queensland congestion-busting traffic technology hits American highways*, GALE|A461442569.
[69] Department of Transport and Main Roads 2016, 'QLDTraffic', *Queensland Government*, Brisbane.
[70] DTMR 2016, op. cit., 'QLDTraffic'.
[71] Transmax 2016, op. cit., 'STREAMS'.

have to navigate through multiple red traffic signals and rely on other drivers to give way, hence reducing the dangers faced by emergency responders themselves.

5 CONCLUSION

Harnessing Big Data brings great potential rewards: Big Data provides a wealth of information that can be mined for high-resolution correlations, relations and predictions and can play a pivotal role in the shaping and development of truly resilient transport systems which enable smart emergency management and response systems that cater to the needs of communities.

5.1 Existing Challenges

Although Big Data can provide key information to evaluate, plan and improve disaster response, the key challenge for its effective utilisation is effectively harnessing multiple sets of data, some historic and some streaming in real time, that are available in a range of formats and platforms, requiring sophisticated data analytics and visualisation software yet to be deployed. Because so much information can be made readily available, especially during unforeseen disasters, software must be developed which can sift out irrelevant or unreliable information and focus on key features which will provide necessary real time inputs into disaster management activities. The real value of these data sets and streams will be the ability to integrate, visualise, analyse and respond to queries at the high speed needed in disaster response.

Current data analytics systems provide limited analysis capabilities with long response times of several minutes, which is an impediment for real-time data analytics. Recently, in-memory computing techniques have been found to achieve significantly higher efficiencies however they are in early stages of development and deployment in Australia. Multiple IT firms are actively working in this field, with researchers currently investigating new methods to improve processor speed and responsiveness. These challenges must be overcome in the future in order for Big Data to be accurately, effectively and efficiently harnessed for disaster management and emergency response. Yet, as a whole, what early emergency responders stand to gain is far greater than what they could lose:

- *Predictive ability for accidents*: Whilst accuracy is still not perfect, the ability for big data platforms to predict likely accidents and identify at-risk areas of the transportation network makes it highly valuable for the preparation and fast response to such events.
- *Efficient and resilient communities in disaster response*: Managing the data created during disaster events is difficult due to the size, complexity and speed at which it needs to be analysed, but big data platforms are now offering the ability to extract the key information to direct and manage the response more effectively and in real time.
- *Assistance leaving danger zones*: Resilient transport systems which can remove commuters effectively from danger zones or unexpected incidents.

The exciting implications of Big Data have yet to be fully realised, especially in Australia, which has not yet integrated its emergency sensing and response platforms. As the case studies presented in this report illustrate, Big Data has the potential to significantly improve disaster preparedness and provide emergency relief. The work conducted and successes achieved internationally from the use of Big Data provides Australia's planners and emergency response organisations with ideas and avenues to harness Big Data to improve the nation's ability to cope with disaster planning and relief.

5.2 Next Steps

The next steps for Australia are to get involved at the forefront of Big Data integration, analysis and application. The software is available (Section 3) and there are already existing 'small data' streams from for-purpose sensors which can be combined with other data streams (eg. mobile phone location data, social media and transport networks) in order to reap significant benefits to communities in both preparedness and response to disasters.

The following sections outline the key factors to be considered:

5.2.1 Better Integration and Standardisation

The decisions related to software specifics lead to the need for standardisation of incoming data streams (such as standardising the format of information being sent). For disaster management in particular, there is need for software to be able to handle multiple incoming types of streams and quickly process them into a single output. This is a particular challenge given audio communication in Disaster events is often across multiple platforms with disaster controllers often having 5 or 6 handsets to communicate with various first responders and support. As this output must be utilised by multiple organisations, further work into who will need the information and the best techniques of collaboration must be identified. If there is a situational need for a flooding or bushfire disaster response system or one that is much more flexible, development of such software systems must be undertaken. Given the serious nature of disasters it is also important that data from unverified sources such as social media is assessed for its legitimacy prior to the deployment of first responders.

5.2.2 Specific Privacy laws surrounding Big Data

Before Big Data systems are implemented there needs to be clarification around open data laws and definitions of what is actually open source and can be used. Especially within disaster response, the selection of data sources and mining of specific, relevant information will often require location technology associated with multi-purpose sensor networks such as social media applications. This is perhaps the biggest barrier to the uptake of this analytics software. Whilst privacy laws already exist within Australia to cover most information, there is need for laws specific to Big Data privacy and usage in terms of emergency response. As public scrutiny into any open source data mining can pose an obstacle to software and platform development, clearly defining the relevant privacy policies is required in the early stages of any platform implementation.

5.2.3 Training

The training of employees in the use of such software will impact the effectiveness of data analytics software. Using and presenting the data in a meaningful way is crucial and thus, appropriate training on not only the platform but also the critical data streams and the mining process is needed to create the most efficient, real time disaster management and response system. A number of the most promising platforms available have already been mentioned in Section 3. Moving forward, the major areas in which such analytics software would need to work within need to be addressed before any final decision can be made. The sources of data, size and speed of data processing and its final applications will all affect the decision on what the most appropriate platform will be. Ultimately, the

best platform to create valuable emergency response optimisation will come down to individual needs and matching them to software specifications.