This book is composed of articles written by A grade students at the School of Computing at University of Derby. These students have been taught the technologies and practices that are at the cutting edge of information security and assurance. The information in this book is aimed at small to medium enterprises that intend to innovate and implement new technology into their business processes.

The topics within cover the benefits and concerns of Big Data and related technologies, software and data management policies. The articles explore questions of ethics, trust, governance, security, audit and provenance as applicable to small and medium enterprises implementing Big Data technologies. Each article is aimed at one aspect of using Big Data for business.

The articles use research from a variety of academic and professional sources, aiming to deliver each topic in an easily digestible format. The book can be read from start to finish, or specific articles can be chosen as and when they apply, as they are written independently of each other. The following articles are published in alphabetical order of the author’s last name.
<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Data for SMEs: Questions of Ethics, Trust, Governance, Security, Audit, and Provenance</td>
<td>Algahtani, Mohammed</td>
<td>6</td>
</tr>
<tr>
<td>Big Data in SMEs - Governance Issues in Operations</td>
<td>Alkanan, Bader</td>
<td>11</td>
</tr>
<tr>
<td>Big Data Analytics: What A Small To Medium Enterprise Should Know</td>
<td>Allen, Gareth James</td>
<td>15</td>
</tr>
<tr>
<td>Big Data for SMEs: Questions of Ethics, Trust, Governance, Security, Audit, and Provenance</td>
<td>Almarri, Abdulhadi</td>
<td>20</td>
</tr>
<tr>
<td>Big Data for Marketing - Security Problems and Challenges</td>
<td>Bensbiet, Abdulmajeed</td>
<td>24</td>
</tr>
<tr>
<td>The New European Data Protection Regulation - Compliance How It Will affect SMEs Regarding Big Data</td>
<td>Cain, Michael</td>
<td>29</td>
</tr>
<tr>
<td>Big Data Monitoring – A Big Data Issue?</td>
<td>Castle, Daniel</td>
<td>34</td>
</tr>
<tr>
<td>Access Control and Authentication in Big Data Storage - How Secure Is Your Data?</td>
<td>Fairbrother, Rachel</td>
<td>38</td>
</tr>
<tr>
<td>Big Data in UK Healthcare</td>
<td>Fowler, Connor</td>
<td>43</td>
</tr>
<tr>
<td>Big Data: Why Aren’t UK SMEs using it?</td>
<td>Fowler, Samuel</td>
<td>48</td>
</tr>
<tr>
<td>Understanding what Big Data Analytics can Provide for your SME</td>
<td>Fox, Jamie</td>
<td>52</td>
</tr>
<tr>
<td>Can SME’s Trust Big Data?</td>
<td>Gamgoum, Youssef</td>
<td>57</td>
</tr>
<tr>
<td>The Two Faces of Big Data - Can SMEs Trust the Accuracy of Big Data?</td>
<td>Holding, Thomas</td>
<td>61</td>
</tr>
</tbody>
</table>
SME’s Securing Big Data - The Attractive Target
Johnson, Ben ................................................................. 66

Big Data and its Location Issues for SME’s
Johnston-Aiken, Hannah .......................................................... 70

Big Data - The Crisis of Information Security in SMEs
Kamoto, Yemurayi ................................................................. 74

Big Data Analytics in Information Security – is it an Effective Solution?
Lovatt, James Peter .............................................................. 79

Can Big Data Be Trusted For Financial SMEs To Invest In?
Malkan, Anish ................................................................. 84

Understanding the Value of Big Data for SMEs - Unlocking The Potential of Big Data and What it can Provide for your Business?
Martin, Adam ................................................................. 89

Storage as a Service: The Hidden Costs
McMahon, Joel ................................................................. 94

A Big Data Approach to Security - An Exploration of Big Data Analytics for Security
Moreton, Nathan .............................................................. 98

Big Data Analytics - An SME and Cyber Security Focused Study
Morland, Sean ................................................................. 102

Should Healthcare SMEs Trust Big Data?
Nadeem, Mohammed Jawad .................................................. 107

Big Data for SMEs - Should SME Managers Trust Big Data?
Nagi, Anwar ................................................................. 112

Big Data for SMEs - Do SME Executives Trust Big Data?
Nagi, Nazir ................................................................. 117

Preventing Big Data Governance Risks: A Collective Action Involving HR and Employees
O’Sullivan, Patrick ............................................................. 122
EU and US Personal Data Protection Laws and their Relationship with Big Data
Pimlott, Ben ........................................................................................................128

Big Data Security and SMEs - Data Governance
Pocock, Andrew ................................................................................................132

Does Security Compliance Help Protect Big Data within SMEs
Rankin, Mitchell ............................................................................................137

Information Security and Assurance - Security Risks and Breaches in SMEs Big Data
Raza, Ghulam Hussain ....................................................................................142

Providing SME’s a Secure Environment for the use of Big Data Resources
Rees, Elliot ......................................................................................................146

SME Security and Governance of Big Data - Big Potential and Big Issues
Searson, Michael ...........................................................................................151

Big Data: Trust Issues with Big Data
Sidhu, Pumandip ............................................................................................156

Big Data – Big Opportunities: Developing Secure Big Data Applications
Silde, Alice ......................................................................................................161

Is Your Big Data Secure?
Sumner, Ben ..................................................................................................166

SME Security - SME Breaches and Security Threats
Tran, Lewis ......................................................................................................171

European Data Protection Reform - Implications on SME’s Big Data
Whittington, Nathan ......................................................................................176

Uncertainty Squared - In Times of Economic Uncertainty, Should Financial SMEs Place their Trust in Big Data Technologies?
Williams, Matthew .......................................................................................181
Information Security and Assurance

Big Data for SMEs: Questions of Ethics, Trust, Governance, Security, Audit, and Provenance

Mohammed Algahtani
Computing and Business Faculty
University of Derby
Derby, United Kingdom
M.Algahtani1@unimail.derby.ac.uk

Abstract—Information system is a system where organization can store, process, receive and generate their important data is called Information System and big data is the process of analyzing vast data sets. This paper discusses the importance of Information System in a Big Data perspective and to the SMEs and how to secure the personal information from unauthorized access. The information is the key to any organization these days and it makes a huge difference among competitors and their business operations. If the data is not secure this mean the organization needs to seriously consider their security measures to make sure their corporate data is safe and secure. This paper discusses those available security techniques by using them a corporate can secure their personal data and also to understand how to identify the risk and threats that can cause damage to corporate data.

Index Terms — Information System; Corporate; Big Data for SMEs; Risk Management; Security Threats.

I. INTRODUCTION

With the introduction of internet and the World Wide Web (www), now everybody can have easy access to the world wide data within their reach. This trend enables them to access information faster than before and more efficiently using modern computer systems and can gain access over Big Data (NHS, 2013). According to Brunel University Report (2014) that the Big Data processing and services will be triple in next few years as predicted to increase by 243 per cent by 2017. Further, Dr Fox added that handling Big Data for SMEs are important and increasingly strategic. With this new advancement any person can have access to information where they can save, print, and access from any location. There are huge advantages of such new development however there are many disadvantages also involved in such capacity when it comes to corporate data or even individual home pc data. There is huge threat from intruders, viruses, or hackers they can gain access to the company core data and can destroy their whole information system which on them they purely rely. In such scenario there is need to make sure the data is secured.

However, at individual level a person can install antivirus on their machine and can secure their personal data. At home or a small network people can establish the Demilitarized Zone (DMZ) as shown in figure below where a person can secure the specific machines from the rest of the world (Goodin, 2012).

![DMZ Security Model](Source: Goodin, 2012)
As the amount of data is growing exponentially in the last few years due to many reasons such as internet growth, social media, cloud computing technology, and mobile devices. Now the problem is how to manage the big data and unlock the economical value out of it through adoption of big data. As shown in figure 2 above, that unique skills are required to handle the big data.

II. SECURE AND MANAGE INFORMATION SYSTEM

The corporate level the Big Data is stored in the form of their product information, manufacturing techniques, customer details, financial data, their employee data and many other important data (QAP, 2013). Now the questions I that is antivirus and setting strong firewall setting is enough to handle big data? The answer is clearly no; not at all. To handle such a big data there is proper policy and guidelines are required which need to be documented all the time for future use. At corporate level, we have to look into what is the information security? It is the process where an organization needs to protect their information from a wide range of threats in order to:

- Ensure business continuity
- Minimize business risk
- Maximize return on investment and business operations.

(Self, 2007)

There are many standards for IT governance, risk assessment, and information security management such as BS 7799-3:2006, ISO 17799:2005, and many more. The use of these standards is to provide a wide range of guidance and policy that can make sure the information that flows inside and outside the corporate environment is safe and secure. This paper will discuss the BS7799-3:2006 guidelines for information security risk management. However, risk management is a key factor to develop an information security management system which meets the fundamental requirements of ISO 27001:2005 standard. The new British Standard BS 7799-3:2006 provides guidance that covers following:

- Risk Assessment
- Risk Treatment
- Management Decision Making
- Risk Re-Assessment
- Monitor and Reviewing of Risk Profile
- Information Security Risk in the context of the corporate governance
- Compliance with other risk-based standards and regulations

(Self, 2007)

The main objective of this standard is to identify, evaluating and managing information security risks if the business wants to keep their information secure as shown in figure 3 below.

III. IDENTIFY RISK AND MANAGE THE SECURITY OF INFORMATION SYSTEM

Management at corporate level should define the policy and guidance that clarify the direction and support for high level of information security management statements which can lay down the key information security directives for entire corporate. This is also called a policy manual which is a set of information security standards, guidelines and procedures. The recommendation for better security management under the standards as follow:

A. Organization of Information Security

There is need to design a suitable security governance structure at corporate internal and external level. At internal level, senior management should provide direction and support in order to approve the policy, roles and responsibilities, and define the information security functions (Whitty, 2013). Contract should be established with relevant authorities (e.g. law enforcement and special interest groups). At external level, the information security should not be compromised at third-party products and services. If the risk is found, there should be proper guideline to reduce them immediately.

B. Asset Management

There should be clear red-line where organizations should know what the core assets they hold and also have the proper guidelines to manage them safely. All important assets must be accounted and have nominated owner who will be responsible for them. All IT related assets such as hardware, software, data, system documentation, storage media, ICT services, UPSs, etc.
must be maintained. There should be recorded locations for each asset and must be identify the owner who is in charge.

C. Human Resource Management

There should be providing a system access to the employee who is working at the corporate.

- Prior to Employment: prior to employment, HR need to properly screening of the applicants, and including in contract signed the term and conditions of the roles and responsibilities of security agreements.
- During Employment: They should be trained, educated, and well informed the security threats and about security procedures.
- Termination or change of Employment: If the employee is no more working with the organization, their access should be closed immediately and he/she should return corporate assets.

(Anderson, 2001).

D. Physical and Environmental Security

The corporate valuable assets should be protected from any natural or manmade disasters. There are many types of incidents can happen at working places such as fire, overheating, accidents, loss of main power, etc. These should be keeping secure where sensitive IT facilities are installed from unauthorized access. Further, the equipment should be kept secure and audited. Power suppliers, cabling should be secure.

E. Communications and Operations Management

There are many components that need to be addressed and managed properly these include:

- All kind of IT procedures, manuals, and policy should be documented and all changes made during the new development should be controlled versions.
- There should be proper control setup for any threats from viruses, malicious and mobile codes.
- Keep routinely back-up for all core services.
- The network should be properly managed and monitor for security controls these includes firewalls, system profiles, etc.
- All kind of information between departments, organizations must be controlled and protected.
- All activities inside and outside organization should be monitor and control.

(Vu, Zouikri, and Deffains, 2012).

F. Access Control

There should be proper controlled to prevent unauthorized access to the corporate core systems. There should be proper user management, define user profiles, rights, special restriction on certain users, network access controls, application access controls, and mobile computing (Greene, 2008). These controls must be ensured that nobody can cross from their access restriction. It is often seen these days many employees brings their PDAs, portable PCS, and connect with the organization systems. These things need to be seriously considered.

G. Information Security Incident Management

There are incidents happens all the time at work place. These incidents, events, accidents and mishaps should be reported properly (Greene, 2008). There should be security alarm system installed at the workplace so it keeps everybody alert if there any incident happens. Further, all employees should be trained and explain about their incident reporting responsibilities. There should be a mechanism where all these incidents properly manage and implement continuously improvement.

H. Business Continuity Management

There should be proper disaster recovery plan and business continuity management and contingency plan. These plans should be properly test all the time because if it not properly tested time by time at the crucial time the system can lead to sever disasters. These controls are designed to minimize the impact of security incidents.

I. Compliance

All corporate who need to make sure their data and information must comply with applicable legislation such as data protection act, copyright laws, cryptography restrictions, roles and evidence (Erkkil, 2007). There are two main issues need to be considered:

- All managers and system owners must make sure that their policies and standards are strictly aligned with complains for example they properly review the system tests, penetration tests, platform security reviews.
- The information system audit should be carefully planned that can help to minimize the disruption of operating systems and also helps to protect against intruders or hackers.

IV. CONDUCT A THREAT AND RISK ASSESSMENT

Threats are described anything that can cause the destruction or interruption of any services that is valuable for organization. There are two kinds of threats one is caused by human and other is non-human threats.

- Human Threats: Hackers accidental, backup operators, electricians, technician, non-technical staff, theft, inadequately trained IT staff, etc.
- Non-Human Threats: Floods, Viruses, Fire, Air pollutions, Heat control, plumbing, electrical, etc.

(Torgler, and Schaltegger, 2006).
Threats are described anything that can cause the destruction or interruption of any services that is valuable for organization. There are two kinds of threats one is caused by human and other is non human threats.

V. CONCLUSION

In this paper we found that big data is new to many companies especially SMEs which mostly concerns about their profit and these SMEs are the major player in the market in creation of jobs and the UK economy. Due to size of the company mostly used outsource their Information Technology infrastructure outside from another IT specialist companies most famously cloud providers. Where big data is concerned it is a high-volume, high-velocity and high variety of information assets that demand cost effective innovation in order to enhance the insight decision making power for SMEs (Beyer and Laney, 2012). SMEs can use big data for their brand promotion, making marketing strategies, and future forecasting but the main problems arises is the ability how they secure and maintain the integrity of such data. This kind of question needs proper security standards and framework where SMEs can apply to assess the risk, threats, ethical considerations and governance. These famous governance and security standards are ISO 27001:2005 new British Standard BS 7799-3:2006.

VI. REFERENCES


Big Data in SMEs
Governance Issues in Operations

Bader Alkanan
School of Computing and Mathematics
University of Derby
Derby, United Kingdom
B.Alkanan1@unimail.derby.ac.uk

Abstract—This article discusses the various governance issues that are faced while using big data for operations management in small and medium enterprises. Big data that involves different textual, audio and video data obtained via internet, mobile phones and other sources, usually exhibits characteristics like volume, variety and variability. It may not only be used in big organizations but may also be adopted by SMEs for finding patterns and trends in data but the use of big data comes with numerous challenges such as lack of skills, high cost etc. Data governance is one of the biggest problems of using big data for operations management in SMEs and can lead to mismanagement, bad governance culture and ineffective restructuring. The adoption of guerilla approach that includes the development of appropriate strategies and policies and regulations can thus help in solving the issues and for managing the application of big data in SMEs.

IndexTerms—Big Data, SMEs, Operations Management, Data Governance, Governance Issues, Guerilla approach

I. INTRODUCTION: BIG DATA IN SMES

Big data refers to the data sets that are extremely large and complex in nature and also generally goes beyond the conventional processing capacity and size of the conventional databases. Russom (2011) says that big data is one of the most emerging technologies in today’s world, the use of which is being highly promoted by different industries, companies and fields. The characteristics of big data are found to be significantly different than conventional data systems. It is not only more complex than conventional systems but is also more difficult to store, acquire, process and analyze. Russom (2011) highlights that with the sudden explosion of data in the world, the use of big data is becoming more and more important for innovation, market research and competitive study.

A. Characteristics of Big Data

There are four main V’s that have been identified to define the characteristics of big data. A brief overview of these V’s is presented below (Mayer-Schonberger and Cukier, 2013).

- Volume: The volume of big data is generally very large and big data is characterized mainly by its volume. Chaudhari (2012) has identified that the volume of big data is continuously increasing and becoming large in nature because the data has been in storage since years and the use of social media leads to the generation of unstructured data, which contributes to the volume.
- Velocity: Another unique characteristic that differentiates between big data and regular data is the velocity or speed at which it’s generated. Because of the use of internet, mobile phones and other technologies, big data is generated at a very high speed. (Snijders et al. 2012).
- Variety: Big data also exists in different varieties, formats and structures. It comes in structured, as well as non-structured manner. The data system can be obtained in different formats, such as text data, data obtained in the form of images or the data feed obtained from internet (Mayer-Schonberger and Cukier, 2013).
- Variability: The inconsistent and variable nature of big data is a characteristic that defines the nature of big data. The data becomes highly variable in nature because the data obtained from social media and that obtained from mobile phones and internet can significantly change with time and never remains the same. This not only makes it difficult to predict data load but the management and analysis also becomes challenging.

B. Sources of Big Data

Big data collected or gathered by an organization may be obtained from several internal, as well as external sources. Internal sources are collected in an organization via transactions, log data, as well as emails. Companies usually record all the information related to its sales and customer transactions, which are then stored and analyzed for making business related decisions. The companies can make various decisions related to their marketing, operations, profit forecasting etc. based on the information based on customer transactions and information. The internal emails are also usually stored by the companies and the information can then
be used for making business related decisions (Boyd et al. 2012).

The external sources of big data include social media, audio and photos and videos. Social media includes the data and information obtained from different social networking sites, such as Facebook, LinkedIn and Orkut. There is a plethora of audio and video data available and shared on different mediums, such as YouTube and Pinterest. All these data combined form a significant part of big data and is used by companies in different forms (Bollier et al. 2010).

C. Common Problems with Big Data

- Ethical Issues: One of the biggest problems observed while using big data is ethical issues. With the availability of information and data via different mediums, one ethical problem that always arises is that of data privacy. It is ethically wrong to share information related to people with a third person and hence it is important to protect and safeguard the privacy of data collected from people to maintain information ethics.
- Trust: Trust is another big issue related to big data in organizations. Lynch (2008) highlights the lack of trust that exists in organizations, which prohibits departments from using big data collected in other organizations. It has also been observed that often management and people do not trust information gathered with the help of big data and hence the credibility issues keep on arising.
- Security: Though the development of internet has connected people all throughout the world, it has also become easier for people to gain access to personal information of people by hacking into their accounts. Hence, keeping big data secured to avoid it being misused is one of the biggest challenges associated with big data (Rabl et al. 2012).
- Audit: Auditing the big data obtained from the internet and other sources is a huge challenge and hence it becomes difficult for companies to make use of the data before ensuring its reliability, validity and originality (Howe et al. 2008).
- Governance: Governance problems arise in big data when the data is not appropriately managed or used well. The issue usually arises because of the versatile nature of big data, which exhibits high velocity and are usually present in huge volumes. The complex nature of data is what makes it difficult to analyze and manage, thus causing governance issues (Rabl et al. 2012).

II. BIG DATA AND SMEs

People are generally under the illusion that big data is meant for big organizations that have millions of customers or have very large operational activities but discusses that big data is also extremely effective and beneficial for the SMEs, i.e. the small and medium sized enterprises. The author highlights that the SMEs have started to identify the significance of big data and are starting to make use of the same for identifying trends and patterns in order to gain competitive advantage in the market. It has been argued that using big data may be an expensive and resource utilizing activity and cannot be feasible for the SMEs. However, they may still incorporate these tools and make use of sustainable IT features to incorporate big data for decision making purposes. The main focus of SMEs is on the collection of data because once the data is collected, analytical techniques can be automatically used to analyze it further. Use of customer relationship management (CRM) systems can help in capturing data in better manner for effectively making use of the resources and available big data (Zikopoulos and Eaton, 2011).

Various applications of big data for the SMEs are:

- Analysis of the trends, as well as patterns existing in their sales, marketing and finance divisions
- Enhancing sales and marketing with the help of collected information on sales and customers
- Development of new products and services
- Prediction of operational demand
- Facilitation of operational practices and activities

A. Problems with the use of Big data in SME Operations

The use of big data in operations management has become very common for SMEs. Building on the business intelligence and different IT expertise internal to a company, SMEs make use of big data for forecasting the demand. This helps in operations management, as the production and manufacturing requirements can be determined on the basis of this forecasting (Bizer et al. 2012). But there are several problems that are faced while making use of big data for operations management in SMEs, which are discussed below.
Lack of skills: One of the biggest problems of using big data for operations in SMEs is lack of skills. As the concept of using big data is relatively new, people are not appropriately skilled to use it well and to apply adequate technologies for analyzing it appropriately according to the needs and requirements of organizations. Lack of skills creates a huge issue for the management of operations in SMEs (Agrawal et al. 2011).

Defining Big data: SMEs are used to adopting traditional methods for the use and analysis of data and the traditional ways of carrying out different activities. Hence, they are not very clear with the method of defining big data and often are not very sure of how to address the question of big data and where to use it. Hence, this creates a huge challenge (Agrawal et al. 2011).

Cost: SMEs work on low operating costs and it becomes expensive at times to gather and use big data in adequate manner. Thus, cost is another big issue that hinders the application of big data for operational purposes in the SMEs (Singh & Singh, 2012).

Integration: Integration here refers to the integration of activities of different functions across organizations. It becomes difficult for SMEs to integrate information obtained from big data to draw effective conclusions (Russom, 2011).

Data Governance: The process of managing and integrating adequate policies and regulations for using big data is complicated and complex in nature and hence SMEs often face the challenge of making use of suitable policies for the purpose of appropriately collecting and interpreting big data (Khatri and Brown, 2010).

B. Governance Problems with Big Data

Governance refers to the processes and activities related to different actions and processes that are related to the establishment of policies that can help in appropriate monitoring and implementation of different practices for organizational management. There are several governance problems that usually occur during the use and implementation of big data for organizational activities and purposes (Rosenbaum, 2010). Some of the common governance problems observed while using big data for operational management in SMEs are discussed below.

- Unrelenting Underperformance: The underperformance of the governance structure for big data management in SMEs is a huge challenge. If the data is not adopted in an adequate manner, it may lead to underperformance of the operations division of the firms, thus causing serious losses.
- Lack of focus: Agrawal et al. (2010) say that the experts and analysts who analyze and study big data from different angles and perspectives need to be focused towards the purpose for which they are using the same. A lack of focus can lead to problems in the adequate use of data.
- Undisciplined capital allocation: The decisions if based on big data may help in the allocation of capital but if the data is not used adequately, it can lead to inappropriate allocation of capital in the organizations (Kleiner et al. 2012).
- Operational Mismanagement: Experts often fear mismanagement of operational activities because of inadequate use of big data. If big data is not used appropriately, it can create several problems in the management of different operational activities. Inaccuracy in data can or its governance can create mismanagement of different processes and the controls used for them (Kleiner et al. 2012).
- Ineffective Restructuring: The ineffective governance of big data in the small and medium enterprises can also lead to restructuring problems. Restructuring involves the re-formation of policies and regulations for the management of different operational activities and if they are not formed in a suitable manner, the restructuring can create additional problems instead of solving the existing ones (Kleiner et al. 2012).
- Abysmal governance culture: Every organization has a governance culture, which defines the governance practices and policies that are used. The generation of bad governance culture is yet another problem that can occur while using big data in SMEs. Existence of abysmal governance culture can not only create management issues but can also affect the performance of the SMEs in a negative manner (Kleiner et al. 2012).

C. Guerilla Approach for Solving Governance Issues

Guerilla approach for solving data governance issues refers to the adoption of the guerilla principles for problem solving purposes. It means that strategies requiring low cost and less
resources may be used by the means of adopting different techniques and processes for effectively managing governance challenges in organizations (Singh, 2009). Different aspects of managing governance problems with the guerilla approach are discussed below.

- **Strategy**: The organizational strategies of the SMEs should be developed in a way such that it specifies the use of big data collected by these organizations for decision making purposes (Singh, 2009).
- **Organizational Policies and Processes**: The policies and processes that are developed and used in an organization should be specific with respect to the use and application of big data for different purposes (Singh, 2009).
- **Monitoring and Measurement**: It is not only important to develop new policies and processes but it is also crucial to develop new ways of monitoring the progress and implementation of them to ensure that they are being effective (Singh, 2009).
- **Training and communication**: If the employees of the SMEs are trained well, they can follow the rules and regulations effectively and hence can use big data in a cost effective manner, thus solving the governance problems.

III. CONCLUSION

The study highlights that the management of big data and knowledge resources is a huge challenge in the operations division of SMEs. In order to manage big data in a cost effective manner, use of guerilla approach is suitable and lead to the development of appropriate strategies for solving these governance issues while using big data. The guerilla approach requires changing the basic structure and factors associated with the use of big data for ensuring that the companies benefit maximum as they can from it.

IV. REFERENCES


Rosenbaum, S.(2010), Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access, Health Services Research, vol.45, no.5p2, pp.1442-1455


Singh, J.P.(2009), Multilateral Approaches to Deliberating Internet Governance, Policy & Internet, vol.1, no.1, pp.91-111

Big Data Analytics: What A Small To Medium Enterprise Should Know
A Look at The Benefits, Risks and Security Issues

Gareth James Allen
School of Business, Computing and Law
University of Derby
Derby, United Kingdom
g.allen9@unimail.derby.ac.uk

Abstract—This report aims to provide a small to medium size enterprise with the information it needs to help make an informed decision about whether big data analytics (BDA) is right for them. It will focus on explaining the benefits to using BDA and also reflect on the risks and security issues associated with big data.

Index Terms—Big Data, Advanced Analytics, Security, Data Protection, Auditing

I. INTRODUCTION

The world of big data analytics (BDA) is a rapidly growing one, Prickett Morgan (2012) highlights that the big data market is expected to grow at an annual compound growth rate of 39.4% and reach a value of $16.9bn by 2015, contrast this to its value of $3.2bn in 2010 it is very clear how quickly this technology is being adopted and the effect it is having on the business environment. Interestingly he goes on to say that the big data market is growing at a rate seven times that of the whole of the rest of the IT market. This report is going to look at the benefits BDA will provide a small/medium enterprise (SME) and what security issues may arise from using BDA, so that you can make an educated decision on whether BDA is the right tool to further build up your SME.

II. FIRST OF ALL, WHAT IS BIG DATA ANALYTICS?

Of course before looking into what BDA can do for you, you need to know what it is. Prickett Morgan (2012) quotes IDCs definition of big data as a data store that contains at least 100TB of data, and companies are storing data in memory and processing it within that memory without pushing it out to disk. IDC also say that big data is about “data in motion”, this is the capture and processing of real time, streaming data. They also go on to say that to be classed as big data, data sets have to grow by at least 60% a year; Also that the data systems are deployed on “scale-out architectures”, this basically means on clusters of servers and storage. Finally they state that systems should be “wrestling” with data in at least two different formats.

“The three Vs of big data (volume, variety, and velocity) constitute a comprehensive definition, and they bust the myth that big data is only about data volume.” (Russom 2011) also supports this definition. So the term “big” in BDA doesn’t just mean you need a lot of data (Volume), it also needs to be of different formats (Variety) and it needs to be moving or in real time (Velocity). These are the factors that makes BDA a powerful tool.

A. How can big data analytics help your SME?

Big data analytics has the potential to aid your SME in taking its next steps towards intelligent business decision making. “Using big data enables managers to decide on the basis of evidence rather than intuition. For that reason it has the potential to revolutionise management.” (McAfee & Brynjolfson 2012). With more effective management decisions achieved through the use of big data analytics your SME will increase its revenue, providing the right decisions are made. Bromehead (2014) states that “Big Data Analytics, including Predictive Analysis, can help companies identify opportunities to boost revenue, service quality, brand image, and even job creation.”

B. What market intelligence does BDA open the door to?

Huster (2005 cited in Trim & Lee 2008) suggests that “the ability to fully understand, analyze, and assess the internal and external environment related to a company’s customers, competitors, markets, and industry to enhance the tactical and strategic decision-making process”. This evidence clearly points out that the more an SME can find out about the factors highlighted by Huster the better an SME will perform. Mayer-Schönberger & Cukier (2013) refer to a case study conducted by Google where they were able to predict the locations of outbreaks of the flu virus based on internet searches. This was done by using the vast amounts of data that Google collects every hour and analyzing it so that these predictions, based on
the evidence BDA had provided them with, could be made. Using this method an SME could predict customer behavior using real time, volatile data and make effective business decisions to increase revenue. It is looking like BDA is a very valuable tool to be using with SAS (date unknown) stating that “big data – through high-performance analytics – could add £216 billion to the UK economy by 2017”. This raises the question of is there any competition to BDA?

C. BDA vs. Data warehousing (DW)

A data warehouse is defined as by Inmon (1993 cited by Atkinson 2001, p.35) “A data warehouse is a subject-orientated, integrated, time-variant, non-volatile collection of data in support of management’s decision making process”. Compare this to the BDA definition where the data has to be volatile, and it is this velocity that allows its users to make informed business decisions in a much shorter time scale this is supported by “One of the key differences between traditional data and Big Data is not just its size but its form. It is constant, live, and flows through an organisation, rather than be a historical lump of data that is not going anywhere” (Anon. 2013). This rapid decision making can be seen in figure.1(Vellante 2013) where the breakeven point is reached much more quickly using BDA then DW.

![Value of Big Data Analytics](image)

Figure 1 (Vellante 2013)

This clear reduction in time scale to breakeven suggests that effective business decisions can be made much more quickly using BDA and in terms of extra revenue generated could be worth $10s of millions to an SME. From this evidence it shows that BDA does still have competition from the techniques that have been used by businesses to extract business intelligence for several years now. However it does appear that BDA provides the greater reward but does it come with greater risk?

III. THE COSTS, CHALLENGES, AND RISKS ASSOCIATED WITH USING BDA

A. Consider the cost

All businesses are constantly looking for ways to improve their situation especially SMEs, where the wrong decision could lead to irreversible damage to the company, and BDA can provide an SME with the information to make the right decision and reduce the risk of catastrophic failure, however it needs to be done cost-effectively. Herodotou et al. (2011) supports this idea stating that “Timely and cost-effective analytics over “Big Data” has emerged as a key ingredient for success in many businesses, scientific and engineering disciplines, and government endeavors”. But do BDA and cost-effectiveness go hand in hand? Cicconi (2006) makes the suggestion that Technical innovation comes with high costs. Bryant et al. (2008) has tried to put a monetary value on the start-up of BDA within a company claiming that the required hardware would cost around $50 million mark and an estimated running cost of BDA is to be around $10 million per annum. Refer back to figure 1, assume that the initial loss is the cost of the software only. Adding the initial cost of the hardware, which most SMEs will need to acquire before starting to use BDA, it begins to paint a new picture. The graph shows that the cumulative cash flow doesn’t break the $50 million mark until around 20 months after the initial investment therefore an SME that has made this investment would not see a return on their investment until around 20-24 months down the line. However this value was assigned to the technology in 2008 and Moore’s law states (Moore 1965 cited by Rouse 2005) that the number of micro components that could be placed in a microchip at lowest manufacturing cost was doubling every year. More recently this has changed to every eighteen months. What this means is that the technology that would have cost $50 million in 2008 would only cost around the $3.125 million mark today, providing the processing power remained the same. Or you could have a substantially more powerful IT infrastructure in place for your $50 million.

B. What are the security issues associated with big data?

As with any new technology there will be new security threats and risks that need to be assessed so that effective governance strategies can be put into place to prevent these risks turning into a SMEs worst nightmare. One security issue linked to BDA is caused by how complex this analytical process is, due to its very nature, the variety and velocity of the data, BDA proves increasingly difficult to keep the whole system secure. This is supported by Audestad (2005 cited by Werlinger et al. 2009) who suggests that a reason one hundred per cent security coverage isn’t achieved is because of the complexity of technology. Coming back to this idea of complexity, MongoDB inc (2014) states the typical company dealing with big data studies many data types from text to video files in order to build a complete picture of their business environment. Combine this with data being stored in clusters of servers rather than one centralized server as with traditional databases the system as a whole becomes very complex. They go on to mention that Facebook ingests over 500 terabytes of
data a day which only highlights the fact that BDA is a technology that will be difficult to keep completely secure, you only have to look at the what happened to Target between 27th November and 15th December 2013 when for this entire period cyber criminals were able to gain access to and export over 40 million customer accounts including credit/debit card numbers and the corresponding security codes for each card. This clearly highlights the need for tight security and regular audits to ensure no breaches like this ever occur because to an SME this could be crippling. Incidentally Skariachan & Wahba (2014) report that retail companies are only spending four per cent, banks are spending five and a half per cent, and healthcare companies are spending 5.6 per cent of their respective Technology budgets on security which is one reason why so many governance failures are occurring in relation to IT security. To effectively manage big data and keep it secure, enough should be spent to maintain high levels of security and consistent (maybe even constant) auditing as a minimum. Now this may seem overkill to some SMEs but with the rise in economic crime, Price Waterhouse Cooper (2014) report that 37 per cent of companies report this type of crime (that’s a three per cent increase from the previous year), they then go on to say that of these reported cases 24 per cent of the companies are victims of cybercrime. This evidence clearly outlines the need for effective security management, which may come at a higher cost, but offset this cost against how much an SME will lose in compensation claims and law suits should the worst happen.

C. Keeping track of your data, is it even possible with big data?

“Today’s business requires that a company’s data are managed in a centralized manner”. (Silvolta et al. 2011) which sounds logical (but is it possible with big data?) the very definition of big data is that it is volatile, it’s constantly on the move, growing in real time and adapting to changes in the business environment, which is good for the decision maker that wants to remain on top of his or her market; the problem arises when the data is not in a centralized location, the data still needs protecting from attack especially when the data is of a sensitive nature. To ensure that BDA is working efficiently for a SME a balance between the freedom the data has to grow and the security strategies put in place to manage the data needs to be met. After all spending that bit more to keep tabs on all of your data should prove to be the right decision, Snow (2008); Batini et al. (2009); Redman (2001) (cited in Silvolta et al. 2011) justify this claim saying that incorrect or mismanaged data costs the retail industry alone $40 billion annually in missed opportunities, failed deliveries, invoicing issues and mistakes. Effective data management and governance strategies should prevent these missed revenue opportunities and keep the data secure from any malicious activity.

IV. WHAT ARE AN SMEs RESPONSIBILITIES WHEN USING BIG DATA?

To quote Franklin D. Roosevelt “great power involves great responsibility”. And big data generates a lot of power for those who can analyze it effectively. With this though comes the responsibility to ensure the data is managed properly and protected against sources of harm, whether that be human threats or non-human such as virus etc.

A. What Does UK law say?

All companies registered in the UK must adhere to the DATA PROTECTION ACT(1998) which states in Schedule 1, part 1, section 7 that “Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data” (Great Britain, Data protection act 1998). Therefore any data stores that fall under a SMEs control/ownership (also known as provenance) need to be protected by adequate security practices to prevent a breach of the law that could lead to substantial monetary penalties and sanctions. The act also states in Schedule 1, part 1, section 8 that “Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data” (Great Britain, Data protection act 1998). Due to big data is so volatile extra efforts need to be put in place so that a SME doesn’t fall foul of this section of the act; especially when dealing with third parties outside the EU or if a decision is made to outsource some data to provide an extra revenue stream.

B. What steps should a SME take to ensure the security of its big data?

Keeping its data safe is vital to any company, and this does not change with big data. If anything it becomes all the more important. The International Organization of Standardization (ISO) states in chapter 5 section 5.1 sub-section 5.1.1 “at the highest level, organizations should define an “information security policy” which is approved by management and which sets out the organization’s approach to managing its information security objective”. (BS ISO 27002 2013). The big data that a SME using BDA keeps becomes that company’s asset and chapter 9 section 9.1 sub-section 9.1.1 suggests that “asset owners should determine appropriate access control rules, access rights and restrictions” (BS ISO 27002 2013). With the volatile nature of big data, information transfer policies should be in place, this is supported in chapter 13 section 13.2 sub-section 13.2.1 which advises that “formal transfer policies, procedures and controls should be in place to protect the transfer of information through the use of all types of communication facilities” (BS ISO 27002 2013). Following these guidelines, an effective governance strategy can be implemented to keep data secure and prevent malicious (maybe even damaging) behavior. Finally a SME should consider Disaster recovery Hawkins et al. (2000) support this idea by saying “The migration from centralized mainframe computers to distributed client/server systems has created a concern on data security. If a disaster occurs to the organization that destroys a server or the entire network, a company may not be able to recover from the loss. Developing an effective disaster recovery plan will help an organization protect them from data loss”. Since Hawkins et al. wrote this disaster recovery has
become known as ‘business continuity’. A SME should have a business continuity plan in effect and guidelines for this process can be found in the document ‘BS 25999’ available on the British Standards website.

C. Is there any protection available to the SME if something should go wrong?

In short the answer is yes, protection of your data is available. The catch is you have to give your data a monetary value. Vladimirov (2014) says that only one in ten companies assigns a monetary value to its data which he admits to being difficult to do due to there not being a tried and tested method in place. He does however highlight that IBM a few potential strategies by measuring the data’s volume, variety and velocity (known as the customer lifetime value) a company can determine how much revenue its data is responsible for and give it a monetary value. Once this is done the data can be insured.

V. CONCLUSION & RECOMMENDATION

To summarize big data clearly has advantages over other data analysis techniques. Allowing companies to make real time business decisions being just one of them as well as the potential earning power it seems to provide. However as with all emerging technologies it has its risks and security issues, the incident with Target in late 2013 being one of the most high profile security failures with regards to big data to date, meaning that substantial running costs are going to have to be endured to ensure the security of the data being processed/stored/outsourced.

From the evidence gathered in this report big data looks like a winner but only if the company has an effective governance strategy in place to ensure its security and efficiency. The start-up costs need to be addressed as well, an SME looking to implement a big data analytics venture of its own must be in a position to cover the costs over a substantial period of time before a return on the investment is seen, do refer back to figure 1 in chapter 1 once a return is seen it is quite a large one. To that end, with the right strategy and controls in place big data could quite possibly make the reward worth the risk.

VI. REFERENCES


Skariachan, D & Wahba, P. (2014) U.S retailers face pressure to raise cybersecurity spending. [Online] Available at:


Information Security and Assurance

Big Data for SMEs: Questions of Ethics, Trust, Governance, Security, Audit, and Provenance

Abdulhadi Almarri
Computing and Business Faculty
University of Derby
Derby, United Kingdom
a.almarri3@uinimail.derby.ac.uk

Abstract—This paper discusses the importance of Information system in the corporate environment and big data technology. Though, security is an important aspect for data and if it breached then it often leads to problems. The data is key to any organization or business process and recently it has been changed from simple data to big data therefore to protect them from unauthorized access is very important task. Therefore, the security issues are serious to information systems because of increasing threats from internet and other form of attacks. Therefore, it is very important that system architecture, or designer need to develop the methods for specifying the information system security. This paper will first discuss the importance of big data for SMEs and then develop an awareness of the available techniques to secure and manage an information system in a corporate environment and then develop the ability to identify the risk and the technique how to manage the security of an information system.


I. INTRODUCTION

Information Technology always helps business and individual works to revolutionize their effort the way they interact. Big data has generated enormous amount of discussion these days. According to IBM (2012) a big data is a data which is handle and process on extreme scale in order to derive maximum value. There are significant opportunities a big data can create for Small Medium Sized Enterprises (SMEs) however, there are huge security concerns and unique challenges which also affiliated with big data concepts. This paper recommends the security awareness plan by implementing security standards into big data environments in order to reduce the costs, risks, and deployment plan. Big data spans three important dimensions such as volume, velocity and variety as shown in figure 1. There are significant use of big data these days, some of the big data projects include: daily million Tweets are sent which turning into terabytes, live streaming programs and CCTV surveillance system where 100’s of live feeds are recorded all the time, and similarly on stock exchange where millions of trade events creating huge amount of data. These all are the example of Big Data.

The Information Technology (IT) Industry is at the heart of development future resilient Information System (Andrew, 2013). The protection of IT and its components are highly demanded than ever before because for an organization the most important asset is their data. Information System (IS) become more critical in human society in all their life style from email to their shopping activities and this does demands more security than ever before especially when high volume variety and velocity makes difficult to ensure data integrity.

This is also because these days most of the data is store on remote location which human cannot access directly if the services are out of reach for example social media networking sites (Larry, 2006). Further, like healthcare organization where the data of patient is stored on their data base and if there is not proper security provided to this system there are chances that medical history of an individual to be breached which could have serious consequences for particular individual.

According to Mouratidis, Giorgini, and Manson (2010) that security is not fully integrated at design level of the information system and mainly considered after the design of the system. The below figure 2 shows the main features and problems that connected with large data sets and explain how big data technologies can solve these problems to the SMEs.
Fig. 2. Features and Problems Connected to Handling different types of large data set (Source: NESSI, 2012)

For this reason Information Security to practice of defending information or data from unauthorized access or users. The main aim of the Information Security, is to provide protect corporate information/data from disruption, destructions, disclosure, illegal use, inspection, recording, or misusing (Lineman, 2013). There are two major aspect of the Information Security which is IT Security and Information assurance. IT security also sometime called computer security which need to protect data that stored in a big organizations and Information assurance is referred to the act to make sure the data is not lost even the critical situation arise (Sweren, 2006).

The most important places that need data protection at high level includes military, hospitals, governments, financial institutes, schools, colleges, universities and many other private and public sectors. Therefore, keeping the data safe is very important because these mentioned organizations data can be used in the wrong way. We all depend on data that affiliate with us in the form of email, personal details, debit/credit card information and many more and these all data stored somewhere remotely.

II. AVAILABLE SECURITY TECHNIQUES

There are many system vulnerability are happening at the present level such as DoS (denial of service attack), computer viruses, hackers, abuse, fraud, and unauthorized access. Where big data environment allows SMEs to aggregate more and more data such as financial, personal, intellectual, and other sensitive data (IBM, 2012). There is a huge concern that the system which handles big data can be compromise to the security threats. Therefore, there corporate should take appropriate measures to avoid this to happen. Therefore, these corporate should ensure of creating a control environment which consists of methods, policies and procedures which measures protection of their assets. Further, also ensure the accuracy and reliability of records and operational adherence to management standards.

Now the need is to develop an awareness of the available techniques to secure and manage an information system in a corporate environment. Big data environment always allow SMEs and big corporate to aggregate more and more data such the data of financial, intellectual, personal and sensitive data related to company. Take the example of outbreak of epidemic where through big data now it is possible to predict such outbreaks by analyzing information on social media such as Twitter or Facebook where people can update or tweet their lifestyle such as “In bed with fever” or “got weird spots on my skin” which an allow health authorities to act faster to identify epidemics (European Commission, 2013). Therefore, big data requires new technology, increase their capacity, new tools and skills (European Commission, 2013). There are security standards such as ISO 17799:2005 and BS 7799-3:2006 that provides guidance on identifying, evaluating, treating and managing the information security and their risks (Lineman, 2013).

ISO 17799:2005 is the standards for security which is a model for creating information security and policies, assigning roles and responsibilities, documenting all the operational procedures and prepare for potential risk and business continuity management, and complying with legal requirements and audit controls (Myler, and Broadbent, 2006). This security framework will provide better management shield to protect big data from unauthorized access.

The ISO standards are something that corporate level security management can trust and recognize without any doubt. It is a best practice on information security management. The scope of the standard is to give the information security management recommendations for those people who are responsible for maintaining and implementing the securities (Doug, 2006).

The new British standard BS 7799-3:2006 provides range of guidance’s to improve the security of the big data at corporate level as listed below:

- Risk assessment
- Risk treatment
- Management Decision Making
- Risk re-assessment
- Monitoring and reviewing risk profile
- Information security risk in the context of corporate governance
- Compliance with other risk based standards and regulations.

(Self, 2007).

The need is to identify, evaluate and managing information security risk are the key risk evaluation for businesses if they want to keep their important data safe and secure (Biery, 2006). The main purpose of these standards is assess the risk involved in the data protection methods, and
then creation of clear management of security policies and then organization of information security.

III. IDENTIFY THE RISK AND MANAGE SECURITY

As discussed in above section that security is important for the corporate level businesses because it helps to align critical business function protection with security effort prioritization (Larry, 2006). Organizations can use these standards not only for protection purposes but also they can standardize their programs, establish distinct guidelines for compliances, and audit purposes (Doug, 2006). Further, these standards are formally guidelines for organizations that allow protecting the vulnerability and protecting from inside or outside threat as shown in figure II-1 above.

Performance and scalability are the most common technical issues in order to deal with the huge volume of data that stored by big data systems and technologies. There are risk and data integrity involved for many SMEs. On other hand such a big scale data there is need to deal with the issues that big data in order to enable cost effective, feasible, scalable storage, and working with huge type of data. The most important thing to make sure is the privacy of big data because without privacy there is no benefit of using big data in SMEs.

The big data provides huge benefits for SMEs and make impact on the social life as well because it creates jobs and growth in economy. There are huge risk and threats involved with such huge amount of data. Therefore, the threat and risk assessment is crucial in order to mitigate the future risk that can create negative impact on SMEs.

IV. THREAT AND RISK ASSESSMENT

This paper has considered ISO 17799:2005 which established 9 useful steps to assess the threats and risk assessment at corporate level. These steps help to calculate the risk that involved with big data usage in SMEs and how to make decisions to mitigate these threats.

A. Step1: Conduct Risk Assessments

At first step, there is very crucial for policy developer that they identify the risk. A risk is defined as the things which cause to exposure of possible destruction or loss (Doug, 2006). There are much risk that involved in at the corporate level these includes the potential hazard that might be internal or external as listed below:

- Natural disasters (e.g. weather, flooding, etc.)
- Human made disasters (e.g. power outage hackers, intruders, viruses, fire, etc.)
- Political (e.g. bomb threats, demonstrations, riots, terrorist attacks, etc.)

By identifying the risk at first step will help organization that which risk organizations feel are likely to occur in their environment. Further, regularly monitoring the threats will help to identify the nature of risk such as IMLogic article (2005) wrote that IM [Instant Messaging] worm was responsible for 90 per cent attacks in 2005. These kinds of attacks utilized social engineering techniques where end user clicks the suspicions links which was embedded inside IM messages. Where user found the malicious code that compromised the security of the user machine.

B. Step2: Establish a Security Policy

When risk is fully assessed the next stage is to establish the policy under the standard ISO 17799: 2005. The fully information security policy should be established and authorized by management and regularly communicated (ARMA International, 2006). This policy should be keep up-to-date regularly which is captured under clause 6.0 for organization information security. A proper responsibilities need to be established, who to given access agreements, responsibilities and problem resolutions measures. These responsibilities will help to gauge the level of threat and those who can handle at each level.

C. Step3: Compile an asset Inventory

This components of the standards address the control, and asset management and how to protect if there is breached found. There organization needs to identify the intellectual property and tools to understand what resources organization currently has and where they are located and who is responsible for them. By adopting this it is higher possibilities that the SMEs will handle better the resources.

D. Step4: Define Accountability

This component is described the human aspect of the security where it applies at the level of accountability that the employees or the organizational need to protect the assets. The roles and responsibilities of the employees are very important because if there is no accountable there is no security at all.

E. Step5: Address Physical Security

This component of the standard required to identify the physical assets such as equipment, environmental, cabling, furniture, etc. An organization building and its premises also need to be secure and nobody will allow accessing the building or business equipments. There are huge possibilities that Big data will not stored at SMEs premises they can use through third-party provider such as Google, IDC and many more. However, securing SMEs own premises can assure the data they use are risk free.

F. Step6: Document Operating Procedures

This component of the standard dealt with the operating procedures and management of audit trail and system log information. Formally preparing documentation for each program and activities is important because this will keep the track of all activities, implementation and development of the program.

G. Step7: Determine Access Controls

This component of the standard establishes the access control of the system administrator to control the access to and distribution of information. These policies should be applied on
all users, equipment, and network services that are directly and indirectly connected with organizations.

H. Step 8: Coordinate Business Continuity

This component of the standard highlights the need of the responding to the event if it occurs. There are chances that there might be security breaches or unexpected interruptions in the business activities. A business need identify the risk and check the impact that happened on the business. Then prioritize the critical business functions and develop the countermeasure to mitigate the impact of the occurrences (Larry, 2006).

I. Step 9: Demonstrate Compliance

This component of the standard provides the IP rights; complain aces measure and legal requirements. It is important for an organization that their business process should match the applicable laws, regulations and legislative requirements for all aspect of their business transactions.

V. CONCLUSION

This big data is mostly a new trend for SMEs and can have enormous benefit to the business process. Most of the time SMEs outsource their IT infrastructure to the third-party providers in order to reduce the upfront cost. It is found that interest of using big data from SMEs has reached to the new level as they need to capture the insight which helps to understand the consumer’s behavior patterns and showcase trends toward their product and services and many more. However, these SMEs need to keep a sharp eye on the cost and execution before establishing the data strategy. Further, security is the major issues relating how to use and handle the size, velocity and high variety of information in order to cost effective innovations. A proper security standards and framework are required to assess the risk, threats and ethical considerations to address these issues. There are many security standards and framework that helps to address such issues for SMEs and maintain the flow of information smoothly. This paper considered ISO 17799:2005 as important future prospect to address these issues.

VI. REFERENCES


23
Big Data for Marketing
Security Problems and Challenges

Abdulmajeed Bensbiet
School of Computing and Maths
University of Derby
Derby, United Kingdom
A.bensbiet1@unimail.derby.ac.uk

Abstract—The study presents a systematic study of the way big data may lead to security issues while being used for marketing purposes in small and medium enterprises (SMEs). While big data helps in getting better customer engagement or customer loyalty, it comes with issues of privacy, management and security. The security issues that arise at all stages of using big data thus need to be managed and eliminated in an effective way for ensuring compliance with security standards.

Index Terms—Big Data, Small and medium Enterprises, Data management, Security, Issues, Data Privacy, Data Management

I. INTRODUCTION

Luo et al. (2013) has said that the world today is a massive collection of data and most of the data that is existent today has come into existence only in the past couple of years. With the rapid change in technologies across the globe, it has become easier to gather data from different sources and in different forms. For example, data maybe collected straight away from the social media sites or from public records or even from satellites that record climate and other information of the earth. Amidst the myriad of information present for the people, there are some data that are huge and complex to process further and such data is called as Big Data. Big data usually refers to the collection of data or information that ranks high in volume, variety and velocity and is also generally complex and difficult to store, analyze and comprehend. These data sets require additional resource capabilities than what is used in normal data processing applications and hence needs the involvement of new and non-traditional techniques (Howe et al. 2008).

II. OVERVIEW OF BIG DATA

Big data is generally considered to be a combination of transactions, interactions and observations because the data is collected from via these activities. Transactions refer to any form of transaction taking place between two entities, which would include e-commerce transactions or any other customer interactions taking place offline. In addition to transactions, interactions refer to the interactions between two or more people such as the interactions taking place on social media, chat rooms etc. The last form of big data refers to observations, including the data collected by general observations of researchers, GPS coordinate findings or demographics observations (Bugnin et al. 2010).

Columbus (2012) state that the searches for big data and content on the internet is continuously increasing and one of the most important parts of big data is social content, which accounts for more than 20% of the total data. There are web tools that capture the behavior of users on a per second basis and hence it may be said that the velocity of big data is also very high (Zimmer, 2010).

A. Types of Big Data

- **Web and Social Data:** One of the most commonly used and obtained types of big data is web and social data. Web data refers to the data collected from the World Wide Web (WWW), i.e the internet being used by people all over the world and the social data refers to the information gathered from social media websites, including Facebook, Linkedin, Orkut etc (Heath and Bizer, 2011). This type of big data is generated and stored every second and because of the existence of large number of users of internet as well as social media, the volume of data keeps on increasing. Such big data may include the demographics of users, psychological profiles of people, likes and dislikes of people, digital images and videos and also the things that people search for online (Cuzzocrea et al. 2011).

- **Machine-generated data:** Machine generated data refers to the different data and information generated by machines, such as the utility data related by machines, the data generated from ATM machines etc. Even the data generated from the machines in hospitals, banks etc all form a part of big data and help in gaining more information and insights into the choices and preferences of users and consumers (Witten & Frank, 2005).

- **Human-Generated Data:** Human generated data is also a part of big data and includes all data and information that are manually gathered and collected
by people. IT departments in companies and call centres associated with different organizations usually provide large volumes of big data based on their observations and recording of responses from the consumers. People also maintain logs regarding different things, which may be used for analysis and interpretation purposes (Witten & Frank, 2005).

**Transaction data**: Another aspect of big data is transaction data, which is gathered from the transactions taking place between the customers and organizations. Transaction data may collect offline or online and helps in providing information and insights on the different attributes and services of customers, which help in analyzing the social behaviour, income levels and spending patterns of consumers (Spiegler, 2003).

Thus, different types of big data help in gathering information and data related to different aspects of general and organizational lives of people and thus is extremely useful for analytical and decision making purposes.

**B. Overload of Big Data and Security issues**

It is found that the amount of data being collected by organizations, irrespective of their domain and size is continuously increasing every year and is expected to keep on increasing in the coming years. The volume of big data is not only increasing but is increasing in a way that it is becoming problematic for people to manage the same. The overload is even causing a deficiency of storage space, as the increase in storage space available for safeguarding the data is comparatively less than the increase in the data volume (Li et al. 2003).

Though the use of big data is multifold and it may be used for numerous applications and analyses, overload of anything is not good. There are several risks involved with the big data overload, which include the need of better resources, need for higher manpower, higher costs, lack of skills and knowledge to tackle data, increase in errors and also increase in security issues. Security issues with big data include the security of personal information of people, misuse of data collected or unwanted access to data. When there is overload of data, it becomes difficult to not only store it but also protect it. The chances of its misuse increase significantly and hence big data overload are causing serious security issues, which are expected to increase further (Schroeder et al. 2006).

**III. RELATIONSHIP BETWEEN BIG DATA AND SMALL AND MEDIUM ENTERPRISES**

One interesting development that has been observed in the use of big data is its incorporation and application by the small and medium enterprises (SMEs). While multinational corporations make decisions related to their strategies on the basis of big data analysis, SMEs can’t do that because of budget and resource constraints. However, there are multiple other ways in which big data may assist SMEs in improving their business efficiency:

- **Customize their offerings**: Data on the current trends and choices of consumers may help SMEs in customizing their product and service offerings (McAfee & Brynjolfsson, 2012).
- **Building new applications**: SMEs may also look at the development of new applications and may even alter their existing products to suit the needs of consumers (McAfee & Brynjolfsson, 2012).
- **Lower cost**: There are large numbers of cost-effective tools that maybe used for analyzing big data and those maybe used by SMEs for identifying areas where costs may reduce by altering operations or other strategies (Davenport et al. 2012).
- **Better Pricing Strategies**: Use of big data may help SMEs in finding better suppliers, which may reduce their costs and they may offer better transparency in their pricing decisions (Davenport et al. 2012).
- **Competitive Advantage**: Because of the large number of competitors, SMEs cannot keep a track of all of them but may gain competitive advantage by being ahead of them in devising better product and marketing strategies (Davenport et al. 2012).

**IV. USE OF BIG DATA FOR SME MARKETING ACTIVITIES**

SMEs may make use of big data for marketing activities to a large extent, as it helps in gaining better insight into the customers and the market conditions and situations (Ratner, 2004). Lohse et al. (2000) have explained the significance of big data for understanding the buying behavior of the consumers by tracking their internet usage analysis. Hart and Saunders (1997) explain the importance of big data for increasing customer engagement. Customer engagement refers to the formation of a close relationship with customer to understand their needs (Carland et al. 1989). Hart and Saunders (1997) explain that the study of big data helps in understanding customers, which then facilitates better electronic exchange between the SMEs and the consumers. Blattberg and Deighton (1991) highlight the application of big data for better communications with the customers. The authors explain that in addition to the engagement, use of big data facilitates communication for informing consumers about the features and benefits of the products and services offered by the company in a way such that they appreciate the same. In addition to the customer aspects, marketing activities also become more efficient and effective with the use of analytical findings and measurements obtained via big data (Gold, 2012).

**A. Challenges of Using Big Data for SME Marketing**

Though big data has lot of potential for improving the marketing activities of small and medium enterprises, there are also various challenges associated with the same. Some of the major challenges associated with the use of big data for SME marketing are discussed here.

- **Data Privacy**: Maintaining the privacy of data collected from the customers is a huge challenge for the SMEs. Especially considering the fact that SMEs have limited resources and money to spend on
techniques and software for protecting the privacy of consumer data, this becomes an even bigger challenge. Managing access to the data and securing the privacy of encrypted data are two most crucial data privacy concerns (Yao-Huai, 2005).

- **Data Management:** Keeping the large volumes of big data, it may be said that the management of data itself is a big matter of concern for the SMEs. They need to devise and identify cheaper and efficient methods of securing the collected data (Madden, 2012).

- **Security Issues:** Security Issues while using big data for marketing occur at two different stages, at the infrastructure stage and at integration stage. Infrastructure security challenges or problems usually arise when the infrastructure or framework being used for computing is not entirely safe or when different non-relational data sources are used for analysis purposes. Security issues also arise during integration either at the time of monitoring data or at the last stage of validation. Detailed analysis of the possible security issues is thus presented below (Madden, 2012).

### B. Security Issues in Using Big Data for SME Marketing

- **Securing Computations** - The SMEs make use of different programming frameworks for data computations. Presence of virus or any third party malware program in the framework may create security issues during computations and hence ensuring that all the computations are carried out in a secured environment is extremely crucial (Subashini & Kavitha, 2011).

- **Non-Relational Data Sources** – Non-relational data sources are the databases that do not make use of SQL (Structured Query Language) for the management of data and the data structure is different from traditional RDBMS. The non-relational data sources currently used do not have strong security measures and hence SMEs also need to be careful of the security issues while making use of these data sources (Subashini & Kavitha, 2011).

- **Data Storage and Transaction logs:** While computation and data source security are important, it is also important to maintain security while storing the collected data. Unauthorized access to these storages and the logs maintained for recording customer transactions may create significant security issues for the organizations (Perrig et al. 2004).

- **Endpoint Validation** - Endpoint validation refers to the validation that is required each node or endpoint of computing. Data while being transferred from one source to another passes through different nodes and hence it’s important to maintain security at each of the endpoints, i.e. the terminal or the PC (Perrig et al. 2004).

- **Real Time Monitoring** - Real-time monitoring is usually used as a web application for monitoring the web activities of users. At times, web browsers are infected with malwares that may transmit personal information of users during real-time monitoring to third parties, which need to be avoided to protect the security of the system (Perrig et al. 2004).

- **Access Control** - SMEs usually consist of a few people working together and is highly labor intensive and hence the chances of incorrect use of access to data storage and mining systems are comparatively less. The SMEs hence need to ensure that only a few selected and trusted people have access and control over the data because alterations in data may not only lead to security breaches but may even lead to the development of ineffective marketing strategies (Furst et al. 2002).

- **Granular Security Issues** - One thing to note while addressing the security challenges for big data is that security challenges not only arise at the bigger application level but may also occur at small granular levels in different categories and subcategories. Hence, it is important to secure data at all levels possible (Denning & Denning, 1979).

- **Securing Data Provenance** - The last and the most important security challenge is to ensure security of data at its origin itself. For example, if the SMEs are collecting data from consumers via mobile phones, then securing the connection and network used for gathering data also becomes extremely important (Denning & Denning, 1979). Thus, it is observed that while the marketing team of SMEs collects and uses big data for different purposes, it is crucial to ensure security of the data at each and every point and step. Security issues must be challenged to protect privacy of consumers and also for ensuring that no wrong data is being used by the organizations.

### C. Approaches for Resolving Security Issues

The security issues that are faced while using big data for SME marketing purposes may be easily resolved by implementing several security measures at all stages of the use of big data. Different stages involved in the implementation and use of security measures while using big data for SME marketing are explained.

1) **Design stage**

This stage refers to the designing of measures that maybe used for increasing security while using big data. The SMEs must always study the design of security measures being used across the globe and then select the one that best suits its big data management and analysis system. The design stage is most crucial because correct design is crucial to facilitate its implementation and outcomes. Few things that
need to be kept in mind while designing the security system are the types of data being used and the impacts may have on user experiences (Chen & Zhao, 2012).

2) Implementation Stage

The security requirements must be clearly identified during implementation so that the security measures are adopted and implemented at all the places that require security. Missing out on even one of the security requirements may cause problems. During implementation of the security measures, one thing that needs to be kept in mind is that the control and responsibilities of accessing big data should be totally minimized because lesser the check points, better it is. Use of more of encryption strategy for data may also be extremely helpful and beneficial for ensuring that unwanted access is not allowed (Chen & Zhao, 2012).

3) Monitoring Stage

Once the security measures are designed and implemented, it is also important to monitor them regularly. Regular monitoring of all the access points and gateways must be carried out to make sure that no undesirable activity is taking place. There are several practices defined for monitoring the security measures and a checklist must be developed and implemented to ensure that all security measures are in place (Sit and Morris, 2002).

V. CONCLUSION

As discussed, the use of big data for marketing purposes in SMEs may be extremely beneficial but it comes with a bunch of privacy and security concerns. It is shown that the security concerns may arise at different stages and during the use as well as analysis of data. As the security issues may occur at any time, it is important to design and implement adequate measures to ensure security at all levels to use big data effectively without any problems.

VI. REFERENCES


Madden, S.(2012), From Databases to Big Data, IEEE Internet Computing, vol.16, no.3, pp.4-6


Ratner, B.(2005), Statistical Modeling and Analysis for Database Marketing, Taylor & Francis e-Library

Schroeder, B. and Harchol-Balter, M.(2006), Web servers under overload: How scheduling may help, ACM Transactions on Internet Technology (TOIT), vol.6, no.1, pp.20-52


Zimmer, M.(2010), “But the data is already public”: on the ethics of research in Facebook, Ethics and Information Technology, vol.12, no.4, pp.313-325
The New European Data Protection Regulation
Compliance How It Will affect SMEs regarding Big Data

Michael Cain
University of Derby,
Derbyshire, UK,
100224312@unimail.derby.ac.uk

Abstract—The value that big data can bring to small and medium enterprises (SMEs) is only beginning to be discovered as smaller enterprises start to adopt business practices followed by the larger corporations. At the same time there is a major change in development with regards to data, as a new regulation is passing through the European Union (EU) and will change the way enterprises approach data in the future. This paper addresses the new changes that enterprises may have to incorporate when the regulation comes into law. It will analyse the proposed changes and what effects they will have on the future in order to aid enterprises in preparation. This paper looks at how SMEs looking to utilise big data will have to take into considerations the future regulation and the added cost this may entail.

Keywords—SMEs, Big Data, Governance, Compliance, EU, Protection, Privacy, Audit.

I. INTRODUCTION

Increasingly big data is being utilised by more enterprises, not surprising as its impact have caused it to be compared to natural resources such as gold and oil, in potential value it could bring to an enterprise (Akerkar, 2013). Although there are still large numbers who have yet to adopt big data and reap the benefits. Public concerns for privacy are at an all-time high (see Fig. 1) due to data breaches, media exposure and with each new story concerns increase (Davis, 2012). The European Union (EU) addressed these concerns in 2012 when they proposed a reform of the EU’s existing Data Protection Directive implemented in 1998 (European Commission, 2012). The new regulation when proposed was said to aid enterprises by reducing administrative costs by €2.3 billion a year, and reducing the burden on processing data by making it easier. The proposed regulation in theory was designed to make it easier for businesses to collect and process data in other countries such as the United States (US). The change from a directive to a regulation means that it will come into law within two years of being adopted. This gives enterprises that currently collect and process data or are thinking of utilising big data two years from adoption to prepare. This article will take the following format: Section II discusses the changes that the new regulation brings that will affect Small and Medium enterprises (SMEs). Section III discusses how SMEs including those who utilise big data are affected relative to each key change in the regulation, followed by the main conclusions.

II. CHANGES THAT WILL AFFECT SMEs

The proposed regulation seeks to improve upon the old directive by addressing current issues of the information age (European Commission, 2012). The new regulation therefore has key changes to address modern technology, data collection and processing. The range of data that is addressed under the regulation has increased such that it applies to any organisations that handles data that originates from an EU resident. If the enterprise resides within an EU member state then they will have to report to a Data Protection Authority (DPA), depending on the headquarters location of the enterprise, or they may have a choice. Changes to notifications require that individuals receive more information regarding their data and give more in terms of consent. The mandatory requirement of Data Protection Impact Assessments is to be carried out if certain risks may occur. The appointment of a Data Protection Officer (DPO) will be needed if the enterprise meets certain requirements. Data breaches have to be reported to the respective DPA within a time limit or face new penalties. The ‘right to be forgotten’ and the ‘right to data portability’ empowers the individual allowing access and erasure of their data on request (European Commission, 2013). Exceptions from rules do occur such that employee information could still be regulated by an individual country of which the employee resides, this could cause some complications. These changes will affect SMEs as they will have to comply with the new legislation and this will incur time and cost, this will be discussed later in the paper.

III. HOW EACH CHANGE AFFECTS SMEs

The EU justice commission’s vice president Viviane Redding (2012) stated that the legislation would benefit businesses in a number of ways. This section shall explore the changes and how SMEs can and will be affected.

A. Range

Article 3 of the regulation covers the factors determining its scope and as stated has increased. The regulation applies to any enterprise that collects, or processes data within an EU member

Figure 1. EU public survey on whether they are concerned about how their data is being used by businesses (European Union, 2011)
state. Any EU resident’s personal data is also protected by this legislation and therefore any enterprise handling this data, regardless of country they are situated in must comply with the legislation. Therefore an enterprise not in an EU member state unaware of their customer location may have to comply with the legislation. Non-EU enterprises will therefore have to start to collect location data, in order to know whether they should begin preparations for the legislation (Ernest & Young, 2013). This brings another set of requirements it must fulfil as they collect more data of a different type. This could cause some businesses to rethink the customers they serve as if they only have a small EU customer base it might be more cost effective to not deal with EU data. This is especially relevant for SMEs seeking to expand their customer base overseas therefore negating growth and limiting the SME until they have the required finance to instigate expansion. The Confederation of British Industry’s (2012) echoed this stating “It will deprive European consumers of new innovative and highly beneficial markets”. SMEs that offer contracted services have a legal obligation to their customers, so could face legal action (Citizens Advice Bureau, 2012), or have no choice but to undergo changes to retain customers and protect their enterprises reputation incurring further costs.

B. Responsibility

Article 35 of the regulation states that a Data Protection Officer is required for enterprises of a certain size but later went onto state that SMEs were exempt. The issue that occurs is that the definition of SMEs differ. The EU defines an SME as “an enterprise with less than 250 employees and no more than €50 million in annual turnover” (The European Commission, 2005) whereas Germany define an SME as “A company that besides generating a maximum turnover of €50 million, also has a maximum of 500 employees” (Deutsche Banke, 2008). As can be seen under the EU definition a German SME would not be exempt from having a data officer. The deciding party should be outlined in the regulations to lessen confusion. The regulation states that a DPO is required to be employed for two years after hiring. The current yearly salary of a DPO is between £35,000 and £70,000 as more companies are required to hire DPO’s salaries will increase with demand (Shah, 2013). In order to stop the loss of employees from big data specialists leaving to SMEs for higher pay, big data businesses will increase employee pay which will affect cost of services. Through association the costs will be made up by an increase in charge for services and therefore cost SMEs utilising big data.

C. The Right to be forgotten

Article 17 of the regulation gives the details of the ‘right to be forgotten’ which allows individuals the right to request that their data be erased from any business, where they no longer have a legitimate reason for keeping it (European Commission, 2012). The effect that this has on SMEs wishing to start collecting data in order to better prepare for using big data can be cumbersome. The customers of the business can request that their data be erased as is their right. The potential that the number of customers do this with “unrealistic expectations” has been predicted to be high by the UK Ministry of Justice (2013). The enterprise may have to deny the customer where data retention is legally required, this could upset the customer and lead to lower customer satisfaction and possible loss of a customer.

D. The Right to Data Portability

Article 15 of the regulation details the ‘right to data portability’ which gives individuals the right to request their data of which an SME must comply. An SME must provide the individual with the data in an electronic structure that is common and allows the individual to use (European Commission, 2013). SMEs that utilise big data in order to track trends in consumers will no longer benefit from the competitive edge this gives them. The right to data portability would allow a competing SME customer to become a customer of another SME and request the data that they have on them. This would be some form of business espionage, as the competing SME now can see what data it utilised by other enterprises and start to collect and process it themselves. This could cost an innovative SME huge loss in profits as other companies mimic its data processing. Similarly withholding information to make it difficult to swap to another SME can actively prevent competition as has been raised by EU Vice President Joaquín Almunia (2012).

E. Breach Notifications

With the handling of big data comes the biggest risk that has been well documented in the media, a data breach. A data breach is one of the worst things that can happen to a SME that isn’t a fire. The repercussions that a breach of security has can be costly and long lasting (Sullivan, 2014). Due to this growing concern the regulation addresses data breaches in article 31 in which it states that any enterprise must on discovery of a breach, inform the supervisor authority within 24 hours and when not within 24 hours have a justification. The next notification must be made to the individuals whose data was involved where it affects the protection of their personal information (European Commission, 2013). This can affects SMEs as the current time constraints emphasise on speed of notification, or face fines. This can adversely affect the investigation process and lead to incomplete information being presented. Incomplete information can leave customers unsure of the SMEs security, to a loss of reputation, a loss in future profit and potential future customers (Campbell, et al., 2003).

F. Policies & Consent

SMEs that have already started to utilise big data from their customers will have to update their current methods due to the new legislation. The way in which they obtain consent from individuals has changed, so that the individual now has to give ‘explicit’ consent as defined in article 4. Explicit consent requires the individual to consciously perform an action such as ticking a box, this means no longer will no action be assumed to be consent (Tene & Wolf, 2013). The SME that collects information will also be responsible for keeping proof that they obtained consent. This will incur costs for storage of proof and for where companies have not kept proof re-
obtaining consent. Article 11 covers transparent information and communication, this means that individuals must have clear and easy to understand polices. If any changes are made to the original privacy policy concerning an individual’s data then those whose data it involves should be notified. Considering the new legislation brings in many changes, many SMEs will have to change their policies in order to comply (Rossi, 2014). This can be costly to contact all those involved especially for SMEs utilising big data due to the volume of individuals involved. It has been estimated that if a SME had to contact an individual by letter it would cost around £15 per customer, when factoring in employee time, and postage etc. This would incur large costs when there are hundreds of individuals involved. Additional costs would also be incurred for having to deal with customer queries related to the change.

G. Penalties

The regulation gives power to DPAs to allow them to enforce the legislation and ensure compliance. Article 79 states that any enterprise that does not comply will receive a penalty (European Commission, 2013), of which two can have big impacts on SMEs. The first of which is regular audits which would cost a SME each time they were carried out. The biggest penalty is the fine of €100 million or 5% annual worldwide turnover, determined by which is larger. This can have a cost of billions to any SME but there are some ways to avoid these hefty fines. If an SME has a ‘European data protection seal’ they can only be fined if the breach of the legislation was intentional or due to inadvertent compliance (European Commission, 2013). The way in which a seal can be obtained is still unknown and could be expensive as if it can save an SME €50 Million. It would be no surprise if the cost was at least in the thousands or millions itself, as companies take advantage in the demand for providing the services required to obtain one.

IV. PREPARATIONS

The Direct Marketing Association has found that 1 in 3 enterprises are failing to prepare for the changes that will be brought by the new legislation (Direct Market Association, 2014) and that 50% of senior executives were not aware of the impacts it could have (Direct Marketing Association, 2013). The following sections will include preparations that SMEs can begin to think about in order to be able to comply with the legislation, especially those that utilise big data (Ashford, 2014).

A. Comply

The most important step in preparation that SMEs can take is compliance with the Data Protection Directive 95/46/EC. The new regulations foundation are built upon the directive and seek to improve it. This is why it is logical to ensure SME complies fully with it in order to make the changes required by the regulation easier, as has been echoed by the deputy information commissioner David Smith (2014).

B. Assess

Carry out an assessment within your enterprise using material already provided by the Information Commissioner’s Office that give guidelines on privacy assessments (Information Commissioners Office, 2013). Find out what, why and how your enterprise collects data. Make sure you know where the data originates and how it is stored, and secured. Think about how you would effectively erase all an individuals’ data and who would be in charge of this. If you don’t have a DPO already think about training someone in house to reduce cost. Assessments like this will become common when the regulation is passed so to carry one out early will aid in adoption of the regulation.

![Figure 2. Survey results on how to deal with misuse or inadequate protection of data (European Union, 2011)](image.png)

C. Plan

The main reason that fines are imposed in the regulation are due to the increase in data breaches and the response of the public (see Fig. 2). Your enterprise should effectively plan for what would happen if a data breach occurred (Photopoulos, 2011) and how you would mitigate the impact. Do you have a dedicated team who would handle this? Or would you have a contractor? Would your enterprise currently be able to gather the required information within 24 hours? By answering these question now and preparing you could avoid a fine of 5% of your annual turnover (European Commission, 2013).

D. Budget

The legislation may not come into effect until 2016, this gives SMEs time to budget effectively (Digital Strategy Consulting, 2014). Although not in effect now surprise costs in the future could be big. Money saved from previous years budgeting will therefore be essential. Budget to include costs for an increase in staff, new software, new hardware and security costs (McLaren, 2014).

V. CONCLUSION

The proposed legislation was supposed to save enterprises €2.3 billion a year. What can be determined is that questions should be asked to how these figures have been calculated. The current regulation may unify the EU enterprises under common law, but at what cost for compliance? As the changes stated seem to incur large amounts of cost which still remains unclear and could negate the supposed savings. The new regulation may benefit the consumer but it is clear that this is at cost to the enterprise. SMEs will have to adapt to the new regulation at a price and this will affect the growth of many SMEs (Ministry
of Justice, (2012) as they budget and priorities for change. SMEs preparing to enter the big data market in order to gain a competitive edge will be disappointed when they find they are at a disadvantage, compared to early adopters due to the new increase in cost associated with delving into big data. SMEs wishing to reduce cost should start preparation early to adopt practices that are pretty much certain to be in the approved regulation such as breach notifications. SMEs updating systems should do so with the new regulation in mind and the potential requirements it may have. SMEs should be careful when choosing to wait as although the regulation hasn’t been fully approved it is clear that it will come into law within time and getting caught with surprise costs will negatively impact business.

VI. REFERENCES


Abstract—the security of big data sources is a highly important topic amongst people in the world of information security, as one small breach in a big data source could result in millions of users’ data being shared with the outside world. Although big data sources may be secured and monitored twenty four hours a day, what happens when the data collected from this monitor becomes a big data source in itself? This paper addresses just that.

Index Terms—Big data, SIEM, Monitoring system, Monitoring, Storage.

I. INTRODUCTION

Big data is a term that is used on a regular basis, whether in the workplace, on the news, or in general conversation. Although a commonly used term, this does not guarantee that the term is being used in the correct context, or to actually describe a genuine big data source. To the vast majority of people, a multi-terabyte data source is considered to be a large amount of data. To some this may be a representation of big data, due to the fact that, as of 2011, the average amount of data stored per household was 464 gigabytes (Gartner, 2012) across all devices, including smartphones, tablets and so on. This average is predicted to grow to 3.3 terabytes by 2016 (Gartner, 2012). If this is to happen, a terabyte of data is likely to be seen as “the norm”, so it is far less likely that people will confuse this with a big data source.

II. WHAT IS BIG DATA, REALLY?

According to IBM, “every day, we create 2.5 quintillion bytes of data – so much that 90% of the data in the world today has been created in the last two years alone.” (IBM, 2013). To put this into perspective, 2.5 quintillion bytes of data is equal to roughly 2220 petabytes (1 petabyte is equal to 1024 terabytes). “That’s roughly equivalent to over half a billion HD movie downloads” (PC Advisor, 2012). This vast amount of data does not just come from people creating documents and files, this data comes from everywhere – “sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few.” (IBM 2013). This is an example of big data.

The definition of a big data source is one that is not clearly defined or globally agreed upon, due to the fact that big data is an ever-evolving area that is much more complex than it first appears. Some may say that one method of defining a big data source would be to set a data storage cap that, once reached by any data source, would define it as a big data source. Although this could be a possible method to define big data, it would only be viable for a small period of time. This is due to the fact that, as stated previously, the amount of data generated and stored in the average household is predicted to grow from “464 gigabytes in 2011 to 3.3 terabytes in 2016” (Gartner, 2012). Owing to this growth of more than seven times the average of 2011, the cap set to define big data sources may rapidly become a commonly seen storage limit. For example, the first 1 terabyte hard disk drive was released in 2007, by Hitachi (Perenson, 2007). Over the following few years, hard drives with more storage capacity were released, increasing the maximum storage capability of a single drive from 1 terabyte. By the end of 2011, Seagate had announced the release of the first 4 terabyte hard drive (Hachman, 2011). In February of 2014, LaCie announced the release of the first 5 terabyte hard drive to be released onto the market (LaCie, 2014). Owing to the speed with which the storage limits of widely available storage media is advancing, it is likely that, sometime in the future, the amount of data stored by small to medium businesses today will be seen in households across the world. With this rapid increase in widely available storage, it would not be reasonable to set a storage cap to define a big data source, as this cap would require adjusting on a relatively frequent basis, so as to remain up to date with currently available storage media, and the amounts of data stored and utilised by both businesses and in households.

Currently, the most common way of defining a big data source, is to identify whether or not it can be described using the “V’s of Big Data”. The V’s of Big Data are dimensions into which big data sources can be broken down (IBM, Unknown). Although the V’s of Big Data are a reasonable method of defining a big data source, they are also an area that is widely dependent on the point of view of individuals or individual businesses. Although many business have their own views of what “V’s” there should or shouldn’t be, there are
three that are widely acknowledged as being the core V’s of big data. These are Velocity, Volume and Variety.

Velocity is used to define the speed at which data is processed in big data sources. Due to the nature of big data sources, data is constantly being added, modified, deleted and accessed, meaning it is never “stationary”. For example, “the New York Stock Exchange captures 1TB of trade information during each trading session” (IBM, Unknown).

Volume defines the sheer amount of data generated and stored in big data sources. For example, the Large Synoptic Survey Telescope Project generates “over 30 thousand gigabytes (30TB) of images” (LSST, Unknown) every night.

Variety defines the varying different types of data stored in big data sources. This data can be stored in a myriad of different data types, such as images, video, audio, plaintext, database, etc.

III. BIG DATA SECURITY – WHY IS IT IMPORTANT?

With such vast amounts of data comes great responsibility on the data source owner’s part. If an unscrupulous individual were able to gain access to a big data source, especially one based around a company’s user-base and their personal details, the opportunities for illegal actions such as fraud would be enormous. For example, if a cybercriminal were able to gain access to Facebook’s big data sources, they would have unrestricted access to the entirety of the 1.11 Billion (Statistic Brain, 2014) individual Facebook user’s details. These details may include user inserted information, such as full name, date of birth, email address, phone number, etc. but it is likely to also include data that Facebook automatically gathers about each user. This automatically gathered data is “usage data collected automatically as the user spends time on the website clicking around” (MySecureCyberspace, Unknown). Although the latter may not seem overly concerning, it is worth bearing in mind that Facebook is not only used to look at the profiles of friends or colleagues – a lot of people “Like” pages for things they are interested in, companies they have dealings with and so on. From this data, it is possible that an unscrupulous individual may be able to build a good character profile of a certain individual. For example, if a user “Likes” the Lloyds Bank Facebook page (Appendices 1) and is also seen to be visiting the page on a frequent basis, it is safe to assume that they have a bank account with Lloyds Bank. Combining this with all the personal details willingly provided by the user, such as date of birth and email address, it may be possible for an experienced criminal to deceive an employee of the bank and make them believe that the caller is authorized to access and change details about the bank account in question. This is but one example of the consequences of a breach to a big data source.

IV. BIG DATA SECURITY – HOW COMPANIES SECURE BIG DATA

The method by which a number of companies go about securing their big data sources, is to utilise security information and event management (SIEM) systems. These systems provide a number of security features, such as “rapid incident response, seamless log management and extensible compliance reporting”, as well as things such as consolidating, correlating, assessing and prioritising security events (McAfee, Unknown).

Although the majority of these features are required to monitor and protect big data sources, there is far more to implementing these features than just simply installing the SIEM and configuring the settings. Other things that need to be taken into account are; the processing capability of the server onto which the SIEM is going to run, the configuration of firewalls and antivirus systems to allow all data and logs from the network to be sent via the SIEM, and many more. One of the most important requirements that may be overlooked is the storage available for the logs passing through and generated by the SIEM in use.

V. MONITORING OF BIG DATA – STORAGE OF THE LOGS

The continuous monitoring of a big data source, or even an everyday network, has a large number of advantages, such as the ability for the monitoring system in place to send an automated alert to a predefined contact within minutes of a potential network breach, or an infected system being detected. Monitoring systems can also be configured to send off alerts when services or servers become unavailable, whether due to malicious intent or simply due to a hardware/software malfunction. These examples are but a small portion of the capabilities of a variety of monitoring tools and security information and event management systems. What the majority of people do not think about, is that this continuous monitoring generates large amounts of data. Although when using a number of generic “off the shelf” security information and event management systems this will not necessarily be an abnormally large amount of data, due to the fact that a number of SIEMs do not offer automatic indexing and searching of the data (meaning users have to manually search through logs to find anything of interest), these systems still store a large number of logs per day, whether they be logs generated by network authentication issues, or due to something more sinister like a distributed denial of service (DDoS) attack. Although at first these amounts of data probably won’t be too much to worry about, especially in small to medium sized companies, after year or two of constant monitoring, the amount of data stored is swiftly going to add up.

Although this doesn’t sound too bad on paper, SIEMs do have their limitations. For example, Barclays tried deploying an SIEM for the monitoring of their network. Although it was an ideal solution for a while, allowing all data from across their multitude of systems to be gathered in one location, to then be shared out to a security operations team for analysis and response in real time, the restrictive capabilities of the SIEM soon became apparent. When it came to adding newer technologies to their network, such as domain controllers and proxy servers, the data being collected became less and less useful. The data ended up being stored away and never utilised, owing to the fact that it contained relatively meaningless data (Gailey, 2013).

In small businesses with a relatively low number of computer systems and servers the amount of data generated is
going to be relatively small at first, but after a number of years, or maybe even months, this data is going to add up. If the systems to be monitored are part of a big data source, the amount of data generated by the monitoring system is going to be vast. At first, the amount of data generated is not going to be much, especially in comparison to the big data source itself, assuming the big data source is continually growing, and that logs are not deleted due to compliance reasons, after years of continuous monitoring, these monitoring logs have the possibility of becoming a big data source in their own right. Even if this data is not utilised on a daily basis by any users, as the majority of intended big data sources would be, there are a number of factors that will identify this as a big data source. The key factors that will define this as a big data source, will be the big data V’s discussed previously. Owing to the nature of consistent data monitoring, the data being processed and stored by SIEMs will be extremely fast moving, too fast to be monitoring by the naked eye in real-time, which meets the “Velocity” V. As already discussed, the amount of data stored and the varying types of data recorded also contribute to these monitoring sources adhering to the big data V’s.

As an alternative to generic security information and event management systems, there are tools such as Splunk Enterprise. Splunk Enterprise is a software tool that monitors and analyses machine data, such as network activity and call records, and turns this data into “valuable insights”. (Splunk, Unknown). Splunk Enterprise is marketed to be able to “collect and index any machine-generated data from virtually any source or location in real time.” (Splunk, Unknown) Going by this advertisement of its capabilities to both monitor devices in real time, and to translate the monitored data into usable information, Splunk Enterprise would appear to be the ideal monitoring application for use with a big data source. This is backed up by the former head of security services for Barclays, who said that their original SIEM “had ceased to be able to cope at about 500 million events per day” (Gailey, 2013). As a result of this, “Gaily implemented software from Splunk – a decision that proved so successful that, two months ago, he left Barclays and went to work as a product evangelist for the supplier” (Glick, 2013).

Although Splunk may sound like the ideal application, the fact that it automatically interprets the data it collects is also its potential downfall. Owing to this added functionality, Splunk requires more data to run effectively. To allow for this automatic interpretation and its improved search facilities, Splunk indexes any data passed through it. The files created during indexing are; the raw data in compressed form, and index files that identify where the raw data is stored, plus some metadata files (Splunk, Unknown). Although this method of indexing makes sense, and will massively speed up the reporting/search processes that can be run through the application, the creation of the compressed raw data means that data is potentially being duplicated. Operating systems such as Windows generate and keep their own event log files (event log storage directory for Windows Vista/7/Server 2008: %SystemRoot%\system32 \winevt\logs) (ServerFault, 2010). If these event logs are then duplicated in Splunk, this essentially means that the logs are taking up double the space they normally would if they were just confined to the Windows system that experienced an issue. As mentioned previously, this may not be a disaster in relatively new companies, but as time passes, this duplication of data is going to become more and more noticeable. This duplication is going to be far more noticeable in established small to medium enterprises, as they are likely to have a lot of network traffic passing through their systems, resulting in more data being logged and potentially duplicated by Splunk Enterprise.

Although some monitoring systems come with the ability to delete data that is classified as “old data”, this threshold is not normally reached for a number of years, despite the fact that some data may never be utilised after it has been logged. With companies such as Barclays generating more than 500 million events per day (Gailey, 2013), having logs deleted after a number of years is simply not enough.

Also, owing to the fact that a large number of companies, mainly within the financial sector, are required by compliance rules to retain their network logs for a long period of time, the storage media for these companies is likely to run out at a rapid pace, especially if the company is responsible for a big data source.

VI. WHEN DO MONITORING SOLUTIONS BEGIN TO BECOME OVERWHELMED?

As mentioned previously, the previous security information and event management system used by Barclays became overwhelmed at around 500 million events per day. By 2013, “the bank generated around 44 billion security events per month – a figure that was set to reach 65 billion by the end of the year.” (Gailey, 2013). This is equivalent to nearly 1.5 billion events per day, three times the number of events that could be handled by the previous SIEM. The above figures were given around the middle of 2013, which would mean that, provided the number of security events did reach 65 billion by the end of the year, this is almost a fifty percent increase in a six month period, and this rapid increase is not likely to subside any time soon. With the constant technological advances of the current age, more and more people are leaning towards the side of unscrupulous activities, such as attempting to gain unauthorised access to company servers and big data sources to defraud innocent members of the public of their wealth.

VII. CONCLUSION

Although big data sources can provide massive advantages to the companies utilising them, they also pose big risks. As discussed in this report, a large amount of security monitoring features need to be put in place to ensure the security of the stored data, but these in turn also generate large amounts of data that may, one day, end up evolving into a big data source. If this were to happen, the company that owns or is responsible for the original big data source, will ultimately be responsible for this new data. There are a number of possible solutions for this issue, each of which will have its own merits and its own flaws.
One of these possible solutions is that the storage of logs generated by monitoring tools is examined on a frequent basis, during which time older or irrelevant data is cleared from the storage media, so as to free up storage for new logs. The key merit of this solution is that it will allow companies to continue to use their existing storage devices, without having to purchase more each time the drives become full. The key disadvantage to this is the sheer amount of time it may take to sort through old logs and to decide which can be deleted and which must be kept. This solution may also not be possible in companies that provide financial services, where there may be compliance rules preventing the deletion of network logs.

Another possible solution would be to continuously swap out storage media one it becomes full. The main advantage of this solution is that no data will be deleted, resulting in companies being able to go back and examine their logs several years down the line, should the need ever arise. The key disadvantage of this is simply down to cost. For example, if a terabyte hard drive is filled up after one week of use, this will have to be replaced with another terabyte drive, which is likely to cost upwards of £50. If this is done on a weekly basis, to at least one hard drive, that is a cost of over £2600 at the end of a single year.

Finally, owners of a big data source could outsource the storage and monitoring of the data to a third-party company. An advantage of this solution is that the owner of the data would no longer have to worry about the storage of the monitoring logs or any of the data. The major downside of this though, would again be the cost. For large enterprises, this may not be an issue, as the money made from the utilisation of the big data source may be far larger than that of the fees of the third party company. For small and medium enterprises, this may not be a feasible solution, purely because their income may not be enough to fund the entire company, and pay for the third party upkeep of their big data source.

VIII. REFERENCES


Access Control and Authentication in Big Data Storage
How Secure Is Your Data?

Rachel Fairbrother
University of Derby
Derbyshire, UK
R.Fairbrother2@unimail.derby.ac.uk

Abstract—With the rapid increase in the size of data that is required to be stored by an organisation in order to take advantage of Big Data Analysis, it can be seen that the next big problem that needs to be addressed in the coming year is to do with the Authorisation. Hayes (2014) states that “The three top security concerns in 2014 will be Authentication, Authentication, Authentication.”

Index Terms—Authentication, Access Control, Volume, Vulnerabilities, Verification.

I. INTRODUCTION

The following article is tailored for the data processors involved in the Big Data process. The way in which data should be stored and protected both from inside and outside of the organisation will be discussed.

As the amount of data that is collected by every organisation on a daily basis increases, there is an increased pressure to discover the most appropriate method of storage for the Big Data that the organisation accumulates. The sensitive nature of the data that is collected makes this process even more important.

The amount of data that was produced and stored on a daily basis in 2008 was only as little as hundreds of gigabytes, however now we are producing hundreds of terabytes (Kaisler, et. al., 2013).

The storage of the Volume of Big Data is such a big issue with the expected growth of data being produced and stored every day being at 7.9 zettabytes by 2015 (Roe, 2012).

The different methods and options for the storage of the expanse of data will be looked into specifically the option of Data Lakes over the traditional option of public clouds (Geer, 2014).

II. BIG DATA

Big Data is not necessarily a new idea, it is just the principle of the way in which the data is transmitted into the organisation together with the varying data types that are being collected by many different organisations. The fact is that the data that is being collected is now being analyzed to such a degree in order to discover the trends of the customers and therefore get ahead of the competition by being able to target the next big thing. It is this analysis that makes this idea of Big Data such an advantageous option for businesses of all shapes and sizes (The Association of Chartered Certified Accountants, 2013).

Big Data is being adopted by more and more organisations around the world because of the advantages that adoption of the Big Data movement has and is still providing for the Big Businesses. The benefits of this movement are just as possible for Small and Medium Enterprises especially with the increasing availability of varying Cloud Services (Dumbill, 2012).

A. Definition of Big Data

“Big data is data that exceeds the processing capacity of conventional database systems” (Dumbill, 2012). Along with the previous definition of Big Data, there are many different aspects of the data that are collected. The main aspects which are going to be discussed are to do with Volume, Vulnerability and the Verification of the data which is collected. The main three aspects that are considered when looking at the idea of Big Data are Velocity, Volume and Variety. It is these three aspects that lead to the security and privacy issues that becoming part of the Big Data movement can pose on the organisation (Cloud Security Alliance, 2012).

B. Definition of Volume, Vulnerabilities and Verification

The V’s discussed below are being looked at from a user focus rather than the traditional data focus.

The Volume of Big Data is to do with the fast moving and increasing growth in the amount of data that an organisation collects on a day to day basis.

Vulnerabilities of Big Data are where the information is susceptible to attack or data breach from either within or outside of the organisation itself.

Verification of Big Data is all about the authentication that is necessary within the storage systems of Big Data (The Chartered Institute for IT, 2013).
C. Volume, Vulnerability and Verification

When it comes to the adoption of Big Data analysis within Small and Medium Enterprises there are a number of questions that need to be asked particularly when looking at the Volume, Vulnerability and Verification of the data that is being analysed.

With the size and scale of the nature of Big Data there needs to be a lot of consideration when it comes to the application of the most appropriate storage solution, this is as the storage system needs to be able to keep up with the continually expanding storage.

Vulnerability of Big Data has strong links with the type of storage solution that is chosen, as if the data storage solution does not contain the required security procedures then the organisation would have a way in which to store Big Data but not to protect it from any vulnerabilities both inside and outside of the organisation.

The Verification of the data that is collected is all about ensuring that the data which is collected is accessed by the appropriate employees within the organisation.

III. SECURITY AND BIG DATA

Due to the very nature of the data that is being collected about customers in order to increase the amount of sales and customers, the key component that must be considered and validated throughout the course of data collection is the security of the data, once it has been collected. There are many advantages to Big Data Analysis but also a lot of disadvantages if the security is not correct.

The three V’s mentioned above, Volume, Vulnerability and Verification are all to do with the way in which the data is secured and governed while in the organisation. All of the processes and policies that are used within the rest of the organisation need to be applied and maintained through the remote storage systems employed for the storage of Big Data. The data that is collected should be protected through the use of Authentication and Access Controls as well as anti-virus and encryption (Woodsworth, n.d).

The biggest security issue with existing data storage systems is that administrators have a lot of power and control over the access control mechanisms and generally have the overriding control over the data in general. The amount of time that it takes for an administrator to be prevented from exceeding their permissions is just under two days (McAfee, n.d). A lot of damage could be done to the data held within any part of the organisation during this time.

The security of the data that is kept within the organisation as well as the data that is stored externally through the use of the cloud services, is very important as the amount of data records that have been breached has seen both a rise and fall over the last four years. In 2010 the number of data records that were breached was at over 3 million, which rose to just over 170 million in 2011 (Verizon, 2013, pg. 47). These figures show the importance of ensuring the security of the data that is collected and especially with the increase in the sheer quantity data that is currently being collected, a breach within one of the cloud environments could do irreparable damage to the business.

In 2013 Target experienced a data breach affecting 70 million customers (Munson, 2014). A breach can also be seen to have occurred within Neiman Marcus, both of these breaches highlight the importance of security and not just external security but security around the data itself (Ashford, 2014).

A. Access Control and Authentication

Access Control and Authentication mechanisms have to be set up and adhered to through the use of user education as a way to ensure that the data is protected. In order for an organisation to be ISO 27002 compliant then access to all information must be regulated and appropriate according to the job role of the employee (BSI Standards Publication, 2013).

The conditions of which the data was collected are also vitally important when implementing and reviewing the security procedures that apply to all forms of data within the organisation, especially that of Big Data.

If the data is not protected from both internal and external forces the organisation could be faced with loss of reputation as well as penalties for breaking data protection. (Information Commissioner’s Office, n.d)

The benefits of Big Data Analysis can only be appreciated if the correct access controls and authentication protocols are implemented within the Storage devices that are used to hold the Volume of Big Data.

If these are neglected then the data and analysis of the data which is collected during the Big Data Analysis Process, can be used incorrectly by employees who should not have access to the data, which would lead to the inappropriate dissemination (The Chartered Institute for IT, 2013).

The number of different security and information breaches that have occurred due to access control breaches of the staff within the organisation can be seen in the table below from PwC (2013).

<table>
<thead>
<tr>
<th>What type of staff-related incidents did respondents suffer?</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unauthorised access to systems or data (e.g. using someone else’s ID)</td>
<td>36%</td>
</tr>
<tr>
<td>Breach of data protection laws or regulations</td>
<td>34%</td>
</tr>
<tr>
<td>Loss or theft of confidential information</td>
<td>32%</td>
</tr>
<tr>
<td>Breach of confidentiality</td>
<td>26%</td>
</tr>
<tr>
<td>Misuse of access</td>
<td>21%</td>
</tr>
</tbody>
</table>

Fig 1: Common staff-related incidents
B. Effect on Small Medium Enterprises

Access control and authorisation is of the upmost importance when it can be seen that 65% of small businesses leave sensitive business data unprotected and unencrypted (Trend Micro, 2012). This issue is even more important when it comes to the type of data that would be stored and protected during Big Data analytics.

The best approach for the protection and security of data with storage systems is “Constant education and training around IT security is necessary to help reduce human error” (Kelly, 2011).

Without this continued training it can be seen from the PwC security breaches survey that 57% of small businesses fell foul to a staff related breach (PwC, 2013, pg. 2).

There are links to be found between the lack of security controls and the increased chance of a data breach within the organization (Munson, 2014).

IV. CLOUD COMPUTING

With the uptake of Big Data Analytics there is the increase in the adoption of cloud services as a way to store and protect the data using both security policies and access control and authentication policies. Till (2013) explains that there is a link between Big Data and Cloud Computing due to the amount of resources that are available with a Cloud environment and that which are necessary with the analysis of the Volume of data which is being collected under the movement of Big Data.

As with every new organisation that goes to the cloud environment it can be seen that although the cloud offers many advantages for the organisation with the benefits of being able to carry out Big Data analysis, there is also the issue of keeping control over the data that is in the organisation control (Cross, 2012). The use of Data Lakes instead of the average public cloud includes different security mechanisms which can be seen to have already been rolled into the Data Lake (Geer, 2014).

A. Data Lakes

The use of Data Lakes helps to protect the data that is stored within the lake from the traditional issues for security which can be found within the public clouds. Data Lakes can be seen to include the need for the implementation of role based access controls which will ensure that even with the Volume of the data which is kept within the Data Lake, it will be fully secured (Geer, 2014).

B. Strengths of Data Lakes

Data Lakes provide the user and the organisation with access to the data held within straight away; there is flexibility when it comes to the data stored within the Data Lake. There is no restriction on the type of data that can be stored within a Data Lake, finally once the data is stored in the Data Lake there is no need for any of the data to be relocated before or after the analysis phase of the data (Eckerson, 2014).

C. Weaknesses of Data Lakes

Data Lakes are seen as the best option to deal with the Volume of Big Data, however there is the problem of the fact that this Data Lake has not be adopted within many organisations and is therefore untested. There is an expense included in the adoption of Data Lakes, however this would mean that the data stored is secured. Lastly the ways in which the data stored within the Data Lakes are governed have not been tried and tested to the same degree as private clouds.

D. Cloud and SME’s

Venturing into the cloud environment can be seen as one of the best and most cost effective ways for Small and Medium Businesses to be able to take advantage of Big Data Analysis, whether from the customer point of view or for the growing improvement of security protocols and policies (Simons, 2013).

The cloud environment appears to be the best choice for SME’s in terms of the Volume of data that they need to store safely on a day to day basis, however a normal public cloud can not necessarily be the most secure option as in 2013 “4% of respondents have detected a security breach or data breach that affected a cloud based service they use” (PwC, 2013, pg. 8). With this in mind a Data Lake is proposed as the best alternative with security protocols being added to the separate pieces of data by the administrator of the organisation. This is without the need for any automation procedures which could lead data open to employees who should not have access to it (Geer, 2014; Cloud Security Alliance, 2012).

E. Cloud and Security

When looking at the security that is included in the private and public clouds systems that are available to businesses, the main problem which needs to be considered is ensuring that the privacy of the users is protected (Kumar et al., 2012).

V. SECURITY IMPLEMENTATIONS

In order for the Data Processors to be able to fully secure the data that is in their possession, there are varying protocols and policies that need to be adhered to.

A. Access Control Implementation

For Access Control to be fully implemented within the storage systems chosen to hold the Volume of data there are several processes that need to be carried out, these include regular reviews over the permissions that are given the members of the organisation (Trend Micro, 2012). Also the processes for access control and authentication that are used within the cloud environment and Data Lakes chosen by the organisation as the most appropriate for the storage and security criteria of the organisation, this can be done through the application of data security rules into an Identity Access Management (IAM) system in order to ensure that the correct data is available to the most applicable members of staff (Geer, 2014). The use of an IAM system means that even with the fast moving Volume of data that is involved in Big Data Analysis, the data is protected in either the Cloud Environment or the Data Lake.

B. Anti-Virus and Malware Protection

The Data needs to not only have correct access controls but also protection from malware and viruses. This can be done
through the use of the correct and most up to date virus protection systems both within the normal organisational network and the Cloud or Data Lake used to store the vast quantity of data used for Big Data Analytics.

C. Data Back-up

Due to the Volume and nature of the data that is handled and stored by the Data Processors the main aspect of data security that needs to be followed is to back up the data on a regular basis. It can be seen through the research by Trend Micro (2012) that Small and Medium Business do not consistently back up their data.

VI. CONCLUSION

The issues addressed within this article to do with the importance of accurate access control and authentication can be seen to be vital for the expected increase of the Volume of Big Data within the next decade, this view is shared by the Canadian Bank who believe that the Volume of data will be 50 times greater within the following decade (Kalyani, 2013).

Big Data Analysis is available to all organisations no matter what the size, especially with the availability of public and private clouds as well as Data Lakes. As long as correct security policies are in place within all storage environments, then the Verification and Vulnerabilities associated with the fast moving data analysis and storage can be seen to reduced and therefore less of a problem. Then Small and Medium Businesses will be able to reap the rewards of analysing trends that exist in their customers’ habits or even in the trends that exist within the world of data breaches and cybercrime.

VII. REFERENCES


are we going to do with all those zettabytes? [Accessed: 28th March 2014].


**Abstract**—Organisations small and large are trying to create a viable Big Data strategy for themselves. Big Data Analysis has the potential to improve efficiency, reduce costs, and provide foresight. The UK’s Healthcare section is one of the many organizations that stand to greatly benefit through carefully implementing Big Data. This paper will look into the potential benefits of Big Data in the NHS, as well as the potential risks.

**Index Terms**— Big Data, Big Data Analysis, Healthcare, UK

I. INTRODUCTION

Big Data is exactly what it sounds like: a large quantity of information. In 2001, the idea of defining Big Data through ‘Three Vs’ was introduced: Volume, Velocity and Variety. SAS, an organisation that works with data, also considers Variability and Complexity as equally defining aspects of Big Data. The importance of Big Data, however, is not in the raw information it holds, but in how it is utilised. (Davenport and Dyché, 2013)

Big Data Analysis has the capacity to scan through vast quantities of data in order to seek out patterns and trends, and then use these findings to seek out anomalies. However, the UK currently only uses Big Data to improve the efficiency of pharmaceuticals, and to try and overcome diseases such as diabetes and cancer. It stands to benefit greatly through the further implementation of Big Data into its Healthcare Section, and mainly into the NHS. The UK is known to have some of the largest and most complete data sets in the world, and the NHS has a much larger than average amount of data that could be used. (HM Government, 2013)

II. THE BENEFITS OF BIG DATA IN HEALTHCARE

A. Financial Savings

The USA is proving the worth of Big Data Analysis in Healthcare. Studies have shown that it has the potential to save their Healthcare Section around $450 billion a year. (Parslow, 2014) Premier, a US healthcare network have utilised Big Data, and through its use have saved 29,000 lives and $7 billion. (IBM, 2013) Using Big Data Analysis in the UK would enable the NHS to see which departments have higher expenditure than others, and see where savings could be made.

It has been proven that by simply improving the analysis of performance data alone, £1.9 billion could be saved. (SAS, 2014) The Government have launched a new campaign, entitled the Quality, Innovation, Productivity and Prevention (QIPP) initiative. Through QIPP, the Government hopes to have saved £20 billion of efficiency savings in the NHS by 2014. Big Data could play a big role in saving money in the NHS in the long run. (Department of Health, 2013)

B. Performance Management

Introducing Performance Information (PI) into the UK’s Healthcare Section would enable the NHS to improve how they manage their resources. It would enable them to set objectives, monitor the performance of staff, and plan management functions more efficiently. (SAS, 2014) On top of this, Big Data would allow the various departments in Healthcare to collaborate more easily, resulting in quicker treatments.

C. Fraud Detection

The UK’s public sector lost £20 billion through fraud between 2011 and 2012. Big Data Analysis could help reduce this figure, as it improves fraud prevention dramatically. (SAS, 2014) Through analysis methods such as high and low values, Big Data Analysis has the potential to discover anomalies in vast quantities of data. (ACL, 2013)

Big Data offers greater control for fraud prevention than internal control systems. The scope that is offered through Big Data Analysis includes testing data against parameters, across applications, and across entire systems from external applications and data sources. This level of analysis simply can’t be handled by internal systems, and could greatly limit the potential for systems to be open to fraudulent activity. (ACL, 2013)

D. Predictive Intelligence

Big Data allows for ‘Predictive Intelligence’. This means that, as results are tracked, links will be discovered linking health conditions to illnesses and diseases. This will give doctors the ability to better foresee a patient’s risk of developing conditions. (Bertolucci, 2013)

In the USA, Big Data is being used to predict which patients are likely to be readmitted within 30 days, and which
remedial actions will be required for these patients. (Parslow, 2014) This is not only saving them money and resources, but is also improving the level of care that patients receive.

E. Personalised Healthcare

Big Data Analysis could be used to personalise healthcare to individual patients, by comparing patients against similar cases. This, coupled with the aforementioned Predictive Intelligence would work towards catering for both their current and potential future healthcare requirements. This would revolutionise the focus of Healthcare away from healing the sick, towards eliminating health risks before they become a problem (Wall, 2014)

The human genome is essentially the code that forms the genetic makeup of the entire human race. Scientists have been working to crack the code of the human genome for many years, and various research projects on it are nearing completion. (Bonsor and Layton, 2014) Upon their completion, doctors will be able to have individual patient’s entire genetic map on medical record, thanks to human genome mapping. This, combined with Big Data Analysis could lead to fully individually tailored healthcare, where decisions are made based on unique patient characteristics. (Data Science Series, 2014) One example of this would be based on the treatment of a cancer patient. The best treatment option could be identified straight away, saving costs of failed treatments, and improving the level of care. (Data Science Series, 2014)

III. EVALUATION OF THE ISSUES

A. ISO 27002:2013

The ISO 27002 framework is based on information security for small to large enterprises. (BSI, 2013) It provides policies and standards that are required to be met in order to ensure that there is minimal risk of a breach in security. Small and Medium Enterprises (SMEs) face similar issues to the NHS in the implementation of Big Data. The issues SMEs and the NHS will encounter regarding safety and compliance when outsourcing analytics are very similar, along with other aspects of implementing Big Data. Because of this, it is a perfect framework to assess the risks surrounding the implementation of Big Data and Big Data Analysis into the Healthcare section of the UK.

B. Care.Data

Care.Data is a database that the NHS had planned to create. However, it has been put on hold for 6 months, due to the public’s opinion on it. (Baldwin, 2014) One of the main reasons for the negativity surrounding Care.Data was the NHS’s failure to properly describe the benefits of what it could offer. (Baldwin, 2014) However, other reasons for the opinion of the public are thanks to a poor focus on the safety of patient data, and governance issues.

One of the biggest concerns of the public was the safety and privacy of their data. This is unsurprising, as the Health and Social Care Information Centre (HSCIC), the holder of all of the data, had the capability of sharing the data with businesses as long as they were financially compensated. (Best, 2014) On top of that, simple anonymity, whilst currently being enough to keep identities safe, will not provide a high enough level of safety in the future. This is due to the advancement of data mining software and techniques. (Glick, 2014) There was also no availability to opt out of Care.Data once you were registered. Due to the level of privacy of the data, this is unacceptable. (Glick, 2014)

There have been many failures in the implementation of Care.Data. These are failures that need to be avoided if a fresh attempt at implementing Big Data into the NHS is to succeed.

C. Internal Vs. Outsourced Big Data Analysis

The requirement for Big Data analysts is increasing quickly. Online adverts for data analysis professionals have increased dramatically, with an increase of 246% between April 2009 and December 2013. (Fogarty and Bell, 2013) While an internal analysis team has less complex compliance and security concerns than outsourced teams, in-house teams are difficult to produce. Big Data Analysis on the scale that the NHS would require a large team of highly skilled data analysts. (Fogarty and Bell, 2013) Hiring data specialists is expensive, and employing a specialist team large enough to utilise the data that the NHS has to offer would cost a very large amount of money.

However, outsourcing analysts has its own issues. Though many aspects of outsourcing Big Data Analysis offer potential benefits, it is something that should not ideally be viewed as a permanent solution. There is a large degree of organisation-specific thinking that is required to provide the models to be used with the data. Without proper insight, the analysis may not answer the questions that the NHS wants to know. (Booker, 2012) Also, many individuals feel uncomfortable with signing off on their personal details being sent to third party companies. (Matthews, 2013) This could lead to many individuals not having their data analysed unless the safety of the data could be proven. This could result in an incomplete or flawed analysis.

D. The Security of Personal Data

There is a growing fear that, as organisations collect more and more information, they will know too much about the individuals represented in the data. Concerns have already been raised in regard to how data in electronic health records that follow existing HIPAA/HITECH provisions do not protect individual’s privacy to a great enough extent. (Kaiser and Armour et al., 2013)

People also pose a threat on the security of data. One mistake made by an employee could lead to a large amount of personal data being leaked, which could be disastrous depending on the volume and value of the data that is lost. (Yiu, 2012)

In order for patients to feel comfortable with the implementation of Big Data, there needs to be a strong focus on the safety of personal data. This will require focusing not only on the safety protocols protecting the data, but also on who has access to it.
E. The Validity of the Data

Validity is one of the qualities that define Big Data. Currently, the NHS collects a large amount of data by having doctors and other medical staff key it in manually. Not only is this time-consuming, inefficient and costly – it is also prone to error, especially in such a busy atmosphere. (Keen and Calinescu et al., 2013) This can lead to false information, and also allows for certain cells of a database to be left blank.

Human error poses the biggest threat to the validity of the data that would be used. Any mistake made would alter the results of any analysis that took place on the data sets, as the data itself would be flawed.

F. The Volume of the Data

The world is generating more data than ever before. Previously, whenever the amount of data that was expected in data sets grew, there was hardware to support the growth – from floppy to CD, from CD to Blu-ray, etc. In this latest revolution, there have not been any new physical technologies. Because of this, the use of private, public, and hybrid cloud computing techniques will become increasingly important in the use of Big Data. (Kaiser and Armour et al., 2013) Cloud computing has many compliance and security issues, as much of the data ends up being stored outside of the physical location of the organisation it belongs to.

As well as storage issues, the frameworks that such vast quantities of data can be used with have not yet been created. Interoperability between sets of data requires continuity in how data are stored across the many departments of the Healthcare section. (Yiu, 2012) Without the correct frameworks each department would need to be analysed separately, limiting the usefulness of Big Data Analysis dramatically.

Finally, as the volume of data increases over time, decisions will need to be made about when old data becomes obsolete. Space issues will ensure that entire archives of data cannot be stored and analysed, and the process used to delete data regarded as unnecessary will need to be carefully planned to ensure the safety of individual’s personal data. (Yiu, 2012)

IV. PROPOSED STRATEGY

A. Internal Vs. Outsourced Big Data Analysis

If there were the financial capacity to create an internal Big Data Analysis team, then there would be many benefits for doing so. The lack of having to send data to a third party would likely help to make individuals feel more at ease with allowing their data to be used as part of the Big Data Analysis.

Clause 6.1 of the ISO 27002:2013 document identifies that in order to ensure information security is maintained, policies should be created and responsibilities should be distributed to individual members of staff. (BSI, 2013) The enforcement of these policies and responsibilities would be much more easily monitored if the analysis team were based in-house.

Alternatively, outsourcing Big Data Analysis to an external organisation has many advantages if done correctly. One obvious advantage would be saving the initial cost of setting up a specialist internal analysis team. (Matthews, 2013)

In order to mitigate the risks of outsourcing the Big Data Analysis, a strong supplier relationship would need to be formed. The security requirements, obligations and contingency plans that are required of the outsourced team would need to be discussed and documented. This would ensure that each party knew what they were liable for, and what was expected from them. (BSI, 2013)

B. The Security of Personal Data

The development of algorithms that randomise personal data could help limit the damage that organisations or hackers could do if they acquired any of the personal data. (Kaisler and Armour et al., 2013) Cryptography could also be used to prevent hackers from gaining access to personal data. Aircloak and other similar organisations could be utilised to acquire this level of protection. Aircloak offer a service where data is made anonymous, but high-fidelity analysis is still possible. They also offer a high level of transparency, which would enable auditors to obtain proof that the software operated as promised. (Aircloak, 2014)

The ISO 27002:2013 document has an entire clause on Cryptography, and how it should be implemented to ensure the protection of data. Given the value of the majority of the information that will be included in the data sets, the strength and quality of the encryption algorithm should be very high. (BSI, 2013) Contingency plans such as backing up data also need to be made, in the event of a successful breach in security. (BSI, 2013)

Staff also need to be made aware of strict policies regarding data security from pre-to-post employment. Issues such as lost USB sticks and laptops cannot be tolerated, and need to be avoided at all cost. (Yiu, 2012) The importance of staff understanding their responsibilities before, during and after their employment is reiterated in clause 7 of the ISO 27002:2013 document. It also highlights the importance of ensuring that members of staff are offered continuous professional development so that their skills remain up to the task. (BSI, 2013)

C. The Validity of the Data

Many technological advances have been made for the collection of data. Better keypads and user interfaces have made it faster and easier for medical practitioners to input data accurately, and minimise the danger of inputting false data. (Keen and Calinescu et al., 2013) However, the ultimate goal should be for data collection in the NHS to be completely automated.

Automated data collection is very difficult and complex to implement. However, it would provide more accurate results, which would result in more accurate and more useful Big Data Analysis. (Keen and Calinescu et al., 2013) On top of this, it would provide more time for doctors and nurses to tend to patients, rather than having to spend valuable time inputting data into programs.

D. The Volume of the Data

One of the most challenging aspects of implementing Big Data into the NHS will be storing such vast quantities of data.
Keeping the data manageable and accessible will prove very difficult, as the NHS does not currently have the necessary facilities. (Keen and Calinescu et al., 2013)

Because of this, at least some data storage will need to be outsourced to a third party cloud computing organisation. There are many issues with compliance and security when storing data in the cloud: data protection, data availability, and the ownership of data all needs to be considered. (Vizard, 2013) However, by setting up the proper policies and standards with a carefully selected cloud provider, all of these potential issues can be addressed. (BSI, 2013) Through techniques such as compartmentalisation, and data encryption, the data stored in the cloud can be kept safe in the event of theft. (Cloud Security Alliance, 2009)

V. CONCLUSION

Big Data and Big Data Analysis have a lot to offer the NHS. By minimising costs, increasing productivity, and providing further foresight into individual patient’s futures, it could revolutionise the way in which medical care is provided.

However, Big Data is not without its flaws. Without specialist analysis teams, the required technological infrastructure and the proper frameworks and policies, it can fail to provide any truly useful information. (Keen and Calinescu et al., 2013) There are also many concerns regarding the safety of individual’s personal data, as well as obvious potential compliance issues.

Despite this, many of these problems can be resolved. Through analysing the needs and wants of the NHS’s unique requirements, proper policies could be created that protect any data involved, regardless of whether the data analysis is managed internally or outsourced. Constant communication with the analysis team, as well as constant re-evaluation of what it is that the NHS is trying to learn will dictate the success or failure of the implementation of Big Data in the long-term. (Keen and Calinescu et al., 2013)

VI. REFERENCES


Big Data: Why Aren’t UK SMEs using it?

Should they be using it? Yes or No?

Samuel Fowler
School of Business, Computing, and Law
University of Derby
Derby
S.Fowler1@unimail.derby.ac.uk

Abstract—Big Data has become something of a “Hot Topic” within the world of computing and business over the past couple of years. It seems like this brand new thing in which will benefit a business greatly. Some think it to be the best new thing, others think it has been around for a while but just re-marketed with a brand new name. Regardless of these fact, this report will focus on why British SMEs are not using it, and whether this is a good thing, or a thing that they should be implementing?

Index Terms—SME, Big Data, The 4 V’s; Volume, Velocity, Variety, and Veracity.

I. INTRODUCTION

Big Data is the term used to describe a large set of data that has been collected, and that data is so large and complex that it becomes a difficult task in order to process any of it. Big data has become well known at defining what Big Data is and whether it is useful or not. Whilst some people believe that it is a brand new revolutionary tool for businesses, others feel that it has been around for a while and marketers have just re-marketed it.

Regardless of opinions Big Data is a term that is frequently being used and will continue to be for some time, at the very least until 2017 according to (E Skills UK, SAS. 2013). So bearing this in mind it brings up many questions for SMEs such as what is big data, is it useful? And how do I use it to benefit my business? Most of these questions remain unanswered as Big Data seems to be something of a hazy topic. There is a lot of attention surrounding it, yet a lot of people don’t understand it.

When referring to Big Data, a common way to separate and describe the qualities of Big Data is by listing into 4 different categories. These 4 categories are called “The 4 V’s of Big Data”. They are; Volume (scale of data), Velocity (Analysis of streaming data), Variety (different forms of data) and Veracity (uncertainty of data). (Ibmbigdatahub.com, 2014) These are the main 4 V’s, however, technically there are more V’s that can be used and analyzed. Most commonly it is either three, four, or seven V’s in total. If referring to the 7 V’s, the remaining V’s not mentioned previously are: Validity, Value, and Visibility (Livingstone, R. 2013). In this report the 4 V’s shall be the main focus and will be referenced to throughout. This is because the main 4 V’s hold most relevance to the topic at hand and are the most useful to compare the type of businesses (SMEs) to.

The main aim of this report is to research why SMEs in the UK are currently not using Big Data, find out the reasoning behind why they aren’t using it, and find out if it is in fact a useful tools for SMEs and ultimately decide if they should be using it or not.

II. WHY UK SMES CURRENTLY AREN’T USING BIG DATA

A study was recently taken out in the UK entitled “Big Data Analytics – Adoption and Employment Trends 2012 – 2017” to determine whether or not SMEs were actually using Big Data, the results that were found are mainly negative. Showcasing that hardly any SMEs were using Big Data at all. The study reported with facts and figures backing up the statistics of the lack of Big Data use.

It was shown that in 2012 a mere 0.2% of UK’s SMEs were using Big Data within their businesses (E Skills UK, SAS. 2013). This is an extremely low percentage. Considering the recent breakthrough of Big Data over the past year or two, and with facts predicting that by 2016 25% of Large Global Companies will be adopting Big Data Analytics (Garter, 2014) It seems hard to believe that more SMEs in the UK wouldn’t be wanting to or already be adopting Big Data in order to potentially improve their businesses.

Big Data is not just a trend that has completely missed the UK as 14% of UK larger organisations have already implemented Big Data, with 7% UK large organisations in the process of implementing at this point in time (E Skills UK, SAS. 2013).

There could be numerous reasons behind UK SMEs not using Big Data, as there are many different angles in which you could view the situation and try to understand why some Small or Medium sized companies would not want to implement Big Data. Is it because they don’t want to? Or is it because they can’t? Whether they can’t get the resources or the staff, or just don’t understand what Big Data is?
A. Lack of Understanding

One of the main reasons why SMEs are not implementing Big Data is because they do not understand what it is. The report published by (E Skills UK, SAS. 2013). showed that 1/5 of businesses claim that they have “Very Poor” understanding of Big Data. This relates to both a lack of understanding of what Big Data is and how to implement it within their business. This is an understandable reason not to use something, if their business is running efficiently and Big Data is complicated to understand, then why should they bother trying to implement and use it.

B. Hiring Specialists

Another reason SMEs are not implementing is because of the data specialist staff that would be required if they wanted to implement Big Data properly (or avoid the hassle of doing it themselves). The E-Skills study found that 57% of recruiters found hiring big data specialists “fairly or very difficult” (E Skills UK, SAS. 2013). The reason behind big data specialists being hard to find and employ is due to how niche and new the term of big data is, there simply has not being enough time in order for there to be a big boom in the job position of “Big Data Specialist”, and because of this small time scale, it means that people working in this profession are very few and far between and most likely being employed by the larger organisations, thus SMEs not being able to employ any. With a serious lack of specialists, it could be argued that the businesses could train and employ from within their own company and take care of Big Data themselves however with the awareness and lack of knowledge surround Big Data it could prove to be a difficult task and could also be too time consuming and costly for a SME to be able to attempt to perform.

C. Security

Security is another big concern when SMEs would contemplate the potential implementation of Big Data in their companies. There are various different security risks that come with big data, some are fairly minor and would just cause inconveniences. However, other could be a large problem, causing major inconvenience and potentially putting even the largest companies in risk, and if it can badly damage large organisations then it could potentially destroy SMEs. A recent study in 2013 found that 58% of organisations faced a security breach whilst using Big Data. Out of this 58% companies were asked how long it took them to detect the source of the breach; 14% were able to find the source in minutes, 33% found it in a day and 16% found the source in a week (McAfee, 2013). Another example that could be used to show potential security risks, is the Spec’s owned liquor stores that lost over half a million of their customer’s data, due to hackers. The Spec’s security breach was estimated to be from 31st October 2012 and lasted up until 20th March 2014, putting many customers of the business at risk. Not only this but the detection of the source was ‘Exceptionally long’ and the security protocols were said to be inefficient (Kaplan, D. 2014). This causes concern as security threats are a natural part of a business using computers to aid their work, however most computer threats can be combatted and normally combated fast. The problem with Big Data is that SMEs will have a security breach and not be able to find the source quickly, even if they have employed a specialist (on their own, this time could prove to be even slower).

D. Staff Numbers

The SAS study also showed statistics stating that companies who have lower staff numbers (which in most cases would be SMEs) are the ones who are using Big Data the least, and the larger the amount of staff the more likely they are to implement it. This is due to the amount of people and whether the implementation and use of big data can be performed.

E. Financial Position

From the various reasons why SMEs are not implementing Big Data, a lot of them stem from the financial position of the company and the companies awareness of their own financial state (Forbes, 2013). Most of these financial problems stem from the reasons already listed; SMEs don’t want to hire extra staff in order to try and implement big data, they don’t want to have to hire specialists (costly in not only finding but also costly when employed), time taken out of the business to focus on big data could be considered a cost as it is taking up company time that could spent elsewhere. These facts and statistics show an insight into why UK SMEs are not using Big Data. It seems fairly logical that they may not be using it or not want to use it. However, even though they are not using Big Data, should they be?

III. SHOULD SMEs USE BIG DATA: HOW THEY COULD USE IT TO THEIR ADVANTAGE

There are many different sources from reports and various that claim that give great praise to Big Data and claim that it is extremely helpful and shows to be a smart move for businesses. Whether these claims are relatable for SMEs is a different matter, and sometimes these claims can just be a marketing technique trying to sell big data rather than actually give people the hard facts.

A. Analysing the basics

The common fear for most SMEs is caused from the lack of understanding of the concept of Big Data, and a fear that their expertise will not be up to the standards needed in order to implement and maintain Big Data. This may be the case if they were large multinational organisations, however, it could be a lot easier than that if they so wished. By just keeping to the basics of analysing data they should be able to benefit from big data without it being too demanding. Basic activities could include; Monitoring online website visits, checking when people check into things such as foursquare and monitoring other various activities on a range of different social media websites, checking when customer interact with the business, and analysing past activity. If they make business decisions from doing this simple tasks, then it could benefit their business greatly. (SME Strategy Management Consulting & Strategic planning, 2013)
B. Basing on Past Activity

One of the most useful benefits of SMEs using big data, is so they analyse their past performance and from this help improve their future performance. This data can be used to perform such tasks as when customer interact with the business or by tracking customer purchases, this will help pin point things like weather, time of year, days bought etc. (SME Strategy Management Consulting & Strategic planning, 2013). The advantage of an SME using them is they can use something basic and easy to use like google analytics that will perform the tasks they need perfectly, whereas bigger corporations will need to use expert tools.

C. No Such thing as Big Data for SMEs? Not Always

“There’s a simple reason why there’s no market in small and medium-sized enterprises for big data – they don’t have big data.”

This statement from Campbell Williams (Six Degrees Group, 2013) may be true as SMEs simply do not have enough data to warrant them calling it big at all. There is simply not a large amount of data to analyse, the important thing is how they analyse and use the data that they have, and make sure they use it efficiently. Not only this, but they could also try ways of expanding their data so they have more to analyse and process. They could do this by collaborating with their supply chain, so they are not only analysing their own company data but also analysing their supply chains data will make their analysis more accurate and more precise and more worth of the title Big Data. (Rijmenam, M. 2013)

Another way in which SMEs could potentially make their data “bigger” or more efficient or accurate is by being more creative. It about thinking outside the box for each individual company. There are many ways that a company could track data, one example is that Sensors are becoming cheaper, companies could place sensors in their shop (if they own a shop) door and monitor and analyse entry and exit. (Rijmenam, M. 2013). This is only one example of the creativity that a business could use, there are plenty more. The key is for the SMEs to be as creative as possible. Not only by taking out extreme measures, some of these things can be more simplistic by just analysing different data such as social data, documents, emails and even voice data. This will all benefit the analysis. (Rijmenam, M. 2013).

D. Make Use of Existing Resources

Regardless of whether or not SMEs have big data, it is wise for any business to analyse their data and use this to their advantage. Even if SMEs don’t have enough data to be able to name it as ‘Big Data’ they should still analyse the data that they have, as it may benefit their business and ultimately make their business run more efficiently, and most importantly (for the business) cut costs and raise profits. (Lid, B. 2014.)

This can be relate to Big Data Analytics and by using some of the techniques and practices that Big Data Analytics uses, it could still be useful to take those skills and use them on a SMEs (even the data isn’t that big). Relating the topic of using existing resources can link back to the four V’s and in particular the ‘Variety’ aspect of the four V’s (Ibmbigdatahub.com, 2014). It relates back to Variety due to the way in which it takes a number of different already existing resources, no matter how different some are and the analyses them in order to benefit the business.

IV. SHOULD SMEs USE BIG DATA; REASONS WHY THEY SHOULD AVOID IT

There may be some upsides for SMEs when they are using Big Data, or just the data they have (seen as it could be considered that they don’t have enough data to warrant it being called big.) However, there are some problems that could potentially come with implementing Big Data. Some of these problems would affect even the largest of corporations, so could potentially destroy SMEs.

A. Problems Existing Companies Find with Big Data

Referring back to the (E Skills UK, SAS. 2013) report issued on Big Data trends, it stated percentages of the different reasons that companies have found whilst using Big Data. The largest percentage was 56% for the problem of determining how to get value from the data. 41% of companies said that the problem was defining the strategy of collecting and analysing the data. 34% of companies stated that their problem was obtaining the skills & capabilities needed in order to implement data, this refers to specialist staff and software that may be required. And finally 33% of companies said that their problem was integrating multiple data sources.

From the (E Skills UK, SAS. 2013) percentages and statistics based on existing companies problems, it becomes clear why SMEs in the UK are choosing not to implement it to their business. Some of the percentages linked with the reasons why SMEs were not currently using Big Data, such as Obtaining skills, determining how to get value, and defining a strategy. Not only are these statistics reasons why SMEs are currently using Big Data but it is also problems that they would face if they did choose to implement and manage Big Data in their own companies.

B. Human Error

One of the main problems that SMEs will find with implementing and attempting to use Big Data is that humans (themselves and their staff) are the most likely to be the ones who cause the most problems and stop it from working efficiently or at all. One of the main problems that occurs from human error is the communication issues that arises. If the people within the company do not communicate enough, then the data collected will be void. For example if the data specialist doesn’t feedback the necessary information analysed from the data found then it will not be effective (VentureBeat, 2013).

Another main human error is what the company actually does with the data that they have collected, if they do not turn that data into real potential then the data collected is essentially useless. A business could collect large quantities of data, yet they could do nothing with it, or make all the wrong decisions with it, causing it to be void. Essentially Companies have to make use of the data and make the right decisions and know
how to use the data given, otherwise it could prove to be useless (TIBCO Spotfire, 2012)

Another problem that comes with human errors is the flexibility and maintenance involved. If the company is flexible enough to move records of data and then analyses them, and then after that maintain this data and the work then it should be fine, however if they can’t they once again Big Data would be useless (DecisionMarketing, 2013).

V. CONCLUSION

To conclude, There are various reasons as to why SMEs in the UK are not using Big Data; things such as lack of understanding, fear of what it might be, lack of specialists in the field and where to find and hire them, security issues and so forth. Not only could this but technically it be considered that SMEs do not even have enough data to warrant calling it “Big Data” at all. However, the fact that SMEs are not using “Big Data” as much as they should could be considered as the business losing out. Essentially there are factors and risks that need to be taken into account when deciding whether to implement it, but with most decisions in business comes these kinds of questions and responsibilities.

SMEs may not be able to technically claim that they have “Big Data” due to them not having enough data, but they can still collect and analyse data for their company and that can show to be a very useful and productive tool for a company. Any company (including SMEs) should be collecting data from and about their company from various different sources and analyse it as it could prove to be extremely beneficial to the way their business is run and could save on costs and help gain profit. This is assuming that they are collecting and analysing the data in the correct manner to begin with.

VI. REFERENCES


Understanding what Big Data Analytics can Provide for your SME

Is your Company Suffering from a Lack of Knowledge Regarding IT Systems Analysis?

Jamie Fox
School of Computing and Mathematics
University of Derby
Derby, United Kingdom
J.Fox5@unimail.derby.ac.uk

Abstract— Do SMEs fully understand the real potential that could be hiding amongst the piles of uncollected data within their systems? The volume of such varying information is growing at such velocity could and should be analysed to retrieve positive benefits including financial and reputational gain. Big data is the next big thing in the analytical IT world and this report will be discussing exactly why SMEs should be taking immediate notice as well as discussing the various factors, both positive and negative within the field of big data.

Index Terms — Big Data, SME, Analytics, Social Media, Volume, Velocity, Variety.

I. INTRODUCTION

No matter what service an SME (Small to Medium Enterprise) intends to provide for its potential customers, its main intentions are to make financial gain and inevitably grow into a more reputable and profitable company. Of course this will rely on various factors that will influence the progress of the SME in either a positive or negative way. The classification of an SME does vary across the world however European Union law defines one as having no more than 250 employees as well as having a turnover of no more than £11.2 million (Ec.europa.eu, n.d.). This report will discuss the potential that is continuously available for SMEs including reputation growth (for example through social media channels) which of course is likely to eventually lead to profit increase. If the concept of Big Data analytics is fully understood and acknowledged as an overall benefit, the possibilities are endless. Simons (2013) explains how big data analytic opportunities and techniques are often overlooked by SMEs due to the concept being seen as the domain of the giant corporations (the potential costs for example) and so they could miss out on various benefits.

This report will also include discussions regarding what issues might occur if these factors are ignored such as security dangers as well as the general risks that will inevitably arise within any SME whether change is made or not. As well as this, the disadvantages of big data will be analysed to see if all SMEs would actually benefit entirely if they invested in it.

Before we progress however, it is important to define the exact meaning of what big data actually is. Big data refers to such large amounts of data that is too diverse, fast changing or large that its current technology and software tools simply cannot process or address efficiently. In other words, the volume, velocity and variety of data are too great (Mongodb, n.d.).

Understanding the entire collection of the 6 V’s in relation to big data is vital when trying to achieve success. Volume, Variety, Velocity, Veracity, Validity and Volatility make up the full set of big data characteristics despite many suggesting that only the first three are the only significant V’s (Normand, 2013). According to a report released by Oracle, the data classified as ‘big data’ refers to 3 different types. These include traditional enterprise data such as customer information or store transactional data which can often be relevant to an SME such as high-street or online shops. Another type of data is machine generated or sensor data for example smart metres, web/equipment logs as well as heart monitoring implants. As explained within a paper published by Cisco (Evans, 2011), this type of data can be classified as the ‘Internet Of Things’ in which data is collected and transferred over a network using unique identifiers called an IP address. The Cisco paper emphasises the importance of this type of data by stating how the lack of human interaction required actually speeds up the transferring and eventual analysis process across wireless technologies. It also highlights the general positive impact the internet has had on education, business and science. Finally Oracle (2013) describes the last type as social data (which will be discussed in further detail at a later stage) such as feedback streams and blogging sites through channels such as Facebook and Twitter. This explanation of what big data is provides a solid foundation for the rest of the report about how SMEs can benefit from implementing this alternative strategy into their own current system.

II. WHY SMEs SHOULD INVEST IN BIG DATA

Big data provides organisations of all sizes with a host of valuable benefits including the most important aspect of an
enterprise; financial and reputation growth. This section will analyse the issues that SMEs face without the use of big data as well as the overall benefits that it can provide for example, decreasing the risk of security issues which may or may not include the use of social media channels.

A. General issues with a typical SME and why they are not currently investing in big data

As mentioned in the introduction, SMEs often tend to overlook big data analytics opportunities due to the assumption that they are the field for large institutions only. Simons explains how this oversight can result in SMEs losing out to rival businesses that incidentally are using data analytics to improve company performance and output (Simons, 2013). The point regarding larger establishments having already capitalised seriously into big data analytics should be used as a perfect example as to why SMEs should provide funding for big data. For example, Google invested $7.3 billion into data centres for storage and examination in 2013 (Circleid, 2014) which does show that international corporations are taking big data analytics seriously.

Data volume is always growing at an estimated rate of a 40% increase per year (and will grow 44 times between 2009 and 2020) according to The McKinsey Global Institute (Oracle, 2013). SMEs will eventually become lost amongst all this data and will be unable to function efficiently without some sort of big data analytics plan to collate the varying information into manageable and useful patterns and trends to keep up with the intense velocity.

A SAS report (SAS, 2013) revealed that the level of understanding of big data is low with 22% of the SME community admitting having very poor knowledge of related concepts and technologies to the topic. The statistic that only an estimated 0.2% of all SMEs have actually implemented big data analytics into their company is also a concern as they are likely to be missing out on potential benefits (which will be discussed more later on). Passingham (2013) emphasises how many SMEs may be left out in the cold due to the in-demand skilled analytic employees heading to big corporations instead with a higher pay being the major pull factor. This adds weight to the idea that SMEs will feel the effects of not implementing big data analytics especially when the larger organisations are doing exactly that. The higher skilled workers will inevitably go to a higher established and reputable company primarily for a larger wage. The SAS report explains how 57% of recruiters found hiring big data specialists very difficult to find due to the increase in demand for such highly skilled personnel (SAS, 2013) which once again emphasizing the difficulty that SMEs face when attempting to implement big data analytics techniques.

Bob Evans (2013), writing for Forbes website, stated that 38% of business managers feel they do not have the right system in place to gather the required information whilst 29% were currently using a system that was not designed to meet the unique needs of their system. Based on the above analysis, it’s clear to see that there are many organisations that have not invested in big data technologies and are wasting money and time on systems that are irrelevant to their current needs.

B. Benefits of Big Data

It is vital for every SME to recognise what big data analytics can bring to their company. If larger organisations are anything to go by, there’s plenty of profit to be made when investing in big data schemes and technology. To emphasise its importance, it is estimated that $300 billion to $450 billion could be saved within the US healthcare sector if big data analytics are applied according to the US health centre reform business technology office report; that’s up to 17% of the nation’s entire healthcare annual spending (Groves, Kayyali, Knott and Van Kuiken, 2013).

With the ever increasing use of social media worldwide on a variety of channels such as smartphones, tablets and personal computers, international communication levels have never been so high. Over 1 billion active users are using some type of social media channel and so the volume of data is quite often unmanageable without the use of big data analytics. In fact, social media now embodies the leading and biggest source of consumer data (Bowden, 2014). Much of the high volumes of data make up the 2.5 Quintillion bytes of data that is created worldwide across each and every system that generates some kind of information (Trustwave, 2013). Bowden goes on to explain how good management of all this big data can lead to positive insights such as noticing the root cause of a problem and failures that can negatively affect the revenue for an SME. As well as this, big data allows relationships to be found to determine business trends and regulate web traffic conditions.

In terms of a marketing point of view, big data would provide a number of benefits for SMEs including cost reductions, time savings and avoidance of security risks. Referring back to the article published by Bowden (2014), here is a list of potential possibilities if particular marketing based SMEs were to invest in big data analytics:

- Define root causes of disasters, concerns and flaws in near-real time, possibly saving billions of dollars yearly (on an overall scale).
- Produce retail coupons (as an example) at the point of sale depending on the customer’s present and past purchases.
- Send personalized recommendations to mobile devices while customers are in the right area to take advantage of offers.

SMEs would be able to collect, collate and analyse the data in order to discover patterns and trends which would help them improve their customer service and provide new and more relevant and specific products and services for individual customers. In more general terms, SMEs invest in big data to understand trends within the market (both offline and online), serve customers to meet their specific requirements and needs and also to unlock corporate values such as company performance, customer service and accountability.

Another benefit that SMEs would experience with the use of big data analytics is the increased security protection that would become available. The vice-president at Splunk (a US data monitoring/analysing corporation), stated that SMEs can make great use of big data if it includes drilling down to the machine data layer: The website Microscope (2013) explain
how monitoring and understanding machine data can enable companies to identify and resolve IT and security issues with an accuracy and immediacy not previously possible. Clearly analysing big data can provide a number of benefits, with preventing security threats being just one major aspect.

III. WHAT IF BIG DATA IS CONTINUALLY IGNORED

If big data is ignored during the coming years then SMEs are almost certainly going to suffer in terms of financial loss and could eventually lose out on customers and sales to rival companies. These rivals could be offering similar services that are using big data analytics techniques and variability as a way of connecting ideas. However, it would still be important to mention that there are a number of enterprises who are already using big data analytics. An incredible 0.5% of the digital universe is currently being analysed by businesses, so clearly there is plenty of room for expansion in terms of companies to work on and take advantage of the potential available despite the current usage of big data.

Evans (2012) explains how companies without aggressive plans for managing and exploiting the explosion in data report losing an average of $71.2 million per year. Evans adds that almost all companies that were surveyed expect data volume to increase following a surge of high velocity data over the past two years (2010-2012).

A. What effects need to take place?

Again referring back to Evans’ report regarding companies ignoring big data, research found that 43% of companies need to improve their ability to manage information and analyse it effectively and to then put it to good use with the customer’s best interests in mind. To add to this, 38% said they needed to upgrade their system and software to be able to collect and analyse more information for more innovative decision making. The same number also believed that extra training and education within their own organisation was required to ensure the best chance of extracting the most insightful amounts of data flowing into its company (Evans, 2013).

Based on the findings revealed above, Oracle offered a number of points to be used by SMEs yet to invest in big data schemes which shows various opportunities available-

- Take Time to Strategise- Work with key stakeholders and business units to understand their data needs. Incorporate their feedback to improve processes across the business as well as finding out the key requirements before and during the process to ensure the final outcome is successful.
- Think Analytically- Improve your analytical support team and ensure managers have the applications and access they need to examine business-critical information first hand.
- Invest to Improve- Arm your organisation with the appropriate technology, staff, and systems/processes needed to optimize information for true business intelligence (Baum, n.d).

IV. DISADVANTAGES OF BIG DATA

Big data, despite having its positive impacting factors on businesses especially larger sized establishments, does of course come with its disadvantages which might hinder an SME at least to begin with.

A. Too expensive for many SMEs

Referring back to the report published on realbusiness.co.uk, SMEs will typically have limited resources as well as smaller budgets which of course is why they’re classified as an SME (Simons, 2013). This would be an initial problem with smaller companies perhaps struggling to invest due to financial struggles or lack of initial investment funds in a new start up business for example. To put this into perspective in terms of hardware pricing, a data warehouse which is the electronic storage of a large amount of information in a secure and easily manageable way and is often used for data mining to reveal patterns of information for analysis (Investopedia, 2010) will set a business back approximately $22.7 million (£13.6 million) for 500TB capacity according to a recent report by Winter, Gilbert and Davis (2013). Of course this is simply giving a guide to how expensive big data analytics techniques and method can potentially cost for large companies in particular and is likely to decrease in price along with an increase in available data capacity as time progresses. Even on the lower end of the scale, data analytic hardware/software such as Hadoop software framework storage can cost around £2400 per node per year which still might prove too pricey for some SMEs (Savitz, 2012). The EU’s classification of an SME’s turnover being no more than £11.2 million (Ec.europa.eu, n.d.) shows how many organisations would almost certainly struggle to invest in such high quality big data technologies.

B. Security Threats

Another issue is that if a big data project is managed poorly, it can lead to security and compliance problems which could potentially cause data breaches, compliance investigations as well as reputational damage for the company itself (Aveksa, 2012). This shows that the data must be protected for only the authorised people have access in order to prevent such threats and issues. However this of course could potentially require further investment which may prove to be too costly or simply too much of a hassle for many SMEs. (Due to high time consumption or a larger workforce may be required).

V. CONCLUSION

Based on the research conducted throughout this report it is noticeable that generally, the benefits of big data analytics outweigh the issues that it could potentially cause for SMEs. Not only will the financial rewards be clearly understood when implementing the new data techniques available, but it will soon become almost a necessity for all businesses including SMEs to introduce big data analytics into their own company for it to be able to grow healthily and sustainably. With such volume, velocity and variety of data constantly building up, it is vital that all companies become organised and use big data
analytics as a foundation for making profit and expanding their businesses into something more than an SME.

Advantages such as financial gain through advanced marketing opportunities including producing specific offers and opportunities for individual customers through the analysis of big data, possibly online through social media channels such as Facebook and Twitter, (Bowden, 2014) will inevitably come into effect. Security enhancements will begin to show once particular data is analysed and understood. Monitoring data will help identify IT and security issues quickly and accurately (Microscope, 2013) which would be particularly useful with the intense volume and velocity of big data in this day and age.

If these guidelines are not followed and taken into action by SMEs, they will be at risk at falling behind rival companies and potentially losing customers by failing to meet their specific needs and requirements. With only 0.5% of data currently being analysed (Evans, 2013), there is clearly plenty of improvement required by enterprises as well as potential.

Very little is perfect and big data analytics is no different. Various drawbacks such as hardware and software initial and maintenance costs can often be too expensive for many SMEs who simply cannot afford to invest for reasons such as having limited start-up money for example. Security issues such as data breaching and hacking from those who are not always authorised to access that particular system is also a potential drawback.

Research shows that big data is inevitably going to be a major part of the IT universe and SMEs will be foolish not to invest as early as possible or potentially face losing out especially on the financial side of a business. SMEs must learn lessons from big data analytics (certainly those already used within larger organisations) and then apply what has been learnt to their own small data to efficiently manage their business in the present day as well as preparing for the future.

VI. REFERENCES


Can SME’s Trust Big Data?

Youssef Gamgoum
University of Derby
Derby, England
y.gamgouml@unimail.derby.ac.uk

Abstract—Big data analytics has been around since for a while though it is not widely used by many companies. It is used to provide companies with accurate information, to allow them to spot trends and therefore make better business decisions and new innovations to improve their business. A major source of Big Data in recent years has been from social media, from hashtags, to users ‘liking’ a company’s page, organisations can now receive instant feedback from their customers which can save the company a lot of time and money.

Index Terms—SME’s, big data, trust, social media, costs.

I. INTRODUCTION TO BIG DATA

Big data is a vast collection of both structured and unstructured information. Usually big data is too large for any one tool to capture, store and analyse the information so companies need to hire analysts to use the big data to spot trends in their industry and products (Smestrategy.net, n.d.). These findings can be used to give great value to an organization helping to innovate products and services faster as they know can find trends in the industry to know what it wants. 98% of stored data in now digital (Devakunchari, 2014), increasing the size of databases at exponential rates. In the digital generation we are now in (Buckingham and Willett, 2006) unstructured information has become more important than structured data. Unstructured data is mostly received from social media, it can be used to spot trends and receive instant feedback from customers on a company’s product or service. Structured data shows more direct feedback on your organization, retrieving information from company sales, formal complaints, competitor’s innovation, etc.

Majority of organizations are not aware of big data analytics because big data can not be fully measured by one tool/software but instead it needs a dedicated team and a group of supporting technologies to properly analyze big data (Macinnes, 2013).

An example of big data analytics can be shown by an amusement park operator at Morey’s Piers in America. The amusement park works with an external analytics company that uses the electronic gate scanners at the start of every ride to see the popularity of each of their rides and what times the rides are most popular (Simons, 2013). This information is essential the Morey’s Piers management team.

The true definition of Big Data can be split into different dimensions. These dimensions are commonly known as the 3 V’s (Sas.com, 2014), though recent researchers now argue that there are many more than the original three V’s in Big Data.

II. DIMENSIONS OF BIG DATA

A. Volume

The volume of big data is huge, especially now we are in the digital generation (Buckingham and Willett, 2006). Devakunchari (2014) states that IBM believes that there are 2.5 quintillion bytes of data created everyday. Majority of this data will be unstructured, taken from social media.

B. Velocity

Data can now be retrieved at exponential rates. Sensors and RFID tags are examples of two tools that give real time data (Smestrategy.net, n.d.). Keeping tracking and manipulating this data to be analysed is a major challenge to most organisations.

C. Variety

Data comes in many different formats; structured numeric data in databases, text retrieved from social media, financial information, emails, video, audio and information from other applications (Smestrategy.net, n.d.).

D. Other V’s

Researchers are continuously adding extra V’s to define big data. Devakunchari (2014) believes that Veracity and Value are attributes of big data. Veracity decides whether the information stored is meaningful to the company or not, it takes into consideration biases and abnormality. Value measures the usefulness of the data. Sas.com (2014) adds Variability to the dimensions of big data. Variability sees if there is a fixed pattern in the data or not, it analyses things like trends of twitter (Sas.com, 2014).

III. SME’S AND BIG DATA

Companies with 250 employees or less are referred to as SME’s (Small and Medium Enterprises). Sas.com (2013) estimates that less than 0.2% of SME’s in the UK use big data analytics and a fifth of all business admitting that they have no or a very poor understanding of big data.
The need for SME’s to use big data is increasing rapidly; with technologies like social media growing in popularity faster than anyone could of imagined. Social media is being utilised by a number of SME’s to communicate and advertise to their customers but so much more information can be obtained to help the companies spot trends and make better business decisions. For such information to be analysed and used to improve the company, the demand for skilled employees and analytic tools is need (Passingham, 2013). Apart from the lack of knowledge on big data, SME’s also do not use big data analytics because of the cost of the skilled employees needed and the technologies. SME’s needed to evaluate the situation to see if it is worth it.

A. “SME’s don’t know that they do not know”

In Europe, the government administrators could save more than €100 billion just by using big data (Mckinsey.com, 2011). Though €100 billion is a lot more money than what SME’s can save using big data, it does show the power of growing data volumes – they don’t have big data” (Macinnes, 2013). Williams believes that big data is confined to big brand organisations, where they can obtain a lot of unstructured data about their organisations from social media and emails.

Dave Becerra (vice-president of strategy and business development at Roambi) SMEs are beginning to see big data as something more than just an enterprise trend. He believes some SME’s are starting to realise that they can identify trends, patterns and gain competitive advantage by harnessing the power of growing data volumes (Macinnes, 2013). Becerra does however caution that before rushing to implement a big data solution, SME’s need to take a step back and remember that bigger is not necessarily better. Analysing big data takes a specific team and resources, which can cost the company more money than they may gain.

B. Little Data

Campbell Williams (group and marketing director at Six Degrees Group) states that “There’s a simple reason why there’s no market in small and medium-sized enterprises for big data – they don’t have big data” (Macinnes, 2013). Williams believes that big data is confined to big brand organisations, where they can obtain a lot of unstructured data about their organisations from social media and emails.

IV. TRUST IN BIG DATA

When working with big data there are many issues that need to be addressed such as ethics, governance, security, provenance, audit, etc., but this report will focus on the trust issues when analysing big data.

As mentioned before, social media has become a primary source for unstructured big data. Social media can be used by SME’s to connect to their customers directly, receive instant feedback, spot trends and patterns in the market, observe their competition and more (Simons, 2013).

Figure 1 shows that over 2.9 million emails are sent every day, over 20 hours of video is uploaded to YouTube every minute and over there are over 50 million tweets posted per day. This shows the popularity of social media and popularity equals power.

1) Facebook

Unfortunately it has been reported that 5% of Twitter accounts are fake and over 7.3% of Facebook accounts are fake too (Business Insider, 2013). Facebook had reported 1.15 billion active users on 31st December 2012 (AllFacebook, 2013) but approximately 5% of those accounts are duplicates, 1.3% improperly classified by the users and 0.9% may have been fake accounts. That means approximately 83.95 million accounts are fake therefore are untrustworthy. These accounts are prone to spamming (AllFacebook, 2013).

2) Twitter

In Twitter’s S-1 filing in 2013 (Sec.gov, 2013), they reported that 5% of Twitter accounts are fake. That is 10.75 million users that are fake (D'onfro, 2013).

3) YouTube

The popularity of YouTube videos is determined by ‘views’ and the amount of subscribers there are to a YouTube channel. Against YouTube’s policies, views can be purchased by companies making a video seem more popular than it actually is. A company in the UK offers 10,000 views for £10 or 500 channel subscribers for £25 (Sparkes, 2014).

4) Conclusions

Though using social media is a great way to analyse big data to spot trends, SME’s need to be wary of the vast amount of fake accounts across all of the platforms. Spam from Facebook and Twitter and the ability to purchase views and subscribers to YouTube may show products and services in the industry and more popular than they actually are. This can badly affect business decisions made on these results.

B. Other Trust Issues

As the demand for big data analytics increases, 57% of recruiters say that filling the big data specialists roles are very hard to fill (Passingham, 2013) and of those specialists
recruited, the employees may not be fully trained in analytics. There is a big demand for the government and the educational system to teach the relevant skills needed in big data analytics.

V. FRAMEWORKS (RESPONSIBILITIES OF SME’S WITH BIG DATA)

For SME’s (as with larger companies), to effectively use Big Data to improve their business; tools, technology and skilled employees need to be employed. Using these assets together can dramatically improve a company’s business, as it allows them to spot trends and have an advantage on the competition. When undergoing these new developments, SME’s need to ensure that all operating elements work in sync to one and other. This is known as Strategic Alignment. This alignment then goes through a management process called Strategic Governance (Kvavik, 2004). Using both strategic alignment and strategic governance will allow the SME’s that want to implement Big Data analytics into their organisation to implement it in an organised manner with minimal risk.

The 3 V’s of Big Data (Volume, Velocity and Variety) show that the collection of Big Data is an ever-growing technique that is technology driven and to stay ahead of the competition, an organisation has to become dynamic and upgrade their systems and technology too (Deloitte, 2012). When implementing a new system in a company, there are many questions that need to be answered to allow the company to move forward successfully and achieve their goals like:

- What are the legal requirements?
- How to plan a new development?
- How to minimize risks

The answers to these questions (and many more) are then put together and become a set of rules for best practices in the organisation. Best practices are defined by a set of governing frameworks (itSMF, 2005).

These best practices (frameworks) need to be communicated to each and every employee to ensure everyone in the organisation know how to mitigate risks. If only the managers know the frameworks implemented by the company, then issues may occur (itSMF, 2005).

There are many different types of frameworks to choose from that can be used to make a set of best practices. The best company strategies do not just use all of one framework but instead uses the best parts of different frameworks to create a tailored set of best practices for the organisation (Gamgoun, 2014).

To develop Big Data analytics into an SME, the recommended frameworks to use is a combination of BS 25999, ISO 27002 and ISO 9000. BS 25999 is a set of practices for business continuity management, ISO 27K is a set of practices for information security management and ISO 9000 is best used for its quality management processes. Using the most relevant practices of each framework for the company will allow the information to be secure, trust worthy and reliable to ensure the business stays on top of trends and ahead of the competition.

A. BS 25999

BS 25999 was developed to effectively manage a Business Continuity Management System (BCMS). BS 25999 instructs organisations to:

- Define responsibilities to their employees
- Continually manage processes such as planning, implementation, performance assessments, management reviews and improvements.
- Develop business continuity plans

These practices are used to ensure that is a system does go down, then the company can still run and their information is not at risk for being either hacked or erased (BSi, 2007).

B. ISO 27K

ISO 27K has been built from the BS 7799 framework to provide a standard for implementing information security management.

Problem escalation is an important practise that should be used by most organisations. When a problem occurs there is a set hierarchy in which the right people need to be notified about the issue depending on their position and the size of the problem. Therefore if a system does go down then the employees will know which manager or technical support team to notify straight away which will reduce the time it will take to resolve the issue.

ISO 27002 also contains practises of business continuity. Like in BS 25999, the business continuity management practises is used to resolve interruptions in business activities and to minimise business risks to stop major failures or disasters.

C. ISO 9000

ISO 9000 is best used for its quality management principles. These principles ensure organisations are customer focused (ensure that the company is meeting the customer needs) and system approach to management. System approach to management allows the company to identify, understand and manage the different system processes in the organisation then bring that information together to retrieve the companies objects (BSi, 2005). Figure II shows the suggest cycle to continually improve the quality of a management system.

Figure 2. Continual improvement of the quality management system (BSi, 2005)
The issue with using set frameworks is that the practises do not work in every solution. SME’s must take the practices best suited to their organisations and tweak them to suit the company.

VI. CONCLUSIONS AND RECOMMENDATIONS

Big Data is an ever-growing collection of information that comes in many different shapes and sizes. The amount of companies that do not understand the importance of Big Data analytics is slowing decreasing yet SME’s think that they do not have enough data available to them for it to be important for their company, this is incorrect. Social media has given every human being a platform to voice their opinion. SME’s searching for comments made by their customers on social media sites will receive instant comments on their company and will be able to spot trends instantly, which will keep them ahead of the competition and will allow the company to improve for the better.

With the amount of fake accounts on social media platforms, companies need to be careful the information they collect. Thankfully there are now more tools and skilled employees that can analyse Big Data.

Information is very useful to any company, no matter on the size of the organisation. The data collected can be used to spot trends and receive instant feedback directly from their customers.

VII. REFERENCES


BSi. (2005). Quality management systems — Fundamentals and vocabulary


The Two Faces of Big Data
Can SMEs Trust the Accuracy of Big Data?

Thomas Holding
University of Derby
Derbyshire, UK
T.Holding1@unimail.derby.ac.uk

Abstract— The amount of digital content available on the Internet is causing SME owners to clamor for access, with experts claiming Big Data will increase profits and improve understanding of customers. This paper explores how Big Data is being used and highlights the caveats that should be acknowledged in order to avoid potentially damaging misinterpretations. Questions of accuracy and trust of Big Data will be raised as well as the pitfalls of mining data from social media.

Keywords - Accuracy, Big Data, Data Mining, SMEs, Veracity

I. INTRODUCTION

Without analysing and interpreting Big Data, what meaning does it have? Adrian Bridgwater (2012) observes that ‘Big data without analysis is, well, just big data’. Indeed, experts and academics are encouraging analysis of Big Data and promoting the benefits of its analysis. According to research published in the Harvard Business Review, ‘we can measure and therefore manage more precisely than ever before’ (McAfee and Brynjolfsson, 2012). Such is the potential, the chief economist for Google states ‘So what’s getting ubiquitous and cheap? Data. And what is complementary to data? Analysis. So my recommendation is to take lots of courses about how to manipulate and analyse data’ (Freakonomics.com, 2008). So SMEs should rush out and acquire as many qualifications on manipulating and analysing Big Data as you can find, right? Wrong. One could not be blamed for acting on such advice from the chief economist of Google, however if such advice is heeded, then great care should be taken when applying the newly obtained skills and technology to Big Data. Some experts are nervous about a foolhardy approach to the analysis of Big Data and warn about potentially seeing patterns and drawing conclusions, where no direct correlation exists. Furthermore, there is the risk of setting out to prove ones hypothesis and manipulating the data in order to do so.

II. INTERPRETING BIG DATA

To fully utilise Big Data, experts are encouraging Big Data analysis (Intuit, 2012). SME owners must err on the side of caution when acting on such advice without consideration of the consequences. There is merit in attaining qualifications in order to better analyse available data, however there are very few courses that are able teach the inherent ability to maintain objectivity and avoid creating inaccurate links, which can occur in order to serve a premeditated hypothesis. It is not uncommon – and this remains true across a host of disciplines– for conclusions to be drawn simply to prove oneself right, even if it is wholly unintended. An analyst could set out to be objective, but the predicted outcome or target which benignly lingers in the subconscious may directly cause an analyst to draw inaccurate conclusions (Anderson and Rainie, 2011).

A. Big Data Objectivity

It would be difficult for even the harshest of critics to deny that within Big Data there is a treasure trove of information ready to be unearthed. It’s no real stretch to say that Big Data is for data analysts what the Galápagos Islands were to Darwin and Wallace. However, the pitfalls of using such data without a commensurate level of objectivity and due care must be acknowledged in order to safeguard against correlation for causation, which result in both poor judgement and ill-informed decisions (McAfee and Brynjolfsson, 2012). Of course, there is also the risk that the analysis of Big Data could be turned to more nefarious usage. Big Data is fodder for those prone to apophenia, and could easily be exploited in order to reinforce biased opinions, an all too common occurrence within the political arena, particularly when a certain political faction seeks to project a ‘preferable’ vision of reality (Anderson and Rainie, 2011). One must also guard against complacency when working with Big Data, it could be quite easy to place your trust in the fact that, given the volume of data, you have an empirical vista of customers and market trends at your disposal. The reality, however, is that the quality of data is not immediately apparent. Greater diligence is required in order to ensure accurate interpretations (Boyd and Crawford, 2011).

B. Big Data Accuracy

Before passing judgement over whether Big Data should be used for analysis, one must also question both the veracity and validity of the data at hand. One simply cannot reach accurate and reliable conclusions unless they are born of data that is in
itself accurate and reliable. Posing a significant threat to assuring the veracity and validity of data is the issue of ‘noisy data’. Knoblock (2007) defines noisy data as ‘any kind of difference in the surface form of an electronic text from the intended, correct or original text’. In other words, any electronic text that has been, or indeed, could be interpreted differently from what it was intended. Noisy data occurs for many reasons, for example; data uncertainty (incorrect user input, incorrect interpretation of language or incorrect spelling) model uncertainty (models are approximate predictions such as weather patterns) and process uncertainty (processes can contain randomness such as delay on travel times) (Subramaniam, 2014). Taking data uncertainty as an example, how can one be certain that the data mined is not noisy data? Consider the many languages, the colloquialisms and the informal chat. Consider that some users are polyglot, and may switch between languages when texting or sending an instant message. Consider spelling mistakes and shortened words, acronyms and absent punctuation, then consider how when interpreting this data that veracity can be assured. Some industry experts argue that it’s a temporary issue, since data analytic tools are improving all the time (Lohr, 2012), but it’s somewhat naïve to believe that a computer could understand and interpret all this data without errors. In short; a computer can understand whatever the programmer has made it capable of understanding, but a programmer who could understand the many facets of electronic text and the context which underpins that text, would be an extremely rare commodity (Manovich, 2011).

C. The Rise of the Machines

There is a general feeling that Big Data analysis is a ready to go and refined solution to the issue of ensuring commerce success (Cheng and Qin et al. 2012). The claims of making more accurate predictions and better understanding consumer buying habits, consumer opinions, political movements, company strategies, marketing campaigns, which culminate in making shrewder decisions (Pang and Lee, 2008) would not fail to capture the imagination of any ambitious SME owner. The confident claims could be interpreted as arrogant, particularly when Chris Anderson, formerly the editor-in-chief of ‘WIRED’ magazine, claims that human behaviour theory, taxonomy, ontology and psychology can now be replaced by drawing correlation within numerical repetition (Ijsselsteijn, 2014). In essence, what the claims are saying is that there is no longer the need to speculate and try to understand why humans behave in the ways they do. Machines with the use of algorithms can now eliminate all the uncertainty surrounding human behaviour and do it in a much quicker and smarter way than ever before.

D. Sociologists Vs. Data Analysts

These claims, however, have caused great concern amongst the sociologist community. The brash statements of those championing Big Data analysis totally ignore the crucially important caveats, which need to be acknowledged to avoid spectacular errors that occur when integrating social science into computer systems (Graham, 2012). There is deep-seated knowledge of human behaviour that machines will never be able to grasp and the threat is that SME owners will naively trust the output of the computers, without truly understanding the fine nuances of why people behave in the way that they do (Anderson and Rainie, 2012). Using mathematical models and algorithms would offer helpful simplifications of the subject, but indeed there are limitations in understanding the wider picture (Lohr, 2012). Big Data can offer great benefits to those who use it correctly, but that does not render theory and contextual understanding redundant. This is acknowledged not only by sociologists, but also by experts in different fields such as player recruitment within football. Football scouts have a remarkable amount of data at their disposal, supplying them with every kind of information about a football player with a touch of a button (Nicholson, 2014), but even with all that data, it does not replace the human brain. It underpins it, it can fortify opinion, but it is no replacement for actually going out and watching the player first hand (Evans, 2014). Analysts need to start thinking outside the box and formulate opinions and decisions on wider and objective views (Anderson and Rainie, 2011). Currently there is too much reliance upon statistical interpretation based on (for example, but not exclusively) online searches. When data is requested, it will go through data cleaning processes to decide which data will be supplied to the end user, which of course is a subjective exercise and can be perceived as discriminatory and inaccurate (Bollier, 2010).

As mentioned, there are crucial caveats when relying heavily on Big Data analysis tools, but issues surrounding actual data analysts should also be highlighted. There is a rallying cry from experts and academic alike urging undergraduates and graduates to qualify as data scientists. They advise their audience to equip themselves with advanced technical skills, which will empower them to provide useful insights and decision-making-support through their understanding of statistical analysis and business issues (World Economic Forum, 2012). However, what is not being specifically promoted, is the need for in depth insight and knowledge into cultures, history and society, coupled with the aforementioned technical abilities. Never before has humanities been integrated with computer programming and furthermore, a stroke of a computer keyboard key will not produce an output influenced by two hundred years of humanities scholarly. On this basis, without incorporating humanities into Big Data analytics, the most one can expect from Big Data is patterns (Steadman, 2013).

III. SOCIAL-MEDIA MINING

At present, there are over two hundred million ‘tweets’ posted each day, from over six hundred million Twitter accounts and over one billion Facebook accounts (Facebook, 2013), so it should come as little surprise that SME owners are being encouraged to turn to social media mining as a source of their Big Data (Salampasis and Pultoglu et al. 2014). The issue remains, however, that there are lingering doubts regarding the veracity, validity and variety of the data that can be mined from social media.
A. Comparing Large Datasets

One prominent issue occurs when analysts apply the same analytic model across data sets; making comparisons and drawing conclusions, believing that the data is homogeneous. Applying the same assessment and forecasting analysis across data sets, for example across both Twitter and Facebook, can result in a loss of intelligible meaning, and in turn, the veracity and validity of the analysis. To illustrate this point, let me provide you with an example; for instance, one might well wish to perform analysis on articulated networks (network defined by the contacts that appear on a mediating technology – e.g. social networking via the Internet or a contacts list on a mobile phone (Boyd, 2004)), and so, for the purpose of comparison, would regard ‘followers’ on Twitter as the same as ‘friends’ on Facebook. This is where the analytic model breaks down, the comparison loses validity, and the conclusions drawn are lacking in veracity, for the articulated network from a Twitter account has a totally different meaning to an articulated network from a Facebook account (Boyd and Crawford, 2011). The motivations for following someone on Twitter may radically differ from the motivating factors for adding a Facebook user as a friend, drawing parallels between and combining the two data sets would be tantamount to comparing an apple to a potato.

Analysis can also be conducted on behavioural networks (networks resulting from social media exchanges and communication patterns (Meiss and Menczer et al. 2008)). Analysts may draw conclusions from the tags attached to a particular photograph on Facebook, the caveat being that if one user gets ‘tagged’ on another’s photo, does that mean they are friends? That they talk to one another on a regular basis? Or they have significant ties or common interests? Take ‘photo-bombing’ for instance, the craze of suddenly appearing in a camera’s field of view just as the photograph is taken. If a ‘photo-bomber’ is tagged, the only conclusion that could be made is that they have been in the same location, or at the very least, at that particular moment. It does not mean they shop in the same places, have the same interests, or even mix in the same circles; there would be very little- if any - value in that piece of data (Boyd and Crawford, 2011). The exercise of seeing patterns in numbers dismisses the reality and complexity of an actual social relationship, which fails short of the aim of the analysis: for SME owners to better understand their consumers. SMEs need to ensure they analyse the data within the context that it is intended and not assume that every bit of Big Data mined from a social networking site is equivalent and interchangeable or that the more a user contacts another users determines how strong their relationship is. Furthermore, a digital social network paradigm should not be mistaken for the same paradigm that befits the physical social network (Boyd and Crawford, 2011).

B. Misleading Data

SMEs may see the value of mining data from social networking sites, indeed they boast hundreds of millions of accounts (StatisticBrain, 2014). However, what does ‘account’ or ‘user’ actually mean? Are there assurances that the ‘user’ is contributing to the social networking site or are they just observing what others say? What assurances are there that the accounts are human and not ‘bots’, which are programmed to impersonate humans for marketing and political purposes (Rouse, 2013)? Twitter has a phenomenal number of users, but which of these users actually participate and ‘tweet’ (a message posted to Twitter (Twitter, 2013)) as opposed to simply listening to what their favourite movie star is tweeting. Very little value can be gained from mining data from Twitter users that do not themselves voice opinions. It is reported that forty per cent of Twitter users are indeed listeners (Boyd and Crawford, 2011), meaning there is at least a two in five chance the data mined will be from a non-participating user. Additionally, over eighty three million Facebook accounts (Protalinski, 2012) and twenty four per cent of tweets (Cheng and Evans, 2009) are generated by bots. SME owners have to question the validity and veracity of the data mined from social networking sites when considering such stats. A fortune could be spent on data that would have zero value to the business.

Social networking sites take measures to prevent bots from creating accounts, however research undertaken showed that Facebook was able to defend against only twenty per cent of bogue accounts (Cohen, 2011). Furthermore, after an eight week period over three thousand friend requests were accepted taking the extended neighbourhood of friends to over one hundred thousand friends (Cohen, 2011). The damage to the validity of ones data by bots does not stop there. A study undertaken by a start up company showed that eighty per cent of advertisements that were clicked were done so by a bot, each click charged for by Facebook, a by-product being that one is paying for advertising that has literally no benefit to oneself (Jackson, 2012).

Let’s not even consider the number of bots that occupy the social networking sites and just focus on the sample of people that are represented. One should consider that five per cent of Twitter accounts generate seventy five per cent of tweets (Cheng and Evans, 2009), thus when mining the data from Twitter, one will more than likely retrieve Tweets from only five per cent of Twitter users. The variety of data is issued a further blow when delving deeper into those five per cent of Twitter users. Over sixty per cent of the most active users were American, with just under seven per cent being from the UK. One should also contemplate that eighty per cent of overall Twitter users are aged between fifteen and twenty nine, with thirty one per cent being between the ages of fifteen and nineteen (Cheng and Evans, 2009). If one is considering how mining social networking sites can help their SME, then the caveat about sample size cannot be ignored. If there were guarantees and control over the sample of data which allowed one to retrieve data from their chosen demographic, then some concerns could be relived. However, Twitter’s ‘Gardenhose’ access only allows access to ten per cent of ‘public’ tweets (Boyd and Crawford, 2011). Who is to say which section of public tweets one is given access to? This lack of sample control coupled with no access to protected tweets paints an entirely different picture for SME owners.
IV. CONCLUSION

There can be no doubts regarding the potential of Big Data. There are eye-catching headlines that tell the tale of the two businesses; the business’ who use Big Data who are top performers and the business’ that do not use big data who are lower performers (Lavalle and Lesser et al. 2014). Along with this however, are the caveats that highlight the need for further research into the veracity, validity and variety of Big Data analytics and Big Data that has been mined from social networking sites. There is unrest when Big Data analytics is being touted as the replacement for human insight and rightly so. The validity of Big Data mined from the Internet is also questionable, as even if one is to avoid the bots and gain great control over the sample of Big Data, how one expresses oneself online and presents oneself to others, which is habitually carefully managed, is often not a direct route into the individuals’ actual life, but a makeup of how one would like to be perceived (Ellison and Heino et al. 2006).

I do believe however, that progress can be made through a marriage of the two, resulting in a hybrid paradigm. There are questions that Big Data Analytics can answer which a sociologist simply cannot. Equally, there are questions that a sociologist can answer, which analytics cannot. Why do we have to replace one with the other? In an ideal system, we would combine sociologist insight with Big Data analytics to produce a more powerful and accurate means of helping not only business’ of all sizes, but also local, national and international governments. As of yet, this idyllic solution is some way off and a coming together of minds, from several pools of experts, needs to take place before the objective can be achieved.

V. REFERENCES


SME’s Securing Big Data

The Attractive Target

Ben Johnson
University of Derby
Derby, UK
B.Johnson4@unimail.derby.ac.uk

Abstract— Big Data is an attractive target for potential attackers who want to harvest large amounts of data efficiently. SME’s typically do not have the security procedures and practices in place that large enterprise can support and implement to add layers of security. This means they are possibly low hanging fruit for a motivated attacker, who may want to steal their data for identify theft, fraud or even corporate espionage. This paper will examine procedures and practices an organisation can implement.

Index Terms—Big, Data, SMEs, Business, Threats, Security.

I. INTRODUCTION

It may be assumed that SME’s are too small to be actively attacked, but based on the fact during 2012, 87% of SME’s taking part in a study identified a security breach (Department for Business Innovation & Skills, 2013) this assumption would be incorrect. This and the 11% growth from the previous year’s results can be seen in Figure 1. This high, growing percentage demonstrates this is genuine risk factor to all SME’s and in order to protect intellectual property and customer’s data, security needs to be taken seriously by all levels of organisations.

Figure 3 – Responses from a survey asking how many respondents experienced an incident (Department for Business Innovation & Skills, 2013)

Big data analytics is becoming increasingly important for businesses to use and influence decision making processes. SME’s are starting to overcome some of the barriers to accessing this type of analytics that large enterprises have been using to influence the business. When SME’s can use big data analytics to examine customer-buying behaviour, it has been shown to have positive impact on the businesses marketing and strategic decision-making process (Donnelly et al. 2013).

In a study conducted by Verizon investigating various breaches they ranked the exploits used in the initial compromise by difficulty. Only 1% were categorised as high and 78% were categorised as being low or very low (Verizon, 2013). This study suggests that the majority of the exploits being used by attackers are not especially sophisticated new attack vectors. Example of which would be a 0-day exploits, which exploit previously unknown vulnerabilities and therefore are highly valued (Bilge, Dumitras 2012).

However as the Verizon study suggests; initial compromise is not typically being gained by 0-day level exploits, but the rather more likely publically known exploits. The majority of which would be preventable by good security practises and procedures, which will be discussed in this paper.

II. BIG DATA

Big data is typically associated by what is known as the three V’s: Volume, Velocity, and Variety. These were the suggested areas of focus in regards to developing data management, by the analyst Doug Laney in 2001(Laney, 2001).

The first of the set, volume is based around the idea there can be increasingly growing quantity of data to collect and stored, so storage and what data is actually collected needs to be managed effectively.

Velocity is based around the fact valuable data can be transported at various speeds and collection needs to be able to collect and process as fast as possible to keep up. The quicker data can be processed, the quicker it can be analysed and the earlier a business has the information it needs.
Variety is about the variance in the structure and format of data that could be collected. To make best use of all the information and analyse it to its best potential means the variation between structures is abridged, so the data set is as extensive as possible when analysed.

Using big data analytics can have huge positive impacts on a business and this is why it can be an attractive tool. The McKinsey Global Institute estimate a business effectively utilising big data can increase their operating margin by more than 60% (Manyika et al. 2011). That is a very significant percentage increase and if anywhere near achievable, as a result it would be hard for any business to not seriously consider using big data analytics.

Growth in big data also has created a new market of opportunities for big data consultancy firms, which include SME’s. This can mean large enterprises are outsourcing implementation of big data analytic solutions to smaller companies (Clawson, 2014).

However, big data can also present many disadvantages and complications. One such example is the retailer Target being able to identify which of its customer are potentially pregnant and then target them with baby related discounts. This particular use case led to an incident with an angry father wondering why his daughter was being sent such discounts, however he later found out she was pregnant (Duhigg, 2012). Target did realise they can be too effective at targeting individuals with marketing, which could unsettle customers. As a result they mix targeted ads with other ads to make it seem more generalised (Duhigg, 2012).

III. VULNERABILITIES

Using big data analytics means collecting and using large amounts of data, typically on customers or a target audience. This means companies are potentially collecting new levels of personal information and storing it, which previously would have disregarded. Stolen personal information is sold regularly in the black market and there are plenty of places to buy and sell this sort of information (Krebs, 2013). This means big data could be an attractive target for an attacker, either to use for identify theft/fraud or to sell for profit.

As mentioned previously a high percentage of SME’s have experienced a breach in the last year (BIS, 2013). This seems likely because they do not have the resources and knowledge that a potentially larger company has in regards to information security. Therefore this would likely make them an easier target for an adversary to attack and extract data from.

This makes the combination of SME’s and securing big data a concerning prospective and raises important questions in regards to what risks have been considered and mitigated.

Data from large corporations can also be passed onto SME’s and so are the security risks. As big data analytics is a growing market, so is the sale of the data from business to business. An example of such is the sale of supermarket chain Tesco’s loyalty card data, to various SME’s (Donnelly et al. 2013). While it is unclear if this information would contain personal details on customers, it does demonstrate this type of big data is sold and distributed among companies. While the security practises of the original organisation may be sufficient, this data can be potentially sold and distributed to where it can then be accessed and stolen.

Another risk factor is SME’s and potential bad security practises providing the entry point for an attacker to then attack a large organisation. A possible example of this is the recent Target breach where 40 million credit and debit cards were stolen. The attackers are understood to have first accessed the network via an SME’s contractor network login (Krebs 2014). While it is publically unknown how an attacker acquired these login credentials, it poses the question would following good security practices have prevented those credentials being stolen.

This story also highlights important factors for important need for network security monitoring (NSM), which is the practise of collecting and analysing data that can potentially indicate compromise or intrusion (Bejtlich, 2013). In the Target breach from the dates provided by the company (Target, 2013), customer data was stolen during an 18-day period from November to December 2013. This is a long time period to have not been able to respond to any of the potential indications, such as the initial breach, unusual network activity or the data exfiltration of 40 million credit and debit cards and up to 70 million personal details (Target, 2014). The amount of data exfiltrated must have created significant outbound traffic that should have indicated unusual network activity, even if this traffic was encrypted to conceal its contents. This raises questions on the effectiveness of Target’s implementation of NSM and if they could have done more in this layer of security, which could have limited the scale and length of this breach.

Another important factor this case study highlights is the problems with current security compliance standards. Target was at the time of the breach PCI DSS approved (Riley et al. 2014), meaning they were found to follow the necessary compliance standards for storing cardholder information. However this attack still led to an incredibly large amount of credit/debit cards and personal information being stolen. While it could have been problems in the implementation, it does highlight the potential this standard is not effective at securing card holder information, which is its intended purpose. This could potentially create a false sense of security to compliant organisations and their customer’s data, when they could and should likely be doing more.

Organisations use risk management as a factor in multiple decision making processes. This generally includes information security and a basic formula for this process given by Zalewski (2012):

“\[\text{Risk} = \text{Probability of an Event} \times \text{Maximum Loss}\]”

This formula means that if the potential risk is judged as low by the impact and likelihood of an event, then the steps taken to mitigate this could then receive lower priority or funding. This can also mean the risk level is deemed acceptable and the potential loss that could be incurred is accepted rather than attempt to lessen this risk further. This
method does make financial sense, an example being if the extra security is required to lessen a low risk/cost scenario is very expensive; the cost will potentially out way the projected loss.

The problem with this method is the potential openness for interpretation and reliance on predictions, typically by management. In a computer network, doing risk assessment on an asset basis could identify specific entities as very low risk, but in the whole picture this could be an entry point for a whole network compromise (Zalewski, 2012). This shows there is high dependency on the technical knowledge and the interpretation of those conducting the risk assessment. This interpretation of risk assessments is an important factor as part of the ISO 27001 standard. It is up to the organisation how and what level of risk it assigns while becoming compliant, this means it could reach ISO 27001 compliance and still be vulnerable (BCS, 2009).

IV. MITIGATIONS

One of the biggest hurdles to overcome in relation to successful security is knowledge and communication. It is important that the relevant parties in an organisation can collectively identify security risks and what security implementations and practises will need to be followed to mitigate them on a regular basis. This requires proactive discussion efforts between management and security employees. This ensures management can make decisions related to security budgets and implementations, with the required technical knowledge to understand the issues.

Another factor is the resources required to implement the multiple layers of security that are essential for good security practise. This could present substantial costs to an organisation but it needs to be considered against the potential costs and loss of business that a public breach could incur. One way some costs can be potentially saved is with the use of open source security software, which can be used to gain compliance while saving on the costs of licensing for closed source software (Whaley, 2010). However this may not be as a straightforward choice as it may seem, as it still needs to be implemented and supported by someone. If this would present additional problems and costs, it could be deemed by the business that a closed source product or outsourcing completely is a more attractive solution.

While exploit and incident prevention are key to a good security model, it is important to remember good incident response strategy is vital. The reasoning for this is excellently highlighted by Kral (2011):

"An incident is a matter of when, not if, a compromise or violation of an organization's security will happen."

This means it is important an incident response plan is created and ready to be followed as soon as a potential breach is indicated. Therefore it is also important some form of NSM is implemented to alert about potential unusual activity at the time or as soon as possible after a breach as occurred. This reactionary effort is so that the scale of the breach can be restricted and reduce/eliminate the amount of data that could be stolen.

As discussed earlier, an organisation who wants good security practises should not limit themselves to only meeting compliance standards. While these are potentially helpful initial guidelines, as the Target breach and multiple others (Robertson, 2014) demonstrate; PCI compliance does not prevent security breaches. It is not alone in this respect, a business can also be ISO 27001 compliant but have bad security practise (BCS, 2009). If these compliance standards are not preventing multiple breaches at different organisations, they need to be improved to protect consumer’s data and not given organisations false senses of good security practise.

V. VALIDATE

Even in organisations that believe they have implemented the best security practise available, it is important validate the security level of the network. This means ensuring regular penetration tests, to test against old and new real world attacks and identify potential security weakness or entry points (Kennedy et al. 2011). This is important as there can be things that organisation are unaware they are vulnerable too or even things they think they are secure against which they are not in a real world test.

A black box penetration test is typically where management arrange a test without informing the employees responsible for detecting and responding to an incident (Engebreson, 2010). This then allows for an as real life scenario as possible, with an arranged test but is still unexpected to certain employees. This then tests for not only vulnerabilities but incident detection and response method as if it was a completely external attack. A lot can be gained from such exercises and testing how it the procedures and practises actually work, which can then be tuned and refined as necessary. It also gives the relevant employees vital experience and keeps them reactive to any potential threats.

VI. CONCLUSION

There are many things organisations of any size, such as SME’s can do to take security seriously and implement good security practises. As we have seen SME’s are attractive targets and big data is becoming increasingly important for organisations of all sizes. This means companies are carrying more and more data, including more customer personal information. This type of personal information is in demand, more and more data, including more customer personal information. Which motivates attackers to obtain and then sell this information on the black market.

This means organisations need to take good security practise seriously not only protect customers, but also protect against financial loss and loss of business a breach could cause. This includes going further than minimum security standards in popular compliance standards, as they are not preventing multiple organisations getting breached.

It also includes good communication between management and employees responsible for the security, to ensure they are clear understanding of issues and the security implementation is of an appropriate level for both sides. This especially
applies in terms of any risk assessments done in relation to information security risk management, as this is an influential step in the security of an organisation.

Appropriate resources also need to be implemented and used to achieve multiple layers of security in relation to a good security practise model. This can obviously incur costs for an organisation and resources need to be evaluated based on this. Multiple options do exist from open source software to outsourcing do provide a multitude of solutions, which provide a variety of options in relation to costs.

Incident detection and response is an important part of good security practise, as it is not realistic to expect to never experience a breach. By ensuring these are in place in advance, it allows for faster reacting response to an incident which therefore can reduce the scale and the impact on the business.

VII. REFERENCES


Abstract: The purpose of this report is to introduce Big Data to small to medium enterprises (SMEs), how it can be used to the advantage of the company but also to list the potential pitfalls of having data stored on the cloud format, based on its location and with the laws and regulations that follow this. This also takes into consideration the Data protection Act 1998, EU Data Protection Directive 95/46/EC and the American Patriot Act.

I. INTRODUCTION

Big data is a buzz word currently being used to describe massive data that is no longer able to be stored on a company’s database or enterprise system. The above statement sounds great, but how does this effect small to medium enterprises (SMEs) and do they even need it? Big data means data that can be analysed in order to find out information useful to the company using this, for instance the company Sweaty Betty used big data analytics to find out where to open a new store. A theme park in America Morey’s Piers, used it to find out which rides were the most popular and thus most profitable, they then changed the online packages based on that information. A restaurant in Boston, America called Farmstead Table used it to find out which dishes were popular then focused campaigns to individuals who ate them, this increased not only profit but also customer loyalty. (Simons, 2013) (MacInnes, 2014)

The cases above prove that big data can be used by SMEs to provide important information, the question is how this works with your own company and what the costs involved are. There is an argument that SMEs don’t need a big data solution that existing customer relationship management (CRM) or Enterprise retail planning (ERP) software would work just as well. This may be true for some but not all SMEs need all the data they have analysed and this is where online solutions come in. They provide the ability to big data analytics without having to install a full big data solution, which will save costs but also provide the tools needed with the flexibility of an on demand solution. (Simons, 2013) (MacInnes, 2014)

The purpose of this report is to make SMEs aware of the potential issues when considering the choice of a cloud based storage solution for the company’s big data and the issues it may face if a security breach occurs.

Today’s data requirements are increasing and the cost of storing all the data a company produces is becoming a challenge for even the smallest of companies. More and more services are moving to an online presence with no central location. The world in terms of data storage is now vast and easily accessible. The need for big data analytics is crucial to more and more businesses because data is no longer collected over a period of time and just stored and unused, but is now central to business growth and success. Being able to understand customers from a variety of sources saved on the company’s database provides a greater understanding into how customers tick. Consequently data analysts are increasing as companies try to gain an edge over their competitors. (Simons, 2013)

With larger amounts of data and storage locations being off site and in the cloud comes the issues of security and how best to make the data stored on these systems safe. With all this comes the cost of upgrading security, tackling data breaches or both. This report therefore also aims to give the SME Company an insight into potential pitfalls and areas that will require more research in order to be successful at big data security. (Schnittl, 2013)

II. LAWS GOVERNING DATA LOCATIONS

The issues concerning the location of where your data is stored isn’t just one of access, speed and price but also involves government regulations and how data breaches are handled.

A. Data Protection Act 1998

In the United Kingdom (UK) the rules and regulations governing personal data is covered by the Data Protection Act 1998 (DPA), cloud storage didn’t exist when the DPA was created but the rules still cover the use of cloud storage. Cloud storage providers (CSP) offer a range of services both private and public which are available in various price brackets. (Roberts, 2011)

Private cloud providers often offer control over the location and how customer’s data is handled giving greater control. (Roberts, 2011) (Widmer, 2009)

Public Cloud providers offer less in the way of control as this is managed by the providers but as with private cloud providers the responsibility of the data’s location and how its stored is for the customers. (Roberts, 2011) (Widmer, 2009)

The problem with cloud storage is knowing where the data is kept, as with most providers their services are often
outsourced out to another providers who again could do the same. Also providers often change who they use because of price, size etc. causing problem when trying to confirm you data’s location and if where it’s stored complies with both DPA and EU regulations. (Roberts, 2011) (Widmer, 2009)

B. ISO/IEC

With ISO/27017 and 27018 both cover cloud computing and will provide guidelines and a standard companies can work to in order to provide security for both the cloud provider and the customer, this is not a duplicate of ISO/IEC 27002 but an updated version, both versions are currently due for release either this year or in 2015. (ISO 27001 Security, 2014) (ISO 27001 Security, 2014)

C. The EU

The UK also falls under the European Union (EU), the laws of the EU have to be adhered to regardless of where the data is stored. This is because the data stored will no doubt contain information on customers or individuals in the UK, and so this comes under the EU’s jurisdiction. (Gorge, 2012)

The EU Data Protection Directive 95/46/EC, is there to protect all the citizens of the EU with regard to their personal data and follows all the key areas contained in an article from the European convention on human rights. This makes the company liable if it should be lost, stolen or deleted. This also includes protection surrounding a citizen’s personal data leaving the EU, which in short means company’s need to follow the guidelines or legal action can and will be taken against them by the government in whichever country the companies location is based. In the case of this report it is the UK government with guidance from the EU, on how best to resolve the issues. Although these rules and regulations aren’t actually based directly on big data, anything that contains personal data falls under this jurisdiction making that issue relevant. (Rouse, 2008)

D. Safe Harbor

As with most company’s today, a lot of the cloud based systems are currently located in the United States of America (USA), and with the competitive pricing and availability they are an attractive option. Companies may think that by using a country outside of the EU they will not be liable to the EU laws however this is untrue as long as the data stored contains information about EU citizens. As the USA are not part of the EU, using an American cloud based company based outside the EU the company using these cloud solutions is breaking the EU Data Protection Directive and the company becomes liable. (Gorge, 2012)

The EU however has also realized the dilemma faced by companies and so has created what is known as the Safe Harbor arrangement. This optional arrangement requires American companies who offer data storage to EU companies the option to sign up to the agreement and follow a list of regulations. That in turn then allows EU companies to use their services without breaking EU laws, and although the agreement is optional once the American company has entered into the arrangement the rules become binding, meaning they will also face legal action if broken. In order to find out which companies are part of this agreement the EU Company can visit the USA’s department of commerce website (http://export.gov/safeharbor) for a current list of compliant American companies. (European Commission, 2012) This agreement also falls under the DPA under principle 8 which allows for the safe harbor agreement to fulfil the DPA regulations. Although the safe Harbor is there to follow EU guidelines the data stored on American cloud servers is subject to the Patriot Act, this allows the U.S Government access to any and all data stored in this cloud server in direct conflict with both DPA and EU laws.

E. Patriot Act

The American patriot act allows access to all EU and UK data which is stored on an American server, which is something that needs to be addressed when looking into American cloud services.

F. Microsoft and the Patriot Act

The patriot act came into force after the 911 attacks that gives access to all online data to government agencies should they need to look at it. A recent issue has come about with Microsoft and their Hotmail email system, The EU email system storage is based in Ireland and as such doesn’t cover American jurisdiction, but a court order Microsoft to give over the information on the grounds that cloud data isn’t subject to international location laws and the company in question is American based. This all comes from a search warrant issued in America in which Microsoft took legal action to refuse, this is still ongoing. (Department of Justice, 2014) (BBC, 2014)

This coincides with the EU reworking the current Data Protection Directive 95/46/EC into a regulation that would stop this from happening. Currently the regulation is looking to come into force in 2017. (Ashford, 2014)

III. PHYSICAL DATA LOCATIONS

A. Redundancy

Once the issue of finding a company that follows the EU rules is resolved, the issue of redundancy becomes the next concern. Redundancy allows a cloud based system maximum uptime by creating more than one backup version of the data stored. This in short means if one backup fails, then the second backup immediately takes over allowing for continual use. Redundant server locations allow for the data to be duplicated in multiple locations, so if the data in one location is damaged for any reason such as power outage, fire or hardware problems, this isn’t felt by the company allowing for data retention to remain. Although this is very practical and a disaster proof option, it in turn causes its own problems when faced with locations. The EU legislation covers all copies of the data in all locations and the potential for data loss through theft is the same whichever backup is compromised. (Meijer, 2012)

The company using the system needs to have a full record of all locations in which their data is stored. This isn’t always easy to find out because each backup may be spread across
multiple locations in smaller sizes unbeknown to the company, and multiple versions of the backups may still be active. Knowing where this data is and knowing the level of security for all backups is vital as the company is still liable regardless of which version is accessed. Although some companies such as Microsoft can guarantee the location of the data stored and give a choice of which data centre, not all companies can do the same, as some cloud companies will probably be using another third party company in which to store the data. (Meijer, 2012)

IV. DATA BREACHES

A. New and Old Storage solutions

Data breaches are increasing year on year and while countries and their governments try to get a handle on this, because of the nature of technology and it’s constantly changing nature, this is an uphill battle. Traditionally, before computers became the main way to store vast amounts of information, it was stored on paper in filing cabinets in large rooms that were under lock and key. To gain access to this information meant a physical presence and a limit on the amount of information that could be stolen due to its physical size. Moving forward to 2014 and data can be accessed anywhere in the world at any time of the day or night. Vast amounts of information can also be stolen and stored onto a device that can essentially fit in a person’s pocket, or in the same way the information is stored virtually on another server, with its location anywhere in the world. (Hamel, 2014)

B. Target

Big data doesn’t relate to physical size in terms of something that can be held, so this makes it very vulnerable to attacks. Big data for many companies contains a rich source of information that is often accumulated over a period of time. An example was given from the recent press releases regarding the company Target, where they were able to determine when women are in their second trimester of pregnancy. This allowed them to focus promotions on these specific groups rather than generic promotions sent to all its customers. (Hill, 2012) Although depending on what data had been used, i.e. if it was just based on till receipts that would be find, but if using data that was bought from an external source was used, then this would break the UK DPA, so it’s easy to see the importance of the protection of this data.

C. UK Data Report 2013

A recent survey conducted on behalf of the UK governments Department for Business Innovation and Skills released for 2013, concluded up to 80% of UK companies outsourced some form of business process to external providers over the internet. That is a huge figure which has increased by 73% from the previous year. Although the figure does seem alarming in terms of scale, it’s important to know 55% of these services are for email and websites. This figure is based on SME’s and other services are growing year on year. In contrast large companies outsource just 15% of email and websites, but use outsourcing for other areas such as big data. (PWC, 2013) (Cooper, 2013)

One thing all size companies have in common is that the data stored is confidential, meaning all companies face similar challenges when securing data. Investment in security is growing year on year, with companies suffering security breaches spending on average less than those that don’t suffer. But is this a false economy? What are the costs to the companies for these breaches? The average cost of a large company’s security breach in 2013 was between, £450k – £850k and for a small company £35k – £65k, although this may not seem a huge amount this is only based on the worst single breach, and with 93% of large companies and 87% of small companies suffering a security breach in 2013, this is a worrying figure. This figure is increasing year on year, and it's no longer an issue just for large companies. Most SME’s will suffer more with the financial cost compared to a large company making this a serious concern. (PWC, 2013) (PWC, 2013) (Cooper, 2013)

V. CONCLUSION

Although big data may be the latest buzz word in the Information Technology (IT) industry, it is here to stay as more and more companies start to look into ways to make use of all the data that has been, or can be, collected to improve marketing and sales. Big data isn’t just for one particular industry, but covers a whole range, from Healthcare, Financial, Manufacturing and Retail etc. Big data provides real potential and as seen with Target’s pregnancy campaign it can be extremely effective and accurate.

There are also many pitfalls not just in terms of security, but also in terms of data protection in how companies use their data.

Location of big data is a clear issue here as the UK is part of the EU in terms of its laws on data protection. This means companies need to be aware of the structure and how to conform to it. All data saved in the UK must adhere to the EU laws, meaning the task of saving companies data needs to be well thought out. Although the Safe Harbor arrangement isn’t standard over all American companies the Safe Harbor list needs to be addressed before any final decision is made as the company not being on the list doesn’t mean the data isn’t secure but rather more of a legal issue with the EU.

Once the issue of location is resolved then comes the issue of redundancy, deletion, and security. Redundancy in the remotely stored data is vital to continuing use and disaster recovery in case the worst should happen. This however comes with its own issues as a company needs the ability to know where all sources of data are stored and comes back to EU law. As with most data stored virtually today, it’s split onto multiple servers with one backup being spread across servers and locations. This all provides the company with a potential nightmare when the need to delete the data arises. Knowing the data can be deleted without leaving a trace is vital as parts of data can be subject to breaches.

This all leads onto security not just on the data itself but also with regard to the company’s ethics in terms of staff
training and policy creation. Having staff understanding the issues faced can be a major contributor in term of security and reducing the potential cost of data breaches. Once these issues have been addressed, constant effective monitoring should ensure conformity to the EU with potential cost issues made minimal.

However is the answer to keeping safe and private data a case of using EU or UK only providers with local storage locations?

VI. REFERENCES


Abstract—there is a significant challenge that is associated with data being so distributed, 'security'. Mike Ferguson in a white paper for IBM 2013 says despite this challenge, there is yet another wave which has come up, which is big data. The question remains, that already there is security crisis in managing data in a distributed computing environment, what then is the impact of data on the enterprise. This paper aims to warn SME’s in financial sectors of the forthcoming problems to their business and the already present security issues being faced by larger organisations who have already totally embraced big data. The world at the moment is at the point where data is heavily fractured thereby making it harder to manage, which is particularly the case for security where very crucial data needs to be protected.

Index Terms—Big Data, Information security

I. INTRODUCTION

The big data era is definitely upon us, evidenced by the amount of data that is being generated, collected and analysed and also given the way data decision making is fast taking over across society. (Dong and Srivastava, 2013) This has given rise to databases building up at very fast rates, and access to this data is through various means and devices, including mobile devices. Given this distributed aspect in data, the most important issue is privacy and security. There is too much personal data out there and linked with other information from various sources could be detrimental (Patel et al, 2012). It is important to understand the security challenges that come with big data and understanding that it is valuable partly because it is often about real people, as Jewett (2014) puts it, it is wise to mind your manners when handling it. Unless the challenges are understood and dealt with, Jordan (2013) emphasizes that we risk turning data that has potential to enhance our business into paralyzing ideology.

II. BACKGROUND

When data sets grow beyond the ability to manage, in terms of acquisition, storage, searching, sharing, analytics and visualization, it is referred to as big data. (Ohlhorst, 2013) However instead of this being a problem, the release and coming of analytics and research has brought a completely different perspective and it has become apparent that more is even better. That is to say, more data means more analysis and more results. A number of large organisations have seen the importance of big data with companies in the industries of pharmaceuticals and energy companies seeing valuable results in drug testing and geophysical analysis. Contrary to this, in a report by Jordan,(2013) he seems to think that there is a possibility of overlooking the challenges that big data poses, in our rush to embrace the benefits of big data. These challenges may include the way companies interpret the information, manage the politics of data and find the necessary talent to make sense of the flood of new information.

III. THE VALUE OF BIG DATA

Once an organisation has established to pursue big data, it is in their best interest to start thinking about the attributes and characteristics of big data i.e. the V’s in big data. So often big data is described and given value by these V’s, which according to Kawaja (2014) were originally 3 V’s but have now evolved to encompass more V’s. These include velocity, volume, variety, variability, value and veracity. In understanding the V’s in connection with big data, they explain and define the value in big data. In relation to security, 4 of the characters seem to matter, which shall be explained below.

A. Volume

Big volumes of data are being generated and stored to the extent that existing traditional systems are failing to handle volumes of data. Every mouse click, phone call, like, text message or search on the internet or even purchase transaction that is done daily, every hour, minute and second creates large volumes of data with the aim to make it useful to companies and consumers (Katal et al, 2013). The current systems have to be tailored in a way that in those large volumes, data is still secure, regardless of the form it is collected. Security becomes very challenging as privacy has to be maintained at every stage of the data storage. Working through large volumes of data in doing analysis may make the data vulnerable to threats of
information security, given this data is greatly distributed (Dong and Srivastava, 2013).

B. Variety

The variety in the forms of data makes it difficult to handle the data. Today’s multi-varied internet culture, the large volumes of data are extremely varied, with data being collected in different forms, some coming in as purchase transactions, website traffic, business reports, twitter, facebook or even blog, which can be defined as raw, structured, unstructured and semi structured. This aspect makes it hard to apply uniform security policy on varying different forms of data. (Pros, 2014)

C. Veracity

The sources of data is dynamic, ever changing but Katal et al (2013) says velocity is not limited to speed of incoming data but also data flow, i.e. data constantly in motion. Information is being created at a faster pace than ever before, meaning even our security techniques have evolve at this pace. However as PWC (2013) explains that our security techniques are a bit slower and behind which has contributed to increase in breaches. Output of content by the varied channels of big data is definitely on the increase and there is need for information security procedures to catch up.

D. Velocity

The sources of data is dynamic, ever changing but Katal et al (2013) says velocity is not limited to speed of incoming data but also data flow, i.e. data constantly in motion. Information is being created at a faster pace than ever before, meaning even our security techniques have evolve at this pace. However as PWC (2013) explains that our security techniques are a bit slower and behind which has contributed to increase in breaches. Output of content by the varied channels of big data is definitely on the increase and there is need for information security procedures to catch up.

IV. SME’S AND BIG DATA

In understanding what big data actually is, one would say SME’s produce data that is relevant and equivalent to the size of their business, such that one may argue and question that ‘Do SME’s actually have big data’. However The Guardian (2014) argues that they may have very little data collected but the questions asked on this data is what is important. In asking these questions, the benefits to big data is realised, by simply identifying certain trends and patterns on the data, exploiting these trends and developing new ideas, products and services to address identified needs from the trends. So to say to SME’s, rather than getting caught up in the big aspect of big data, they should instead focus on using big data techniques to their existing information stores, linking and analyzing, identifying trends in order to benefit from big data. For SME’s the benefits to big data may not seem apparent now., (Davenport, 2014)

Today SME’s currently face the challenge of having limited resources. Chardonnens, (2014), says as smaller organisations they may find they are mostly unable to afford infinite data storage facilities and even have a smaller budget for big data, making big data irrelevant. However Tien, (2013), sees it differently saying while he agrees with the shortage of resources being a limitation, he reckons smaller organisations have more flexible IT infrastructure, with fewer legacy system issues, therefore meaning an ability to change practices quickly. By using specialist software or services, smaller companies can combine existing data with external data to uncover insights.

Pantelis and Aija (2013), revealed that the demand for big data processing and services will almost triple over the next few years.

Figure 1: Pentaho 2014, Operationalizing the Buzz: Big Data 2013 pg. 14

Chardonnens, (2014), has predicted that by 2017 the demand for big data specialist will increase by 243%. It is believed, according to The Guardian (2014), that even the government acknowledged that it needs to work with academia and business to tackle the problem. This is amongst the challenges being faced by organisations in regards to big data. Information security is non the less another and very crucial crisis in big data. The sheer size of big data repository brings with it a major security issue. So having established, that SME’s may not currently possess big data in the same way as larger organisations do, it is only correct to predict the same challenges the larger organisations are facing now, will come to them in only but a few processes away. Pentaho (2014), shows a survey that reveals that the financial services sector has embraced the big data era more, and given the nature of their business, financial services are more prone to accumulate large volumes of data, especially personal data. See figure 2.

Figure 2: Pentaho 2014, Operationalizing the Buzz: Big Data 2013 pg.

V. ISSUES IN SECURITY

Since a main concern of big data comprises access to data from multiple and diverse domains, security and privacy will play a very important role in big data research and technology. (Schell, 2013). There is no doubt that there are limitations on the current standard IT security practices and this leaves a big
question, what security and privacy technology is adequate for controlled assured sharing and efficient access to big data. In a recent publication by The Guardian (2014), the American president was being urged to open up a review into big data and privacy threats; this was after a general consensus that the bulk collection of personal data by services, companies or organisations is putting consumers at risk. However as the PWC, (2014) crime survey of 2014 has shown, this is equally the same concern the organisations need be having. Security issues have given rise to economic crime, both in the large and the small size organisations, says the survey. It reveals that economic crime is at its highest in the financial services, see figure 3. Regarding the nature of business in this sector, crimes of this type are bound to be high and therefore security needs to be a great concern. Good security ensures business is carried out efficiently and without interruption.

In so many ways, there seems to be a general agreement in that instead of information security being a crisis, the rise of big data has brought about more and better ways of security monitoring. Chardonnens,(2014) seems to explain it saying security monitoring generates big data, which he believes is a means to an end, the end goal being improved risk-based information security decision making based on analytics performed on the data. Tienbou et al (2013) also agrees to this fact, saying big data could in fact mean big security. There is emphasis on how big data is a solution to security, saying the data collection, integration, analytics coupled with new emerging tools created for data analytics greatly assist as a foundation for threat detection, security analysis and log data management Although this may be factual, Kaisler et al (2014) agrees with the promised benefits for information security while also sighting the increased risks presented by big data. As big data changes the face of business, the security risks have become much greater. Hasan et al (2013) agrees with this fact that as they highlight the issue of user profiling, which is normally associated with financial services sectors, saying while collecting user personal information for the purpose of creating profiles is essential in financial services, there is that problem of privacy. In all aspects, big data has presented increased security risks and a number of key issues can be pointed out surrounding big data.

**A. Data Breaches**

With more data now available online, due to more transactions, conversations and interactions, cybercriminals have more ground and companies have more to worry about since breaches involving big data could have far-reaching consequences and reputational damage, legal liability and financial ruin. (Durbin, 2012)

**B. Data in the Cloud**

In a bid to support big data storage and big analytics, businesses have rushed to adopt cloud services, which come with unforeseen risks and consequences. The cloud no doubt is a high target and market for harvesting data and poses more risks and this then means businesses need to get their security right on cloud sourcing strategy. (Dong and Srivastava, 2013) In the process of inputting data in the cloud data store, this may result in privacy and confidentiality restrictions being removed.

**C. Consumerisation**

With the current age of big data, more and more devices are being used and being manufactured that gather, store, and access and transfer data. (Durbin, 2012) this has brought rise to a big challenge to organisations on how they control the personal devices brought into the work place. Durbin further emphasises that failure to manage and secure personal devices brought into the work place by employees and balance the need for security with productivity will give rise to threat in information security by means of data transfer, malware infection and data theft. (ICO, 2014) This is confirmed by Slideshow.com (2014) who says there is a high percentage of data loss due to data transfer, mainly as a result of employees bringing in varied devices into work.

**D. Interconnected Supply Chains**

It is quite rare for an organisation to run independently. Somewhere, somehow, in this edge of big data, organisations become part of rather complex, interdependent and global supply chains. With SME’s being quite unable to afford grasping the technology of big data as earlier cited, they may resort to outsourcing and having business relations with cloud providers. In this regards, Durbin (2012) highlights that the key role in information security becomes the coordination of the contractors and the business in forming a well understood platform of information security.

**E. Privacy**

This becomes an issue as huge amounts of data are generated and stored in distributed places, to such an extent that this is being done has effect on the data protection laws over varying jurisdictions which may be different, thus privacy becomes a concern.

**VI. GUIDANCE FOR MANAGING THE INFORMATION SECURITY CRISIS.**

According to PWC (2013), there is a significant increase in data breaches as cited by the Information security Breaches Survey of 2013, which is both in SME’s and large organisations. This, the report claims is due to the advance in technology which includes the big data edge which has obviously come with a lot of insecurities in issues of information security mentioned earlier. SME’s need to grasp and understand the security issues discussed, highlights Hasan et al (2013)and thus the guidelines given below advise on possible ways to cap and or minimise the risk of the security crisis.

• Cyber resilience and preparedness is crucial. SME’s need to maximise on the data they have to carry out big data analytics to identify trends in security breaches.
• There is need now for management to acquire personnel with skills, develop the skills in individuals to research, analyse, interpret and be creative in big data analytics – hence creating training programs.
• Use big data analytics to establish any misuse, unusual access to systems through remote login, mobile or other personal devices, thus dealing with consumerisation.
• High risk companies can be established by analysing interconnected suppliers, business data and comparing them across different dimensions of information security risk.
• Implement privacy best practices, making sure to design them to fall in line with the analytics programs being used for big data, thereby building transparency and accountability.

VII. CONCLUSION

This paper described the concept of Big data, its importance and the existing information security issues surrounding it. To accept and adapt to this big data era, many challenges and issues exist which need to be brought up right in the beginning before it is too late. All those issues and challenges have been described in this paper, explaining the information security issues surrounding the attributes of big data to the security issues around big data technology. These challenges and issues will help SMEs which are moving towards big data technology for increasing the value of the business to consider them right in the beginning and find help in the guidelines given to counter them. (Hasan et al, 2013)

VIII. REFERENCES

Ohlhorst,F.J. (2013), Big Data Analytics - Turning Big data into Big Money. John Wiley and Sons, USA.


Dong, X.L. & Srivastava, D.(2013), Big data integration: Data Engineering (ICDE), 2013 IEEE 29th International Conference Page(s): 1245 - 1248


Tien, J.M. (2013), Big Data: Unleashing information Service Systems and Service Management (ICSSSM), 10th International Conference Page(s):4


Katal, A and Goudar, R.H. (2013), Big data: Issues, challenges, tools and Good practices, Sixth International Conference on Contemporary Computing (IC3), Page(s): 404 – 409


Kaisler, Stephen ; Armour, Frank ; Espinosa, J.Albert. (2014) introduction to Big Data: Challenges, Opportunities, and Realities Minitrack. 47th Hawaii International Conference on System Sciences (HICSS), 2014


Big Data Analytics in Information Security – is it an Effective Solution?

James Peter Lovatt
University of Derby
Derby, United Kingdom
j.lovatt1@unimail.derby.ac.uk

Abstract— Big Data is a term for the recent phenomenon of data sets that are difficult to store, analyse and visualise; due to their massive, complex and varied nature. The process of wielding these data sets to reveal meaningful correlations is known as Big Data Analytics, a process that allows for faster analysis of big data compared to conventional methods. This paper explores the importance of big data analytics and the value it can bring to the information security problem.

I. INTRODUCTION

Big Data Analytics is changing the way that information technology professionals look at approaching the analysis of vast sets of data we know as Big Data. Due to the massive amounts of data that are being produced by organisations, conventional methods of querying data can provide little coherent results in a short space of time. When time is of the essence – which is key to information security – big data analytics solutions are necessary to make sense of the overwhelming flow of data that companies are presented with (Baseline Magazine 2014).

There are three core aspects to information security that information technology professionals should adhere to in order to protect their data from outside malicious sources. These are confidentiality, integrity and availability. Data breaches conducted by malicious individuals can spell utter destruction for an organisation’s credibility and chequebook alike; and with high-profile data breaches that span over long periods of time, such as the recently reported eighteen-month incident with Specs Liquor which resulted in a serious loss of confidential customer details (Finextra.com, 2014), it is clear that organisations must take proactive measures to locating and preventing data breaches in a timely fashion.

However, the Specs incident is not an isolated event; information security breaches are proving to be a great problem for organisations. It takes on average months or more to identify the source of a breach for the majority of companies (Verizon 2013), a fact that has direct correlation to the lack of steps taken by organisations to process the overwhelming security data generated by today’s systems.

II. DEFINING BIG DATA

The Internet has expanded into a key centre for commerce and communication over the last decade. With this expansion comes great benefits to those who adopt it. As more and more interaction with the Internet is happening, massive amounts of data are being generated at an exponential rate; therefore a problem is posed where the scale is simply too grand to be processed by all but the largest of organisations that have sufficient resources to realistically fund such endeavours (Robinson 2012). Modern data generation is on such an enormous scale that the sheer volume of it can be overwhelming to the data warehouses of smaller organisations as a result of the magnitude of its generation (Lee 2013). The term used among information technology professionals to define this phenomenon in data generation is Big Data.

As the sheer volume of big data is daunting to even consider processing by the majority of organisations, there exists a viable and effective approach to managing such data. The technological challenge is handled by classifying the four dimensions of Big Data, Volume; Variety; Velocity; and Veracity (Dumbleton 2013), and analysing the problems they pose to the relevant discipline (e.g. information security) so as to propose effective solutions.

When the challenges posed by the collection and processing of big data are overcome, the proper wielding of big data by organisations can be extremely valuable to further their interests in any given field. For example, security specialists can utilise big data analytics in the interest of information security to manage security logs, correlate security events, monitor network activity and monitor user activity so as to gain a deeper insight into network and data security incidents. (IBM 2013) Through extension, the analysis of the massive quantity and variety of data can lead to effective and timely data breach resolution.

III. THE FOUR DIMENSIONS AND DATA SECURITY

The four aforementioned dimensions of big data, Volume; Variety; Velocity; and Veracity each correlate to the modern state of security data generation. It is important to understand
the problems posed to data security in respect to the dimensions of big data, so as to integrate effective solutions.

A. Volume

The problems posed by the volume of big data are the most immediate challenge presented to IT structures. The auction website eBay stores roughly ninety petabytes of data related to customer transactions and behaviours (Tay 2013) – and Windermere Real Estate uses anonymously reported global positioning system signals gathered from nearly one hundred million drivers to predict traffic jams (Wingfield 2013).

In respect to security data, Cisco index almost one terabyte of log data per day; billions of NetFlow records; millions of intrusion detection alarms; and millions of host security log records (Bollinger 2014). Without the use of big data analytics, this massive volume of security data would be extremely difficult to analyse using conventional methods and it is probable that crucial information would be overlooked as a result. Another example is proposed in the Internet Security Threat Report 2014 by Symantec, which states that 568,700 web attacks are blocked daily (Wood et al. 2014). Data generated by this volume of web attacks would be astronomical, and this refers to but one avenue of attack. These examples serve as a testament to the volume of data that organisations are presented with.

Simply storing the data is no small task as big data can quickly overwhelm conventional storage methods, therefore handling the massive amount of data requires scalable storage. Furthermore, processing and querying the immense amounts of data requires the computational power of nothing short of a supercomputer – so a distributed approach to querying the data emerges. In the scenario that conventional methods of database management cannot cope with the volumes of big data, solutions arise in the form of parallel processing architectures and also in the form of Apache Hadoop (Dumbill 2012). Hadoop provides a unique service with which organisations can query their big data across multiple servers, which is scalable to the user’s needs (Kalakota 2011).

B. Variety

The diversity and variety of big data is broad and rarely falls into rational structures that can be queried by normal database structures. Data is gathered from every source involved in any process in regards to a relevant field in information technology. Different browsers send different data; versions of software provides different data; users may choose to withhold information at their discretion; and all communication via computer based mediums produces a whole slew of different data. Applying a rational mind-set when querying such data is impossible in most scenarios, so a distributed processing method such as Apache Hadoop is required.

With conventional security solutions, the analysis of select events, logs and security alerts is possible. Nonetheless, crucial data may fall through the cracks and get overlooked, further compounding the problem by delaying the process of identifying relevant data so as to produce security solutions. With big data analytics, every avenue of data generation can be utilised to provide solutions - including system audit trails; network flows and anomalies; business process data; configuration information; full packet and DNS captures; network and virtual activity; database activity; and application activity (IBM 2013). These examples merely scratch the surface of the security data that is generated by modern systems, but they provide insight into the wide range of structured, semi-structured and unstructured data that big data security solutions can process.

C. Velocity

The velocity of big data refers to the speed at which it is generated. It is the necessity of speed in relation to analysing big data when a business objective such as preventing a security breach is concerned. With the emergence of data streaming generating huge amounts of data quickly, it is very important that companies process and parse this information in a timely fashion in order to maintain an advantage by using the structured data to provide solutions in a timely fashion (Dumbill 2012).

D. Veracity

Veracity refers to the general trustworthiness of information in big data analytics. IBM states that poor data quality annually costs the US economy $3.1 trillion per year, and one in three business leaders do not trust the information they use from poor data to make decisions (IBM, n.d.). From this it clear that the veracity of big data is a serious problem. The volume and velocity of big data generation makes it extremely difficult to determine which data is useful and trustworthy – considering that making decisions based on inaccurate data can be devastating.

IV. INFORMATION SECURITY

Organisations that host critical information such as confidential customer data and trade secrets on their servers are growing more and more at risk due to data breaches. It is becoming much more difficult for organisations to protect their confidential assets against malicious individuals as technology progresses and the Internet is being increasingly used as a forum for communication and e-commerce.

Computer security threats are on the rise, and organisations are being attacked continuously by outside sources. The annual Internet Security Threat Report 2013 by Symantec documented that a 42% increase in targeted malicious cyber-attacks from the previous year (Haley and Wood 2013). The consequences of a breach in data security can result in significant cost to an organisation, with the magnitude varying with the type of data loss. Furthermore, losses in confidential consumer information can result in extreme litigation costs and loss of trust between customer and provider, but on the other hand losses in financial records can usually be reconstructed fairly quickly (Liu and Kuhn 2010). The scope of losses can be broad and devastating, and as the risk of data breach is higher than it has ever been – and is continuing to grow – companies and organisations must approach computer security with much more attention and resources than ever before (Symantec 2009).
A widely used benchmark for risk assessment and ensuring the security of information systems is used by information technology specialists. This benchmark focuses on three core aspects of security – confidentiality of information, integrity of information and availability of information (Chia 2012).

A. Confidentiality

The confidentiality of data ensures that access is only granted to authorised users. The primary goal of confidentiality is to prevent against unauthorised access to data, resulting in unwanted individuals or organisations viewing, editing or deleting data. The focus of maintaining data confidentiality extends to data at rest and data in motion, where data at rest refers to data stored on a medium such as a hard disk or a flash drive, and data in motion refers to data that is conveyed over a network connection (Gibson 2011).

There are a few approaches to ensure that data confidentiality is not breached:

- The first and foremost approach to ensuring confidentiality is access control. Access control is the ability to assign access rights to specific users of a computer system, where a system administrator can assign appropriate rights to designated users in order to access, modify and delete stored data (Lakshmi et al 2013).

- The second approach is encryption of data in transit. The eventuality of stored data is that it will get transferred for use elsewhere; be it over network connections or by physical transfer (e.g on a flash drive) (SANS Institute 2006).

B. Integrity

The goal of integrity is to ensure that information has not been accessed and/or changed by any unauthorised individual. The reason that maintaining the integrity of data is important is because data that has been tampered with could prove costly to an organisation.

As with data confidentiality; cryptography and encryption play a crucial role in preserving the integrity of data. A commonly used technique to enforce data confidentiality is the use of hashing. Hashing algorithms such as SHA1 and MD5 can be used to verify the integrity of data by assigning a fixed sized bit string to a piece of data, known as a hash value (Gibson 2011).

C. Availability

Availability in information security refers to the necessity for authorised individuals to be able to access data when it is needed. Individuals with malicious intent have taken notice of this key requirement for data and attacks designed to deny the availability of data are common place. These are known as denial of service (DoS) attacks (Chia 2012).

Data availability can also be compromised in multiple other ways, denial of service attacks are but a single method for achieving malicious goals. Many methods of unauthorised intrusion exist where an outside party could gain control of an organisation’s database and steal or outright delete data. The fact is that data availability is at constant risk, so it is very important that organisations take steps to ensure it (Clemmer 2010).

These are but a few of the problems faced by organisations on a daily basis. Ensuring these three aspects of information security is an extremely important step in ensuring that data remains secure (Chia 2012).

V. BIG DATA ANALYTICS IN INFORMATION SECURITY

Verizon reported that in 2012, 66% of questioned organisations took months or more to discover the source of a data breach. Compared to their earlier statistic of 2010 where 41% of organisations took months or more (Verizon 2013), a worrying trend is revealed among data breaches that links to the inability to handle big data efficiently and in a short time span. Furthermore, a recent study by McAfee on security breaches revealed that almost two thirds of organisations take much longer than a few minutes to detect a security breach, with just 35% detected within a minute. Organisations on average store approximately fifteen (15) terabytes of security data on a weekly basis, and it takes roughly ten hours to detect a security breach (McAfee 2013). It can be deduced from these statistics that the exponential growth of security data generation is creating a problem of considerable importance, where organisations lack the means to effectively analyse big data.

Considering the dramatic increase in computer security threats – a 42% increase in targeted malicious cyber-attacks were recorded by Semantec in 2013 from the previous year (Haley and Wood 2013) – and the systematic failure of identifying security breaches quickly, it is imperative that big data analytics are a focus of information technology professionals; so as to implement effective solutions for identifying and preventing cyber-security breaches.

Organisations need to seriously consider implementing a solution for handling big data into their security protocols, as often by the time a breach has been identified and detected the damage is unable to be prevented and has already been done. This is where big data analytics comes in.

With the outlined threats to information security constantly looming, and the fact that cyber criminals are improving their arsenal of ingenious criminal methods, it comes as no surprise that large organisations are adopting big data analytics to combat the oncoming storm (Barrett 2014). Companies are employing the solution to analyse log files, transactions and network traffic in order to weed out anomalies and suspicious activities – which would take extreme amounts of time without the use of big data analytics (Cloud Security Alliance 2013).

The four dimensions of big data as outlined prior; volume; variety; velocity; and veracity impact the efforts of cyber security specialists determining that conventional methods of analysis are very ineffective at sorting and querying big data. Big data analytics however makes possible the analysis of the immense volume and variety of data, and so prove to be the key to identifying security breaches that often take months to resolve using conventional cyber-security methodologies.

It is widely recognised by organisations that big data analytics could prove essential in the battle against information security breaches; a recent survey (Baseline Magazine 2014) states that 56% of respondents agree that big data analytics
technology will prove vital in protecting against cyber-attacks, however 29% state that their companies have no plans to implement big data technology whatsoever. Evidence of the success of big data analytics is demonstrated by Intel IT, who claim that by implementing a big data analytics solution the data collection analysis process of log files was reduced from two weeks to twenty (20) minutes, all while selectively monitoring 1.5 billion directory service events daily to focus on an approximate count of one thousand five hundred (1,500) lockout events indicative of a brute force attack (Intel IT 2013).

VI. CONCLUSION

Generation of massive amounts data is a phenomenon that impacts every organisation that uses the Internet to conduct its business. In the interest of information security - where the consequences of data security breaches are immense – it is more important than ever to adopt effective solutions to handling big data so as to interpret the massive data flow and develop efficacious security solutions for data breaches.

The foremost advantage of applying big data analytics directly combats the greatest current problem with vast sets of security data – the inability to process it quickly enough so as to identify and prevent data security breaches before the confidentiality, integrity and availability of information is compromised.

In conclusion, the use of big data analytics to identify correlations and patterns between the data sets yielding immense volume, variety and velocity is an extremely important solution to the modern information security problem of quickly analysing big data to prevent data breaches; which a number of organisations are not taking seriously.

VII. REFERENCES


Can Big Data Be Trusted For Financial SMEs To Invest In?

Message to Consultants of Small and Medium Enterprises

Anish Malkan
University of Derby
Derbyshire, UK
a.malkan1@unimail.derby.ac.uk

Abstract—Trust in Big Data is a top priority to ensure that it returns a meaningful outcome adding value to produce positive results. Maintaining this level of trust in Big Data is a constant challenge as it relies entirely on how the information is governed for its data model to be successful. In this paper, it has been observed that the term ‘trust’ can be broken down into various factors that contribute to some of the growing concerns that SMEs could potentially face when dealing with Big Data systems.

Index Terms—SME, Value, Volume, Veracity and Trust.

I. INTRODUCTION

With companies rushing into collecting customer data without careful consideration to what data is actually required; companies are reducing the trust that people have in the services being provided to them. Ensuring trust in Big Data is not always about doing the right things at the right time but also making commercial sense for having the right information needed. As Big Data offers many new opportunities to how data is analyzed, exploiting its risks also has an impact to how its users can mistrust the system. Such exploitations bring in more concerns to how privacy and transparency is controlled to ensure trust in the data (Little, 2014).

Unlocking value from Big Data has created a debate on how the volume of data plays a key role amongst the C-suites in the company; however discussions on how to gain such information (most of which are personal data) required is often left out. For large companies and SMEs to gain this sort of data, consumers need to have trust that their data will be used only for the purposes required and for purposes where such data is needed for (BCG, 2013). There are many benefits to using Big Data as it provides opportunities for companies to maximize their potential through analytical tools allowing them to analyze data better, faster and to gain the competitive edge (Dyché, 2013). It allows the company to understand its clients and requirements better therefore improving the business efficiency and producing effective results. A major factor when value is created is the trust in the information that is used to make decisions. A lack of trust in the information’s source can derail the success of any analytical project. Therefore establishing trust in Big Data is important to value creation when the variety and volume of information grows (Capgemini, 2013a). This paper breaks down the subject of trust in Big Data systems in order to identify its main concerns for SMEs.

A. The four ‘V’s to Big Data

There are four fundamental ‘V’s that SMEs need to consider about the data type properties in Big Data systems:

- Volume - As there is a large amount of data available, it is crucial to understand and know how to turn that data into a credible, readable and structured format within a short period of time. As there could be time constraints on this, without appropriate resources to process the volumes of data and information, SMEs may struggle to know which aspect or areas of Big Data needs to be focused on and analyzed (Capgemini, 2013a).

- Velocity - This is the rate of change of data coming at different speeds (Sicular, 2013). In financial sectors, it is important to process large amounts of data as quickly as possible as the rate of change is rapid and the results are only valuable for that period of time. Appropriate resources are required to process this data quickly as possible. (Dumbill, 2012).

- Variety - Data can be gathered in both structured and unstructured from various sources such as text, audio, logs, video and files. Unstructured data creates problems for storage and data mining for analyzing data (Normandeeau, 2013).

- Veracity - As data can come from different sources, it raises more questions as to where does this information comes from? How accurate is this information? Can it be trusted? Without appropriate validation and verification processes, such data could be invalid therefore increases the uncertainty with the trust element of Big Data for its information held (Lukoianova & Rubin, 2014).
B. Architecture

In order to get into the world of Big Data, companies need to consider three necessities; the first is having large quantities of data in a format that can be accessible easily for analysis. Most large companies already have this. Some have more data than they can utilize (Parise et al., 2012). The second necessity is having advanced analytical tools such as Hadoop and NoSQL. These are both open source tools that support the database and scalability storage required for large scale processing techniques needed for Big Data systems (Oracle, 2011). With the advanced analytical tools available, this brings about the third necessity which is having employees with the right skills and experience that are capable of maintaining such systems from data science to the privacy law with an understanding of the business with its relevant sources of value (Pearson & Wegener, 2013).

II. FACTORS AFFECTING SMES TO INVEST INTO BIG DATA

A. Awareness

Companies that are implementing Big Data technologies and analytical tools are observing the positive impact it has on their businesses; however there remains various challenges to SMES perception to Big Data implementation. As some businesses are not fully convinced about the use of Big Data in company goals for long term projects, they are more interested in producing short term results with less overhead costs therefore leaving other related technology problems to be solved in the future. Although Big Data can be thought of as a new concept for data handling techniques, some organization struggle to understand how does it apply to their business, and with less awareness or knowledge on the area this brings about less investment into such systems (Mashood, 2013).

B. Skills

In a report by O’Neill (2013) “just 9% of the 257 respondents rate their organization as extremely effective at identifying critical data and using it to make decision.” The statement in itself raises more questions, was the remaining 91% less skilled to identify the data that was needed to be analyzed? Or did they not have the correct data for analysis? The survey continues to suggest that “38% says that Big Data expertise is scarce and expensive” and that data warehouse appliance platforms required for setting up Big Data systems are expensive.

E-Skills UK (2013) has indicated that 57% of Big Data recruiters state that it is difficult to find people with the right skills, qualifications and experience to employ for Big Data analytics. The skill set in this market is to increase by 243% over the next 5 years; from this statement it would appear that it is better to invest into Big Data at a later period as there will be more support available for it. The survey continues to state that within the SME community, there is very poor or little knowledge about the technologies and processes involved in the implementation of Big Data systems.

C. Technology & Costs

Given that Big Data architecture is expensive to invest in due to the storage required to accommodate the large data set size, but the expense also relates to the time required to be able to load the data into multiple systems. The challenge here is not only to build a system for processing different tasks, but to build a system architecture that can be flexible to allow other systems to be built on top of it and would allow all the tasks and processing to work efficiently, simultaneously and be able to handle the workload (Agrawal, 2012). For some SMEs this works to their disadvantage if they are still a start-up and the capital for such investment is not readily available.

In a survey conducted by Online Business Degree (2013), “39% agreed that Big Data could cause more problems than it solves between now and 2020”, with potential for more problem arising from Big Data, it would appear SMEs would rather not invest in such troublesome technology and rather wait for a time when it is more stable for investing into.

SAS (Passingham, 2013) had identified that an estimated 0.2% of SMEs in the UK are using Big Data while a fifth of all businesses have admitted that they have a poor understanding of the issues and technology surrounding Big Data. The reason being for poor development of Big Data within SMEs is because the amount of data collected by them is significantly less than large organizations therefore most of the skilled analysts head to big corporations leaving smaller businesses behind to deal with the gap in the skills market.

As Big Data grows over time, it brings about analytical challenges that questions how should this large volume of data be dealt with? Does all this data need to stored and analyzed? Most of this data can be unstructured, semi structured or structured that require the need to be analyzed which can be time consuming to produce results in a short period of time. Some SMEs may soon realize that Big Data technology is not for them and would rather concentrate on gathering only the information they require and storing it in the way they want as it works better for them with less costs involved (Katal et al., 2013).

D. Experience over data analytic tools

Hesse (2012), states that according to a research by SAP, 91% of UK SMEs rely on their experience rather than the reporting tools for decision making and that more than 50% of UK SMEs rely on their gut when making business decisions. Within the same research it was also reported that less than one thirds of SMEs are using Big Data reporting tools for sales forecasting and for financials. Evaluations such as these tend to make SMEs shy away from the concept of Big Data technologies when experience is relied upon for a better decision making process.

III. ISSUES THAT AFFECT TRUST IN BIG DATA

A. Data inconsistencies.

Some of the many economic sectors benefit from Big Data technology such as healthcare, retail, banking and financial services amongst many others. They all rely on accurate data in order to create additional value and data for analysis results.
This raises one of the most important questions for what happens when there is inconsistency in Big Data. Once any data is gathered, inconsistencies can occur in various forms; from the data collected to the knowledge and expertise required to analyze that data. If this data is not handled accordingly following certain guidelines (governance) or protocols, the inconsistency of the data can cause a return of poor or invalid result which is then followed with a question; how trusted is this information to use? (Zhang, 2013)

Any analytical model only produces effective results based on the quality of information that is put into it. This mean that sources from where this information and data is from needs to be reliable and trusted. This dictates the accuracy of the information. If this is not done correctly, it then produces “garbage in, garbage out rule” (Capgemini, 2013b).

B. Human factors.

A survey conducted by IBM (Bertolucci, 2013) has identified that a lack of trust in between departments is one of the preventions from Big Data being unlocked to its full potential. It was stated that only 40% of its respondents feel a strong sense of trust between them and the executives at their firm. This was mainly due to age-old human conflicts that occur especially in the finance sectors where employees avoid sharing customer data across different areas of the business due to competition amongst themselves.

C. Security concerns

Security within unstructured data is a top concern for consultants that are thinking about an investment into Big Data analytics. Both large scale firms and SMEs are concerned that data held within Big Data systems are not protected with the same sophisticated encryption and level access controls used in traditional databases. For Big Data platforms to be trusted, it must consist of sufficient security provisions for recording access control and logging capabilities (Forrester Consulting, 2013).

A report by PwC (2013), has stated that only large companies used to be targeted by significant outsider attacks, however there are reports that these attacks are also increasing amongst small businesses. The research indicated that 63% of small businesses were subject to outsider attacks, while 23% of small businesses were affected through denial-of-service (DOS) attacks. 9% of small businesses confirmed that the outsider attacks were able to steel their intellectual property and confidential data. With more content available in Big Data systems, this enables a greater risk and loss during security breaches. Trust in the system is gradually lost as more incidents are reported through Big Data implementation.

Ensuring appropriate algorithms for protecting sensitive information is important to prevent data breaches or ensuring data is not leaving the organization unintentionally (Anon, 2013). One common process for securing information in Big Data is by anonymizing it through which Capgemini uses a process for master data that anonymizes the information when it is loaded in the system such that names and ages are anonymized and places the information in a demographic format (Capgemini 2013b). This too requires the right skill set and expertise to implement the correct policies for data protection.

D. Big Data protection

As Big Data grows to become more accessible, the security and privacy concerns also grow together. Collection of large scale information is becoming more routine amongst businesses that implement Big Data however the tools and techniques used to manage these large data sets are not designed to incorporate sufficient security and privacy protocols. These include security policies, Intrusion Detection Systems (IDS) and measures for data compliance. Businesses have to invest further into additional costs for tools and business models to accommodate such functionality to increase their security and to reduce chances of data breach to maintain trust in the system (Schmitt, 2013).

Real-time security monitoring has been a challenge due to the number of alerts created through false positives generated by different devices. Some of these false positives are ignored as IT managers or analysts are unable to cope with the amount of alerts that is being reported. Within Big Data this problem will increase exponentially as the volume and velocity of data streams increase. Without investment in the right threat detection software, Big Data is under a constant threat from both outsider and insider attacks (Cloud security alliance, 2012).

IV. GOVERNANCE & PRIVACY FOR BIG DATA TRUST

A. Achieving Governance

According to Jones (2014), information governance in Big Data has been one of the poorest elements within the IT department due to that fact that companies concentrate more about Technology than Information that brings about issues to how the business actually works. In order for Big Data to work, less information needs to be governed in order to succeed. The goal in Big Data is to find the minimum amount of meaningful data and information that will allow for effective collaboration and produce analytical benefits that can be delivered to companies. As information comes from different sources, there should be governance in each area so as to validate which information is credible for use and how it should be stored.

In order for uncertain data to be turned into trusted information, companies need to enforce governance procedures for how the information is created, shared, cleansed, consolidated, protected, maintained, retired and integrated within the enterprise. All of these elements are vital for information to achieve positive results to meet the company’s goals and objectives (IBM, 2013).

As Big Data relies on vast amounts of information collected from various sources for company insight, companies are choosing to store everything as they feel that it is all crucial and do not want to archive or delete meaningful data. One of most important questions to this is, without appropriate information governance policies in place, how can a company know what piece of information is meaningful or not to them. Most organizations are concentrating more on trying to manage
structured information and less attention has put forward on policies to process unstructured contents (Murphy, 2012).

Partnoy & Eisinger (2013) have indicated in their report that ever since the 2008 financial crisis, due to the lack of transparency within financial sectors and Big Data, “79% of investors have no trust in the financial systems” and that the public too have lost faith with financial organizations. With evaluations such as this, more and more individuals are reluctant to provide information that financial services require as an input into their Big Data model.

A report study conducted by Navetta (2013), has identified an instance where brokers using Big Data systems to collect detailed employment information including salary information for 190 million individuals were sold to other debt collectors and financial institutions. With privacy and data protection laws not fully complied or consent from the public, there are more chances that individuals are less likely to have any trust in the financial sector if they feel their information is not secure.

PwC (2014) has identified that amongst retail and communication sectors, the financial services are an industry where economic crime is reported at with a risk of 49%. With constant scrutiny from cybercrime, fraud and other threats, would SMEs have the financial backing for investing in the latest threat detection and security services to ensure that Big Data is protected at all levels?

V. CONCLUSION

Big Data is continuously growing as the volume, variety and velocity increases; however it faces challenges to the veracity of the information held. Big Data can only retain its value when the veracity of the information can be trusted (Schaeffer & Olson, 2013). In this paper, some of the key concerns surrounding its technology, security, governance and privacy have been identified, and how it affects trust in Big Data. All of these aspects are important in order for SMEs to consider prior to opting for such a transition to implementing Big Data systems in their business. Such evaluations are important to achieve a positive outcome for company goals.

VI. REFERENCES


Understanding the Value of Big Data for SMEs
Unlocking The Potential of Big Data and What it can Provide for your Business?

Adam Martin
University of Derby
Derbyshire, United Kingdom
A.Martin12@unimail.derby.ac.uk

Abstract— With the volume of data in the world growing every second big data can provide valuable insights, that can help transform your business. With every large organisation now jumping on the big data bandwagon, isn’t it about time your organisation did? In order to do so you must first of all understand it. This paper will aim to provide your business with a knowledgeable insight into the world of big data, discussing both the positives and negatives, allowing you to determine if it is the right option for you.

Index Terms — Big Data, SME, Analytics, Volume, Velocity, Variety.

I. INTRODUCTION

Big data is the next big thing, or that is what we keep hearing at least. Big data analytics is the process of collecting and analysing data sets (Gartner, 2013). With the volume of data created everyday as high as 2.5 Quintillion and the volume of business data doubling every 1.2 years (Trustwave, 2013), big data is big business. Big data has been making waves of late within large organisations, question is can SMEs (small to medium enterprises) also share this same success? Capitalising on the ever increasing amount of data that is being produced, by effectively implementing big data analytics allowing them to gain a strategic advantage. Big data analytics has many benefits for businesses of various sizes, and has already proved to be a great asset to many large organisations. With Google investing a massive $7.35 billion in data centres in 2013 (Circleid, 2014), this figure alone showing the sheer value of data. However it appears that many SMEs are reluctant to embrace big data analytics, with one of the main factors being simply a lack of knowledge and understanding of it. If SMEs are to successfully adopt it and use it to its full potential, they must first of all understand it. This paper will begin with a definition of the term big data, then on to assessing why your SME should invest by discussing the benefits it can provide, while taking into account potential issues. The aim of the paper, to begin to give an insight in to big data and how it can fit in with your SME, demonstrating the potential opportunities an effective big data analytics strategy could unlock.

II. WHAT IS BIG DATA?

As said previously big data analytics is the process of collecting and analysing vast data sets. The two part process consists of first collecting the data, and is then followed up by analysing the data. While a more in depth definition is that big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making (Gartner, 2013). With SAS saying that big data may be as important to business and society as the Internet has become (SAS, 2013). While Oracle describes big data as the electricity of the 21st century, saying that it is a new kind of power that transforms everything it touches in business, government, and private life (Oracle, 2014). With claims as strong as these can SMEs really afford to ignore big data?

Big Data in its most simple form consists of three V’s volume, variety and velocity, these being the three defining properties of the term big data. Volume simply refers to the amount of data, while variety refers to the number of types of data and velocity refers to the speed of data processing (Rouse, 2012). While IBM data scientists add another dimension veracity, this being the uncertainty of the data (IBM, 2013).

A. How big is Big Data

The amount of data in the world is growing by the second, in both the forms of structured and unstructured data. It is estimated that a staggering 2.5 quintillion bytes of data is produced everyday with 90% of the data in the world being produced in the last two years alone (IBM, 2013). While it is also estimated that by 2020, there could be four times more digital data than all grains of sand on Earth (IBM, 2013). This data coming from everywhere, social media sites, digital pictures, videos and purchase transaction history being just a few. Facebook alone having 70 billion pieces of content shared a month (Statisticbrain, 2014), such a staggering amount of data from a single site. While Cisco predicts that global mobile traffic alone is set to reach 11.2 exabytes (An exabyte is 1 billion gigabytes of data) per month by 2017 (Cowin, 2014). With data volume estimations like this, it shows the sheer amount of data there currently is and just how big the growth
of data is set to be in the future. With all this data available it would be foolish for your business to ignore this.

III. WHY SMEs SHOULD INVEST IN BIG DATA

Big data is able to provide a whole host of benefits. This section will look at the issues SMEs may face without the use of Big Data including some of the issues which are currently causing them to struggle to utilise it, then going on to discuss the potential benefits its use can provide.

A. Why SMEs are currently not Investing in Big Data

Adoption rates for big data analytics are also extremely low or at least this was the finding from a recent SAS report. As part of their study, from 541 SMEs they had been in contact with, none of them had implemented big data analytics at the time of questioning. Meaning the proportion of SMEs that had implemented big data analytics was estimated to be less than 0.2 percent (SAS, 2013). This statistic being quite concerning, why such a small amount?

From a SAS conducted study it found that the SME community had very little, if any knowledge of big data analytics, with very few likely to implement it. The SAS report highlighted that the level of understanding of big data is often low, within the SME community 22% of respondents admitting they had poor to very poor knowledge (SAS, 2013). While another survey conducted by Harris Interactive, on behalf of software vendor SAP showed that only 25% of the SME executives surveyed could agree on the definition of what exactly big data was (Thebigdatasignithgroup, 2012). These statistics showing just how poor the overall understanding of big data is within SMEs.

Lack of skills and resources is another problem. SMEs currently feel there is a lack of big data expertise, another SAS report shows that just under 57% of recruiters found hiring big data specialists very difficult (SAS, 2013). With SAS (2013) predicting that the demand for big data specialists is only going to get bigger, with it set to rise to 69,000 jobs over the next five years in the UK alone. While an Oracle study found that lack of resources is another potential problem, with 38% of business managers stating that they do not have the right systems in place to gather the information they need (Evans, 2012). Lack of the required expertise and resources seems to be just another potential barrier preventing SMEs from embracing the big data initiative.

B. Issues that can arise from no Big Data Strategy

The volume of data is constantly growing. The McKinsey Global Institute estimates that data volume is growing 40% per year (Oracle, 2013). If SMEs choose to ignore developing a big data strategy they may end up becoming lost in this wealth of data.

Information security is an important aspect for every business regardless of size, with the wealth of data organisations now store, a lack of sufficient security measures can lead to data breaches with serious detrimental repercussions and not utilising big data analytics can mean SMEs are missing out on a range of security benefits. A Trend Micro-sponsored Ponemon Institute study, found that an alarming 78% of organisations have suffered from at least one data breach over the past two years (TrendLabs, 2012). While a PwC report confirms the significant continuing impact of cybercrime on businesses, with one in four of the respondents reporting they have experienced it and over 11% of these suffering financial losses of more than $1 million (PwC, 2014).

With another recent PwC report on security breaches, it shows that in 2012 an alarming 76% of small businesses within the UK suffered a breach, with the average cost of their worst incident being £15,000 to £30,000 (PwC, 2014). With this alarmingly high rate it seems that SMEs are still under the misconception that because they are just a small organisation, and do not have comparable amounts of data when compared to larger organisations, they are not at threat from cyber hackers. A study done by The Hanford also confirms this finding that 85% of small business owners think a data breach is unlikely; thus they don't often implement simple security controls (Nachreiner, 2012). This is really not the case; SMEs are just as vulnerable as larger organisations, if not more than.

In adopting a big data analytics strategy it can provide your business with a number of security benefits, each of which will be discussed further in this section.

C. Benefits of Big Data

Big data can provide your organisation with a whole range of benefits, be it increased efficiency in business processes, reduced spending or improved security. One large scale example was the US health care system. It was estimated that use of big data could have provided a saving of $300 billion to $450 billion in reduced healthcare spend in the United States according to a US health system reform report (Groves, Kayyali, Knott and Van Kuiken, 2013).

Big data can be used by organisations to gain a better understanding of their customers, their behaviour, needs and preferences. A highly publicised example of this was from U.S. retailer Target, they were able to use customer data to predict that a teenage girl was pregnant before her father had any knowledge of it, allowing them to target her with relevant advertisements consisting of coupons for baby clothes and cribs much to her father's dismay (Hill, 2012).

Cases such as this show the phenomenal power big data analytics can harness, but also the potential controversy it can create as well as the questions of ethics it raises in regards to the privacy of the customers. With some potentially feeling very uncomfortable that these types of predications can be made regarding their personal lives. Westpac an Australian bank says it has achieved a $22 million increase in revenue by using big data techniques which allow them to provide targeted offers to their customers when they interact with the bank (Corner, 2014).

Big data can also be used to optimise business processes. It can allow retailers to generate predictions from social media, web search trends and weather forecast, which can then be used to optimise their stock.

D. Big Data to Improve Information Security

Big data can be used to improve information security to detect cyber attacks before it is too late. Security consultant
Mischel Kwon warned that times have changed and we cannot afford to wait for an antivirus or having malware trip a firewall (Rush and Shaw, 2013). Visualisations of the data can be made allowing patterns to be indentified which wouldn't usually be visible. Security officers are able to see patterns leading up to an attack which wouldn't usually be apparent, as a result they can see when the attack begins rather than finding out when it has already happened and is then too late to act upon. However it can require a large amount of data to produce these visualisations, with Kwon saying that it can be a billion events a day meaning this could lead to up to 24 terabytes of data a day (Rush and Shaw, 2013). Ultimately it will depend on the size of the SME as to whether this is a benefit, depending on the resources such as storage and processing power they have available to them.

Big data is now also being used to aid in the prevention of fraud, most commonly in financial and insurance firms. An example of this was one U.S. insurance firm, they suspected there was something more behind some seemingly unrelated but fishy insurance claims, but they just couldn't seem to be able to find a connection. However with the use of big data and using data on social connections they were able to reveal that the claims were coming from related parties that were attempting to defraud the organisation. Fraud is big business and is costing businesses a lot of money. Krishnamurthy at Bank Tech stated that in financial services, credit card fraud has reached some $5.5 billion, banks and financial services have had to resort to aggregating data from multiple sources, including social media, to profile customers behaviour and detect fraud (Rush and Shaw, 2013). From a Global Economic Crime Survey conducted by PwC it found that fraud detection through suspicious transaction reporting and data analytics increased by over one third, going from 18% to 25% (PwC, 2013). This showing the effect big data analytics tools can have in helping to detect and mitigate against fraud. Research firm Gartner stating that by 2016, more than 25% of global firms will adopt big data analytics for at least one security and fraud detection use case, increasing from a current 8% (Kar, 2014).

Big data analysis can help boost network security by detecting anomalous network behaviour that is indicative of a security breach. Once the anomalous behaviour has been detected, it can then be established if it is actually indeed a threat or whether it is just an unusual harmless threat. However the benefits are greater the larger the organisation is as the more data generated allows for a more accurate view.

IV. THE DISADVANTAGES AND POTENTIAL CONSEQUENCES OF BIG DATA

A. Too Expensive for majority of SMEs

Despite the many potential benefits big data has to offer it also has its disadvantages, some of which may initially affect some SMEs looking to adopt a big data strategy. One obstacle that may prevent some SMEs from adopting big data is cost, SMEs will not have the same budget or resources as larger organisations and may therefore struggle to implement it. With a typical data warehouse with a 500TB (Terabyte) capacity costing approximately $22.7 million (Winter, Gilbert and Davis, n.d.) and the mean annual recorded turnover for a small business £2.7 million and £11.4 million for a medium-sized (BMGResearch, 2012) it is just not feasible. However there are more affordable options that would be better suited towards SMEs. The low-cost scalability option of Hadoop can allow effective management of large volumes of data, and does not require the same initial investment as data warehousing. With a commonly quoted figure for a Hadoop cluster less than $1,000 per TB, several times lower than the average list price for a data warehouse platform (Winter, Gilbert and Davis, n.d.).

B. Security Threats

Big data projects include certain data which can be catastrophically damaging to enterprises if it were to be leaked. This data often consisting of customer data containing personally identifiable information, purchasing habits or credit card information. From a study conducted by McAfee it found that 58% of organisations using big data had suffered a security breach, with only 35% of them actually able to detect the security breach within a matter of minutes. In fact of those that suffered a security breach within the last year just 24% had recognised it within minutes, with the average detection time a staggering 19 hours (McAfee, 2013). A rather alarming example was Spec's a Texan chain of liquor stores, they recently discovered a data breach lasting close to 17 months with more than half a million customers financial information being leaked (Hawes, 2014). Target was another high profile example with approximately 40 million debit and credit cards being compromised in a data breach over a two week period (Finextra, 2013) leading to an extra $61 million in expenses to cover the damage (McGrath, 2014). Figures like these showing the astronomical costs data breaches can potentially incur. According to a Global Economic Crime Survey conducted by PwC it found that 56% of the respondents suspected the main perpetrator was internal (PwC, 2014). With Forrester Research stating that amassing larger volumes of data increases the risk of cybercriminals or insiders readily compromising sensitive information. To ensure big data environments are as secure as possible, the number of people that can access the data should be limited, with users access levels continuously monitored throughout their employment (Aveksa, 2012).

Enterprises should ensure they adopt comprehensive identity and access management protocols to ensure employees are seeing data that is only suitable and relevant to them. In doing so it should prevent them from seeing unnecessary data and therefore help minimise the risk of data leaks.

V. WHAT IF YOU CONTINUE TO IGNORE BIG DATA

SMEs often tend to overlook big data analytics due to a number of reasons, however if they choose to continue to ignore big data analytics, it could have a severe impact in the coming future. It was found that on average companies not adopting big data analytics were losing $71.2 million per year on average, bearing in mind that was for each company alone (Evans, 2012). Evans (2012) then goes on to state that from almost all of the companies surveyed, each expect huge increases in data volume to continue following a surge in data volume of 86% over the past two years. With this potential loss
in profits and the predicted growth in the volume of data only set to increase, it would be wise for SMEs to begin looking at adopting a big data analytics strategy sooner rather than later.

VI. CONCLUSION

From the points discussed in this paper it is quite clear that big data is a big opportunity for SMEs, with the advantages far out weighing the disadvantages. Many large organisations are already reaping the benefits big data analytics has to offer. However at the moment it currently seems that the vast majority of SMEs are ignoring the big data initiative, with it appearing that one of the main reasons simply being a lack of knowledge.

In this day and age in such a competitive world, it seems foolish not to harness the power of big data to help get that edge over your competitors. Big data analytics can not only improve the efficiency of business processes and allow better understanding of customers, but it can also provide a range of security benefits, although it can potentially create certain security issues. However despite this, if the necessary steps are taken the risk can be minimised. Cost was another limiting factor; with big expenses it can pose a problem to some SMEs, although now it seems there are more viable options allowing businesses of all size to begin using big data.

In order for SMEs to begin using big data they must first understand it, once they have begun to do so they can then begin to develop an effective big data strategy allowing them to really see its full potential. With the volume, variety and velocity of data constantly growing it almost seems like the only option and with the amount of money organisations not utilising big data are losing it would be foolish for SMEs too simply ignore it, especially when it can provide businesses with better insights and increased knowledge.

VII. REFERENCES


Storage as a Service: The Hidden Costs

Joel McMahon
School of Business, Computing and Law
University of Derby
Derby, United Kingdom
100169047@unimail.derby.ac.uk

Abstract—Many SMEs believe that cloud computing is the way forward and are turning to it for their storage needs. This is including storage for big data which in simplistic terms is large or complex sets of data which can allow an enterprise to have an unprecedented insight into a market and improve decision making. However, they are frequently rushing into the migration to cloud computing. This paper examines the drawbacks that are commonly overlooked or not considered by SMEs when deciding to switch to cloud computing. Issues that affect both public and private clouds are included to provide an accurate picture, however the focus is primarily on the use of public clouds.

Index Terms—Cloud computing, SME, STaaS, SaaS, Costs, Rogue Cloud

I. BACKGROUND

Storage as a Service, commonly abbreviated to either SaaS or STaaS, is an architecture model in which a provider will provide a client digital storage on their own infrastructure, usually on a pay-per-use basis. This service is also commonly known as a public cloud. A 2013 survey polling 5,000 enterprises of all sizes concluded that 94% of enterprises were at least discussing the use of cloud services (Symantec, 2013).

When considering the use of cloud computing in any situation, an SME must first ask a number of important questions which will determine whether they will end up going for a public or a private cloud. There are, however, three questions which are at the forefront of this decision:

- How much storage space is required?
- What growth is expected?
- How much would it cost to set up internally?

If the enterprise is able to afford the amount of storage they would need as well as having the technical expertise required within the enterprise to setup and maintain a private cloud, they may end up doing this. However, if they are expecting a large amount of growth it may still be cheaper to just rent storage space and increase their requirements as and when needed.

A large factor in the growth of enterprises turning to public clouds is due to larger enterprises taking advantage of economies of scale. To take an example, for a provider such as Amazon it will be significantly cheaper to acquire and maintain storage space than it would be for a smaller enterprise due to a combination of bulk purchasing and the expertise required to set up a cloud already existing within the enterprise. As such, it is often cheaper in the long term for the smaller enterprise to rent storage space off Amazon. Furthermore, there are no complications such as predicting growth. When the SME needs more storage, they can simply pay more and it will be made available to them. Likewise, if they are renting far more space than they need, they can rent less.

However, in spite of the financial benefits there are a number of issues surrounding ethics, governance and trust which are seldom considered when deciding to implement cloud architecture.

II. HIDDEN COSTS

One of the main issues surrounding storage as a service is that of rogue cloud implementation. A rogue cloud implementation is where a senior member of staff within an enterprise enters into an agreement for cloud services without the approval of the appropriate IT staff within the enterprise. As such, these rogue clouds are generally a huge risk to the business as there are a significant number of factors that would normally be looked at before trusting sensitive data to a third party for storage as well as a potential breach of the UK Data Protection Act or the EU Data Protection laws. An example of such an implementation would be the use of Dropbox to share sensitive information. A 2013 Survey found that 70% of SMEs polled had experienced a rogue cloud implementation (Symantec, 2013).

Common issues stemming from rogue cloud implementation are listed below and will be looked at in turn.

- Exposure of confidential information
- Loss of data
- Stolen goods or services
- SSL certification
- Data Protection Act e.g. requests from the public to change out of date information
- Storage of redundant data
- Excessive downtime
A. Exposure of Confidential Information

The exposure of confidential information is an issue that while it does exist for private clouds, the risk can be significantly greater when using a public cloud and exponentially higher where rogue clouds are concerned. Of the 70% of SMEs that experienced a rogue cloud implementation, 40% of them experienced the exposure of confidential information. (Symantec, 2013)

The storage of sensitive data with a third party adds in a number of points of weakness when considering data protection. The first point of weakness lies in the transmission of the data to the third party. While in transit, there is the possibility that data may be intercepted through a number of different techniques such as man-in-the-middle attacks or spoofing.

Once the data is in the possession of the third party, the next points of weakness lie with the third party itself. It is likely that little to no research will have been carried out into the third party storage provider.

The issue of the location of the storage provider may not have been considered. The laws surrounding data protection differ greatly across the world. As such, it is very likely that if the third party was located outside of the EU they may not adhere to the same security standards as an enterprise located inside the EU. Similarly, if the third party was located in the US they may not adhere to the optional Safe Harbor scheme and may not have the necessary security in place to prevent the exposure of the data – as an example only 3511 of the 4503 organizations who have at some point adhered to the scheme are currently certified as of March 2014 (Export.gov, 2014). Slightly less than 1000 companies involved in data handling have not reaffirmed their compliance and are out of date. As such, it is illegal for data currently held within the EU to be transferred to any of these 1000 companies within the US due to breaking EU data protection laws.

Furthermore, the country in which the data is to be stored may have laws enabling law enforcement or other governmental agencies the right to access and view any data stored within their country. An example of this would be the USA Patriot Act. Section 215 of the act added “The Director of the Federal Bureau of Investigation or a designee of the Director (whose rank shall be no lower than Assistant Special Agent in Charge) may make an application for an order requiring the production of any tangible things (including books, records, papers, documents, and other items) for an investigation to protect against international terrorism or clandestine intelligence activities, provided that such investigation of a United States person is not conducted solely upon the basis of activities protected by the first amendment to the Constitution.” (Cornell University Law School, 2012)

What this means is that any company based in the US may be forced to hand over confidential data to the FBI, even if the data is presently stored in an EU data centre. Similarly if data is held by a subsidiary based outside of the US but owned by a US parent company, this law still applies.

There is a possibility that the screening and training undergone by the staff at the third party may not have been checked along with the procedures required for a staff to access client’s data. If the third party was to have insufficient screening techniques for their staff, the risk of a staff member accessing confidential information or even going so far as to take copies will always be present. This threat can normally be negated by having a set procedure for a staff member to access data and by having sufficient security measures protecting the physical storage location.

There is however one point of weakness which lies within the clients company. Data stored in the public cloud is naturally going to be more accessible from outside world than with a private cloud. It is nearly impossible to secure a public cloud in such a manner that data leakage is impossible by employees as if they wish to access data from outside the company, they will be able to. Similarly, the rise of Bring Your Own Device (BYOD) means that there are more ways than ever for data to be leaked. Before committing to a public cloud a risk assessment should ideally be carried out based around this, although it is rarely done.

B. Loss of Data

Loss of data can be split up into two distinct categories. The first relates to the actual loss of data itself.

It is common for most enterprises to have disaster recovery plans which will be put into effect in the unlikely event that as an example, all data is wiped from the companies’ servers. While the use of a rogue cloud would in fact prevent the loss of the data stored with the third party in this scenario, it is likely to interfere with any backup procedures in place within the company. As such, unless IT staff are aware of the existence of this rogue cloud they will not back up anything stored on it and the risk of losing this data is significantly increased. It is important to note however that most third party providers will normally add data redundancy so that any data lost from location will still exist in another.

A 2013 survey found that nearly 50% of all businesses that have experienced a rogue cloud implementation lost their data and had to resort to backups. Of these companies, 66% of them had their recoveries fail (Symantec, 2013). With offline storage, it is very likely these same companies would not have lost their data as it was found that they were using on average 3 different offline backup methods. In addition to this, the majority of companies estimated that a full recovery from a catastrophic data loss would take over 3 days to achieve which for an SME is no small amount of time.

Alternatively, the data loss may be on the part of the third party provider through breach of contract. In 2011 Amazon’s web services crashed completely despite having numerous countermeasures such as data centre isolation in place as well as backups being taken periodically and stored at numerous data centres. The downtime lasted for several days and Amazon reported losing a small percent of their total data stored covering a space of 11 hours. This data wasn’t recoverable through forensic techniques and as such, customers were left without 11 hours’ worth of data (Business Insider, 2011).

The second category of data loss relates to stolen goods or services.
It has been found that 25% of SMEs who experienced a rogue cloud implementation faced issues relating to account take over, website property defacement or stolen services (Symantec, 2013). One of the main reasons behind this is primarily due to carelessness on the part of employees. At a 2011 Black Hat hacking conference, researchers at Bishop Fox demonstrated how with knowledge of Google and cloud service authentication, they were able to find thousands of access codes and passwords for public clouds (Bishop Fox, 2011). Once an attacker has access to this information, they are free to do whatever they wish.

In the case of an SME looking at storing big data in a public cloud, this could mean a number of things in terms of data loss. Leaked data could mean hefty fines from the UK Information Commissioner’s Office (ICO) while lost or stolen data could mean greater costs for the company particularly if the data was bought in e.g. customer purchasing habits.

C. SSL Certificates

While generally an issue when using the cloud to store content such as websites or applications, it has been found that only 40% of enterprises believe that the third party they entrust their data to have SSL certificates that meet corporate standards (Symantec, 2013).

If for example the third party is using a self-certified 1024 bit SSL certificate while industry standard is now 2048 bit as of December 31st 2013 (CA/Browser Forum, 2013), this means that the potential for fraudulent transactions or for a man-in-the-middle attack is far greater than if their certification was up to date.

As of April 2014, this issue has become significantly larger with the discovery and publication of the Heartbleed bug for OpenSSL. Recent surveys have shown that around 17.5% of all websites and providers using SSL were vulnerable to the bug which allowed attackers to access up to 64 kilobytes of memory on the targeted server and ultimately recover private keys to decrypt any encrypted server traffic (Netcraft, 2014). While a number of large companies were warned about the exploit before it was publically announced, a significant number weren’t. As a result, in the period of time after it was revealed the companies who were slow to apply the fix were more vulnerable than ever.

In the case of STaaS it is likely that an affected provider could allow an attacker to gain access to sensitive information as it is being transmitted to and from the SME.

D. Data Protection Act

Within the UK, one of the main principles of the Data Protection Act states that an enterprise must be able to provide specified information relating to a person by a set deadline or face fines or even legal action depending on the number of missed deadlines. Where clouds are involved, this becomes a significantly more complex task, particularly in the case of rogue clouds. 66% of polled enterprises missed the deadline in a 2013 survey (Symantec, 2013) with 41% being unable to meet the request outright due to being unable to find the requested information.

Alternatively, an organisation may find itself in a place where it is unable to prove their compliance if information is split across multiple locations or even stored in locations that most are unaware of. This was a major concern for more than half of the organisations that responded.

With regards to big data this could quite easily be a significant issue. To take the previous example of customer spending habits, if a customer requested to see this data through a Subject Access Request (SAR) under the Data Protection Act and the SME was unable to locate all of this data, they can face a fine of up to £5000 and the person may seek compensation as well as have the ICO carry out a search for the information on their behalf.

E. Data Duplication

Another issue that is rarely considered is that of duplicated data. An enterprise that is moving from local storage to cloud storage is likely to have a large amount of duplicated data for backup purposes. An enterprise will commonly neglect to deduplicate this data (Remove duplicate copies to ensure only unique data is stored) before migrating to cloud computing, especially where rogue clouds are concerned as it is unlikely that the members of staff involved will even be aware of the duplication of data if they work outside of the IT department. Consequently, this leads to significantly higher costs than necessary. A 2013 survey found that just 11% of enterprises had virtually no duplicated data stored in the cloud with 51% of enterprises having between half and all of their data stored in the cloud duplicated (Symantec, 2013).

Additionally, duplicate data in the cloud can cause data protection issues. A UK based enterprise must delete or rectify any and all inaccurate data stored about a customer if requested to do so. With a large amount of duplicated files in the cloud, there is a large risk of not updating all the duplicated files and thus non-compliance with UK data protection laws. Similarly, principle 7 of the Data Protection Act requires that any accidentally lost data must be recoverable. With numerous copies of duplicated data in the cloud with potentially differing content, the risk of recovering an out of date version of a file is higher.

F. Downtime

As previously mentioned, any unnecessary downtime for any enterprise can have potentially disastrous consequences. As such, when entering into an agreement with a third party there is usually a service level agreement negotiation in which acceptable downtime will be specified by the purchasing company. The third party will then have to either compromise or meet their requirements. When rushing into an agreement, neglecting to state downtime requirements beforehand can of course lead to lost revenue. It is also important to research any potential third party providers to ensure that they can actually deliver on promised uptime percentages as the difference between 99.999%, 99.9999% and 99.999999% uptime is a very costly one.

There are however far greater concerns than revenue lost for X minutes of downtime per month/year. If the third party provides storage space to a large number of clients, the risk
faced by all the clients is increased. The provider may face a data breach with one of its clients and be forced to suspend services for all clients until the breach is investigated as an example.

In the case of big data storage however this issue is not as significant as the storage of important day-to-day files in the cloud. If the provider has downtime, it is unlikely to be too costly for the SME in normal circumstances.

III. CONCLUSION

It is clear that the main threats posed to an SME when considering public clouds for big data storage are improper planning and rushing into agreements in the form of rogue clouds and the impact of this is a number of issues concerning ethics, governance and trust. As such, it is actually not that difficult for the average SME to re-evaluate their own situation and lower their costs in a storage as a service agreement as well as resolving the majority of the issues posed in this paper. When evaluating the situation it is essential to involve IT staff in the process to gain an accurate picture of the businesses needs as well as the currently in place solutions as well as alternative possibilities such as hybrid clouds where sensitive customer information is stored inside the business while non-sensitive information is stored in a public cloud. Only after all of this has been done will it be possible to decide which the better option for the company is and then it is possible to proceed.

IV. REFERENCES

Bishop Fox, 2011. Pulp Google Hacking. Miami, Stach & Liu, LLC.


A Big Data Approach to Security
An Exploration of Big Data Analytics for Security

Nathan Moreton
University of Derby
Derby, United Kingdom
n.moreton1@unimail.derby.ac.uk

Abstract—Security has always been an issue within Information and Communication Technology. The recent emergence of Big Data has opened up a massive area for research and exploration. The use of Big Data analytics can provide a serious advantage to a business or organization, whether it be to enable new products or for the potential security benefits. Thanks to Big Data analytics data is being processed and analyzed at a faster rate than ever before. Big Data has only recently entered the security field but it looks like it is here to stay.

Index Terms—Analytics, Big Data, security.

I. INTRODUCTION

This report will provide a detailed exploration in to the concepts of Big Data, and current data security issues and solutions. After which, the possibility of using Big Data analytics as a security solution will be examined with the ultimate aim of finding how they can be used and if they are effective. Section 2 of this report will begin with an explanation of Big Data. This is includes what it is, who it effects and its impact on enterprises in general. Section 3 provides an analysis of the current issues within data security and explores how effective current security measures are. Finally, Section 4 provides a study into Big Data analytics and examines their potential and effectiveness as a security solution.

II. WHAT IS BIG DATA?

In around 1970 there originated a computing term known as Moore’s Law. Simply put, Moore’s Law states that approximately every two years the overall processing speed and power of the computer would double (Mooreslaw.org, 2014). This law has proven to be true as in the past ten years there has been much significant advancement within Information and Communication Technology (ICT). It is estimated that 2.5 quintillion bytes of data are created every day and that ninety percent of the data in the world has been created in the past two years alone. This data comes from everywhere, whether it is posts on social media websites, enterprise content, videos posted to the internet or transaction records of online purchases (Gross, 2011).

With regards to business, Big Data is a collection of data from both digital and traditional sources; it provides a collection of information that can be used for analysis and discovery (Arthur, 2013). This data exceeds the processing capacity provided by conventional database systems. This is due to the data being too big, moving too fast or not matching the structure of database hardware. To truly gain value from the data a business must choose an alternate way of processing it. The value of Big Data to an organization comes in two forms; analytical use and the enabling of new products. Big Data analytics provide a way to reveal data that can provide an insight in to previously unknown areas, which were often too costly to explore, this could be something such as peer influence among customers, which can be revealed by analysing the transactional, social and geographical data of a consumer (Dumbill, 2014). Companies are able to use data to adapt their services to better meet consumer needs, as well as optimise their operations and infrastructure and enable the discovery of new sources of revenue (IBM Big Data Hub, 2014). Section 4 of this report will provide a more in depth explanation of Big Data analytics and how they can be used for security purposes.

A. The V’s of Big Data

Big Data can be broken down into four foundational defining elements, these are: Volume, Velocity, Variety and Veracity and are known as the V’s of Big Data.

1) Volume

Volume refers to the enormous amount of data involved. As previously mentioned around 2.5 quintillion bytes of data are created daily and it is estimated there will be as much as 40 zettabytes of data in the world by 2020, this is three hundred times the amount of data that existed in 2005 (Reliant-technology.com, 2014). Enterprises are constantly collecting information; this information comes from machines, networks and human interaction with systems. The main attraction of Big Data analytics is the ability to process this data into useful information. It is the sheer volume of data that provides the most immediate challenge to conventional IT systems as it calls for scalable storage and a distributed approach to querying (Dumbill, 2012).

2) Velocity

Velocity refers to the speed at which data is being created today, which is a huge contributing factor of Big Data. The
flow of data received by an organisation is continuous and massive. By using Big Data analytics this data can be analysed in real time by a business to make valuable decisions that may provide a strategic and competitive advantage. However, as with the volume of data, the velocity at which data is being created could create significant problems for current storage methods. The velocity of data creation is not the only issue; the main problem lies in analysing such a vast amount of information within a timeframe that the data in question is still useful.

3) Variety

Variety signifies the ever increasing array of data types. This can include both structured and unstructured data ranging from audio, image and video data to retail transactions, text messages and genetic codes. While traditional analytical and database methods were exceptional at handling structured information that could easily be organised in tables and rows, Big Data introduces the problem of a broader range of data types which cannot be easily structured or analysed by conventional software (NoSQL, 2011).

4) Veracity

Veracity, while it may not be as large of a defining factor of Big Data as Volume, Velocity and Variety. Veracity refers to the biases, noise and abnormalities in data. It isn’t just about data quality but also the ability to understand the data in question. For an organisation, it proposes the question of “Is the data that is being stored, and mined meaningful to the problem being analysed” (Normandeau, 2014). When analysing Big Data an organisation must strive to keep the data clean and error free if they wish to use it as a successful analytical tool.

Big Data provides an opportunity to find insight in to new and emerging types of data and content. Until recently there have been no effective methods of harvesting and analysing this data. Due to current technological advancements it is now possible to process this information in a way that it becomes useful. The next section of this report will analyse the current state of security for data protection, looking at what security is, current data security methods and their limitations with the ultimate aim of discovering how effective Big Data analytics are for security purposes.

III. DATA SECURITY ISSUES

The world of security in the Information Technology department is an ever changing environment. Cybercriminals are taking advantage of the gigantic attack surface found in today’s technological age. This is an age where any individual is able to access business applications and environments from a plethora of different devices no matter where they are in the world. Existing enterprises may be unprepared for what this means for their current data security solutions. As they gather more and more sensitive data such as social security numbers, credit card information, financial records and intellectual property the potential they will be targeted for an attack increases. With the evolution of the internet and storage media, confidential information may be compromised on a larger and faster scale than ever before (Bennett, 2014).

A. What is Data Security?

The fundamentals of data security lie within the CIA triad; confidentiality, integrity and availability. This is a widely used benchmark for evaluating the effectiveness of information systems security.

1) Confidentiality

Confidentiality refers to limiting information access and disclosure to authorised users and preventing it from being accessed by unauthorised users. Basically ensuring data remains hidden from those who are not supposed to see it. This is achieved by certain authentication methods such as user ID and passwords that will uniquely identify a systems user and access control methods which limit the information a user can access depending on their identification. Encryption also plays a large role in the confidentiality of data. Encryption ensures that data cannot be intercepted or accessed during transmission or transport (Clemmer, 2014).

2) Integrity

Integrity refers to the trustworthiness of information resources. It includes the concept of data integrity, meaning that data has not been changed, either accidentally or by deliberate malicious activity. As well as source integrity, meaning that the data has actually come from the entity the receiver thinks it did. Integrity also includes the concepts of validity and reliability. Validity meaning that the data received is correct while reliability implies that the data would be the same when generated under identical circumstances (University of Miami, 2006).

3) Availability

Availability simply means that data must be available when it is required. There is a big problem with keeping data available as attacks or accidents can often bring down systems, this can cause data to be overwritten, modified, deleted or destroyed. An organisation needs to ensure that its data is constantly available and there are a number of ways to do this. This includes load balancing, fail-over and facilities for data backup and restoration (Clemmer, 2014).

B. Big Data and Security Issues

With the emergence of Big Data it is predictable that the number of successful attacks and activities of a malicious nature will increase. The 2013 Check Point Security Report demonstrates this as they found that 63 percent of organisations who participated were infected with botnets. A botnet is a piece of software that infects numerous computers on a network; it can remain hidden and undetected and may allow an outside intruder to take control of the system. The report also highlights that in 75 percent of organisations who participated, a host had accessed a malicious website and that this occurs, on average, every 23 minutes (Checkpoint, 2013). These statistics are further reinforced by the 2013 Cisco Annual Security Report which states that, while all organisations face some risk, larger enterprises have more than 2.5 times the chance of encountering web based malware than smaller companies. It also shows that in 2013 83 percent of this malware was made up of malicious scripts that were designed to compromise system security. The most staggering fact from this report is
that in the last year alone malware on Android platforms has increase by roughly 2577 percent (Cisco, 2014). This amount of cybercrime has a serious cost, in a recent study it was estimated that cybercrime costs the U.S. economy up to 140 billion dollars a year (Los Angeles Times, 2014).

These statistics demonstrate that current data security measures are not enough and that another form of data security is needed. With the recent emergence of analytical techniques for Big Data it has the potential to provide another much needed layer of security. The next section of this report will analyse potential techniques for using Big Data analytics as a security solution, with respect to its overall strengths and weakness as a form of data protection.

IV. BIG DATA ANALYTICS FOR SECURITY

As was reported in Section 1, Big Data analytics provide a way to reveal data that can provide an insight in to previously unknown areas. These techniques can be used to analyse the massive amounts of security related data that are routinely collected by enterprises on a day to day basis. Due to technological advances and the rapidly depleting cost of storage and CPU power as well as the development of frameworks which allow users to take advantage of computing systems storing large quantities of data, Big Data analytics now have serious potential to be used by organisations to improve their current security solutions (Cloud Security Alliance, 2014).

By incorporating Big Data analytics into security programs, an enterprise is able to receive richer content for assessing risk as well as learning what is normal for a particular user, group, or environment. As an enterprise gains more information on both its systems and users it is able to enhance its ability to detect and prevent malicious behaviour or activity which could lead to more severe problems.

A security model driven by big data will have a number of differences when compared to conventional security models. A Big Data security model relies on diverse data sources, both internal and external that when combined can create a synergy that reveals new security related information. It will also contain automated tools that can collect a diverse amount of data types and prepare them for use by an analytics engine. These analytical engines are capable of processing extremely large amounts of data in real time meaning that discoveries can be made instantaneously. A Big Data security model will also include advanced monitoring systems that continuously screen systems and resources that are deemed to be of high value, it will then be able to make assessments based on behaviour and risk models instead of static threat signatures. It should include active controls such as user authentication and be able to assist in the decision making when an activity of high risk is detected. Finally, a Big Data security model must have a high degree of integration with current security and risk management tools, allowing for detailed investigation of potential security problems by analysts and to trigger automated defensive measure in the case of malicious activity (EMC, 2014). Big Data based security models can provide a much wider array of protection when compared to conventional means, this is mainly in the form of threat detection and risk assessment, both of which are required for Information Technology systems to run efficiently and unhindered in a business environment.

An example of the potential use of a Big Data driven security model comes in the form of Advanced Persistent Threat (APT) detection. An APT is a targeted attack against a high value asset that often goes undetected. The most common goal of the APT is to steal Intellectual Property from an enterprise. Over the years APT’s have become more diverse and sophisticated than ever before which means their detection relies heavily on the expertise of human analysts. Big Data analysis provides a suitable solution for APT detection as originally massive amounts of data from many different sources needed to be manually collected, correlated and analysed which was a hugely time consuming process. With the introduction of Big Data analytics a wealth of information can be stored and analysed quickly, leading to improved information about user behaviour that will provide better means of detecting attacks (Cloud Security Alliance, 2014).

The sheer number of uses for Big Data and its reliability and effectiveness have had a serious effect on how it is perceived on the global scale, so much so that Gartner.com (2014) believes that “By 2016, 25 percent of large global companies will have adopted Big Data analytics for at least one security or fraud detection use case”.

However, Big Data analytics have not been met with the same enthusiasm everywhere, it was predicted at a UNITED Security Conference that there will be a large breach from one of the analytics providers in the next 12 months (Baker, 2014). This prediction may stem from the disadvantages of Big Data analytics.

All of the advantages Big Data analytics provide do not appear from nowhere, which is often overlooked by a business or enterprise. There must be some form of interaction with the data for it to be processed at maximum efficiency, and the right queries and security related questions must be asked of the data to receive the most efficient results. After implementing a Big Data based security model an enterprise will often expect it to work without any outsider input, while this is quite often not the case, as to be efficient a data analytics tool will require monitoring on a day to day basis. This can obviously cause unexpected expenditure to organisations that were not expecting to hire new workers. Another potential problem faced with a Big Data security system is the fact that many sensitive files and collections of data are correlated into a single location (Chickowski, 2014). This provides a grand prospect to an attacker who is prepared to infiltrate the system with the intention of accessing restricted data.

V. CONCLUSION

Big Data has had a big impact within the realm of Information and Communication Technology. It is the conclusion of this report that, if implemented carefully and correctly a Big Data security model can provide great benefits to current data security methods. This comes in the form of both improved risk assessment and early threat detection. This
allows an enterprise to run uninterrupted in an age where cyber-attacks are an increasing problem. These security measures are imperative for the future, as was discussed in Section 2, the number of security threats is rising and is showing no signs of slowing down. Big Data security analytics provide a new approach to an area that needs improvement.

In a world that is producing more and more information and increasing in technical prowess every day, there needs to be a solution to the problem of categorising and analysing the large amounts of data that have previously been unknown and inaccessible, the answer to this is Big Data analytics.

VI. REFERENCES


Abstract— Big data has become the buzz word of all large organizations, tempting some into believing the world is at a data revolution. The benefits of big data can be enormous for large organizations but the small and medium enterprises (SMEs) are lagging behind. With the majority of the EU’s GDP produced by SMEs, this paper explores what value can be obtained from big data analytics, with a focus on cyber security and the current barriers to uptake.

Index Terms— Big data, analytics, cyber security, SME

I. INTRODUCTION

This paper explores the meaning and foundations of big data. Delving into how big data can be used to support data analytics, by focusing on the types of analytics available and in use today, and looking at the benefits they bring. It considers the issue of big data for SMEs, looking at how SMEs can overcome these issues. The paper then focuses on how big data analytics can be utilised to counter cyber threats.

II. WHAT IS BIG DATA ANALYTICS?

A. Big What?

Big data is the accumulation of large amounts of data from various sources, in various forms. Just because there is a large amount of data, does not automatically qualify it to be big data. There are other requirements to earning the big data label, these have been summarised as the V’s of big data. There are claims that up to eleven V’s exist but there are three in particular, which form the foundation of what makes data, big data (Hurwitz and Nugent et al., 2013). Veracity is not part of the core three but it is an underpinning requirement in analytics (Claverie-Berge, 2012). A Venn diagram showing the components of the core three V’s can be seen in figure 1.

1) Volume

The amount of data that is being collected needs to be large enough that patterns can be found through statistical analysis; the size of data sets also reflects how the variations of data can be represented (Minelli and Chambers et al., 2013, p. 5). Although large volumes may be labelled in terms of their collective digital file size, it is important to understand that both a 1 Gigabyte(GB) data set and 1 Terabyte(TB - 1000 GB’s) data set can both be considered big data. What defines the right amount of volume is not its total size but the number of records (Wessler, 2013). In the first example, each record could be 1KiloByte (KB) and in the latter example, each record may be 1GB. It is fair to say that there would be a considerably larger amount of 1KB records in the collective 1GB big data file compared to the number of 1GB files in the1TB data set.

2) Variety

The variety of big data forms a key part of what makes big data analysis a technical advancement, over standardised data analysis. In the digital age, there is overwhelming amount of data being passed via user to machine, user to user and machine to machine (Machina Research, 2014). In standardised data analysis, the analysis would take a singular source of data for its results (Wessler, 2013). The overwhelming amount of data that has come from the digital age has changed this, as organisations no longer rely on a single server, service or line of feedback as their analysis source. Social media, clouds, physical infrastructure and smart devices all provide data. The variety is not just in terms of source but also in terms of structure (Singh and Singh, 2012). Data is stored and shared across countless formats, media and structures, it can therefore be considered as heterogeneous.
3) Velocity

In the digital age, data is an evolving and changing form of information, unlike non-digital documentation, such as books, the information is not static. This fluidity of data can be seen in various core structures, which underpin our way of life, such as the New York Stock Exchange that captures over 1TB of data every trading session (BBVA Innovation Center, 2013). The fluidity of data and velocity at which it is created could be seen as a problem for non-digital record keeping but big data thrives on this. Through high volume capacity, organisations can be sure statistical analysis is up to date, to empower smart decisions.

4) Veracity

Veracity of data can be considered the method by which data is scrutinised to ensure validity, if validity is not guaranteed then allowances in the analysis of big data need to be allowed for (Boyd and Crawford, 2012, pp. 662–679). A perfect example of where the accuracy of data is questionable is social media. An estimated 7% of all Twitter accounts are bots (Taylor, 2013).

B. Defining Big Data Analytics

Big data analytics, as it is described in its title is the analytics of big data sets. In a survey completed by TDWI in 2011, only 28% of organisations surveyed said they understand what big data analytics is and have given it a name. In the same survey, 65% of respondents said they know what big data analytics is but had not set a term for it. The lack of consensus in naming big data analytics is one of the barriers to understanding its presence in organisations (Russom, 2011).

C. The different types of analytics

Traditional data is stored in large data warehouses, in a common structured form, which are commonly databases (Oracle, 2013). Analytics of traditional data would rely on constructing complex queries to target and highlight data, which would then be used in an analysis tool or scrutinised by human analysts (Davenport and Dyché, 2013).

The main difference between traditional data analytics and big data analytics comes back to the core 3 V’s. Big data is a variety of data structures, not all homogenous; this creates larger volumes of data, which are supplied by mixed data inputs at different velocities (Ghazal and Rabl et al., 2013, pp. 1197–1208). Traditional analytics relies on carved and crafted data sets, which have followed strict structuring and storage to work with homogeneous tools (Rajan, 2011).

Data analytics can be broken down into four sub-types, as shown below in Table A. These analysis types have been developed and constructed, using traditional data analytics; big data creates a new variable to this engrained ecosystem and provides potential for these data analysis types to evolve (Hurwitz and Nugent et al., 2013). The evolution of historical data analysis types is not without merit, embracing change through big data can help unlock its full potential.

<table>
<thead>
<tr>
<th>Analysis Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic analytics for insight</td>
<td>Slicing and dicing of data, reporting, simple visualizations, basic monitoring.</td>
</tr>
<tr>
<td>Advanced analytics for insight</td>
<td>Complex analysis such as predictive modelling and other pattern-matching techniques.</td>
</tr>
<tr>
<td>Operational analytics</td>
<td>Analytics become part of the business process.</td>
</tr>
<tr>
<td>Monetized analytics</td>
<td>Analytics are utilized to directly drive revenue.</td>
</tr>
</tbody>
</table>

Table A. Types of data analytics (Hurwitz and Nugent et al., 2013)

A. The state of big data analytics in SME’s at the moment

Cyber security analytics are a necessary requirement in the connected age, to counter ever-changing threats. Traditionally the majority of log information from security technologies, such as firewalls and server logs were skimmed at a high level and stored for a limited amount of time (Cybenko and L et al., 2012, pp. 0005–8). Security analytics falls under basic and advanced insight; with the introduction of big data, these analysis methods can provide powerful new tools for cyber security. Removing previous limitations of storage and processing, provides a more granular and extensive exploration of data.

The current uptake of big data analytics by SMEs is far behind that of large organisations. A report by SAS (2013), found that of the 541 UK SME’s surveyed, only 0.2% were using big data analytics. This is in stark comparison to the 14% of large organisations who said they were already using big data analytics in the same survey (SAS, 2013, pp. 13-21). Although the data set is representative of a small portion of the UK SMEs, with an estimated 4.8 million SMEs in the UK alone (Department for Business Innovation and Skills, 2013), it is a realistic reflection of the uptake of big data analytics by SMEs in a modern industrialised nation.

1) Value Exposure

A key issue with SMEs is understanding what is meant by ‘big data analytics’, without fully understanding what the terminology means, many SMEs will not then be able to fully understand the value that big data analytics may have (SAS, 2013, pp. 13-21). It is important not to generalise all SMEs as a potential benefactor of this technology, it is fair to argue that some SMEs lack the core Vs to support big data analytics. The lack of these foundations may be due to the nature of their market and business. SMEs may also not reach the core Vs in the near future. This sector specific benefit is reflected in the current knowledge of big data.

1 Automated accounts, not representative of a real person.
analytics by larger organisations. SAS (2013) found that the lack of understanding of big data was more prevalent in businesses operating in construction, extractive and production sectors.

Logical approaches to big data analytics will be needed by SMEs, to gauge whether it would be value add. Once a thorough understanding of big data analytics has been gained, SMEs can explore what aspects of their business may benefit from big data analytics.

2) Cost

The cost of technology is one of the core factors, which may deter the use of big data analytics by SMEs (Bowman, 2013). The average IT operation budget of an SME is increasing at 1/4 the rate of large organisations (Computer Economics, 2013, p. 6). The current barriers may change, as the price of data storage, cloud storage and open source technologies reduce the overhead costs of big data.

3) Staffing/Skills

Staffing is a major issue for big data analytics regardless of business size, with an estimated shortfall of 140,000 staff in the United States by 2018 (Chen and Chiang et al., 2012). SMEs already have a smaller amount of staff and many are subject to having multiple functions. Many SMEs may struggle to find the right employees as larger organisations create a vacuum effect on the specialist work force, with the promise of higher wages and more promotion opportunities.

4) Lack of right technology

Traditional data analytics took advantage of data warehousing infrastructure, SMEs that have implemented such solutions may struggle to adopt big data using this infrastructure (Devlin, 2011). The limitation of data warehousing is due to the velocity and variety that big data provides. In an attempt to either evolve or replace unfit infrastructure, technologies such as Hadoop and NoSQL have been developed (Searchbusinessanalytics.techtarget.com, 2012).

The fluidity of SMEs infrastructure is also limited due to revenue, skills and scale (MeriTalk, 2012). The current catalogue of big data products are focused around larger organisations, with little speciality or focus on SMEs thus far.

IV. UTILISING BIG DATA FOR CYBER SECURITY

A. The Evolving Threat Landscape

Cyber security is a cat and mouse game, with attackers constantly evolving their techniques and tools to outwit defensive measures that are in place (Ponemon, 2013a). Cyber attacks are a growing phenomenon and as more processes and tools are made digital, such as the internet of things, the number of attacks are set to continue (Proofpoint.com, 2014).

The cost of cyber security attacks can be devastating, even more so for SMEs, as they have less revenue to absorb losses due to data theft or damage. The estimated cost of a single record being lost or stolen in the UK during 2012 was £86 (Ponemon, 2013b). There are three cyber security threats to an organisation, which may be prevented or mitigated by using big data security analytics.

1) Malware

Malware can come in many varieties, with varying complexity and purpose. Some malware may find its way onto organisations networks through user action, hacking or espionage. Malware can cause a range of damage, from data leakage, destruction or physical system damage (Mahmood and Afzal, 2013).

2) Botnets

Botnets can provide issues to an organisation in two forms, the systems of an organisation may become part of a botnet through malware or the organisation may be impacted by the attacks of a botnet (Bailey and Cooke et al., 2009, pp. 299–304).

3) Denial of Service

Denial of service attacks seek to disable or greatly reduce the performance of an organisations internal or external system. The impact can be total loss of business, system or suppression of systems. Denial of service attacks are mainly formed from distributed networks of systems like the botnets discussed previously (Lee and Kim et al., 2008, pp. 1659--1665).

4) Malicious Insiders

Malicious insiders are a difficult cyber security challenge, as it is not purely about defending the outer perimeters. Inside malicious activity was the second highest priority threat in a survey by Ponemon 2013a.

B. How Big Data Can Help

Big data provides the ability to move a step forward with security analytics, from the reactive and signature based defence to proactive and intelligent detections. Big data allows data streams to be inspected and analysed, to look for malicious patterns such as: a user trying to gain access to systems they do not normally use, a system sending beacons requests to an external address, an abnormal amount of external website requests or a sudden change in server settings (Tien, 2013, pp. 127--151). All of these patterns form some of the signs caused by the cyber security threats discussed earlier.

The amount of data organisations are collecting to support big data security analytics is set to increase, 83% of the organisations surveyed by RSA in 2013 are collecting more information security data now, than in the last two years (RSA, 2013).

Big data analytics is allowing high velocity, high volume, and veracious information security data to be quickly stored, sorted, examined and acted upon (Ponemon, 2013a). This allows for new cyber security techniques such as advanced persistent threat modelling, network infiltration detection, behavioural anomaly detection, malware footprint analysis, reactive isolation and real-time forensic capabilities (Mahmood and Afzal, 2013).

In the real world, big data security analytics could have a real impact. A recent network infiltration and data theft from the retail company Target resulted in the theft of 40 million
credit card details and personal data loss of 70 million customers (Elgin, 2014). The data in this hack was taken over a period of months, ironically, the malware detection system installed by Target did alert the security team during the attackers attempt to export all the stolen data, but they did nothing (Riley, 2014). Big data analytics could help prevent attacks coming to this stage, by using intelligent pattern analysis on the data sources provided by big data, it might have detected the attack sooner.

C. Current barriers to successful big data security analytics for SMEs

The current barriers to SMEs adoption of big data security analytics are parallel to those seen with big data analytics in general. It is reasonable to suggest that big data security analytics creates an even more niche and succinct case for adoption, of which only a certain sector of SMEs will benefit from. Such sectors may include defence suppliers, financial providers or research organisations.

V. CONCLUSION

This paper has explored the foundations of big data analytics, unfolding what the current uptake of big data by SMEs is. In doing so, the issues behind big data adoption have been explored, focusing on specific issues with big data security analytics adoption. The cyber security threats to organisations will not cease to cause issues but the inclusion of big data security analytics into the defenders toolset, should allow for more advanced and resilient preventative techniques. Data will continue to grow and by being able to use it effectively, organisations will be better equipped to deal with the threats of tomorrow.

“If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle”

— Sun Tzu, The Art of War

VI. REFERENCES


105


RSA. 2013. The Big Data Security Analytics Era Is Here. [report] ESG.


Taylor, J. 2013. 7% of Twitter users are not human | Our Social Times - Social Media Agency, Social Media Training. [online] Available at: http://oursocialtimes.com/7-of-twitter-users-are-not-human/ [Accessed: 12 Mar 2014].


Should Healthcare SMEs Trust Big Data?

Mohammed Jawad Nadeem
University of Derby
Derby, United Kingdom
M.Nadeem2@unimail.derby.ac.uk

Abstract—This paper examines the usage of big data by healthcare SME’s and questions whether it is currently an appropriate tool to be used based on data quality. Data quality issues occur quite often in regards to healthcare whether it is incorrect patient records or misleading drug trials. Whilst there are views by companies that data errors to some extent can be allowed, as imperfect data can still deliver valid results during analysis. This reports findings go against them and show that with healthcare data errors existing across a wide spectrum of data types, and having an impact on the end analysis, healthcare SME’s should consider carefully what work they use big data for until data issues are resolved.

Index Terms—Healthcare, Big data, Care.Data, veracity, variety

I. INTRODUCTION

Data is one of the most important strategic assets for an organization as it helps them to define and plan what they can go onto do. In order to gain competitive advantages companies can collect and analyse data to start to see trends, which helps them to plan for the future. However the methods and software used to analyse this data are rapidly changing simply due to the amount of data we generate (Macdonald, 2011).

Nowadays around the world we as people are generating around 2.5 quintillion bytes of data a day with companies themselves generating around 300 terabytes of data a week (MIT Technology Review, 2013), the data itself also takes numerous forms from structured data within databases to unstructured data in the form of Facebook feeds. This data can be highly useful for companies as it shows current trends and patterns and in order to take advantage of this before the competition it is imperative for them to be able to quickly collect, analyse and use the data (Storage, 2011).

This is where the term big data comes in, it simply refers to using computing power to analyse large datasets of complex information. Using big data can help companies to increase their competitiveness by driving efficiency, solving problems, predicting trends and gauging customer feedback (Oracle, 2013).

Whilst big data therefore sounds very beneficial there are some drawbacks to it that as we shall examine, revolve around whether you can trust the end results of the big data analysis. Data unreliability is a large issue in companies of any size with IBM saying that by 2015 around 80% of data will be unreliable and that a third of business leaders do not trust the data they have to make decisions on (Easton, 2013). These simple facts on the unreliability of data show that data on a wide scale is considered inaccurate; and if the data is inaccurate then surely big data which analyses it will lead people to make the wrong decisions?

II. HEALTH CARE & BIG DATA

Healthcare SME’s are a natural sector to fit in with using big data simply due to the volume of medical data they have access to ranging from patient records, drug trials and sensor information. Case studies of hospitals and healthcare companies that have used big data have shown that they can use big data to save money, make better medicine or help reduce readmissions. For instance a collaboration between McKinsey & Company and BeyondCore into hospital patient admissions showed 19 patients that were costing hospitals in the area $100,000 each. With this data their plans and visitations were rearranged to save the hospital money (Eastwood, 2013). As well as big data being used to solve problems that healthcare SME’s face it is also revolutionizing the way they work. A hospital in Boston called Beth Israel is pioneering a new approach to the way its emergency department works. Rather than a doctor using a checklist to try and analyse what the patient is going through, they access an app that has data from 2 million patients. Using this they can compare symptoms and reach a quicker analysis about what is troubling their patient (CIO, 2013).

However even here with this pioneering approach that Beth Israel is taking they are hitting problems regarding their data. Mainly their problems stem from the fact that they can’t trust it, for example a doctor will be looking to compare a symptom that he calls “high blood pressure”, but he misses key records related to high blood pressure because they are called by different names such as “hypertension”. This problem with data is causing Beth Israel problems when trying to use data analytics and the data unreliability issues spread further than this.

Data analytics when used by healthcare SMEs pulls on a wealth of data ranging from patient records to drug trials. However each one of these data types can have issues that can result in incorrect data eventually being used, which can go onto have an effect on patients’ lives. For example the drug
Vioxx was released in 1999 and soon had over 80 million people worldwide using it. However, the drug was pulled from the market in 2004 after causing over 90,000 cases of heart disease. It turns out that the drug trials that were used to support the drug through the approval process before release, were both falsified and had key information withheld about the drug toxicity (Mercola, 2012). From here we can see a good example about how incorrect data can go on to have such a large detrimental effect, and is something that big data needs to be wary of.

III. VERACITY

Veracity can be defined as examining how much you can actually trust the data you have. Veracity points out that data you have can be duplicated, be biased or have abnormalities within it, which during analysis can throw up problems and lead to the wrong conclusions. John Easton (2013) an employee at IBM sees veracity as a large issue affecting companies stating that 1 in 3 business leaders don’t trust their information, with 27% of employees being unsure on how accurate their data really is and veracity costing the USA alone around 3.1 trillion dollars a year.

In general when examining data a number of different factors can come up that effect veracity, for example data could be affected by input errors, missing fields of information and wrong calculations (Greenfield, 2012). However because veracity examines the trustworthiness of data you could expand it to look at numerous other issues that can arise such as analyzing the integrity of the data collection methods and people who analyse and use it.

Healthcare companies as mentioned previously rely on a wide variety of data types that can be analysed via big data. However not all of them can be equally trusted, for example external wireless medical sensors can give unreliable data up to 20% of the time (GSF, 2013) and a study showed that the findings of an MRI scan were disputed and unclear nearly 40% of the time (Long, 2012). Some companies such as Deloitte effectively argue these claims saying that even though clean data is very important, you don’t need perfect data in order to receive useful information (Deloitte, 2013). However in healthcare where decisions affect human lives, shouldn’t data be held to higher standards?

IV. DRUG TRIALS

To understand veracity clearly we will be looking in depth at the reliability and trustworthiness of drug trial data.

The procedure of carrying out clinical testing for drug trials is a long and costly procedure that can have a large number of issues that can alter its outcomes. Richard Chin (2009) CEO of a Biotech firm stated that the reasons most clinical trials fail and give unreliable results stem from logistical issues when the trial is occurring that can give unreliable data such as applying the wrong randomization codes so bias occurs. Other issues he says arise from the trial testing the drugs in the wrong way for the wrong goal with the wrong population type. Finally he points out that most clinical trials fail or give unreliable results as they test the wrong dosage of the medicine in the wrong people, thereby giving rise to bad data.

Issues such as these can go onto affect the final drug as inaccuracies in how the drug is made and tested can cause lapses in the understanding of the chemicals that make it up; thereby making it likely that side effects will arise that are not understood.

From Chin’s analysis of the failures of clinical trials we can see that drug trials that are used to support a drug have a chance of being flawed. So therefore even though we can go onto use these results for big data analysis we cannot expect a total guarantee that the results will be accurate; which as we have seen with the drug Vioxx where it was released to the public, with all the research behind it not being understood can cause harm to patients.

However there is meant to be a safeguard to stop this from happening where bad clinical trials can then be weeded out by the research and academic community by undergoing peer review. However here lies a large issue, most companies will only publish results that put their drugs into a favorable light and even trials that do look bad can be manipulated into showing some form of positive result.

Other factors that go onto affect drug trials include that of favorable results being submitted multiple times and companies and universities pressuring their staff to try and make marketable trials to improve sales. A study carried out in Toronto into 164 cancer drug trials showed that the majority of the trials had no actual positive results but overwhelmingly (60%) emphasized small factors to make it look like a positive outcome was received. They also found evidence of bias in the way results were reported with side effects being downplayed (Ubelacker, 2013). The head researcher Tannock pointed out that he believes the staff were pressured into doing these actions; and when you look at the high costs of making drugs with some drug failures costing the company up to 30 million dollars, you can see why people are willing to alter their work (GEN, 2014).

These factors of manipulation and bias heavily affect the reliability of such data being used in big data analysis. As we can see from here drug trials can be flawed and manipulated; and with this incorrect data being used in big data analysis incorrect results can be generated which can go onto affect actual lives. Coupling the inaccuracy of the drug trial data with what the consultancy TCPInnovations (2012) said was a lack of big data talent in the healthcare industry leads to a high chance that the results from analyzing drug trials data will not produce accurate or reliable results, which will ultimately harm both patients and the healthcare industry.

V. VARIETY

Moving now from looking at the flaws of drug trials to examining flaws in data sets, by examining a factor of big data known as variety. Variety is a word used to describe big data that defines the variety of data that people will be examining. For healthcare this can mean how people will examine differing types of records such as drug trials, X-rays or patient records. Or variety can also examine the variety in the data sets
that are used, for example in the USA currently there are at least 44 different large healthcare data sets that examine everything from patients records for Utah to analyzing hospital discharge data. Each of these data sets also comes with its own set of standards and acronyms to use, which can confuse matters during analysis. The Minnesota department of health published an article on the use of standards for e-records stressing that they were needed to combat issues that were rising with interoperability of sharing and sending data (Minnesota Department of Health, 2011). This lack of standards in data and terminology can cost businesses more to analyse and clean the data and can also risk a patients health and safety. Until these issues in data sets are fixed restrictions are effectively being placed on the gains of new information.

Now moving on to examine some of the issues surrounding some of the most used data types in big data from these healthcare data sets, such as patient records and unstructured data (EHDP, 2012).

VI. PATIENT RECORDS

Patient records can be grouped together and analysed by companies to help identify health trends, identify people for research purposes or even improve performance.

However there are some key drawbacks to trusting the information in patient records, whilst electronic records are now preferred over paper based ones, staff still say that the amount of errors have remained consistent (Scientific American, 2013). These simple errors occur more often than we think according to a study done by the Pennsylvania Patient Authority who found that the number of errors were rising on digital records. These digital errors ranged from incomplete fields, text in the wrong places, incorrect option selections and software bugs that would frequently delete or move data. These errors effects were more likely to be amplified compared to paper errors and affect a larger group of people; they found that in their own hospital over 8 years they had caused harm to nearly 4000 patients due to errors on medical reports (Baum, 2012). Electronic records software tend to use the fact that they will create fewer errors as a selling point (Hsieh, 2014), however from this study we can see that not only are electronic records causing more inaccuracies; but they are also due to software bugs capable of changing and deleting correct information that was stored previously.

Errors in medical records can also occur when transitioning to an electronic based system. Doctors within the UK spoke out when they were transitioning records to an electronic system to say that they had found at least 1 in 10 of the record’s being incorrect in some way. They stated that this could lead to problems with treating patients and risk their health (Smith, 2010). These errors found in electronic medical records raise huge questions about how much we can trust the data. As we can see a variety of errors from missing to incorrect data affects the record as a whole and is a widespread issue. How can you trust this data then to analyse and pull information from it if like the above example there is a possibility that over 10% of the information you have analysed is incorrect? As well as this the Scientific American (2013) pointed out that there are worries about the lack of training and regulations that manage these transitions, stating that the US government was encouraging people to move to electronic systems using subsidies as incentives; and in the rush to move to a new system there was a lack of training to make sure the data was cared for. This only raises further issues about the ongoing reliability of data as if people have not been trained into how to manage it effectively, how can we expect to trust the outcomes we gain from it in the future?

VII. UNSTRUCTURED DATA

Other medical data types when used in big data also show varying levels of unreliability such as examining medical image data, unstructured notes or behavioral data.

All three of these data types see their usage rising in the medical field with medical image data expecting to represent 60% of all medical data by 2015, however they are also considered to have medium to high levels of unreliability (Siam, 2013). Siam (2013) set out to analyse these date types and found that due to the processes the data had to go through before it could be analysed by big data; they were being rendered more unreliable. All three data types where being analysed by software that would attempt to detect the kind of data they had, pull the data out if needed and file the record so it could be analysed. However Siam (2013) found that due to the algorithms used, incorrect detection and filing was occurring over 15% of the time for medical images and over 10% of the time for written records as well as key information would be missed from the records.

As we can see from this study then using big data to try and make an understanding of unstructured data is currently not a valid idea simply due to the algorithms being inefficient, any conclusions that are pulled from such an analysis will not be reliable enough to use. This is a point that was also stated by the SIIM organization that works on medical images, they stated that more advanced algorithms were needed as well as more metadata collected on unstructured data before it could be of any use in big data analysis (SIIM, 2014).

From looking at an overall perspective at the variety in medical data types we can see that there are multiple issues that exist for each different data set which can reduce its effectiveness. And when you are trying to create a reliable plan of action to improve your healthcare system or analyse how to create a better medication, having to rely on information that can be incorrect over 15% of the time is not a worthwhile endeavor or efficient use of funds.

Moe Alsumidaie (2014) a healthcare strategist agreed with these viewpoints that big data within healthcare was being used incorrectly with most healthcare organizations unable to deal with data quality and volume issues and unable to use or find appropriate talent to turn any findings into actions. He stated how in his experience he saw how healthcare companies unable to use big data efficiently posed a greater risk to their patients, as they could end up concentrating on the wrong areas to improve or innovate.
Researchers however such as those at SIIM (2014) say that advancements are occurring frequently that will allow for more advanced algorithms to be used in unstructured and structured data that will improve the data quality and analysis; however until then the big data analysis from these data types will not be reliable enough.

VIII. CONCLUSION

Healthcare companies being able to use big data is an amazing opportunity, as it can offer not only revolutionary ideas to help monitor standards but allows for savings and advances in everything from designing medicines to helping patients.

However the issues that currently exist around the reliability of the data and its analysis, whether it is at least 10% of patient records being incorrect to big data talent not being in the pharmaceutical industry; are too large to ignore. Using big data analysis on the data we have now will pose risks as eventually the aims of analyzing patient data is to go onto impact a patient’s life; and if companies are making the wrong decisions or creating the wrong medicines because of incorrect data then they are not only harming themselves financially but also harming the people they set out to help.

In order for big data analysis to work well and provide reliable results, data reliability issues need to be addressed within the healthcare industry through increased regulation and training. And until this happens the usage of big data analysis within the healthcare industry needs to happen cautiously, with the results being treated with a degree of caution.

IX. REFERENCES


Big Data for SMEs
Should SME Managers Trust Big Data?

Anwar Nagi
Business, Computing & Law
University of Derby
Derby, England
100146980@unimail.derby.ac.uk

Abstract — Big Data is relatively a new phenomenon and SMEs are dealing with colossal amounts of information generated from it. Should SME managers trust this for organisational gains, with all the stigmas attached to it? 34% of SMEs already express that large volumes of data captured is not even looked at. The veracity of it is questionable too as evidence shows that not all of Big Data is accurate, credible or has been authenticated. Employees analysing data do not seem to have the right skills set or correct procedures to follow. This paper highlights awareness to SME managers regarding trust issues of Big Data that they do not know, but they need to know.

Index Terms — SMEs, Trust, Security, Data Scientists, Big Data, Managers, Volume, Veracity, Value, Variety, Velocity

I. INTRODUCTION

In this day and age, organisations are obtaining tremendous amounts of data from various sources, which is vastly growing rapidly, and overriding storage facilities. Becoming difficult to control and manage, hence producing Big Data. Big Data is everywhere and is completely anything from emails, retail transactions to social networking data to patient records within healthcare surgeries. Organisations such as small to medium sized enterprises (SMEs) need to analyse these various chunks of data in order to make effective turnovers by looking for trends within the data to target potentially new customers, and to improve overall as an organisation. On the other hand, is the overriding question of trust and whether SME managers should rely on this availability of Big Data? Thus, concerns have previously been voiced regarding the data accuracy, the tools used for analysing the Big Data and trusting the credibility of individuals analysing the data. This paper will aim to address why SME managers should be aware of these questions of trust regarding Big Data and importantly illustrating the reasons as to why they should be aware of these issues. This paper will explain the benefits and drawbacks of SMEs that make use of Big Data by addressing examples to several SME industries as the issue of trust is common theme in all SME sectors. To conclude this paper will bring together the points to address the question in issue.

II. BIG DATA

Various sources have illustrated Big Data in many ways and this is put simply as being large quantities of data, which surpasses conventional databases as illustrated by Dumbill (2012). The data is so widely spread that it is both structured and unstructured fragments of information that requires analysing and manipulating so that the data is meaningful. Big Data can be boiled down numerously from various locations by people and machines, which can be further categorised under social media data, machine data and transactional data. As an example Onepoint IQ (2014) states that 88% of transactional data usually consists of data records such as payment orders, delivery records and storage records. Currently being over 1 billion users of Social media such as Facebook, YouTube, LinkedIn, Instagram and Twitter, this accumulates to 43% of big data as stated by IBM (2012). Without a shadow of a doubt social media is one of the main tools of communication amongst people today and their activities generates an overwhelming amount of data. Data from various sensors such as heart rate monitors and parking sensors on cars, naming just a few are examples of everyday life machine devices that also collect and provide data, (Pritchard, 2012). All of these numerous types of data amass to total 2.5 quintillion bytes of data created each day, (Desouza, 2014). Big Data is categorised under the ‘V’s’ which are variety, validity, value, veracity, volume and velocity. Which make organisations such as SMEs aware about the data that has been compiled together by data scientists, (Massey, 2012) and (Russom, 2011). These particular V’s link in with the question of trust. The benefits of Big Data for SMEs in sectors like healthcare make use of this by improving the efficiency of patient care and improvements in health cost reductions, (SCC, 2014). Medium sized retailers also make use Big Data by analysing their prices of items in relation with customer sales to see whether marked prices require adjusting, (Ohlhorst, 2013).

III. DRAWBACKS OF BIG DATA WITH REGARDS TO THE V’S

The credibility of Big Data has to be questioned despite its advantages and whether does it really help SMEs, and if SME
managers should trust the data which it provides. To consider this, some V’s of Big Data that relate to trust will be assessed in depth to provide a clearer picture of trust issues which SME managers may not be aware.

A. Volume

This amounts to the masses of various unstructured data sets coming into organisations at any one time. This is usually in terabytes (TB) or even petabytes (PB) which is left to data tools or data scientists analyse, (Mobertz, 2013). To provide a clearer indication of the scale of volume of data Pros (2014) illustrated that by 2015 there will be approximately 7.9 Zettabytes (ZB) of data and by 2020, this will have risen over 4 times the amount to 35 ZB. The implications with volume is that before any of the data can be used it is firstly required to be scrutinized and cleansed by removing any redundancies, which takes up a lot of time of SME’s as it is a time consuming procedure, (Doyle, 2013). HP (2013) explains that 34% of SMEs express that due to large volumes of data captured, it has resulted in half of the data not even being looked at. The question of trust has to be addressed here and made aware to SME managers whether if the large volumes of data out there is accurate, reliable, suitable, and up to date for them to make use of. For example, Boyd and Crawford (2011) illustrated that on social media websites such as Twitter remain many bogus accounts that do not represent anybody, and information like this is being collected in large volumes. This clearly questions the reliability of data and the situation is further made difficult as there is no way in distinguishing between legitimate and illegitimate accounts, which questions if SME managers should trust analysing large volumes of Big Data from social media because of this.

B. Veracity

The veracity of data refers to the abnormality within the data that can be anything from errors, inconsistencies or common mistakes, (Normandeau, 2013). As there are large amounts of data that has not been authenticated, how can SME managers trust Big Data to make business decisions? This can be supported from the fact that already 50% of SME’s from various sectors rely on their gut feeling rather than using Big Data to make business decisions as stated by Hesse (2012). It can be assumed that SMEs are doing this as statistics show that poor data from big data immensely costs SMEs between 20-35% of their total revenue (Luckie, 2012). The greater the volume of Big Data the more issues of veracity are in existence; there is a direct correlation between these two V’s as one directly complements the other, (Subramaniam, 2014).

This brings back the question that if there are issues of Big Data through veracity, is it advisable for SMEs to be using this if the quality of data is lacking? LaValle et al. (2011) portrayed that 1/5th of SME industries such as retail and financial institutes are already concerned about the quality of data that is being used. This pinpoints the subject of trust and greatly indicates that maybe SME managers need to implement a new business strategy in place when it comes to analysing Big Data to remove any bad data inconsistencies that may hamper their potential growth.

C. Velocity

The velocity of data is an important trust factor concerning Big Data and SMEs because this refers to the quickness in which data is created as stated by Zaslavsky et al. (2013). As previously explained by HP (2013), there is an enormous amount of volume of data out there and the velocity at which it enters SMEs. The data in most instances is required to be analysed and processed in real-time in order to make rapid, critical business decisions. SME managers need to take into consideration that even though all this data is coming in, the speed at which it should be processed is far too greater as stated by Massey (2012) and Bietha (2013). To put things into perspective regarding velocity, an example of the scope of data generated from just simply a smart phone can be everything you key, browse, send is all logged which provides valuable data information, ScaleDB (2012). This shows that employees who analyse data need to be competent in order to keep up with the rate of velocity from the volume at which data that needs to be analysed. Regarding financial SME managers, they need to be aware that data such as bankcard transactions need to be analysed in real-time. So therefore, can SME managers trust their employees to deliver the correct analysis of data when it is required from the large chunks of volume? (Agrawal et al, 2011).

D. Variety

The trust issues of data velocity is closely linked to another ‘V’ of Big Data called Variety, and this is something which SME managers need to give a thought to regarding the question of trust in big data. Evidently, by now the reader

Areas of Most Big Data Difficulty

Volume and velocity combine to be the two most challenging attributes of big data.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume: 46%</td>
<td></td>
</tr>
<tr>
<td>Velocity: 41%</td>
<td></td>
</tr>
<tr>
<td>Variety: 27%</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. The effects of Volume and Veracity of Big Data (Subramaniam, 2014).

Fig. 2. Velocity and Volume are similar in terms of Big Data Challenges, (Vizard, 2013).
should be aware that Big Data covers a range of data from many locations, and sources that are compiled into data sets (Mongodb, 2013). The question of trust rings alarms here, as SME managers need to consider if the strategy being adopted to analyse numerous data sets is dependable? Grimes (2013) explained that instead of analysing a large variety of data it would be productive to just solely focus on a small, correct percentage of the data set. SMEs like those in the manufacturing or retail sectors would benefit from knowing this so that valuable company time is saved. As we all know in the world of business time equals money.

IV. BIG DATA - TRUST IN EMPLOYEES

The question of trusting Big Data has to be further examined. Factors such as looking at employees who deal with the data needs to be addressed to SME managers, so that they are aware of the trust implications of employees dealing with Big Data.

Employees within SME organisations such as retail, healthcare, finance, sales; contain various skillsets and people have different ways of working, (Green and Martinez Solano, 2011). Considering this, the question needs to be addressed to SME managers whether they are aware that with all this availability of large amounts of data; do they have the right people who can correctly analyse the data in a consistence fashion? The Avanade survey already shows that in excess of 60% of SMEs admitted that their employees have to advance by developing new or existing skills to interpret Big Data so that it can be used for business value properly, (Avanade, 2012). If this is not already carried out by SMEs, how can managers trust their data that has supposedly been ‘verified?’ This is also supported by LaValle et al. (2011) who emphasised that over 30% of SME managers want to be able to trust their data and listed it as one of their top three priorities.

Another employee factor regarding Big Data trust which SMEs managers must be aware of is how the volume and velocity of data affects the productivity of their employees? Avanade (2012) illustrated that 56% of SME employees cannot cope with the quantity of data that requires to be processed. Hayes (2013) who advanced on this issue of trust by explaining that employees begin to change the interpretation of the data by giving it different meanings if they do not understand. As a result, the credibility of the data is being questioned. How can SME managers trust the Big Data if their employees are cutting corners within the analysing process? Evidently, this questions the trustworthiness of the employees too and raises doubts about their lack of professionalism which SME managers should also be aware of.

V. BIG DATA - TRUST IN DATA

As already illustrated by the V’s from above it can be seen that there are many trust related issues attached to Big Data which SME managers should be made aware of. Now looking at this from a trust and security perspective it is clear that volumes of unstructured data collected from various sources has possibility of not being secure, (Intel, 2013). SME managers must consider this security related trust issue because it could be catastrophic if foreign data being analysed happens to be infected; which could potentially harm the SMEs infrastructure, (Allouche, 2013). Another factor relating to trust of data that has to be addressed is the relationship of data regarding volume and storage. 17% of SMEs admit that more data is stored but not analysed, Petersen (2013). This was further considered by Ohlhorst (2013b) which justified that if the data is not analysed, could the value of it be trusted due to being stored for a prolonged period?

VI. BIG DATA – TRUST IN PROCEDURE

SME managers need to be aware of the procedures that are being used to deal with Big Data. If the correct strategy has been implemented to analyse Big Data, SMEs within the retail sector could correctly analyse both unstructured and structured pieces of data, to gain benefits from this, which is pioneer for organizational development, (Smithers, 2013). It is known that some SMEs have not even incorporated a clear strategy to analyse big data as stated by Bughin et al. (2010). This questions if SME managers should trust Big Data because without a clear, structured strategy how are managers able to trust the analysis of the data? This suggests that SME managers should seriously consider implementing this.

VII. CONCLUSION

This paper has considered several areas regarding Big Data and the trust implications associated to it, which SME managers need to be aware of. To re-cap, it has been illustrated by Desouza (2014) that Big Data consists of 2.5 quintillion bytes of data that is created daily and the sheer volume of this is immense. SMEs like the healthcare industry can benefit from this and make use by improving health cost reductions and by making improvements within patient care if the variety and velocity of data is handled the correct way, (SCC, 2014). As
explained by HP (2013) 34% of SMEs do not even analyse all of the variety of data due to the great scale of it. The question has to be asked if SME managers should trust and use this Big Data if it is not analysed properly? Big data veracity is another important issue for SME managers to know, and SME sectors such as retail because it is clear that there are inconsistencies and errors within Big Data as being outdated, as stated by Zaslavsky et al. (2013). Therefore, how can SME managers then use and trust this to make business decisions? Despite the initial trust of the actual data itself, it can be argued that employees need to develop or learn new skills properly to handle, and interpret Big Data for business use as stated by Avanade (2012b). It is also clear-cut that SME managers should be aware that productivity of their employees is hindered, as they cannot cope with the quantity of data that is required to be processed. This puts into question how can SME managers trust the use of Big Data if their employees are not even giving 100% commitment towards their job? It is also known that some SMEs do not have a clear strategy in place to analyse Big Data in the first place, (Bughin, et al, 2010). SME managers need to be made aware to all of these trust limitations addressed regarding big data. If they implement a clear strategy to handle big data correctly, re-train employees, analyse all the given data professionally and adequately, then they could trust Big Data and use it towards their advantage in the business world.

VIII. REFERENCES


**Big Data for SMEs**

**Do SME Executives Trust Big Data?**

Nazir Nagi  
University of Derby  
Derby, England  
100166319@unimail.derby.ac.uk

**Abstract** — Big Data benefits SMEs and that is no secret. The question of trust is a stigma that remains attached and whether SME executives can trust Big Data due to its inaccuracy and incorrectness. 46% of SMEs made wrong decisions from unreliable data, which also questioned if employees analysing the data could be trusted? It has been found that SME executive do not know about certain trust issues from the Big Data in relation to the human factors of trust and the Big Strategy. The information that is highlighted in this article is aimed at SME executives and to make them aware if they can trust the Big Data based on asking questions.

**Index Terms**—Big Data, SMEs, Trust, Challenges, Volume, Veracity, Velocity, Human, Factors, Strategy, Executives, Errors

**I. INTRODUCTION**

It is by far no secret that in today’s business environment, the evolution of Big Data has enabled many organisations to use the captured data that is available from various sources, in order to make important business decisions based on looking at trends and patterns on generally how people live their life. Big Data is everywhere around us. From data on social media, such as Facebook and Twitter to credit card transactions, meaning every little information captured about a certain individual can be then linked together; hence creating a vague profile of that individual, which would then be used by organisations such as small to medium sized enterprises (SMEs) for advertisement or marketing campaigns to benefit the growth of the company. With every positive, comes a negative and with Big Data there is no exception to this. Even though SMEs benefit from Big Data, there are still concerns from the SME executives regarding trust issues with the data itself, trust in the people who analyse the data and trust in the strategy that is used to analyse the data. This paper is going to look into these trust factors and question if SME executives should trust Big Data, relating to point they may not be aware of and why they need to know this. This paper will further discuss the positives and negatives of SMEs who use Big Data, relating back to the issue of trust. Due to the issue of trust, it relates to various SME sectors so this is why this paper will not focus on a specific sector. In conclusion, this paper will justify whether if SME executives trust Big Data or not.

**II. WHAT IS BIG DATA?**

There are many ways to define Big Data but SAS (2013) explains, Big Data as being huge quantity of unstructured information, which cannot be examined using ‘traditional’ systems and tools. Villars (2011) further explains that Big Data is more to do with how organisations such as SMEs in particular need to deal with keeping up with this large, growing data. When explaining the origin of the Big Data, we have to realise that what we do on a daily basis evidently results in humans or machines leaving behind traces of information, consciously or subconsciously. This is why 90% of the world’s data was created in the last 2 years, (Sintef, 2013). Activities such as using social media, online shopping, emailing, videos and images; are all examples of common sources as to where some of this Big Data is coming from, (Gopalakrishnan, 2013) and (IBM, 2012). As you can imagine, vast amounts of this data is being generated daily, which means the overall quantity is going to be huge. HP (2013) explains, that 2.5 Exabyte’s of data is generated every day. Big Data has defining properties related to it in order to understand what the data is about. These properties are known as the ‘V’s of Big Data’, (Hurwitz et al., 2013). The V’s in relation to trust in Big Data are Volume, Variety, Velocity, Veracity, Value and Validity, (Sagiroglu and Sinanc, 2013) and (Normandeau, 2013). SMEs within various sectors make use of Big Data for their business benefit. For example, small retail organisations use Big Data to monitor customer-buying patterns, (CE, 2014). Small delivery businesses use Big Data from the GPS to monitor routes to lower fuel prices (Sahholk, 2013). Medium sized financial companies use Big Data to tailor their services to meet their customer’s needs. This is why 21% of financial SMEs have invested in over $1m worth of Big Data, (Versace, 2012).

**III. CHALLENGES OF BIG DATA V’S**

It needs to be understood that even though Big Data is beneficial for some SMEs, the executives still need to know how the V’s of Big Data can cause trust problems for SMEs.

**A. Big Data: Volume**

Schaeffer and Olson (2013) and Russom (2011) explain that the volume of Big Data refers to the size of the various information data sets, that organisations receive which needs
analysing by employees or systems. The main problem with this data, is that it needs to be stored and then analysed in due time. Microsoft (2013) illustrates that SMEs currently store a quantity of about 100 terabytes of unstructured data. The problem for SMEs is that it takes a lot of time and resources in order to clean the unstructured data first before it can be analysed. Bitidy (2012) and Ohlhorst (2013) illustrate that SMEs now use various means to clean the data and it can be conducted using specialist tools or manually by employees. In relation to trust and the volume of the big data, the big size of the data does not always mean better data for various SMEs. Boyd and Crawford (2011) portray that SMEs who are analysing data from social media such as Twitter and Facebook are not aware if the data they are analysing is accurate or not. The problem with this is, there are many accounts on social media that are ‘bot’ or many individuals have multiple accounts, which means, this will give inaccurate analysis due to not knowing what is reliable as it does not represent the whole population, (Ferguson, 2013). This is why 46% of SMEs made wrong decisions from unreliable data. Evidently, this raises serious questions of trust regarding the volume of big data especially that has been extracted from social media. Therefore, this problem needs to be addressed to various SMEs for example in marketing who farm and then analyse social media data; so they are aware of this trust problem.

B. Big Data: Veracity

Walker (2012) and Liliendahl (2013) inform that veracity of Big Data refers to the inconsistency and uncertainty of the data. Inevitably, if the data SMEs are analysing contain various errors and incompleteness, then how can they trust the data? This is why 33% of business executives do not trust the data they have in order to make informed decisions, (Petersen, 2012). The issue of trusting Big Data is becoming a serious concern now especially when statistics show that the rate of volume of Big Data is continuing to rise tremendously. This is why by 2015 80% of all the data gathered will be uncertain. (Subramaniam, 2014) and (Marin, 2013).

C. Big Data: Velocity

ScaleDB (2012) makes clear that the velocity of Big Data relates to the volume and veracity because the data in this era is being generated in large quantities and at a very fast pace, from different varieties. We have to keep on reminding ourselves that all this new data needs to be stored, analysed and then processed, (StateTech, 2013). This is why the velocity of Big Data, refers to not only the speed at which the data is created but also the swiftness at which this data is needed to be processed. There is urgency for some SMEs who need to have this data to be analysed in real time or as soon as possible in order to make critical business decisions, (Rijmenam, 2013). Modern cars have over 100 sensors that measure the changing data, (Pritchard, 2013). All this data needs to be analysed so that the unstructured data then has some sort of value given to it. The issue of trusting the employees who deal with the velocity of Big Data is a concern for SME executives. Rogers and Sexton (2012), explain that 39% of the data analysed is not in real-time enough. This means that this raises serious questions regarding trust and if SME executives can trust their employees to deal with the velocity along with the volume of big data in order to deliver real-time information. Therefore, this problem needs to be addressed to SMEs for example in online retailing who need statistical data to be analysed efficiently and require rapid feedback.
IV. HUMAN FACTORS OF TRUST WITH BIG DATA

The issue of whether SME executives trust Big Data can be looked at from many angles along with viewpoints. The competence of employees who analyse the data comes under scrutiny too. This then raises questions regarding Big Data and trust for various SME executives. SME executives need to ask themselves and question if they can trust the skills of their employees who are analysing the data. Buguin, Chui and Manyika (2010) illustrate, that some employees from C-level executives do not have the right skills along with the talent to capture the main value from the data. Shah (2011) further elaborated that 40% of the employees do not have the right skills along with the idea when it comes to analysing the data. SMEs from various sectors need to be aware of this trust issue and should ask themselves, “Why is this the case?” Clearly if the employees do not have the right skills to analyse the data then surely this raises serious doubts within the data that is being interpreted and then being used by SMEs. The issue of training employees to be able to analyse the data effectively is a sticky topic for some SMEs because as we all know, training costs money and this is something some SMEs are trying to avoid in order to save costs, (Evans, 2013). We can then question the SMEs as to what is more important, providing adequate training for data analysis or losing revenue from poor data analysis. Luckie (2012) illustrated, poor data analysis costs SMEs between 20-35% of their revenue. Hesse (2012) explained this is why 91% of SME executives in the UK are relying on experience when it comes to decision making. Poremba (2013) noted, some employees use guess work to analysis data. This is why SME executives need to be aware of this trust issue with Big Data and the employees who are interpreting the information. If it is analysed wrongly, this can have a negative impact on the value of data, for the SME and that is why SME executives need to be aware of this.

V. TRUST WITH BIG DATA

As previously mentioned with the V’s of Big Data, there is a trust issues that need to be addressed by SME executives. This is why data governance is an important factor in data mining, especially when value needs to be extracted from the volume and velocity of data, (Howard, 2013). If the data needs to be analysed in real time, due to the velocity will SME employees assume the data they have is accurate, if they only have time to remove just the minor duplicate errors? SME executives need to address this situation, as they cannot afford to make decisions based on assumption from the data. In addition, when it comes to Big Data, SMEs store the data so that it can be analysed when the time is right, (Ohlhorst (2013)). SME executives need to be aware of this because the longer the data is stored, the value of the data depreciates which then becomes outdated, (LaValle et al, 2011). This is why SME executives need to be aware of the quantity of data being stored because the question can be asked, how can SME executives trust the value of the data if it has been stored for a long time?

VI. TRUST WITH THE BIG DATA STRATEGY

Before SMEs analyse the big data, there needs to a strategy in place regard who or what is going to analyse the data, (Zhang, 2012). Labrinidis and Jagadish (2012) illustrate that up to date systems need to be implemented in order for SMEs to be able to process the large data. Specific analytic tools such as software and cloud services are available in order for SMEs to process and analyse the data, (Dumbill, 2012). The issue of trust and Big Data strategy is questioned with the use of cloud services because Schaeffer and Olson (2013) point out that the issue of the veracity of the data becomes an issue with cloud service providers over the ownership of the data that is analysed on the cloud. SME executives need to be aware of this trust in analysing Big Data because then they have to question if they can trust data analysing tools such as the cloud.

VII. CONCLUSION

Overall, this paper has looked into SMEs use of Big Data and the trust factors SME executives need to question in relation to the use of Big Data.

The volume of the unstructured data affects SMEs, and executives need to be made aware if they can trust the employees to clean the data properly as well as if they can trust the sources as to where the data is coming from, (Bitiday, 2012) and (Boyd and Crawford, 2011). The Veracity of the data questions SME executives on the credibility of the data, especially when the volume of the Big data is ever expanding, resulting in major inconsistencies and errors within the data, (Liliendahl, 2013) and (Petersen, 2012). How can SME executives trust and use the data, if there are unnecessary errors? In terms of velocity along with human errors, due to some data being needed in real time, can SME executives trust the employees to deliver the right data without errors? Due to the volume and velocity of the data, some data is stored and processed later or never becomes processed at all, (EIU, 2011) and (TIBCO, 2012). SME executives need to question if they can trust the use of the data once quality of it has been
outdated. The skills of the employees need to be addressed because as Shah (2011) explained, 40% of the employees do not have the right skills or knowhow when it comes to data analysis. SME executives need to seriously question the trustworthiness of the data they receive from their employees after the ‘analysis’ because if the employees do not have the knowledge then how can the data be accurate? Maybe SME executives need to provide efficient training along with encouraging employees to collaborate when it comes to analysing data analysis. SME executives need to question if they can trust the data they use based on the strategy that is being adopted. Schaeffer and Olson (2013) explained that based on the data being used on cloud services, can SME executives afford to trust the service provider, as the data can cause ownership problems. SME executives need to be aware of this before adopting strategies of the cloud service. It is clear, SMEs do benefit from Big Data however, it cannot be ignored that Big Data does have its drawbacks and the issue of trust is a serious factor for SME executives. By looking at various points from the trusting Big Data, the trust of employees who analyse Big Data and the trust factors from the Big Data strategy, SME executives need to be aware of these serious issues before making any conclusive business decisions based on the data. There is no clear-cut answer to the given question because SME executives need to be aware that changes are needed within SMEs from employees to the data itself. If these changes are made then maybe SME executives can gain trust in Big Data.

VIII. REFERENCES


Preventing Big Data Governance Risks:
A Collective Action Involving HR and Employees

Patrick O’Sullivan
School of Computing and Mathematics,
University of Derby, UK
Student ID: 100241074
Derby, UK
P.OSullivan2@unimail.derby.ac.uk

Abstract—The purpose of this paper is to provide SMEs with an awareness of Big Data. There are a variety of challenges that they will face and they face these challenges because they lack the understanding and an awareness of Big Data. However, because of this lack of awareness and understanding strategic members of HRD (human resource department) may need to determine solutions that prevent and minimize the risk to the business. This paper provides a discussion and evaluates a governance matrix that emphasizes the provision of a structure that will create an effective awareness and understanding of Big Data.

Index Terms—Big Data, Challenges, Employees, Awareness, Human resource department, Governance matrix, SMEs

I. INTRODUCTION

Information Governance is the specification of decision rights (Logan, 2010) combining structures and processes that ensure the effective and efficient use of information that will enable an organisation to achieve its goals (Logan, 2010). However, for a process to be effective the employees must follow procedures and work policies (E.P.A, 2003). They must attempt to find the right balance between the management, employees and HRD (human resources department) (Cotton, 2012). Historically the size of a company has always required specific members of the HRD to play a strategic role within the business. However, Big Data processes are posing many challenges for organisations due to their lack of understanding, knowledge, attitudes, behaviour, training and expertise in this field (Boyd & Crawford, 2012). This paper points to factors of governance failure and justifies implementing a secure framework which may mitigate some of the risks.

II. FAILURES OF INFORMATION GOVERNANCE

This section looks at some factors of information governance failure that may have an impact on SMEs. Some parts of ISO 27002 principles are discussed within as they inform SMEs of practices necessary for the protection of information security.

A. The failure to perceive

Many senior managers fail to perceive accumulated data as they are accustomed to information overload (Logan, 2010). This is due to poor management skills, attitudes and behaviour (Secure Thinking, 2011). SocITm (2014) suggests that senior management within many companies are mainly to blame for information security breaches. However, they need to use best management practice and processes (Secure Thinking, 2011) as this is vital for good ‘information security practices’ (Secure Thinking, 2011). Including the principles and practices from ISO 27002 may also reduce business risks while providing real benefits to the customers.

B. The failure to act

Different groups and many individuals have different views of information management (Logan, 2010). Information has to be managed in such a way so that users realise what has been accumulated (Normandeau, 2013). However, this process takes time away from what they require in order to get their jobs successfully completed, meet their objectives and secure promotions (Logan, 2010). As such a clash of interest may occur between lower level staff, IT staff and higher level senior management which can result in bad behaviour towards governance processes because different groups might act in ways that allow them to put their own interests first (SANS Institute, 2014).

C. A lack of accountability

As seen from both Enron bankruptcy case and (Oppel Jr. & Sorkin, 2001) and the WorldCom bankruptcy case (Romero & Atlas, 2002) the issue of trust within a business is an important
priority. Previous research on governance of IT suggests that there are conflicts of interest and IT governance issues need to be resolved before leading to overall success (E.P.A, 2003). By developing policies and by communicating them and adhering to these policies can prevent risks and mishaps.

D. Employees lack of awareness

Cobb (2014) states that Big Data security is this year’s hot topic but Oltsik, (2013) suggests because of how Big Data is being used security infrastructures are no longer adequate and there are many risks to information security. SMEs need to look at the rapid change within the IT industry to determine if there is a security skills shortage in the organisation. If so they need to implement staff skills awareness training that involves using SIEM (security incident and event management systems) (Wood, 2014). By paying careful attention to how they use secure information and being discreet with media files and other online users might prevent security issues. The SANS institute (2004) has suggested that many employees lack real functional experience that is necessary within their roles and as such they create security incidents. This is because the way employees access information and how this process has now changed over the past two decades needs to be thought about (Kenexa, 2013). SMEs and employees need to develop their knowledge of security incidents and information governance by following the practices and policies shown within ISO 27002.

E. Skills-attitudes-behaviour

A recent report has suggested that there seems to be a shortfall of prospective Big Data experts globally (Price & Wilkinson, 2013). Recruiting Big Data experts who have the knowledge or training employees within the company would compensate for a shortage in Big Data staff. By altering the way in which their information security processes are handled and by involving effective information governance practice in the organisation might prevent many security issues (PWC, 2014). A report by PWC (2014) states many industries have become the victims of crime due to some cultural attitudes and behaviour and attitude of many senior managers and CEOs (chief executive officers). There are many cyber-attacks, fraud and threat to information security due a lack of fundamental knowledge in employees (Hern, 2013).

F. Privacy & Capacity

Information sharing should be a shared decision making process that includes open discussion, decisions, action and outcomes (Boyd & Crawford, 2012). Some SMEs may be suppliers to bigger companies or rely on support organisations from payroll services to marketing professionals (Cavoukian & Jonas, 2012). As such they need to look at the volume of their data that is accumulating daily (Dumbill, 2012). This is because volume calls for "scalable storage and a distributed approach to querying" (Dumbill, 2012). There are many SMEs who do not have the capacity to process any of this information due to their relational database infrastructure and might decide to use an external agency or an organisation who structure their data because they don’t have the capacity or the skill sets required (Dumbill, 2012). To overcome privacy issues they need to have a systematic approach and implement a privacy and security policy that reflects the needs and risk of the business and make sure they are consistent with national standards such as ISO 27002 (Cavoukian & Jonas, 2012).

III. EVALUATING A GOVERNANCE MATRIX

The following section evaluates a Governance matrix with the intention of analysing the parts that might prevent some risks. While there are many different Governance matrixes used in business process and principles of Governance are the same. Governance must now be addressed by many SMEs due to data volumes rapidly increasing and data environments becoming more and more complex (IBM, 2013). As such, a matrix may be used to mitigate some of these concerns.

1) Governance formalities

Basic Governance formalities need to be in place to cover a board of directors, shareholders and shareowners. To overcome issues involve a corporate board member or a senior executive with the responsibility for improving corporate practices (IFC, 2014).

2) The Board of directors

The board of directors need to meet regularly and consult one another on a regular basis. The board of directors should have a different composition of competencies and skills. As such they will have a range of duties that should allow them to deal with any oversights (IFC, 2014).

3) Control environment and processes

Use adequate internal control systems that are in place and make sure they are reviewed regularly by external auditors in accordance with national standards (ISO, 27002). Following and enabling consistent practices might prevent any risks.

4) Disclosure and transparency

Make sure adequate accounting and audit systems are in place that include any financial reports that have been prepared by an internal accounts and any yearly financial statements need to be audited by external auditors. Following national standards enables good auditing practice (IFAC, 2014).

5) Treatment of minority shareholders

Shareholders need to be kept informed of company policies, any company strategy that is used and results of operations (IFC, 2014). Include best practice code as this will mitigate some risks and leave the company in a better position when dealing with shareholders.

6) Business value

Business value is important for customers and stakeholders. Using ITIL best-management practice allows many companies to succeed because business value is measured in the overall success of the business (ITIL, 2013).

IV. EVALUATING BIG DATA PROPERTIES

The following section evaluates some properties of Big Data because having an understanding is important as they include volume, variety, velocity and validity of data which are now seen to be crucial to SMEs in how they handle their data.
A. Volume

Big Data implies an enormous volume of data (Normandeau, 2013). Over the years employees have been creating data but due to the advance in technology there is a reliance on data that is generated by computing machines and computing networks. The volume of Big Data is a growing concern because of how much needs to be analysed. Schaeffer & Olson (2013) suggest this is because of the high volume of data sets that might range from a few dozen terabytes to many petabytes of data that is located in a single set. Schroeck, et al (2012); IBM (2012) have stated that 2.5 quintillion bytes of data are created daily. Given that SMEs can differ in size and the number of their employees it is understandable that how they adopt to using Big Data and interpret its use will vary across each sector.

B. Variety

Variety refers to the different types of Big Data (Rouse, 2014). Different types of data can include emails, videos, photos audio and pdfs etc. The wide variety of data is causing concerns due to the nature how it is deployed why it is being deployed and for what purpose (Boyd & Crawford, 2012). This is because a lot of data is unstructured which is creating many problems and concerns for data storage mining and analysing (Normandeau, 2013). There may also be constant change in the variety of data depending on the size of the SME (Lavastorm, 2014).

C. Velocity

Velocity refers to the speed at which data is now generated and the speed of how that data is constantly moving around (SAS, 2014). In this age of technology how the Internet and mobile phone products are currently consumed and how many products are now delivered are increasingly interconnected and instrumented (Dumbill, 2012). How SMEs are reacting to deal with data velocity presents many challenges (SAS, 2014).

D. Validity

Big Data is now becoming a part of everyday life for many companies and as such the validity of data must also become a critical component of analysis (Johnson, 2013). It is important that content of any data is checked and measured correctly. The criteria and the results of data analysis needs to be understood (Johnson, 2013) so it can therefore be used effectively.

V. GOVERNANCE AND SKILLS PROPOSAL

The following section proposes and justifies implementing a skills competency framework. The framework is based on best practice models such as best management practice and it will include components of ISO 27002 because Big Data processes are now bringing a range of security and governance concerns. This is creating challenges for SMEs and by involving various components that are relevant may help to mitigate the risks.

A. Management-Employees

A Business has to plan carefully to ensure they have the right number of suitable employees and that management attain the necessary skill set (Needham& Dransfield, 2000). As such they need to look at the challenges posed by Big Data and recognise a growing need to support the management and the employees (Needham & Dransfield, 2000). This can be done by including a training program that focuses on investing in the work force that will balance employee and management skill sets, improve capacity and enhance the business (IKM, 2008).

B. Ethical responsibilities

Business ethics are seen to be good for company performance and ethical responsibility should stem from the board of directors (Dando, 2013). The board of directors should involve ethics within the business because these are relevant to how their business is directed. By incorporating business ethics it allows for values and standards to be followed and enables the board to structure and set examples which involve leadership. Casson, (2013) suggests that if more companies followed a code of business ethics it would enable them to be more focused on the purpose of the business.

C. Information security

To overcome many issues management and employees need to develop their knowledge of security and governance incidents. Determine business requirements for information security as this allows the implementation of commonly accepted security controls which aims to prevent many accidents or risk of fraud (ISO, 27002).

D. Compliance

To overcome many issues of compliance companies need to identify legislation that is applicable to the business in order to meet business requirements (ISO, 27002). IT compliance and IT issues need to be resolved before leading to overall success (E.P.A, 2003). By developing their compliance policies and by adhering to policies should prevent many risks. While enabling compliance with legislation allows the opportunity for value to be added to the business as this is vitally important for customers and stakeholders.

E. IT & Business Value

IT is seen as a commodity (Carr, 2003). However, De Vera & Murray state in their report the Art of Listening (2013) that ‘you need to be intentional and have a larger vision’. Include ITIL best-management practice allowing companies to look at customer and business value. In doing so business value can be measured in the overall success of the business (ITIL, 2013) and if proper ICT strategies and goals have been implemented customer and business value can also be recognised (ITIL, 2013).

F. Setting direction

It is the responsibility of all managers in companies and the HR department to fully recognise the needs of their employees (Needham & Dransfield, 2000). This can be done by including a sufficient training program that focuses on investing in their employees. When employees feel valued and they understand their worth to the company they may attain skills and expertise in Big Data knowledge. In doing so it may allow for them to
continue to improve the overall capacity of the business (IKM, 2008).

G. Making employees aware

SMEs need to look at the rapid change within the IT industry to determine if there is a security skills shortage in the organisation. If so they need to implement staff skills awareness training that involves using SIEM (security incident and event management systems) (Wood, 2014). By paying careful attention to how they use secure information and being discreet with media files and other online users might prevent security issues. The SANS institute (2004) has suggested that many employees lack real functional experience that is necessary within their roles and as such they create security incidents. The way employees access information and how this process has now changed over the past two decades needs to be thought about (Kenexa, 2013). SMEs and employees need to develop knowledge of security incidents by following the practices and policies shown within ISO, 27002. Many SMEs are still unaware of how to address specific aspects of Big Data and they need to determine the area of initial focus.

VI. CONCLUSION AND RECOMMENDATION

SMEs need to decide if their staff development processes are sufficient enough to enable opportunities, allowing employees to develop an awareness and understanding of Big Data. To prevent many risks they need to understand Big Data as the amount of data volume and variety increase the greater the risks become (Normandeau, 2013). This is because some SMEs lack the basic fundamentals in security and governance practices. In order to compensate for these SMEs need to have good governance practices and also determine if their security practices are adequate enough and that they follow the national standards of ISO, 27002 (Gurbaxani & Kraemer & Melville, 2004). This paper therefore recommends the implementing of a structure involving HRD, management and their employees because this will allow them opportunities to work together to develop strategies so that may prevent the risks that Big Data poses.

VII. REFERENCES


Public sector information security breaches are a significant concern, with a growing body of literature addressing the issue. For instance, Logan and Debra (2010) discuss the challenges of information governance and why it is so hard. Additionally, Lavastorm Analytics (2014) highlight the need for a strategic approach to data variety and volatility.

IBM (2013) has emphasized the importance of ensuring that information governance is adequately addressed, as it is a cornerstone of corporate responsibility. The company has developed frameworks and tools to assist organizations in managing their data assets effectively.

IFAC (2014) has published guidelines on auditing standards that are crucial for maintaining the integrity of data governance. These standards serve as a baseline for assessing the effectiveness of data management practices.

ITIL (2013) has outlined best management practices for information governance, offering a roadmap for organizations looking to implement robust data management strategies.

Kenexa (2013) has conducted research on the impact of redesigning work processes and has concluded that such changes can lead to a more efficient and effective workforce.

Navetta, D. (2013) has explored the legal implications of big data, addressing the challenges and opportunities that arise from the collection and analysis of large datasets.

Needham, D & Dransfield, R (2000) have authored a book on advanced business management, which includes discussions on the integration of information technology with organizational performance.

Normandieu, K. (2013) has written about big data failures, an issue that highlights the importance of rigorous data governance practices to prevent breaches.

Oltsik, J. (2013) has discussed the role of white paper reports in the context of big data security, emphasizing the need for actionable insights.

PWC (2014) has released a report on global economic crime, emphasizing the importance of robust data security measures to prevent economic losses.

Price & Wilkinson (2013) have explored the implications of big data analytics, underscoring the need for data scientists to possess a comprehensive skill set.

Riazi, S (2013) has investigated public sector information security breaches, noting the failures of management and the need for more robust data security protocols.

Rouse, M (2014) has defined the concept of the 3Vs (Volume, Variety, Velocity) in the context of big data analytics, highlighting the need for a comprehensive approach to data management.


EU and US Personal Data Protection Laws and their Relationship with Big Data

Ben Pimlott
School of Computing
University of Derby
Derby, England
B.Pimlott1@derby.ac.uk

Abstract – A look at the current EU and US data protection laws and the relationship they have with Big Data, especially where one undermines the effectiveness of the other, and what may change going forward

Index Terms—Big Data, Data Protection, Safe Harbor

I. INTRODUCTION

In the modern world a ensuring the privacy and security of an individual’s personal data is considered highly important but the most effective way of doing this was and remains a matter of considerable debate. The differences between the European Union’s centralized regulation and the United State’s self regulation lead to the creation of the Safe Harbor Framework in order to minimize the difficulty this difference caused to organizations in both regions. However these laws were set in place well before the Big Data was a subject of consideration and the clashes between the nature of Big Data and current data protection laws

II. EUROPEAN UNION DATA PROTECTION REQUIREMENTS

The current EU legislation on personal data protection is Directive 95/46/EC, also known as the Data Protection Directive, which came in to force in late 1995, with a deadline for its implementation by EU member states of late 1998. The Directive covers all processing and filling of personal data which takes place within the EU, regardless of whether the controller of the data is located within the EU or not.

Personal data is defined as any information related to an individual who is or can be directly or indirectly identified, whether by an identification number or via specific factors in the data relating to any aspect of the individuals identity (European Parliament, Council of the European Union, 1995). Processing of personal data is any action or set of actions taken upon personal data and filling is any storage of the same. The controller is whatever company or individual is responsible for the data and the Directive applies as long as either the controller or anything that is used to the process the data is within the EU’s territory. (European Parliament, Council of the European Union, 1995). The Directive also specifies that when any personal data is collected or processed this must be for a specified purpose that the individual the data relates to is made aware of and consents to.

Equipment used to process the data includes the personal computers of users so if a user in the UK purchases something from an online store in the US any personal data such as name or bank details technically falls under the directive, as would the reverse case. If a European company purchased data from a company in the US it would also fall under the directive, regardless of if it was ever stored or processed in Europe as the controller of the data is located within the EU.

The regulations of the Directive allow for the free transmission of personal data between EU countries. The transfer of personal data to non-EU countries is only permitted when the non-EU country provides an ‘adequate’ level of protection for the data, judged based on the legislation of the country and/or any applicable professional codes of practice enforced there (European Parliament, Council of the European Union, 1995). The definition of what counts as adequate data protection is flexible depending on the sort of data being transferred and the circumstances in which it is being transferred.

III. UNITED STATES DATA PROTECTION REQUIREMENTS

The US has no overarching law on data protection but instead has a collection of sector specific laws which are combined with a variable selection of state level laws on various aspects of data protection to produce a complex selection of data privacy and security laws (Mousedale, 2012). These laws include the Health Insurance Portability and Accountancy Act (HIPAA) which applies to healthcare institutions and the Fair Credit Reporting Act (FCRA), applying to data collected by consumer reporting agencies, among many others. In total the US has around 20 national...
level data protection laws and hundreds of state level laws (DLA Piper, 2012). It is important to note that the definition of what qualifies as personal data varies depending on which set of regulations it is being considered under.

While the US has few restrictions on the transfer of personal data outside the US the position of the various regulators is that the applicable laws still apply after the data leaves the US and US regulated companies remain liable for any exported data or subcontractors (Jolly, 2013).

IV. SAFE HARBOUR

The mixed approach to data protection in the US makes it difficult for a European company to determine if a US company provides an adequate level of privacy protection under the standards of the Data Protection Directive, which can have limiting effects on EU-US business due to the complexity and costs involved in ensuring adequacy. In order to provide a efficient and straightforward way for US organizations to meet the adequacy requirement the US Department of Commerce and the European Commission worked together to develop the US-EU Safe Harbor Framework (Trade Information Center, 2013) which was approved by the EU Commission as being sufficient for the purposes of the Data Protection Directive in July 2000 (European Commission, 2000).

The main advantage of the Safe Harbor Framework for US companies is that a compliant organization is automatically judged to provide adequate privacy protection and the requirements for compliance are designed to be streamlined and cost effective (Trade Information Center, 2013). This also benefits European companies by providing an easy way for them to check if a US trading partner is safe to transfer personal data to. It should be noted that the Framework is a self-certification scheme for meeting EU data protection standards and not in any way legislation (Connolly, 2008).

Although the Safe Harbor Framework was jointly produced by the EU and US the views held on it by the two differ. In particular the US views it as highly regarded and successful while it is viewed by the EU as more of a practical compromise (Connolly, 2008). A 2004 review by the European Commission raised several concerns about the implementation of the framework, chiefly that a number of participating organizations were either publishing no privacy policy or one that did not comply with the whole of the Framework and that the methods of recourse being made available were not fully compliant with or capable of enforcing the Framework (European Commission, 2004).

Additionally not all of the organizations that claim to participate in the Safe Harbor Framework are actively doing so or in full compliance with the standard. A 2008 study found that of the 1597 records on the Safe Harbor list at that time only 1109 were current and only 348 of those were fully compliant with the enforcement principle of the Framework (Connolly, 2008). Additionally of the 348 companies that the study found to be compliant only 54 were compliant for all categories of data (Connolly, 2008).

V. IMPACT OF AND ON BIG DATA

Big Data Analytics involves the collection of an absolutely massive amount of information and the processing of that data in order to produce valuable and actionable information. If the purpose for which this information is being gathered is to build up a profile of an individual in order to target advertising and such more effectively then this means that automatically becomes personal data under the EU definition as the entire purpose of its collection and processing is to allow for the identification of an individual.

Since the data being handled is classed as personal data it falls under the jurisdiction of the Data Protection Directive, with several impacts on the collection of data for use in Big Data Analytics. The requirements of the Directive that impact the use of personal data in Big Data are as follows:

- That an individual be notified when their personal data is being collected.
- That they have the choice to opt out of having their data collected.
- They must be clearly informed of the purpose for which their data is being collected. It is permissible to use the data for a purpose other than the stated only if consent is acquired from the individual again.
- They must be clearly informed if it will be transferred to any third parties.
- That only that information which is necessary for the stated purpose of collection is retained.
- That data is only retained for as long as needed to complete that purpose.
- An individual has the right to request of an organisation that their personal information be deleted from the organisations records at any time and the organisation must comply.

The directive also requires that any organization holding personal data store ensure it is stored securely but this is not specifically impacted by Big Data. Companies compiling with the Safe Harbor Framework are also required to meet these requirements.

Compliance with the first four of these requirements is generally achieved at the same time through a notice to solicit consent, such as the privacy policy of an online service or the user agreement of a piece of software or a device. This means that someone shopping online, using Facebook or any similar activity which provides a wealth of information that can be used in Big Data Analytics should in theory be aware of where their personal information is going at what it will be used for, satisfying the letter of the law on personal data protection.

Unfortunately there is a significant gap between theory and practice in this case meaning that although the law is being met the intent behind the law is not as in practice people often have no idea who has their personal information or what is happening to it despite having consented. The reason for this is that most people do not read though items like privacy policies properly and instead just skip to agreeing, especially online (Cate, et al., 2013). There is a very good reason for this,
namely the sheer volume of such documents people encounter online mean that it is highly impractical for users to read through all such documents they encounter. A 2008 study found that if American users read through every online privacy policy they encountered in a year in full they would spend 244 hours a year doing so (McDonald & Cranor, 2008) and the number of such policies users encounter is likely to continue increasing.

The reason this method is considered sufficient despite the volume of data involved rendering it impractical in the current circumstances is because the laws governing data protection where created based on guidelines published by the Organisation for Economic Co-operation and Development (OECD) in the 1980s, long before the idea of Big Data was even a remote possibility, or indeed before the internet of today had even begun to form. When the Data Protection Directive was drawn up it was entirely reasonable for someone to read and understand all notices of consent they encountered and an organization could cover all the uses they had for any personal information collected in a relatively straightforward fashion so they were effective as part of the data protection framework. In the current environment however there are a number of issues with the use of notices of consent, especially regarding Big Data. Firstly, as mentioned earlier the volumes of such notices people encounter vastly reduce their usefulness as a tool as it is no longer reasonable to expect someone to read through all of them. This is exacerbated by the second issue, which is that in order to cover all the uses of collected information and third parties that it may be passed onto notices of consent become increasingly broader and more permissive to ensure they cover as many eventualities as possible (Cate, et al., 2013). The final issue is that with the rise of Big Data it has been shown that there are many possible uses of data that are not apparent when it is collected which means that either notices of consent must be made even broader to cover these future uses before anyone has any idea what they are, which could easily make them vague to the point of uselessness or companies must be able to contact individuals to acquire further consent to the use of their personal data which would increase both the volume of data companies must store and the amount of consent people are asked to give greatly.

The requirements that only the necessary information for the described purpose be retained and that the data is retained only as long as necessary for that purpose do not mesh particularly well with Big Data. This is because they are intended to keep the amount of personal data an organization is retaining down to the bare minimum necessary while idea behind Big Data is to analyze absolutely massive amounts of information and see what can be learned from it. While this can be done by looking at a very small amount of information from an immense number of individuals it is more efficient the more information the analyst has on each individual. This problem can be answered either by increasing the breadth of the purpose for which information is gathered but this runs the risk of making the description of purpose so wide-ranging it becomes meaningless as a protective measure.

Finally the volume of data involved in Big Data poses an issue with the right of individuals to request that all data an organization holds on them be deleted. The sheer amount of data involved means that it may be very difficult for a company to delete all data they know relates to an individual and impossible to ensure they have no information from any source relating to a given individual (Kuner, et al., 2012).

It is difficult to make any statements regarding the relationship between American data protection laws and Big Data as the laws themselves are so varied. An advantage however is that because they are created in response to issues as they occur they are more flexible than the centralized EU system and are more able to compensate for newer concepts like Big Data.

VI. FUTURE CHANGES IN RESPONSE TO BIG DATA

As the principles underlying current data protection law where drawn up when the shape of the online world looked very different from now it is no surprise that revisions to those principles are needed. Among the recommend changes are that focus should be more on protecting collected personal data than limiting it in the first place as it is becoming increasingly unfeasible to do so (Cate, et al., 2013) and that the permissibility of using personal data should be considered based on an evaluation of the potential benefits and harms that use could cause (Cate, et al., 2013). Other ideas that are also being considered are if it is still meaningful to regulate personal data protection based on national boundaries in the current global information environment (Kuner, et al., 2012) and if the right to have personal information forgotten is still meaningful in light of Big Data (Kuner, et al., 2012).

The European Commission is currently planning to implement a new set of data protection law, the General Data Protection Regulation (GDPR) in response to the impact the advance of technology has had on the scale of personal data collection and processing, with implementation intended to begin at the end of 2014 (European Commission, 2012). The GDPR is based on the same principles as the current framework but is intended to correct for the legal uncertainty and fragmentation that has developed within the EU as different countries responded to changes in the online environment (European Commission, 2012).

The key changes within the GDPR strengthening an individual’s right to be forgotten, easy access to your on data, a right to data portability, the requirement for explicit consent to certain types of data processing and increased responsibility and accountability for data processors (European Commission, 2012). A particular focus is on social media sites and search engines and the GDPR will require that users of such be clearly informed when their personal data is processed, that their explicit consent will be required before a company can process their personal data and that they will always have the right to have such a site erase their personal data immediately upon their request (Schmitt & Stahl, 2012). Additionally companies will be required to delete any personal data once it is no longer necessary and the company has no legitimate reason keep it (Schmitt & Stahl, 2012). The GDPR also
expands the territorial scope of EU data protection law so that in addition to applying if the controller or processor is within the EU it also applies if the subject of the data resides within the EU if the data will be used to sell to or monitor subjects within the EU (Schmitt & Stahl, 2012).

VII. CONCLUSION

The relationship between Big Data and data protection law is a complex one, especially when considered internationally. It is further complicated within the EU by the fact that the principles on which the current EU legislation are based were drawn up long before the technical capacity for Big Data existed so they include no provisions for it and part of the hope for the GDPR is to address this.

There are three main issues to be concerned with about the interaction of Big Data and personal data; that of consent, the right to know what data is held and request its erasure and the international regulations involved. The issue of consent is straightforward but still easy to misstep in. If a company is collecting personal data directly they must make sure that the subject is aware and consent to all uses of data is put to. This means that if a use for the data emerges later that was not explicitly consented to when it was collected the organization must either acquire further consent relating to the current data or collect new data with the new use as part of the consent. When acquiring data from a third party care should be taken to ensure that they have proof of the appropriate consent as well.

In terms of an individual’s right to know what data an organization holds on them and request its full erasure at any time the chief issue is volume. Depending on the scale and organization of data involved it can be very difficult to comply with this in a timely fashion, which can be considered a breach of regulations. Finally it is becoming increasingly common for countries to feel that their data protection regulations apply to their citizen’s data wherever it is processed so it is important to know where all personal data being processed comes from and which laws apply. Taken together these all mean that when processing personal data a company must be aware of where it all comes from and its purpose, regardless of the scale involved.

VIII. REFERENCES


Abstract - This paper discusses what big data is and how it can effect an SME in the future. It also explains how an SME can protect its information systems so that the data extracted can be safely secured. Big data is a very valuable resource as it contains data that can effectively increase a company’s customer base, target a certain demographic with a promotion and also monitor current products that the company might sell. SMEs do not know that they do not know Big Data is important to them.

Index Terms - SMEs (Small, Medium Enterprises)

I. INTRODUCTION

Big data is becoming the new big venture for many large organisations. It allows a company to explore a new demographic through social media, for a retailer or supermarket chain it can help them identify new lines or remove unwanted lines. This will all help cut costs in the long term.

Big Data can also be useful for an SME, however with big data comes big responsibility, there are standards such as ISO 27002 that need to be followed and Government Acts such as the Data Protection Act 1998 that need to followed. To an SME reading this it will seem that big data is out of their league and budget however this paper will show that any size company can make use of the new big data trend and the security aspect to the data need not be a major concern with the current data governance in place. To be able to use big data an SME first needs to understand what it is and how the 5 Vs, Volume, Variety, Veracity, Velocity relate to big data. Soon big data will be everywhere and available for all different business sectors, data can already be taken from mobile devices and social media sites. Once the data is extracted it needs be processed and put into a structure so it can be understood clearly and all the value from the data can be taken. The correct security measures need to be in place in order store the data securely, such areas as IT policies, security policies and virtual/physical security need to be covered to not only keep the data secure and safe from threats but also to adhere to the Data Protection Act 1998.

II. WHAT IS BIG DATA?

What is big data? This is the question that needs to be answered first in order to gain an understanding of what this paper is about. Big Data is simply vast amounts of data, big data is information that cannot be processed using normal methods (Eaton, et al., 2011). Big Data is a growing concern for many SMEs mainly because they have access to massive amounts of data but are unable to use it due to it being in a raw, unstructured format (Eaton, et al., 2011).

It’s important to understand even though big data can be very hard to analyse it is still very crucial and can hold information of great value. Take the health care sector as an example of how big data can be useful. Doctors and nurses will look through the NHS patient database to find out family health history, current medical conditions of individual patients and all sorts of medical information revolving around a patient. This database is huge but there are other databases that are used in partnership like the online medical journals that detail diseases, symptoms, causes and cures, these are perfect examples of big data. Each of these databases contains a wealth of information that doctors and nurses can use, the information will used to connect the dots about a patients ailment and hopefully to a treatment plan. Data security is paramount in the health care industry and the NHS is no exception, employing over 1.3 million staff the NHS is one of the largest employers within Europe (Smith, et al., 2010). Even though data security is a massive subject area for the NHS there are still data leaks, reports such as a USB memory stick containing patient data being found in a car park, an average loss of 835 patient records every day and spending six weeks trying to find a memory card from a medical photographers office in the Sick Kids Hospital in Edinburgh (Smith, et al., 2010). These data leaks show that even with the tightest of IT governance security policies data leaks can still and do happen, however an SME that follows similar IT governance techniques to the NHS will most likely see very few data leaks due to the amount of data they hold compared to the millions of records currently held by the NHS.
Another example of an organisation using big data is Target. Target is a superstore chain in America, the UK equivalent would probably be Asada. Target uses big data analytics to help promote products using a technique called predictive analytics (Wigley + Company Solicitors, 2013), for example this allows Target to predict when a woman is pregnant through the woman’s buying habits. This situation came to fruition in the form of an angry father, according to the original article published in the New York Times called “How Companies Learn Your Secrets” (Hill, 2012) the customer daughter had received some coupons promoting baby products, the only problem was she was still in high school, the father believed Target were trying to encourage the young teen to get pregnant. The manager dealing with the complaint apologised in store and then called the father a few days later to apologise again however the father stopped the manager in his tracks and explained that it turns out his daughter is pregnant after all and apologised himself (Duhigg, 2012). An SME could use big data in a similar way for predicting sales for example some fruit might sell better than others in summer in which case the company would be able predict when to order more stock to meet public demand.

So now we have an understanding of what big data is and how it is used by large organisations we now ask how SMEs are involved with big data? SMEs (small, medium Enterprises) made up 99% of businesses in the UK during 2013 (Rhodes, 2013) this makes the SME market the largest employer in the UK, so how does an SME use big data? Well according to a report done by E-Skills UK the simple answer is, they don’t. E-skills contacted 541 SMEs and none had implemented big data analytics (E-skills UK, 2013). Though the number of SMEs contacted is very small it is still a shocking statistic, the report continued to explain that the results from a larger survey found that only 9% of SMEs who employ between 100 and 249 staff had actually implemented big data analytics. Big data is something that SMEs need to invest in to be able to succeed in current and future markets. In 2010 over 1ZB of data was generated worldwide and by the end of 2014 we will be generating 7ZB a year to put that in context 1 Zettabyte (ZB) is equivalent to 1073741824 Terabytes (TB), IDC states that organisations that use big data will thrive whereas those that don’t are at a disadvantage to competitors (Villars, et al., 2011).

One of the reasons behind SMEs not making full use of Big Data Analytics could be the data security and governance risks involved. Storing vast amounts of data whether on customers, suppliers, colleagues or products requires not only a large amount of financial resources but also the ability to follow Data Protection policies set in place by the company itself but also the Data Protection Act 1998.

III. WHAT ARE THE BIG DATA VS?

There are 5Vs of big data these are volume, velocity, variety, verification and value (Beulke, 2011). Volume, this is simply the volume of data available. As previously mentioned it is estimated that nearly 7ZB of data will be created worldwide this is a lot of data and such volume has to be accommodated. An SME would most likely find it simpler to store their big data off site at a data centre though this would not be cheap, it would still be cheaper and easier to manage than having it all on site. Outsourcing the data would aid the SME in terms of data security, they would no longer have to worry about physical security and access to the server that would all be monitored by the data centre, however the SME would still need to provide some level of security such as user privileges and passwords to protect certain confidential data such as customer and colleague information. The SME would also need to make sure the data centre they use complies with the Data Protection act of 1998 for example one of the data protection principles highlights “appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.” (House of Commons, 2014). This basically means that the data centre need heavy round the clock physical security and also cyber security such as firewalls and passwords but the SME also needs to prevent unauthorised access or loss of data. This can be achieved through backups, passwords and strong data governance such as IT policies stating what employees can and cannot do with the data.

Velocity is not only the speed but also the number and frequency at which data flows (Beulke, 2011) from sources such as businesses processes, machines, networks and human interaction (Normandeau, 2013). Big data is enormous and the data comes from all sorts of locations such as social media sites and mobile device this means that the flow of data is continuous (Normandeau, 2013) and in order to handle this amount of data a company needs to have the correct infrastructure in place. An SME would need to have the correct storage strategies to cope with the influx of continuous data, a few TB hard drives and a few servers would not be able to handle the task. This would be a costly venture for any company and most likely out of the price range of most SMEs, this is where a data centre would have to be employed thus eliminating the cost of building a warehouse to store enough servers to hold the data themselves.

Variety is the different types of data both structured and unstructured and the sources that it came from. As discussed data can come from a variety of sources such as social media, mobile phones and other mobile devices, the data from these comes in many different forms for example emails, photos, videos, PDFs and audio (Normandeau, 2013). Unstructured data creates problems for storage, mining and analysing the data (Normandeau, 2013), unstructured data is a piece of data that cannot be easily scanned for certain pieces of information, for example an email is a form of unstructured data because an inbox might be filtered by data, time or size but all the emails would be about various subjects. To make an email inbox a form of structured data then all the emails would have to be sorted by subject and the contents of those emails would have to be constant and not deviate from the subject. So why does this cause a problem? Well the data collected could hold 1000s of emails and photos all containing different subjects and at the same time discussing other subjects as well this would make it
extremely hard to almost impossible to make the data structured in any way.

Verification refers to trust (Self, 2013), can you trust the data? With the vast amounts of data being used there is bound to be some quality issues. These quality issues could mean that some of the information is incorrect or biased making any conclusion drawn from the data, wrong. SMEs need to be able to verify the data however to test such data is a problem faced by many organisations and companies around the world. One solution to test the data is through using HDFS (Hadoop Distributed File System) (Gudipati, et al., 2013). HDFS involves taking data from different sources and loading it into the HDFS the next step performs map reduce operations which involves processing the input file and applying the map and reduce operations to get the output. This output can be extracted and loaded into a downstream system such as a data warehouse for further processing (Gudipati, et al., 2013).

The last V is value. Value is considered one of the most important aspects of big data. The value of big data is basically asking is the information extracted from the data of value. Does it aid the project? In an SME where cost could mean make or break for the business, investing in big data is a massive step, as previously mentioned, investment in IT infrastructure is extremely expensive especially when big data is going to be used. An SME would have to make sure that the data collected has value, would be of use and could aid whatever project is being trialled. In terms of the big data analytics used by Target the valuable data would have been areas such as date of purchases, quantity purchased and name and address, over a period of time Target would be able to use this information to send promotional coupons to the customer for their manager life goal such as having a baby or moving house.

There are many more Vs of big data however the 5 that have been mentioned above are possibly the most important due to what is covered under them. Where security and SMEs are concerned only a few Vs come into play and these are vulnerability and verification. Verification comes into play with security because verification is all about trust, can you trust this data? Can you trust the quality? Can you trust the source? Vulnerability is important to security for the obvious reasons, how at risk are the systems from the data? Is the data safe?

IV. BIG DATA SECURITY WITH SMEs

As SMEs begin to invest in big data, security risks will arise due to the company networks being more open and extended to allow suppliers and partners access. This will open the networks up to more complex attacks, these complex attacks are not normally detected until after the damage has been done (Curry, et al., 2013). This means SMEs need to ramp up security in order to benefit from big data in a safe manner. SME’s can do this in various cost effective ways such as applying new and improved company policies, improving data governance, adhering to the Data Protection Act of 1998 and also following the ISO 27002 guidelines on information security.

In 2013 63% of small businesses where attacked by an unauthorized outsider, 23% were hit by a denial of service attack, 15% detected successful security breaches and 9% know that data was stolen (Department for Business Innovation & Skills, 2013), these figures are astonishing. These cyber “pirates” have realised that SMEs are a very easy target, the attackers are able to retrieve confidential data such as personal information and banking data. As Winston Churchill once said “To improve is to change; to be perfect is to change often” (Churchill, 1925) in many ways this saying can relate to SMEs and security, an SME has to constantly stay up-to-date with the latest ISO standards and also keep company policies up-to-date in order to keep networks secure.

A. Company Policies

Company policies are a big part of data security in any business or organisation. A policy is a list of objectives or rules that apply to certain areas of an IT system. For example a company IT policy might rule that USB storage devices are not allowed or company documents are not allowed to be taken out of the business, these are fairly common examples, however if the security policies are too strict then users will find a way to get around them thus opening up weakness in the system (GFI , 2010). Where big data is concerned a company policy to help prevent the risk of a security breach could be that all data vendors be vetted (Gray, 2013). Some vendors could exchange data with other third party companies or not manage the data effectively this basically means that sensitive data could be at risk from various methods of attack or even leaked onto the internet ruining the company reputation forever. A way to combat this would be to get the vendor to agree to follow the standards you would expect and also request to see their data security policies (Gray, 2013). An SME can very easily protect confidential data such as customer details and financial data through implementing a good policy that highlights all the rules and objectives, when dealing with data vendors it is important that each one is vetted to ensure compatibility with the company’s security policies.

B. ISO 27002

ISO 27002 is an information security code of practice it highlights potential controls and control methods which could be implemented (ISO 27000 Directory, 2014) to help prevent a cyber-attack. ISO 27002 works with ISO 27001 it basically helps an organisation or company find ways of improving their IT system security through the use of documents such as a risk assessment. The ISO 27002 standard is very useful as the latest version talks about 144 different ways of controlling IT security these controls will also improving data security which means the big data being imported and extracted would be a lot safer, certain controls could be like what we have already discussed for example policies and user restrictions.

C. Governance

IT governance is becoming a massive deal in every business sector, data governance is a discipline that covers areas such as quality, management, policies and business processors. IT governance can influence the benefits of an
SME’s IT investment (Devos, et al., 2012), this basically means that if good IT governance is implemented in an SME then all the benefits of having an IT infrastructure will become apparent making the entire system easier to use and more welcoming to new additions. IT governance revolves around three main issues, one is determining what IT decisions are to be addressed through governance, two is who has the decision rights and what the nature of decision rights are and three how decisions made will be orchestrated so that all the correct individuals are involved and understand the implications (Huang, et al., 2010). Effective IT governance should involve the entire company with both IT and non-IT stakeholders at every level of the company. If a company has an effective IT governance system in place then they stand a chance of getting 40% more value from IT (Fonstad & Robertson, 2006). In terms of an SME using IT governance it is important to understand the value, as mentioned above a company could see 40% more value from their IT systems, this is a massive amount no matter what size network. What’s important to note is for an SME there could be opposition to any new policies or procedures being put in place, this could be due to any number of reasons but the main being people afraid of change and being stuck in their old ways, for an SME this could be a hard issue to overcome as with many small businesses some staff may group together and block any decisions made.

V. CONCLUSION

Big Data is a valuable resource to any size business but it could be instrumental to an SMEs success. Big data could allow an SME to target new demographics by offering personalised offers. However the data that is collected needs to be secure, this can be done through good IT governance. By using IT policies, codes of practice like ISO 27002 and by following the Data Protect Act 1998 an SME will be able to keep on site data secure and safe, a data centre could be used to help protect the data at an off-site location and also saving money in the long term.

VI. REFERENCES


Churchill, W., 1925. London: s.n.

Curry, S. et al., 2013. *Big Data Fuels Intelligence-Driven Security*, s.l.: EMC.


E-skills UK, 2013. *Big Data Analytics*, s.l.: E-skills UK.


Does Security Compliance Help Protect Big Data within SMEs

Mitchell Rankin
University of Derby
Derby, UK
M.Rankin1@unimail.derby.ac.uk

Abstract — Small and Medium Enterprises (SMEs) are using technology and large amounts of data more than ever and, as a result, experience an average of 17 security incidents a year. Consequently, organizations need to look at solutions and frameworks, such as ISO 27001, to help to safeguard their companies and increase their security. This should be done in such a way that it covers employees, clients and organization infrastructure.

Index Terms—SME, big data, security, incident, management, framework, threats.

I. INTRODUCTION

A popular aspect of IT since 2012 has been Big Data; which is defined as large amounts of structured and unstructured data, often growing too fast to fit in traditional database architectures (Dumbill, 2012). It is very attractive to SMEs as it allows them to conduct Big Data analytics for patterns and make better business decisions, thus becoming superior to their competitors. McGuire, Manyika and Chui (2012) state that some businesses are able to increase their operating margins by 60% by using big data.

However, big data introduces challenges that SMEs are often not prepared for; in order to understand these challenges, it is best to look at what is often referred to as the ‘V’s of Big Data’.

IBM Big Data Hub (2013) have 4 V’s to define Big Data, these are Volume, Velocity, Variety and Veracity. However, it is not limited to these, as Richard Self presented to IBM (2014), adding a further eight to this, including: variability, value, validity, volatility, verbosity, vulnerability verification and visualization. Volume and Veracity are often the biggest issues for SMEs as they can lack the technology to keep up.

There is greater complexity in determining what security compliance is, as there is no compulsory standard for SMEs or any company to comply with. However, the frameworks most commonly adhered to are the ISO/IEC27001:2013 guidelines. Frameworks such as these are needed to help protect and maintain companies’ assets to provide security over their data.

Because of the increase in the use of big data, there is much greater concern over Information Security; in 2013 87% of SMEs surveyed reported they had suffered at least one security incident (Willets, 2013). This statistic shows that currently SMEs are struggling with data security and shows that they are failing to implement security frameworks, such as ISO/IEC27001.

Studies undertaken by Potter and Waterfall (2012) have also shown that even with security policies in place, poor understanding and lack of training also lead to breaches. As a result SMEs are facing costs of £35k-£65k for a serious breach.

As part of understanding if security compliance is playing a role within protecting data, especially within SMEs, this report will look at the risks facing SMEs, where security failures are most common, and why SMEs are not doing more to protect their data.

II. SECURITY THREATS TO SMEs

Whitman and Mattord (2011) define a threat as an object, person or other entity that could represent a danger to an asset. These threats come in a multiple forms, for example: from outside cyber criminals attempting to penetrate an organisation’s infrastructure, internal staff-caused incidents, both accidental and misuse, or the theft of computer equipment leading to the theft of company data. In 2013, on average SMEs suffered 17 breaches (Willets, 2013); a suspected increase of over 50% in the last year alone.

A. External Threats

In the past, it was only larger organisations that were targeted by cyber criminals and outside threats, as they were seen to have the most profitable data, and many SME management teams still believe this to be the case. However, Willetts’ (2013) survey shows that 63% of SMEs were attacked by an unauthorised outsider in 2013; a 22% rise from 2012. In addition to this, cyber criminals have been hitting 23% of these SMEs with Denial of Service (DoS) attacks, demonstrating a rise of 8%—targeting assets such as web servers in an attempt to disrupt business.

From these attacks, it is reported that 9% of small businesses’ have confirmed they have had their confidential data stolen from external threats, showing a rise of 3% in a year.

Data has also been acquired via hardware theft from companies, including mobile phone / tablets and old hardware (servers/workstations) that has not been decommissioned.
correctly with 12% of SMEs reporting this happening to their business. (Potter and Waterfall, 2012)

B. Internal Threats

While SME’s management are not able to control whether their company becomes a target of cyber criminals, there are steps that can be taken to minimise the potential impact. To an extent, SMEs are able to control their staff and training; however 36% of the worst breaches reported were caused by human error, with 10% of these caused by deliberate misuse of companies IT Systems (Willetts, 2013). Overall, 57% of SMEs reported data breaches due to staff-related misuse, showing a rise of 13% from 2012. Additionally, 17% of these staff had broken Data Protection Laws in the process. Breaches such as these are often caused by employees’ lack of knowledge of security policies.

Along with the more traditional problems already discussed, SMEs are now also being affected by both BYOD (Bring Your Own Device) and the use of social networking on organization equipment.

Scott (2012) conducted a survey of two hundred IT companies: over 60% of them were aware of employees using personal devices while connected to corporate networks, but only 23% had policies in place to manage them, such as segregated guest networks. The threats from BYOD include employees risking contracting viruses or malware from their personal device, which could affect their company’s infrastructure; or employees are able to download sensitive and confidential data onto their personal device, which in turn increases the risk of theft/loss of that data.

Social networking while at work is also a rising problem: in their annual report, Cisco (2014) discusses the dangers and likelihood of downloading malware from social networking sites such as www.facebook.com or www.LinkedIn.com; which can easily affect entire infrastructures. The findings of Baldin (2010) support this by discussing how social networking contributes towards the staff misuse breaches, as employees can leak confidential data onto social media sites via social engineering. With only 11% of SMEs monitoring social media for staff breaches, there is the potential for that percentage to be much higher than recorded.

C. Cost

Willetts (2013) goes on to look at the resulting costs of data breaches; in 2013 the worst breaches cost SMEs between £35,000 and £65,000, depending on the severity of the breach. In a time of financial difficulties, this is extremely damaging to an SME.

Financial costs are not the only consequence SMEs face from a security incident; Markose (2010) demonstrates the impact of data breaches to SMEs by looking at other factors affected, e.g. damage to a company’s reputation. A public company could experience a decrease in shareholder value, loss of customers as a result of a breach and, in certain cases, fines and civil penalties. These factors combined could cause greater damage to a company than the financial cost.

D. Case Studies

The following are two cases studies in which big data has been lost or stolen from a company by the threats discussed in Section II.

1) Her Majesty’s Revenue and Customs

A case study which clearly demonstrates staff-related breaches involving Big Data can be found by looking at the Her Majesty’s Revenue and Customs (HRMC) case from November 2007. BBC (2007) reports that two CDs containing personal and financial information on twenty five million people regarding child benefit were sent to a third party auditor, but did not arrive.

Investigations found that junior staff at HRMC had ignored security guidelines and sent the disks with inadequate levels of encryption via general post (instead of recorded delivery) (Anon, 2007). Fortunately, the data did not reportedly come into the possession of cyber criminals, however the incident did result in the chairman of the HRMC resigning, and serious damage to their reputation.

2) Sony Computer Entertainment

In 2011 Sony became the victim of an external hack, and as a result their online PlayStation Network services was offline for 7 days and a substantial amount of customer data was stolen, including customer names and account details, addresses and credit card information (Seybold, 2011). While Sony have never officially released the details behind the type of hack experienced (because of the scale of attack) there was a lot of speculation, with the most likely explanation being a combination of continuous DoS attacks and SQL Injections to access Sony’s sensitive databases (Anthony, 2011). Tassi (2011) discusses the resulting cost of this intrusion; Sony calculated that the breach cost them $170 million dollars, with an additional fine of £250,000 for serious breach of the Data Protection Act (BBC, 2013).

III. CURRENT PROBLEMS

Even if SME’s do recognise all the threats against them, unlike larger organisations, they continue to face a multitude of problems.

A. Expense

A main factor of not implementing any security framework is cost. In a report published by the European Commission (2013), 15% of SMEs reported their most pressing problem is access to finances. Consequently, SMEs do not have the budget for IT security, and are unable to invest in current technologies that the frameworks require; often compromising for equipment that does not meet all their requirements, thus increasing the risk of external threats. Also, due to financial constraints, SMEs often outsource their IT and IT security, as it can be more cost effective. However Ashford (2012) states that this can often increase risk to an SME, especially when it is confidential data that is being handled, as it can easily be lost or stolen.
Introducing SMEs to ISO: information on security policies, human. Smaller firms bec
s security policies education or T job,
Exome vulnerable as managemen
t problem failing to adhere to security guidelines (conduct security audits to see where employees are 
assess possible vulnerabilities they may have, and also do not carry out any type of risk assessments to
D. (Passingham, 2013).
Failing these, SMEs are unable to, so SMEs therefore lose out. This issue may continue to worsen, as IT
security is growing at 3.5 times faster than any other IT job, with an increase of over 73% of total IT security jobs between
2007 and 2012 (Rosenbush, 2013).
B. Education
MacInnes (2013) continues to look into the lack of education SMEs have into the current methods used to deploy
good security policies. Smaller firms become vulnerable as they lack IT expertise to understand the whole picture; they are
often reactive to security incidents that happen, instead of being proactive to prevent them from occurring.
Passingham (2013) discusses how SMEs lack security and the ability to deal with the issues around Big Data, due to a low availability of new IT-skilled employees. Larger enterprises are able to offer higher wages that SMEs are unable to, so SMEs therefore lose out. This issue may continue to worsen, as IT security is growing at 3.5 times faster than any other IT job, with an increase of over 73% of total IT security jobs between 2007 and 2012 (Rosenbush, 2013).
C. Management
Many SMEs also have poor or no IT Service Management (ITSM) systems in place, such as ITIL or Prince 2. They therefore lack proper incident management, problem management and patch management techniques. Failing these, they are neglecting and evading security incidents. Without correct problem management, SMEs are not learning from their previous mistakes and are unable to develop solutions to help prevent incidents of the same nature from happening again. Without the correct patch management techniques, SMEs are potentially missing vital patches to software and operating systems, again leaving them vulnerable to external hackers (Passingham, 2013).
D. Assessment
The final issues currently plaguing SMEs are that 23% of them admittedly never carry out any type of risk assessments to assess possible vulnerabilities they may have, and also do not conduct security audits to see where employees are currently failing to adhere to security guidelines (Willett, 2013).

IV. ISO/IEC 27001:2013
ISO/IEC 27001:2013 is the most common security framework available, which most SMEs should eventually aim to be certified in. It provides companies with a framework about information security management systems (ISMS). Any company is able to become ISO 27001 certified by following its best practices, and complying with the guidelines set out by ISO/IEC 27002. However it is not currently compulsory for any company dealing with data to comply, as long as they do not break Data Protection Laws (ISO 27000, 2013). The scope of ISO 27001 is about identifying what the information assets are within an SME, and from there it allows management to assess the risk to the assets, design risk management strategies, and then begin training and awareness with their staff.

As previously mentioned, SMEs current problems are the lack of education about security frameworks available. Fig 2 shows that the amount of SMEs that have implemented (or that are planning to/ in the process of implementing) ISO27001 have risen slightly since 2012. However, over 43% of SMEs are unaware of these guidelines. Introducing SMEs to ISO 27001 frameworks would be a start for many to look at security risks, and hopefully to become security compliant. However, while Brewer (2013) discusses the benefits of ISO 27001 to SMES he also clearly states that it is not always suitable for all SMEs, and some should only become fully certified if there is a contractual requirement from customers or clients. One reason for this is that an SME has a large amount of sections to cover, and due to issues discussed in section III, implementation is not always possible. He therefore suggests that smaller SMEs should begin by looking at ISO 27002 and ISO 27005, and thus begin to create a suitable security framework.

V. ISO/IEC 27002/27005
Unlike the formal specification of ISO 27001, ISO/IEC 27002:2013 is more of an advisory document on how to govern a company’s security (mainly IT), and replaces the redundant ISO 17799 standard. ISO 27002 is a code of practice for the implementation and control over a company’s information assets. This includes: information on security policies, human resource security, asset management, access control, cryptography, physical and environmental security, incident management and compliance. (ISO 27002, 2013)
ISO 27005 is similar to ISO 27002 as it provides guidelines and advice for companies’ IT risk management, with its main

Figure 1. Percentage of SMEs that have formally documented security policies (Willets, 2013)

Figure 2. Percentage of SMEs that are of ISO27001 and have implemented it (Willets, 2013)
purpose to assist in implementing information security management systems (ISO 27005, 2013). The downside of all the ISO frameworks is that they are frequently updated; currently they were most recently updated in 2013 and before that in 2005. In these updates, new methods of practice are often announced and new threats introduced. While eight years might be considered a long period of time, to an SME every update requires further investment into new technology, and as previously established, can be very expensive (Calder & Watkins, 2012).

VI. CONCLUSION

In conclusion, it is clear that many SMEs are either not well-informed in IT Security, or not doing enough to secure their assets and protect Big Data. A large percentage of SMEs are failing to implement frameworks, however not always due to negligence; these frameworks can be expensive and resource heavy to implement, especially for smaller businesses. However, in the long run implementing ISO27001 can secure their data and help prevent breaches, which would be an attractive quality to prospective clients. Overall, if companies take security compliance seriously and follow guidelines such as annual auditing or employee education then this will help to lower the amount of annual breaches we see in SMEs and thus, bring greater security to Big Data.

VII. REFERENCES


Information Security and Assurance
Security Risks and Breaches in SMEs Big Data

Ghulam Hussain Raza
University of Derby
Derby, UK
G.Raza2@unimail.derby.ac.uk

Abstract—This report contains the definition of big data and the types of big data there is discussion on the SME's. It contains the role “Big data” play in “SMEs”, and the role “SMEs” play in business world. The statistics have been mentioned for both big data and SMEs. The risks and breaches were discussed along with the frameworks and methods could be used to minimize these risks and breaches.

Key words—Big data, SMEs, Statistics, Breaches/Issues, Framework, privacy.

I. INTRODUCTION

Technology is growing faster day by day, the more technology we have the more information or data the enterprises have to deal with, whether its texts, surveys or emails. Presently, all the data being generated by businesses is increasing. Even the large enterprises are getting targeted by the growth of big data. Managing big data can be very cumbersome and compound. This data can be very overwhelming for SMEs if they are short of the appropriate IT resources. Applying various frameworks will increase business productivity and find new opportunities. (Jayme, 2013).

II. What is Big Data?

Big data is described as a massive amount of data which is both structured and unstructured and is very large making it difficult to process it using normal software and database techniques. In most enterprise circumstances the data is very large or it is very quick or it goes over the current processing capacity. The term is mostly used for volume of data which isn’t always the case. When this term is used by the vendor, it might mean the technology which includes different methods and tools that an organisation requires to deal with large volume of data and storage facilities is utilised. (ITBusinessEdge, 2011).
Approximately about 2.5 quintillion bytes of data is created by the world every day. There is big amount of data that about 90% of the data today was alone generated in last 2 years. The data which has been gathered comes from nearly everywhere such as sensors that are used to collected climate data, transaction records and posts to social media websites. All of this information is known as Big Data. Big data contains 2 sets of information. One of them is known was structured data which is 10% the whole big data. This kind of data is already stored in database using multiple networks. The second type of data is unstructured data which is 90% of the big data information. Information in unstructured data known as “human information” for example; tweets, videos, posts on Facebook, conversation in call centre, TV footage, texts, emails etc. (Technology Advice, 2013).

A. The Four Vs of Big Data

1) Volume
Defined as the total number of bytes associated with the data. Unstructured data are estimated to account for 70-85% of the data in existence and the overall volume of data is rising. Benefits of this category includes: Tweets which are 12 terabytes and are generated usually everyday will be converted into an improved product sentiment analysis.

2) Velocity
Defined as a pace at which the data is to be consumed. As volume increase, the value of individual data points tend to more rapidly reduced over time. Usually even few minutes could be too late. The process which is very important such as catching any fraud activity, big data is used to flow into SME to increase the value. For example analysing about 5 million trades generate every day to find any possible fraud.

3) Variety
Defines as the complexity of the data in this class. This complexity avoids traditional means of analysis. Big data can be classified as any structured and unstructured data for example it could be as small as a text or log files, it can be an audio or video file. This section accomplishes 80% of the data development in, video, document and images to improve and increase customer satisfaction.

4) Variability
Defines as the interpreting the data in different ways. There are difference interpretations for different questions. In business such as SME’s out of 3 there
is always 1 leader who does not trust the information used to make business decisions. Which raises the question how these leaders consequent to the information even if they don’t trust it? Creating a trust in big data makes a massive challenge as the amount of sources and variety increase. (Technology Advice, 2013).

These are the recent statistics of Big Data. 90% of the world’s data has been created in last two years. In 2017 big data is planned to grow into $53.4 billion market, which is more than last year $10.2 billion. Out of all the data enterprises store 80% of it. Creating of website has increased and now more than 570 new websites are created every minute. Cloud will be very useful by 2020 because at least one-third of the data will be stored or passed through the cloud. 51% of the people say having a lack of talent is a big issue in order to make better decisions with big data. 28% of the UK enterprises are currently using big data. 48% of them are thinking of using it.24% do not have any future plans of using big data.95% see the data as very critical to success in next few years.37% of the enterprises are familiar with it.(Raconteur, 2013). (McCafferty, 2014).

III. WHAT IS SME’S?

In the European economy Small and medium sized enterprise play a central role. There are one of the main sources of industrial employment, innovation and skills. In a big European union which contains 25 countries, around 23 million SMEs provide approximately about 75 million jobs and represent 99% of all enterprises; But SMEs often face the markets imperfection. SME’s often find is difficult to obtain credit or capital especially on their early stages. Limitations to resources can also decrease the access to new innovation and technologies. Therefore, it is European Commission’s propriety to support SMEs with their economic growth, social/economic cohesion and job creation. (European Commission, 2012).

The report of Department for Business innovation & skills 2013 states that at the start of 2013 there were estimated about 4.9 million private sector businesses in the UK. At the beginning of 2013, this report highlights the estimation of 4.9 million UK private sector business employed approximately 24.3 million people and with an estimated combined annual turnover of £3,300 billion. The total difference was 127000 of these comparing to the start of 2012. The growth was balanced by the number of employing businesses decreasing by 26,000. (National Statistics, 2013). There are more SME’s in private businesses sector which is 99.9%, employing approximately 14.4% million people, 59.3% of private sector employment. So the approximate all together annual turnover of £1,600 billion accounted for 48.1% of private sector turnover. Out of all the businesses about 99.2% were small containing 0 to 49 employees, only 31,000 which is 0.6% were medium-sized containing 50 to 249 employees and 7,000 which is 0.1% were large containing 250 or more employees. The SMEs delivered the largest population since 2000. Their estimated numbers increased from 3.5 million to 4.9 million which is 41.0%, this happened from the start of 2000 to start of 2013. (National Statistics, 2013).

IV. BIG DATA IN SME’S

SME’s are often slow at implementing and improving their technology systems. An example being big data, which is data used to find out the behavior of the customer which enables the business to make offers according to customers need. Big data is any type of data which can’t be organized easily and it can’t be enquired using normal relational database methods. Before, the big data only used to be collected by large business enterprise, SME’s had to face big challenges such as having lack of sources to gather and store big data. Without being able to store this big data, SME’s couldn’t be successful, however now SME’s are also involved with big data system. (Eric, 2013).

What is “big data” in SME’s? It is a recorded growth of data from a range of sources. Big data comes in both organised and unorganised formats. The organised data contains transactional data which is usually allocated in “Enterprise Resource Planning” or “Customer Relationship Management”. A perfect example of organised data is spreadsheets; SME’s could be in trouble if they don’t have a plan in place to manage the data. Unorganised data could be really difficult to search and retrieve, such as data being generated by machines which could be sound, digital videos or images. This kind of data can be lost forever because of its overwhelming abundance. IDC predicted that the data which is generated by machines could be increased 15 times more by 2020. (Jayme, 2013).

V. SME’S & BIG DATA SECURITY RISKS/BREACHES

Many people have stated the issues linked to big data analytics privacy, it raises questions such as: should organisations for example SME’s have access to vital information about groups or individuals?. The biggest concern about big data in SME’s, is security which is very critical but a very less debated issue. It makes is very interesting for hacker when SME’s collects and store big data because unstructured data stored by SME’s is valuable. The big data contains petabytes of data which is often a large data for hackers to steal. (Clark, 2014).

Small and medium sized enterprises are facing increasing number of security threats, however various number of SMEs use suitable controls to secure their sensitive information. For the SME’s who don’t apply enough controls to secure their sensitive data believe security as “someone else’s problem”, which results in increasing risk of security issues or breaches of very sensitive customer or business data, also an exposure for various large enterprises that authorised SME’s to carry out work. SME’s don’t have access to skills to improve their security, and usually avoid unnecessary expenses. (Shan, 2013). (QUIGLEY, 2013).

The sales pitch for new technology has encouraged the endemic attitude which promotes business profits, but areas like security which could obstruct the sale is hardly mentioned. SME’s face number of factors and difficulties
associated with this which includes lack of resources and money, more serious priorities, unfamiliarity and anything which is not related to business progression and growth is ignored. Various SMEs find security confusing, complex and technical, also they are not convinced about the business case. The SMEs don’t have enough knowledge or expertise in order to identify and deal with the risks. (Lacy & James, 2010). There is another big issue when this type of data is gathered; the internet search terms can be misunderstood. Poor enterprises would use internet search term to estimate product value or maybe to target potential customers. In a house there are often multiple computer users and for any reason they might search for information on internet that is not fully relevant to them. Collecting this type of data, analysing it, and using it can result in wrong analytic and usually leads to bad decision making. (Goodendorf, 2012)

When it comes to big data, the types and volume of data in business and IT are too big to process in an ad hoc manner. Besides, it is very difficult to secure such vital information from the data being gathered. Even with big amount of money has been spent on information security the hackers or attackers seem to be getting the upper hand. The report by “Verizon Data Breach Investigation” (2012) states that about 91% of the breaches result in data compromise under a day or less, but 79% of the breaches too a week or more to determine. These are the very few factors:

- Enterprises with IT-enabled systems maintain to grow more complex. Organisations now days ask for much more agile and open system, producing incredible new opportunities for innovation, communication and collaboration. This results in a susceptibility that hackers, nation states and cyber criminals learn to use.
- The funding to attackers is increasing and they are more organised. While the attack has become more active the defenses are static. Attackers create an attacking method to use the weakness of enterprises hyper-connected and user-centric infrastructures.
- Observance is now even harder to reach. Managers and legislators are becoming more prescriptive. Enterprises especially the one with various lines of international or business operations, have difficulties to keep the track of present controls in place, controls that are managed accurately and the ones are important or needed. (EMC, 2012).

VI. FRAMEWORKS

There are various security risks SMEs face associated with big data; there are few frameworks which could be used to counter these problems. The possible framework to avoid all the issues is ISO/IEC 17799:2005. This framework includes a guidelines and general principles for maintaining, initiating, improving and implementing information security management in SMEs. There are main security domains:

- **Security policy:** This method will provide SME’s the information security system.
- **Organisational/Enterprise security:** producing a management framework for enterprise and management information security, and will address information security responsibilities.
- **Assess management:** SMEs will be able to protect all the sensitive or critical assets from attackers.
- **Personal security:** using this SMEs will able to reduce the risk of theft, error, misuse or fraud of the big data inside the system, also providing awareness and training about threats to big data.
- **Physical and environmental security:** This will help SMEs to prevent any unauthorised access to information regarding big data and stop unauthorised people to damage enterprises premises.
- **Communication and operations management:** this method will decrease the risks failure and its concerns by safeguarding the suitable and secure use of big data information processing facilities. (Saint-Germain, 2005).
- **Access Control:** The SMEs will control the access to big data information to safeguard the protection associated with networked systems and awareness of any unauthorised activities.
- **Information systems acquisition, development and maintenance:** This will prevent any misuse, modification or loss of information in operation system and any kind of data information in software.
- **Business Continuity Management:** This will improve and develop SMEs capacity in order to respond quicker to interruption of critical activities from natural disaster, incident, catastrophes or failures.
- **Compliance:** Using this method SMEs will make sure that all the laws and regulations are followed properly and the current policy works with the security policy to ensure the objectives set by management are achieved.

Another possible framework could be used to minimize risks using risk assessment is BS ISO/IEC 27005:2011. It consists of these activities:

- **Risk Identification:** using this method SMEs could identify the risk which causes potential loss.
- **Risk analysis:** This could be undertaken by SMEs depending on the critically of assets. This can be qualitative and quantitative depends on circumstances. SMEs can use qualitative to identify the level of risk if its low, medium or high
and quantitative will used as a scale with number values.

- **Risk Evaluation:** after using the risk analysis then SMEs will be able to tell how often the threats occur along with the knowledge of how easy they can be exploited. (BSI, 2011).

**VII. CONCLUSION**

In this study the correct meaning of big data and SMEs were discussed. SMEs play a big role in present business world however they are limited to various sources. Tasks are given to SMEs assigned by large enterprises. From statistic it shows there are more SME’s in private businesses sector which is 99.9%, employing approximately 14.4% million people, 59.3% of private sector employment. Big data used in SMEs comes in both structured and unstructured format. Big data can be really difficult to handle if the right tools are not used, security and privacy of big data can be a major issue. SMEs should be given more funding in order to build a security system or should consider using frameworks such as ISO/IEC 1799:2005 and BS ISO/IEC 27005:2011, their methods or steps could be really useful for SMEs in order to protect their data, such as “Risk Assessment” discussed in study in order to stop the data from getting hacked, also stop the unauthorised people to access unnecessary data.

**VIII. REFERENCES**


Providing SME’s a Secure Environment for the use of Big Data Resources

Elliot Rees
University of Derby
Derby, United Kingdom
E.Rees2@unimail.derby.ac.uk

Abstract — this paper examines the importance of Big Data as a resource for a Small or Medium Enterprises (SME’s). The issues surrounding what can happen if the big data held, is lost or stolen. How can a SME secure the use of Big Data and protect itself from a ruined reputation or expensive fines or both. How a combination of security frameworks can provide confidence to SME’s stake holders.

Index Terms— ISO27001; ISO27002; Big Data; SME; Compliance; Governance; Security

I. INTRODUCTION

The term Big Data is used to describe ‘the exponential growth and availability of data, both structured and unstructured’ (Davenport, 2013). The use of big data analytics is a fundamental driver for business growth. ‘Consumer Data will be the biggest differentiator in the next two to three years. Whoever unlocks the reams of data and uses it strategically will win’ (Grill, 2013). It is important to protect this data whilst it is being analyzed.

The paper takes a look at the previously developed descriptors of Big Data, the V’s, in an attempt to conceptualize them for Small or Medium Enterprises (SME) management.

The use of Big Data as a resource is something SME’s are developing. In using this resource certain pitfalls are lying in wait. This paper discusses how business may benefit from the use of Big Data. It also shows an example of the moral issues that have arisen in using Big Data. It looks at the global security threat that exists to all sectors of industry, when using and storing Big Data.

An Information Security Management System (ISMS) is described in this paper with specific theoretical and practical guidance that is directed toward an SME management team.

II. WHAT IS BIG DATA AND WHAT ARE ITS USES.

Small and Medium Size Business’s inform decisions by using Big Data to maintain business growth. In informing business decisions Big Data Analysis can assist by harnessing social media content of customers; for example individuals may tweet about your product, you may then have questions about the individual that will affect your marketing decisions; these questions could be, How old are they? What sex are they? Where do they live? What industry do they work in? Where do they spend their social time? Big Data Analytics of Social Media claims to answer these types of questions and more.

In using the power of Big Data your business may be or want to be exploiting a loyalty card scheme. Large businesses have been using loyalty schemes for over 30 years with Tesco being one of the most successful. Loyalty cards are often linked to your customers Card or Debit Card. In loyalty card schemes you may be able to gain insights into your customer’s life for instance? How old are they? What is their sex? What is their weekly spend? What items they buy? What day of the week they shop? Do they take up offers? What items are they likely to buy in the future?

There are many companies that fail to realize the need for a cohesive social media structure. A report compiled by The Altimeter Group (Li, 2013) found that only 34% of companies interviewed felt that their social media activity was in line with their business outcomes. The data from this report was drawn from 125 companies with over 100 employees. It shows that even the larger companies are lagging behind in harnessing social media and the Big Data it creates. The companies that are capitalizing on the harvesting of social media Big Data like Shell are able to measure social media content which concern their brand reputation (Li, 2013).

To understand the challenges of handling Big Data it is valuable to look at the 3 V’s.

A. Data Volume

The important issues on Big Data handling come from the shier amount of Data. The growth in the use of Big Data as a resource has grown at a rapid rate in recent years and it is becoming a vital factor in business planning. In a report by Manyika, et al, (2011), ‘nearly all sectors in the US economy had at least an average of 200 terabytes of stored data’.

B. Data Velocity

The speed in which the data is produced and accumulated is known as Data Velocity. The important factor is how quickly you can act on the Data produced. A scenario showing the difficulties of data collection comes from Dumbill (2012), ‘The Large Hadron Collider at CERN generates so much data that scientists must discard the overwhelming majority of it…’
C. Data Variety

Data Variety describes the types of data and its source. The majority of social media data comes in the form of unstructured data; in fact 80% of Big Data comes in unstructured form (Pritchard, 2012). This data can take the form of diagrams, documents, videos but the ‘tipping point’ described by Pritchard (2012) is the data formed from what companies are thinking on the web.

There are many more V’s associated with Big Data like Value, Variability or Virtual (Philip, 2014). It is also important to consider Vulnerability as well, because ‘The volume of customer data that is collected via online point-of-sale and e-commerce sites is a pot of gold for hackers’ (Hewlett Packard, 2013).

III. THE THREATS AND FEARS

The advantages and risks to your business are both huge if you are handling Big Data. If miss managed you potentially put at risk; Customer Confidence, Business to Business Confidence; Brands Reputation and face a large fine. A prime example comes from the beleaguered American retailer ‘Target’ in December 2013 they revealed that 40 million card holder details had been compromised. This breach could equate to huge losses through having to update Information Security Infrastructure, potential legal suits brought by customers and Fines from the Payment Cards Industry (PCI) Council. The fines can be as much as $90 per cardholder for Target this could mean 3.6 billion dollars (Williams, 2013).

In the more recent security breach ‘Heartbleed’ which has revealed vulnerability in a widely used encryption tool, open SSL. This breach could affect every SME that operate over the web. In fact the Federation of Small Businesses is reporting that fraud and online crime is costing its members up to £4000 per year (Tufnell, 2014).

There are further risks to using Big Data analytics as the American Retailer Target found out again. Target employed Big Data analytical techniques to great effect by predicting if women were pregnant. In taking stock of what products women were buying they found out what stage of pregnancy their customers were in. This use of analytics has led to huge growth in Targets sales figures, Big Data analytics has led to a 23 billion dollar growth in 8 years (Hill, 2012). The problem was Target sent vouchers to expectant mothers and in one case the expectant mother was at school age and hadn’t told her family of her pregnancy, the family only found out when the vouchers came through the door. The data your company holds can have huge value but it must be used correctly.

The Greek myth describes Pandora opening a jar, ‘Which she was not to open under any circumstance…and then all evil contained escaped and spread over the earth.’ (Lindemans, 1997). This could be a fear that businesses may have of opening the lid on BIG Data as a resource.

IV. DETAILS OF RECENT DATA BREACHES AND LOSSES

In examining the figures we can begin to see what is vulnerable; where the issues are arising in the Protection of Big Data, more specifically how employing good Governance; managing Security and becoming Compliant can help protect your business and its customers. A report published by PWC (2012) shows that 76% of small business that responded had a security breach in the last year. Looking at the report in closer detail 20% of small business has lost sensitive customer data.

The type of data most frequently targeted by criminals not surprisingly is Card Data. This Card Data along with other types of customer data like; Personally Identifiable Information (PII); and Email Addresses made up 96% of all data targeted by thieves in 2012 (Trustwave, 2013). In terms of which industry is most at risk from thieves, it is the Retail Industry. The targeting of The Retail Industry grew 15% from 2011 to 2012. The Retail Industry accounts for 45% of breaches in the market place with other Industries like The Food and Beverage Industry breached 24%; the Hospitality Industry breached 9% and The Financial Industry being breached 7% (Trustwave, 2013).

The issue surrounding Card Fraud is clearly a massive one. The cost of Card Fraud has doubled in 5 years and in 2012 the losses stood at over 10 billion dollars worldwide. These losses still effect business that are PCI DSS certified. Certification doesn’t make your business bullet proof. The issue comes when ‘organizations viewing PCI compliance as a single annual event, unaware that compliance needs to have a 365 day-a-year focus’ (McRae, 2014).

V. GOVERNANCE THROUGH INFORMATION SECURITY MANAGEMENT SYSTEMS

To control the data that flows through a business, good governance protocols need to be implemented. In employing an Information Security Management System (ISMS) business can help protect against data breaches and help defend the important integrity of their brand. When employing such a system it is in the interests of the business and its policy makers to insure ‘The Information Security Management System preserves the confidentiality, integrity and availability of information…..’ (BSI, 2013).

This can be achieved by deploying the key features of an ISMS framework. There are frameworks for the setup of ISMS. These frameworks identify the areas that need to be addressed to securely use and hold customer data. The frameworks like BS-7799, ISO27001 and COBIT are commonly used by business. The implementation of ISMS is not to be overlooked when addressing Data security.

The difficulty of frameworks such as ISO27001 is the implementation process is lengthy; a drain on resources and a drain on the time of senior managers (Mookhey, 2006) and in SME’s management time is at a premium.

A. Pre Compliance Requisites

The steps of compliance for ISO27001 are described: The initial stage of ISO compliance requires scoping the current needs of the business. An assessment of the current situation needs to be carried out. The information technology systems and frameworks that are currently used need to be compared by high level G.A.P. analysis against the ISO27001 regulations. The G.A.P. analysis can inform you on how far you are away.
from ISO27001 compliance it doesn’t tell you about possible problems or risks to the business (Kosutic, 2014).

A plan needs to be prepared with a timeframe and all key stages of development. The plan needs to understand the context of the organization and the interests of third parties (BSI, 2013). It is critical that senior management understand why this is being implemented and that constant support needs to be maintained not only for implementation but for maintenance and continual improvement (BSI, 2013).

The phase of selecting credible and attainable controls needs to be completed. The controls are the starting point for a credible security policy (Sayer Vincent, 2006).

B. Risk Assessment

This operational phase of the ISO27001 standard is closely linked to the planning phase in particularly the guidance surrounding Information Security Risk Assessment and Information Security Risk Treatment. The Risk Assessment provides a sound basis for ISMS. In operating SME’s this phase will help to provide a basic level of Information Security Governance but it will obviously not lead to compliance. The purpose of the Risk Treatment Plan is discussed by Kosutic (2011), “The purpose of Risk Treatment Plan – to define exactly who is going to implement each control, in which timeframe, with which budget….”

C. Certification

In the final stages of application for ISO27001 certification, the business has to prepare Statement of Applicability (SOA). The document shows the relevance of the controls selected and the risk treatment plans in place.

Following the SOA and the Risk Treatment Plan the process of continually improving ISMS, including corrective action and internal audit documentation needs to be kept up to date. The final stage of pre certification visits, 1st and 2nd stage audits can be conducted followed by certification a pass or fail on reaching certification. If the certification is passed continual surveillance visits within 6 months to a year needs to be carried out. If the business fails certification correction and non-conformities, record corrective action and follow up action, re-apply for the stage failed.

VI. AN ALTERNATIVE VIEW OF AN ISMS

In view of these stages of the ISO27001 and the in-depth nature of each stage; SME’s owners and management may feel it is not necessary, at this stage of the businesses life to gain compliance with the ISO27001, but making steps toward compliance is a result of developing basic ISMS.

A first step would be to examine and design a strong security policy along with conducting a risk assessment. The idea must be to identify what is critical to the business or where are the vulnerabilities?

The 27002 standard ‘puts the meat on the bones’ (Kosutic, 2010). The controls are laid out as in 27001 but have detailed implementation guidelines for the control. In an example from 27002 ‘Access Controls’ a policy control asserts that ‘formal authorisation for access requests’ (BSI, 2013). The policy for formal authorisation for access requests is a key policy to protect critical data and to help SME’s have answers to who has access?

There are further controls in the policy that should be included in a risk assessment. It must be clear who has access to the site where Big Data is stored. A chain is only as strong as its weakest link and data storage is the same. There is no point having up to date patches and well configured firewalls if you are going to leave the doors open to the safe. In ISO27002, section 11 all the precautions needed from checking the security badges of a delivery person to securing access to patch panels and cable rooms are addressed.

SME’s have to view how they are going to insure the availability of their internally held data. It has been reported that 43% of Small Business in the UK spend more time on changing passwords than backing up data (Tufnell, 2014). Invoking a Risk Assessment, critical business data should be highlighted and the outcome should be to regularly back up that data.

It is important to constantly review the externally facing data. In using 3rd party security firms to provide vulnerability scanning and cyber exposure reporting gives an assurance against breaches.

The 2 standards differ because ISO27001 addresses ISMS at management level, it draws a framework for business to work toward and it can be certified. ISO27002 on the other hand addresses the technical aspect of Information Security, not at the management level. The combination of the 2 policies has been suggested but Kosutic (2011) believes the size of the combined documents may disengage management.

If a SME attained ISO27001 certified this will help protect the data they hold and it would provide business to business confidence. This is not the full story however; ISO27001 certification is a mechanism as discussed by Standards Consultants (2013), ‘An auditor doesn’t really care that you have specified a minimum of 16 characters for passwords…..The auditor is concerned that there is a policy in place and that there is a mechanism to review it.’ It suggests that if you have a policy, you are secure. As discussed this is not the case for example the retailer Target is PCI DSS compliant this has not prevented a break down in their ISMS.

It has to be the case that a solid base level of security is of primary concern for SME whether that means a compliance certificate or not. To demonstrate that your business has a hold on data security it must be able to show that it has comprehensive policy documentation in place to guard against breaches and that senior management understand and are committed to their ISMS.

VII. CONCLUSION

It is clear that there are huge challenges that face SME’s in trying to harness Big Data resources. They have to understand the complexities in the Data by using Volume, Variety, Velocity and Vulnerability. It has been shown in this paper that Big Data can provide great insights into customer’s habits and thoughts, although evidence is there to show that mismanagement of Big Data can incur damage through financial costs and reputational damage.
The governance controls and management SME’s have over Data use and storage is vital, the overriding factor to secure this Data is to implement an ISMS. ISMS could be in the form of ISO27001 with certified status. There are strong technical details present in ISO27002 which point the way to IT security without certification.

Alternatively SME’s may find it more appropriate, due to time or financial constraints to form a set of structured documents that make up ISMS. This form of ISMS with a security policy and risk assessment can form a base for certification whilst providing a level of security and peace of mind. In merging the basic governance controls in ISO27001, and then identifying the risk areas for the business, backed up by the technical controls of ISO27002 to form a strong security framework. The framework created makes for strong governance and technical security, without adhering to the strict rules of certification compliance, although it can provide confidence against the threats and breaches to Big Data that can be so costly to a SME’s.

VIII. REFERENCES


Abstract—This report looks at two key issues linked to big data; security and governance. It will define what I consider big data and what it can do for an SME but also how it can be harmful to an SME. Security and governance are two ways to combat big data management and data breach costs and safeguard against future breaches and the monetary costs associated with it. Future proofing in preparation for more stringent EU regulations will also provide key benefits in the long and short term.

Index Terms—SME, Security, EU general data protection regulation, Governance.

I. INTRODUCTION

The amount of data being created by the world population has exploded over the last few years as the number of smart phones, computers, networked devices and social media interactions increases rapidly. Combined with the maturing of broadband networks and increases in network operators’ speed there is more data than ever before. We send over 144 billion emails a day, 340 million tweets a day, upload over 70 hours of new video a minute to YouTube and nearly 350 new blogs are uploaded to WordPress every minute (Conner, 2013). The US mobile operator AT&T boasts a single database that contains 312 terabytes of data on its customers (Compare Business, 2014). The amount of data is only going to grow exponentially as the fact that 90% of the total digital information in the world was created in the last 2 years (IBM, 2014).

II. DEFINING BIG DATA

How big data is defined can vary between institutions and companies. Gartner in 2001 determined that big data can be thought of in terms of three Vs, volume, velocity and variety. Meaning that big data is ever expanding in size, increasing in the rate at which it comes at you (speed it is being created) and the variety of sources that are creating it (MIT, 2014). A more simple definition seems to be that big data is the storing and analysis of big data sets (Barker & Ward, 2013). Others believe there are more than three Vs of big data and validity and vulnerability also need to be considered as key issues when considering the implementation of big data.

III. BENEFITS

Big data has been around for a few years now and it can have big benefits to all sizes of business and help them achieve greater success. Some see big data as the new gold rush of the information age as gaining lots of information to analyse becomes a more important asset in a business’s arsenal when trying to increase profitability and success (Peters, 2012). It can help increase efficiency, improve customer service, identify new market opportunities, reduce go to market time and help to improve targeted marketing (Tankard, 2012). Big data analytics can allow a business to accurately profile customers, using social media and data mined from other sources (DataScience, 2012). With this specific customer data a fitting product or service could then be offered with an increased chance of the customer being receptive. Figure 1 shows some of the ways a company could use big data analysis to its advantage.

Fig. 1. Making use of big data (ISACA, 2014)

Big data can also increase return on investment and half of companies that used it more than 50% of the time exceeded their goals 60% of the time and conversely companies that used big data less than 50% of the time only met or exceeded their targets 33% of the time (Forbes, 2013) A study commissioned by SAS found better use of big data and analytics could add up to £216 billion to the UK by 2017 (SAS, 2014).
Using big data analytics smartly can have real world tangible benefits to the everyday person too and not just the profit margins on a financial spreadsheet. Using big data analysis to improve and more accurately predict helped reduce energy usage by 35% and decrease ICT costs by 50% in a metropolitan city in Spain (DataScience, 2012).

Another key benefit which can provide competitive advantage to a SME in a crowded market is real time analysis of data allowing management to make faster and better informed business choices (IBM, 2012).

IV. QUANTITY OVER QUALITY?

It seems the more big data a company has at its fingertips the more informed its business strategies and decisions can be (e-DBA, 2014). This has created a gold rush to mine and analyse data but this big data can have many issues and pitfalls. The big data is of limited benefit in some cases due to internal politics, poor reliability/validity of data and poor analysis of the data (Wolpe, 2013). The big data being captured and analysed is of little value if the sources and capture points it comes from cannot be relied upon. It is predicted that by 2015 80% of all data will have low certainty and therefore not be so valid and reliable (Easton, 2012). If the data being analysed is not certain then a company could make a big error making business models and marketing strategies based on incorrect data sets. If this were the case in a large company it could simply bite the bullet and accept its mistake and absorb the costs and take a different direction but an SME making the same mistake from its big data analytics could be crippled by the same bad mistake.

A survey carried out by TEKsystems found that 50% of IT leaders didn’t have complete faith in their data and often when a cloud storage and analysis scenario was used the data was not cleaned for errors and correctness but simply captured in full and uploaded as in its entirety (Shacklett, 2014). This means any big data analytics performed on this ‘dirty’ data is less than reliable.

V. SECURITY COSTS AND ISSUES

Using big data analytics means having a large amount of data together in one place. This pooling of data is a big target for malicious parties such as hackers and the risk of industrial espionage from rival companies or just human error. A study conducted by PwC on behalf of the department for business innovation and skills revealed that the cost of a data breach to an SME was £35000 to £65000 and to a large company was £450,000 to £850,000 (PwC, 2013). A study conducted by Symantec and the Ponemon Institute revealed that the most common cause of a data breach was down to human error and system problems (Symantec, 2013). These two issues should be a high priority to any SME that wants to use big data and minimise the risk of high costs from breaches.

VI. ISO 27002

One of the important internationally recognised standards for data security is ISO 27002. It is a good practice guide that any company dealing in sensitive data and so big data would be smart to follow (Computer Weekly, 2008). While the standard applies to information security in general it is still applicable to digital data as big data can be considered physical as it has to be stored on some digital medium (typically on hard drives as part of a single or multiple servers).

While the ISO 27002 is simply a best practice guide it is not mandatory for a company to attain the standard. It could be said that acquiring an ISO certification could improve public perception of an SME’s data security and data protection policies and practices it may miss some key issues revolving around the use of big data.

VII. EU TURNS UP THE HEAT

It is very important for an SME to take its security and governance seriously. There are a number of negative consequences of bad big data security and governance including, loss of customers through negative PR, loss of
earnings and the cost of any fines for data breaches. If your SME is using big data as one of its competitive tools then it must ensure that it is secure and stays private in accordance to government regulations (nationally and internationally depending on where the company trades) and regulatory bodies.

The EU is bringing in a more stringent general data protection regulation regarding data privacy rules which will replace EU directive 95/46/EC and supersede ISO27002. It is designed to toughen up on how companies handle their data (EurActiv, 2013). The most significant of the changes applies to personal data (which could be argued the bulk of big data is). Companies now have to use stronger safeguards for their data in terms of protection and privacy.

While the intention of the EU general data protection regulation is to promote greater, security, privacy, accountability and simplicity amongst European businesses using big data the International Chamber of Commerce points out some concerns in the proposed regulation in its current state (ICC, 2013). The ICC points out that some requirements of the new regulation may not be achievable; the right to be forgotten is something that can be tricky to do with modern always connected technology. It is no longer as simple as deleting a row in a spreadsheet because attempting to remove a customer now could mean having to scour the cloud for all of the little pieces of information scattered around the cloud (which could be across different servers in different countries) even then it could be possible to un-delete the information which would then breach the proposed regulation and an SME could expect to face fines for non compliance which could be up to 2% of total turnover (London Economics, 2013), however a company that obtained a European data protection seal could avoid penalty except in the most negligent cases (Out-law, 2014).

It is worth noting that the right to be forgotten is nothing new and has been in previous privacy and data protection regulations it is the advancement in technology and specifically the reliance on cloud data storage that makes it a big issue when it comes to a company being compliant with it (Mantelero, 2013).

The new proposed EU regulation for data protection also requires a company to notify the data protection authorities and the individuals whose data has been accessed of any breach within 24 hours of it being discovered (Cabral, 2012). This could provide difficulty to an SME who do not have a large number of staff dedicated to information security that a large company would have this again could mean fines for not being fast enough detecting the breach of reporting it. Such fines could soon add up and affect the financial stability of the SME’s as well as its customers losing faith in the brand due to bad PR from data breach reporting.

The proposed regulation also has implications on how an SME or larger company uses any data it has obtained from individuals. The regulation implies that when an individual consents to give his or her personal data it can only be used for the purpose that person was informed it would be used for (De Hert & Papakonstantinou, 2012). This means that there could not be a repeat in Europe of the American Target style fiasco in which Target used big data analysis of customers and their buying habits to predict if they were pregnant and then send appropriate coupons for a pregnant woman to use in Target (Duhigg, 2012). This helps to alleviate concerns about the unethical use of big data when personal data that has been collected from customers, in this case purchase history, that is then used for other purposes which the customer never consented too and crosses a line of acceptable use in the public eye.

One perspective of the proposed regulation change is that it may have a negative impact on the number of companies making use of big data. The increased sanctions against the misuse of it may deter small to medium sized companies from big data acquisition and analysis because of the increased fines and costs of maintaining a big data capability (Cumbley and Church, 2013). It would be a shame if European businesses lost some of their big data analytical abilities and therefore some of their international competitiveness because of a fear of a new EU regulation.

Any SME needs to take the increased costs in to account when producing medium and long term forecasts. The regulation is expected to come in to effect in 2016 (Pearce & McDonald, 2013) at the earliest it is important for a company to be compliant with them before that. This will mean they are prepared with best practice for the protection and privacy of their big data. It also means that they have less chance of being hit by a data breach and the associated cost of time and money. Anticipating and implementing compliance will also provide an opportunity to spread the potential costs across fiscal periods in the most cost-effective way, this is particularly important to an SME that might not be able to absorb surprise costs like a large company could which may include new database, e-commerce and new management systems (Travers Smith, 2013).

The EU general data protection regulation also mandates that a company dealing with sensitive big data needs to have an information security officer named as part of their data security and privacy measures. This should obviously already be happening but if a company did not already have a dedicated information security officer named the creation of the dedicated position could cause considerable set up costs as well as potential issues in recruiting a suitable candidate from a declining number of qualified individuals due to the rapid increase in the number being needed as part of this new regulation.

The management also needs to be aware that they need to take in to account the opinion and expertise of an information security officer and not see their position as a hindrance to the making fast decisions from their big data analytics. The security officer is there to ensure the security and integrity of the sensitive data which is also making the data itself more useful in terms of its validity and reliability (IBM, 2012). A multitude of other benefits from having an information security officer was highlighted by research conducted by the IT policy compliance group which showed they can lead to improved productivity, reduced risk of cost from data breaches, less data loss and increased profit (ISACA, 2010). Obviously any small
to medium sized business dubious to the necessity of an information security officer would be wise to take that in to consideration when looking at the bottom line.

It is also important for companies large and small to be prepared for the worst case scenario in which they do suffer a significant big data breach. The company would need to have appropriate insurance measures or cash in reserve incase fines need to be paid, extra man hours are needed to secure and clean up after the data breach.

VIII. CONCLUSION

Using big data can have massive benefits to any company increasing their competitive edge through improved targeted marketing, increase efficiency and improve customer service amongst others. It is important to keep in mind that quality and checking the validity of information being used for big data analysis needs to be taken seriously. Existing ISO standards and the incoming tougher EU general data protection regulation mean that an SME looking to take advantage of big data also needs to be prepared for the cost of implementing appropriate big data privacy and security as well as being prepared for the costs and fines if they get it wrong and suffer any data breaches. The notion that with big data comes big responsibility is something that should be kept in mind of any SME’s chief executive officer and top level management. It is also abundantly clear that appointing a chief information security officer is also something that should be dealt with utmost urgency as not only will it provide some future proofing compliance but will have its own numerous benefits a company would receive before the new EU regulation comes in to effect.

IX. REFERENCES


Abstract—Big Data has hit the world by storm; it has brought advantages as well as disadvantages for many SME’s (Small to Medium sized enterprises). Big data is being portrayed as the new phenomenon for SME’s today. However the veracity of big data is proving to be a problem which is leaving many SME’s with problems and money loss due to the errors that occur from big data. The sources of where big data is being gathered from are not totally reliable and it is clear that SME’s should most likely hold off from joining this new phenomenon due to the new overload of data and the uncertainty of the reliability of the data being used. The purpose of this article is to inform SME’s who are looking to invest into big data the pros of what big data might bring, most importantly to give them enough information about the issues of the reliability and trusting of the data they receive and analyse.

Keyword—SME, Trust, Reliability, Cost.

I. INTRODUCTION

Small to medium size enterprises (SME’s) over the years have grown excessively. Out of the 3.7 million businesses in the UK, 99% of small businesses account for this and they account for 58% of employment as mentioned by Hillary (2012). In order to be accounted as a SME the European Commission (2009) adopted on 6th May 2003 a definition which defines small and medium-sized enterprises. These definitions come under three size criteria which are shown in the below image:

<table>
<thead>
<tr>
<th>Company category</th>
<th>Employees</th>
<th>Turnover</th>
<th>or</th>
<th>Balance sheet total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium-sized</td>
<td>&lt;250</td>
<td>≤€50 m</td>
<td></td>
<td>≤€43 m</td>
</tr>
<tr>
<td>Small</td>
<td>≤20</td>
<td>≤€10 m</td>
<td></td>
<td>≤€10 m</td>
</tr>
<tr>
<td>Micro</td>
<td>≤10</td>
<td>≤€2 m</td>
<td></td>
<td>≤€2 m</td>
</tr>
</tbody>
</table>

Figure 1: Definition of SME table.

Due to this growth it has meant that SME’s have stated to create, gather, generate and analyse a lot of data. In 2012 the digital world of data was expanded to 2.72 zettabytes (1021 bytes) (Sagiroglu, 2013), which is more data than what we had stored in the entire Internet 20 years ago. The information volume doubles every 18-24 months for SME’S as mentioned by Zheng, Zhu and Lyu (2013). IBM (2014) indicated that every day 2.5 exabytes of data is created and 90% of the data on the internet has been produced in the last two years. From this it is clear that data has started to take over rapidly. This is why this data has started to be used by many SME’s across the world to allow them to gain better understanding of customers and services. This data to the SME’s world is known as big data and shall be look at in greater detail within this paper, looking mainly at whether SME’s can trust the data that is received by the different analytic tools used by todays SME’s.

II. WHAT IS BIG DATA

Big data is one byte more than you are comfortable dealing with as stated by Rogers (2012). A formal definition of big data states that big data is sets of data that are so large and complex that it becomes impossible for it to be handled by standard database management systems tools or traditional data processing tools. The data that is gathered is no use to anyone and is worthless, but once this data is analysed using the tools created for big data, it allows the data to be presented in a clear and digestible format allowing business to gain great opportunities. This why it has taken the SME world by storm not due to its very large irrelevant data (Katel, 2013) but due to the ‘relationally to other data” (Boyd and Crawford, 2011) and the different opportunities it ‘might’ bring SME’s.

The value of big data comes from the different patterns that can be formed by making different connections between sets of data about the individuals and groups of people. Boyd and Crawford (2011) argues that SME’s see big data as a pure opportunity that allows them to use big data to target advertising towards it consumers and make business decisions as each of us are now a walking data generator in the eyes of business analytics.

Big data initiative span four unique dimensions which are:

- Velocity: referring to the speed of which data is being generated at and the speed of how it is analysed.
- Volume: referring to the size of the data files that have been gathered.
• Variety: referring to the different sources that the data is gathered from.
• Veracity: the understanding of the reliability and unreliability of data and being able to conduct trustable analysis.

Often the variety of big data is a major issue for SME’s like retailers, for instance a lot of SME’s use location data from mobile phones and social networking sites like Facebook and Twitter (sources that are interesting and powerful sources of data), which provide a great insight to peoples thoughts which is difficult to gain from other sources as stated by Balar et al. (2013). Take for example; on black Friday in the USA, Alex Pentland (Corbitt et al. 2013) used peoples GPs data to calculate how many people were in the shopping retailer Macy’s parking lot. This was to give the business an idea of how many customers would be shopping with them along with the most important advantage of how much revenue they were most likely to make. A problem here that the retailer probably didn’t consider was how they could for certain make sure that the people in the car park would actually walk into their stores as well as how they would know that the consumers would purchase something. Taking into consideration that a lot of man power, time and equipment is used to analyse these different data sets, the data at times does not give accurate information that SME’s can use to make decisions on or get facts.

This then is where the problems of veracity start to begin. As the traditional systems for velocity are not acceptable enough on performing the analytics on data which is constantly in motion (Katel, 2013). Due to the raw structured semi and unstructured data, the velocity is difficult to be handled by the existing traditional analytic systems (Katel. 2013). Not only this but with the increased automation of data collections and analysis, SME’s are now dealing with complex structures as the data that is generated from these systems are ‘heterogeneous’ of multiple data types as mentioned by Zheng, Zhu, and Lyu (2013), showing that the key difference between traditional and big data is not just data velocity but its form (Anon. 2013). Creating and bringing about problems of trust, due to the vast volume of generated data. This is where veracity, the trust issues of whether big data can be trusted start to be questioned.

III. TRUST

Trust is a major problem when it comes to big data. Many SME’s need to consider questions like how can the data gathered and analysed be trusted. If data is relevant, how much data would be enough? This is where the issues of trust spiral in, if SME’s consider using big data in order to get trustworthy information they need to gain the trust of their customers first. This is an important aspect and something that can never be 100% achieved. Research by Rose (2010) states that SME’s like retailers that manage to create trust between them and their customers manage to gain “five times more data” than other SME’s that do not have their customers trust.

However, this is how majority of SME’s lose ‘trillion of dollars’ promised to them by investing into social and economic valued big data due to the trust not received by its consumers but most importantly trust being broken. Mainly due to the fact that trust can be broken between consumers and businesses at a click of a button. For example if a customer’s purchase automatically gets posted and shared on Facebook without them knowingly having authorized it, this is a means of how many SME’s are most likely to lose their customers trust due to them using different social media sites to gain data about their consumers.

Keeping this in mind, Facebook and Twitter are two different sources of where SME’s use different variety of data, which produce terabytes of data every day, as well as keeping in mind that this amount of data is difficult to be handled using the existing traditional systems as mentioned by Katel (2013). Also keeping in mind that 20% of these accounts are fake as stated in the research by Cohen (2011), allowing it to be seen that already two sources of where majority of SME’s collect their data have reliability issues, bringing the problem where a lot of loss could be made if SME’s rely only on this data to make decisions.

Also remembering that in order to get customers details it can be an extremely challenging situation as 75% of customers privacy of personal data remains a top issues in the world of SME’s (Rose. 2013). So once they do give in to giving their information it is quite easy to lose their trust if SME’s misuse their information. This can then lead to many consumers giving false information which can affect the veracity of the data.

Some sort of control is required over the volume, velocity and variety of data that is being generated due to these SME’s. What SME’s need to ask is where is the evidence to say that due to big data, businesses are improving their performance there is a lot of information and case studies to show the value of being big data driven (McAfee and Brynjolfssonm, 2012) and business experts and academics are trying to move and convince all SME’s into investing into big data (McAfee and Brynjolfssonm. 2012) however no one has currently tackled the big question of how companies that use big data have successes and if the data gathered is trustworthy

This then bring the second issue of trust, not knowing if the information received is still accurate or even useful as various problems that contribute to trust of the big data management includes scalability, unstructured data, accessibility, real time analytics and importantly fault tolerance (Patel, Birla and Nair, 2012). The Web is a major source of gathering data for many SME’s, this is where the main challenges hit which are linked with using the different data applications required to gather the data to determine which data should be treated as trustworthy (Bizer et al. 2011). The quality of data is not just a problem that is caused due to the analysis of data by the IT department or business department it is an overall organization problem. The reason being that many SME’s don’t state specify roles and responsibilities pertaining to management and delivery of data quality expectations. Not only this but many businesses avoid the use of data quality processes at time and don’t apply data quality processes as frequently as ideal. All of these results in data quality issues that
compromise the results of reporting, analyzing, forecasting and other key businesses requirement or they stop businesses processes thereby reducing the timeliness and accuracy of decision making.

This then brings another issue where big data becomes misleading due to the trust being lost by consumers who might have lost their trust in the SME generating misleading information for many SME’s.

IV. COST OF MAKING BIG DATA TRUSTWORTHY.

Over the years tools have been invented to handle the volume, velocity and variety of big data. This is where big data analytics have needed to be employed in order to help handle the enormous quantity of data. In a survey for the business analytics (Anon. 2012) it was founded that 97% of companies with revenue exceeding £100 million (Winter, Gilbert and Davis. 2013) were found to use some form of business analytics (Chen, Chiang and Storey. 2012). The collection of big data and storing it comes at a massive cost for SME’s especially small SME’s due to them not having the capital to invest in this. However when SME’s do decide to invest, it is not just the cost of gathering and analysing the data but the cost of making sure the data is trustworthy which affects many SME’s as if the data is not trustworthy then the rest of the process has no use to them.

It is clear that IT leaders in SME’s will take all technical aspects in mind before storing all the data, which brings conflicts between IT departments and businesses leaders. Research conducted by IBM (2014) show that 47% of business feel strong level of trust issues between the IT departments due to the collection. This brings problems when it comes down to the use of the data as it can cloud judgment.

The fact is that it is impossible for devices to be absolutely 100% full proof from errors (Katel, 2013), in order to get a 100% reliable fault tolerant machine or software is impossible as stated by Katel (2013). The only possible action that SME’s can take is to reduce the probability of failure to a minimal and acceptable level. This then brings the issue that no matter what some sort of loss of money and consumers shall always be made as the entire process of big data cannot be totally reliable.

If SME’s do consider in investing with big data they shall need to invest in some sort of technology, making SME’s take into consideration ‘do they have the technological capability to take advantage of big data’. It is not worth investing millions of pounds into technology that could not bring accurate information. Highly due to the fact as Stonerbraker (2013) states the simple analytics that data warehouses can apply to the systems are no longer up to scratch. Due to the unstructured volume of data spouting up from everywhere.

This then takes us back to the main point of trust. SME’s need to consider the costs of errors that are made due to the data analysed but most importantly to where the data has come from, which have become enormous as studies show that knowledge workers waste up to 50% of time hunting for data, identifying and correcting errors and seeking confirmatory sources for data they do not trust (McAfee and Brynjolfsson. 2013). Usually when data is scarce, too expensive to obtain or not even available it is usually to someone who is in charge to make the decision. However, when big data is involved, this isn’t this case. People would think that having a HiPPO (Highest-paid person’s opinion) is enough to overcome inaccurate data as they have the final say. The problem here is that with big data it has been proven in the big data community that the HiPPO can usually be changed if the data he/she receives does not agree with them (McAfee and Brynjolfsson. 2013). So they end up agreeing with what the data tells them rather then what their own opinion is. Again showing that even when people are faced with this type of data they always seem to assume that the information is correct, this however in majority of cases is not the case. Creating a problem for when SME’s use big data as it is clear that the veracity of big data is not up to 100% and that even thought this is the case business executives are still trying to promote the use big data.

The collection of big data and storing data comes at a massive cost. SME’s will always want more data storage whereas IT leaders will take all technical aspects in mind before storing all the data which again brings conflict between the IT departments and businesses leaders again. Problem occur in all forms with big data especially when the data becomes misleading, which brings problems for many SME’s that need to make decisions based on this data.

A. Misleading of data and skills required.

As stated previously Facebook and Twitter is a top source for SME’s to use to gather data to improve their business. How do SME’s however know what is being said on these sites is due to what consumers feel or due to what their peers think should be said. Millions of pounds are usually spent on data that has zero value from these different sources and SME’s need to consider the veracity and validity of these different sources, as external sources is what makes big data. Many SME’s are not ready for this type of data. Boyd and Crawford (2011) discuss in their article the different data that can be generated from different social media sites. Many would think that the ‘friends list’ on an users Facebook account and Twitter account are similar, however this is never the case as the reasons for why users have certain people on their accounts can differ, showing that these different sets of data cannot be combined to give a clear indication of what users want or feel. Again showing that unreliability of different data sets that are used for big data.

The second challenge that is affecting the trust of big data is the people that handle it. There is a growing skill gap for people who know how to manage and mine big data. SME’s feel that new technology that promises pattern recognition and machine learning (Lohr, 2012) can offer them the reliability of data they require. However, the cost of these and the expertise that this type of technology require is lacking, this is why SME’s are struggling to provide their original task of giving excellent customer experience. The problem is also with small SME’s being unable to afford the top technology and the expertise required in order to make the data they have reliable and trustworthy. There are ways that SME’s could get around
this problem however it would mean relying on other businesses data (Manovich, 2011), which can never truly be relied on as the questions of variety arise as it is never clear if the data has come from a reliable trustworthy source. SME’s have reasons to believe that this type of data is helpful however at times it is not the case as the veracity of this data is always questioned.

V. CONCLUSION

There are reasons for why many SME’s should believe what their executives and businesses analytics are telling them about big data bringing them more business opportunities. However they are not being informed of the trust problems behind the data that is being collected. SME’s are not yet ready for this type of change, as there are many reasons for why they should wait. Not much information is out in the market to help SME’s get reliable trustworthy data from their consumers. On top of this not many people in the market have the appropriate skills to carry out the analysing of the data that is collected, meaning that not many SME’S shall be able to trust truly on the analysing that is made from the data.

The lack of trust that consumers have is bringing veracity problem also as not many consumers are providing the correct information, as they fear of their data is being misused. SME’S need to wait and be able to take great advantage of big data only when they have the ability to understand the risks and have more knowledge on how to prevent the problems that are associated to big data. Only then could it be possible for big data to be used correctly and even then there shall still be problems of veracity. However if SME’s are aware of the different trust issues that can arise then SME’s can be prepared for the disadvantages that may arise and deal with them in a manner in which SME’S shall have arranged beforehand due to the knowledge which at the moment is lacking for many SME’S.

VI. REFERENCES

Anon. (2012) Big data needs to be a big headache, Strategic Direction, 28(8), pp. 22-24


Katal, A. (2013) Big data: Issues, Challenges, tools and good practices, Contemporary Computing (IC3), India, 8-10 August, Nodia: IEEE


Abstract — This essay considers the security of existing BD platforms, focusing on Big Data storage, to advise small and medium businesses developing Big Data applications and tools based on these platforms, regarding the issues they are likely to face in terms of security assurance.

Index Terms — Big Data, Application Development, Security, NoSQL, Cassandra, MongoDB

I. INTRODUCTION

As the world is becoming more reliant on technology and the Internet, our daily activities produce more and more data through the use of various interconnected devices. The amounts of information produced double every year (Monshaw, 2012) with 90% currently available data having been produced in the last 2 years (Selinger, 2013). Furthermore, it is estimated that information production will grow 40% every year with consumer data doubling every 18 months (Gantz & Reinsel, 2012). This exponentially-growing data is referred to as Big Data (BD). Based on its general sources, it can be divided into three categories – process-created, machine-generated and human-sourced (Rogers, et al., 2013). BD is characterized mainly by its volume, variety and velocity (Dumbill, 2012), however, there are more V’s emerging in relation to BD as it becomes a greater part of everyday life and business (Normandeau, 2013). As it turns out, vulnerability can be one of them.

Having access to all this information enables smart and interconnected devices to become even smarter, using data from various sources to inform their operation and decision-making. In the same way, businesses are catching up to using BD to their advantage (Brynjolfsson, et al., 2011). 36% organizations are already implementing BD projects, and over 50% organizations predict that BD will become a strategic priority in the next 5 years (Varonis, 2013). BD has increased data usability, resulting in over $2 billion increase in annual revenue, as well as a $65 million increase in net income from improved data accessibility (Tyagi, 2013). 72% organizations successfully use BD to improve their operational efficiencies (Tyagi, 2013).

The problem is that there are no uniform standards to the format and structure of such data. Due to the wide variety of sources, from car sensors to social networks, the manifestations of BD are also different (Normandeau, 2013) and hence hard to manage in a uniform way. To address this, new technology has been introduced to collect, process, store and analyze BD (Dumbill, 2012).

As the demand for BD and related technology grows (Zeidenstein, 2013), more companies will be offering applications, based on existing BD platforms, or even entirely new tools. Small and medium sized businesses (SMBs) are also increasingly making use of BD opportunities (Valsh Velkat, 2014). Software and systems development can be a challenging industry for SMBs due to large and experienced competitors in the market. Providing secure and functional solutions for BD can become a significant competitive advantage. When developing tools for Big Data storage, processing or analysis, developers need to consider security in design of these tools, as well as the security of any underlying components.

II. AGE OF VULNERABILITY

BD technology has not been around for as long as traditional data processing tools, yet even these traditional tools often prove ineffective in securing sensitive information (Frei, 2013), such as credit card details, addresses and identifying information. Alarming, 90% organisations have had at least one breach of their systems in 2013 (Ponemon Institute, 2013). On the whole, more than 0.5 billion records containing sensitive information were leaked in 2013 (Gideoni, et al., 2014).

Furthermore, cybercrime is on the rise, especially the financial subtypes (Parton, et al., 2011; Ponemon Institute, 2013). This means that attackers are more actively seeking to profit from poorly secured data. This also means that data stores are a prime target (IBM, 2013). As systems security develops, the attacks are also becoming more advanced and persistent (Ayrapatov, 2013; Gideoni, et al., 2014). Data leaks and breaches can result in phenomenal loss on the part of the breached organization. Ponemon Institute (2013) estimates the average cost of a compromised data record is $136. Globally,
the loss from financial cybercrime reached $11 billion in 2012 (Simonetti, et al., 2014). Security vulnerabilities can cost a company not only financial loss, but also damage their reputation by significantly weakening consumer trust (Gideoni, et al., 2014).

BD is no different in containing sensitive information that should not be accessed by unauthorised parties. In fact, the fraction of BD that needs to be protected is growing faster pace than BD itself (Gantz & Reinsel, 2012; Securosis, 2012). While businesses are increasingly adopting BD technologies (Vesset, et al., 2013) to make better decisions, increase profit and expand market opportunities, it is also likely that they are putting themselves at a greater risk of a costly data breach, by rushing to implement new technology and applications based on developing or non-existent security models.

### III. NEED FOR BIGGER STORAGE

The volume and variety of data has lead to the development of new storage techniques. Where relational databases are no longer capable of managing the expansion of data, and where data is largely unstructured, non-relational NoSQL databases, such as CouchDB, MongoDB, Redis and Cassandra (Urbanski, 2012; Chow, 2013), are used. Three major types of NoSQL are represented by key-value stores, column stores and document stores (Leavitt, 2010), however there are many other types and subtypes available to fit various tasks. This new way of storing information provides agility and flexibility unavailable with structured data stores (Dumbill, 2012; Zeidenstein, 2013). Simple, schema-less data models, high scalability, reliability and use of primitive query languages make NoSQL ideal for storing and managing BD (Okman, et al., 2011; Urbanski, 2012). Additionally, NoSQL allow for changing data and data types on the fly (Chickowski, 2012), providing for the real-time response capabilities sought by BD users. This is not merely new technology, but a completely new approach to databases. High demand, speed of development and deployment (Kirkpatrick, 2013) and availability of open source solutions are bound to make these databases attractive to application developers.

While providing numerous benefits, NoSQL tend to trade off features of traditional databases, such as integrity and security (Zeidenstein, 2013; Rogers, et al., 2013). In fact, Brewer’s CAP Theorem (Brewer, 2000 in (Han, et al., 2011; Sullivan, 2011)) states that from consistency, availability and partition-tolerance, only two qualities can be achieved simultaneously. This applies to NoSQL storage. Availability and partition-tolerance are crucial for these systems, due to the nature of BD, hence consistency of records will often be sacrificed to comply with core BD requirements.

### IV. NoSQL – NO SECURITY

Unfortunately, compromised consistency is not the only disadvantage of NoSQL. Most of these databases have no security model in their design (Okman, et al., 2011; Securosis, 2012; Chow, 2013; Kirkpatrick, 2013; Stevens, 2011; Sullivan, 2011; Urbanski, 2012; Rajan, et al., 2010). In addition, while a degree of control was enforced in relational databases, the schema-less, unstructured nature of NoSQL provides no mechanisms to ensure sanity or validity of data (Chickowski, 2012).

NoSQL security vulnerabilities are a concerning example of why BD application developers should be extra vigilant when it comes to security. However, they also demonstrate that there is wide scope for developing tools and applications that could make up for these weaknesses.

NoSQL were developed for use cases of operation in trusted environments (IBM, 2013), or environments where security would be provided solely by middleware (Rajan, et al., 2010) or perimeter security (Securosis, 2012). Looking at breach statistics, it is easy to see that infrastructure security often fails, and trusted environments are extremely rare. In addition, users of BD technology often focus on obtaining a working implementation as soon as possible (Kirkpatrick, 2013), sacrificing proper setup of security measures in the process. Inexperience of BD users due to relative novelty of the technology also adds to the problem.

Cassandra and MongoDB can be used to demonstrate some of the most prominent vulnerabilities. In both databases, data at rest is unencrypted and there is no provision for encryption (Okman, et al., 2011). Furthermore, data is transmitted between nodes unencrypted, and no SSL or TLS encryption is provided by default for communicating with external clients (Okman, et al., 2011). In addition, MongoDB makes use of standard RESTful interfaces that send data in HTTP packets (Kirkpatrick, 2013), hence providing a well-known exposed API. These vulnerabilities show how NoSQL fail to protect data at rest and in transfer. Another major set of vulnerabilities lies in provision of authentication and authorisation. Cassandra does not require authentication by default; when authentication is enabled it uses a flat Java file to manage users, passwords and password properties, and stores passwords in plaintext or as MD5 hashes (Okman, et al., 2011), while MD5 is not a cryptographically secure algorithm (Wang & Yu, 2005). MongoDB also does not enable authentication by default, and does not support it in distributed mode; if enabled, passwords are stored as MD5 hashes of a known-format string (Okman, et al., 2011). Both databases only allow for read or write authentication, applying the permissions to the whole database, and using a flat file to manage permissions (Kirkpatrick, 2013; IBM, 2013). This can result in inconsistencies across clusters and inability to change permissions when in use. Most importantly, there is no access segmentation – every authorised user will have access to the whole data store. For the attackers, this increases the value of obtaining user credentials, which can be done by sniffing network traffic, for example.

Furthermore, the assumption that NoSQL means no SQL injection is also wrong. Most NoSQL solutions are vulnerable to injection attacks (Chow, 2013; Urbanski, 2012; Chickowski, 2012). The various query languages used for NoSQL are very similar in syntax to SQL (Chickowski, 2012), and hence open to similar attack vectors. The number of XSS and SQL injections is decreasing (Gideoni, et al., 2014), while overall security breaches are on the increase, which could mean that
attackers are shifting their attention to new, poorly secured technologies, such as NoSQL data stores.

Some of these weaknesses can be remedied by additional configuration of the system, infrastructure improvements (e.g. putting the data store behind a reverse proxy (Okman, et al., 2011)) and writing custom scripts for the data stores – a benefit of their open source nature. However, there are many organisations that will not have the skills, time or funding to achieve this. End-users also lack awareness of many of these security issues (Kirkpatrick, 2013), when deploying their NoSQL stores. Add-on security solutions from vendors, such as HP, Accenture or IBM, are also available for improving overall security of BD tools. However, the cost and potential performance and flexibility reduction from introducing such middleware limits its applicability to BD solutions. Furthermore, every additional element in the system introduces new security concerns (Securosis, 2012). NoSQL data store vendors are also becoming alert to these security issues and working on improvements (Securosis, 2012; Chickowski, 2012; Zeidenstein, 2013). However, these are all post-design solutions that patch known vulnerabilities. The lack of a security model is a design flaw, making many of these post factum attempts at security seem somewhat futile. In fact, in their attempts to provide helpful tools for end-users, NoSQL vendors can simplify some of the possible attacks. For example, MongoDB provides a “MongoSniff” tool that allows to trace real-time activity of the database, which can also be exploited by perpetrators (Chow, 2013).

V. WHAT MAKES A SECURE APPLICATION

A. Asset Protection Requirements

The level of application security is largely dependent on the type of data it deals with. Data can be divided into levels of security associated with its ascending sensitivity: privacy only, compliance-driven, custodial, confidential and lockdown (Gantz & Reinsel, 2012). BD tools can be involved in handling any of these types of data. Hence an important part of requirement gathering should be establishing the necessary level of protection for the information handled by the company. This does not always concern only the data handled directly by the application to be developed, but also any data that could be exposed through the application. Guidance on evaluating the risk and hence the necessary protection is abundant. ISO 27001 and 27002 (British Standards Institute, 2013) provide guidance on Asset Identification and Risk Evaluation for organisations establishing their security infrastructure. The same principles can be applied to estimating the value of application data and hence the necessary levels of protection.

B. Big Data Performance Requirements

The requirements that BD puts on non-relational data stores will also be a factor in designing security of the applications that make use of these databases. Rogers, et al., (2013) state that response, economic factors, workload, scalability and structure flexibility are crucial in providing BD solutions. Hence developers need to make sure that the security architecture of their applications does not slow down BD processing and response times, and that it scales with the data, as necessary.

C. Consideration for NoSQL Vulnerability

Application designers and developers need to be aware of NoSQL security vulnerabilities. Most applications working with BD will have to use a non-relational data store in the back end, especially if they want to ensure sufficient scalability. Most of all, applications working with no NoSQL need to ensure protection from injection attacks, port scans, buffer overflow and man-in-the-middle attacks, from which these databases are poorly protected by default. Given the weaknesses discussed, developers need to ensure that their applications provide for and manage encryption, as well as integrate some form of user roles and access control. Most importantly, there are many different NoSQL distributions, and the issues they face are vendor-specific. Application developers need to make sure they fully understand the specific data store their application will use.

Awareness of existing vulnerabilities and other possible attack vectors also provides the opportunity for software developers to produce their own storage solutions, integrating security into its design, offering a competitive advantage over existing NoSQL data stores. However, this implies having strong security skills in the development team, and allowing for enough time for research, design and implementation of security within the development lifecycle.

It is impossible to provide complete security, same as it is impossible to fix all issues with NoSQL (Stevens, 2011). Attackers will always be able to find new ways to infiltrate a system, provided there is profit involved. However, making it too difficult to break in can be sufficient (Securosis, 2012). ISO 27002 (British Standards Institute, 2013) guidelines for Acceptable Risk can help determine the reasonable extent of security controls.

D. Understanding the Client

When embedding the tools and applications built for Big Data into existing client infrastructures, the restrictions, limitations and vulnerabilities of these environments need to be considered. Their security setup needs to be considered as much as the security the developed application, as the application will not be isolated from the rest of the client environment in the vast majority of cases. Security is not a purely technological issue, affected by various organisational and human factors (Kraemera, et al., 2009; Gantz & Reinsel, 2012). One of the aspects of client infrastructure and operational environment is the human factor. Developers need to consider the competency of end users of their applications so that the functionality or security of these applications is not easily compromised. Many companies taking on BD initiatives have no experience or awareness of the common security issues and technicalities (Chickowski, 2012), which needs to be addressed by training provisions and comprehensive documentation. Most of all, when designing applications, the security architecture needs to be as non-disruptive of business-as-usual as possible. People will sacrifice security for convenience, as short term benefits tend to be more gratifying.
than the more abstract long-term ones (Schneier, 2013; Schneier, 2009), hence disruption of workflow would result in a natural reluctance in users to comply with security procedures.

VI. CONCLUSIONS

BD can offer many benefits to organisations in increasing profits and reducing operational costs. However, it introduces new technology, which often sacrifices speed and performance for security and integrity of the data it deals with. The vendors of BD tools and platforms have placed a large portion of responsibility for security on application developers and middleware providers. The novelty of this technology also means there is a lot we do not know about it, including unacknowledged security vulnerabilities.

Developing new applications that take advantage of BD can be very profitable, as more companies begin to invest in BD strategies (Zeidenstein, 2013). However, security issues of BD tools stretch beyond NoSQL data stores (Securosis, 2012; Zeidenstein, 2013), and developers building applications, using BD components or platforms need to be aware of these issues when designing security architecture for their applications. Furthermore, developers need to make sure they have sufficient skills to understand not only the existing vulnerabilities, but also other possible attack vectors. Finally, the client’s infrastructure should be considered as an operational environment, accounting for competence of end users and aiming to enable cooperation in not compromising application security.

New technology can and should be implemented, as it offers new opportunities and allows to compete in the market, but the risks need to be acknowledged, carefully considered and addressed. In the end, “cybercrime is not just a technology problem – it is a business strategy problem” (Skalak, et al., 2014, p. 4).

VII. REFERENCES


Parton, T. et al., 2011. Cybercrime: Protecting Against the Growing Threat, s.l.: PricewaterhouseCoopers LLP.

Ponemon Institute, 2013. The Post Breach Boom, s.l.: Ponemon Institute LLC.


Is Your Big Data Secure?

Ben Sumner
Department of Computing and Law
University of Derby
Derby, United Kingdom
B.Sumner1@unimail.derby.ac.uk

Abstract— Big data has grown a great deal over the past decade, and with the increase in volume of this big data comes the problem of a breach. The benefits of big data are well known and documented, yet the risks are less in the spotlight. This paper will bring the less noticed risks and consequences of big data breaches to the spotlight in order to allow companies to see both sides of the big data coin.

Index Terms— Big data, data breach, SME, Data breach protection

I. INTRODUCTION

This paper will explore the definition of big data. Inquiring into the costs of big data, specifically to companies which fall into the SME bracket. Exploring commonly overlooked aspects of a data breach, and the unforeseen costs that are insured by a breach.

II. WHAT IS BIG DATA?

Big data by definition, is a collection of data which has been accumulated. However large quantities of data does not necessarily define big data, there are other defining factors known as the V’s of big data. (Hurwitz, Nugent, Halper and Kaufman, 2013). There is discussion over the number of V’s in big data, up to as many as eleven V’s on occasion, however there is agreement upon the three core V’s. (Berman, 2013) These are constituted of the following:

- Volume – Large Quantity of data
- Variety – The amount of different forms of data, through complexity or data format
- Velocity- the ever changing nature of the data

These three core V’s must apply, “it is the size, complexity and restlessness of the Big Data resources” (Berman, 2013).

III. SECURITY

A. Big Data loss and its cost

Big data however has many risks associated with it, as the cost of a data loss to the average Small Medium Enterprise (SME) ranges greatly. According to a recent report by PWC (PWC, 2013) the cost of the worst annual breach for a small and medium enterprise can be as much as £450,000, this shows the potential risk in big data. The initial monetary loss is far from the largest impact felt by the company on the wrong end of a data breach, there are numerous other consequences that can cause catastrophic damage to the company. As defined by Zurich (Zurich, 2012), there are numerous hidden costs which come into account in the wake of a data breach, these costs are:

- The cost of a financial examination to determine the severity of the breach
- The cost of notifying third parties to the breach
- The cost of handing customer complaints and customer reimbursements
- The cost of maintaining credit and identity monitoring to prevent customer identity theft
- The cost of public relations, and the damage it may do for the company image
- The cost of legal defenses to defend claims
- The cost of fines from regulators

These forms of damage can ultimately lead to the company no longer able to hold its own in the marketplace. SME’s especially are especially naive to the costs of a data breach by estimating the breach at an average 68% of the cost of an average breach (£94,750 estimated - £138,700 actual) (Faronics.com, 2012).

B. Are you taking a risk?

With 87% (NCSA/Symantec, 2010) and growing small and medium enterprises are using the internet for essential day to day activities and that 73% (NCSA/Symantec,2010) of small and medium enterprises believe that a “trusted and safe internet is critical to their business success” (NCSA/Symantec,2010) . It is surprising to find that as little as 31% (NCSA/Symantec, 2010) of these companies even has an informal internet policy to govern how they access the internet.
The figure that just 2% (NCSA/Symantec, 2010) of small and medium business owners are concerned about internal data breaches caused by employees or former employees shows that many small and medium business owners are looking the wrong way for data breaches. These companies should be looking to be far more scrupulous about their internal affairs and their internal security posture as opposed to the external risks, given the pie above 37% of all data breaches occur from within from human error or maliciously from the inside of a company. (Symantec, 2013).

There are a number of vectors of attack that companies suffer resulting in data breaches; with six out of ten of the top ten breach vectors of SME’s being from an internal vector (Watchguard.com, 2008), shows that the ignorance to the internal threat is a very large issue to SMEs.

As many as 65% of SMEs say that their organisation’s confidential business data is unencrypted (trendmicro.com, 2012), this is a heavy risk today’s world, where the attack vectors are not just from the outside but are from within. This lack of encryption provides any and all employees unfettered access to data which if publically available could be highly damaging for the company.

IV. AUDIT

A. Have you been breached? Are you sure?

Big data breaches are a major event in any company but in a small and medium enterprise the event of a big data breach can be critical, it can be the end of the company if the scale of the breach is large, but how it is handled can affect how and whether the company continues to operate in the future.

Surprisingly big data breaches take a long time to discover. In as many as 62% (Verizon, 2013) of profiled breaches, a breach was left unnoticed for months and in some cases (4%) (Verizon, 2013) years, yet these breaches were conducted in a matter of hours (or less), and 1% (Verizon, 2013) of those breaches are discovered by auditing, unfortunately this is due to the lack of information auditing in companies of all sizes but most notably in smaller businesses without the capital to afford such teams. A notable example of where finding a breach was severely delayed was with the 2014 Target breach. Although target would not be classified as a SME it provides a very important warning to ensure that a data breach is discovered swiftly to avoid catastrophic circumstances. One of the critical failures of the target breach can be attributed to human intervention, whereby security warnings which arose were ignored leading to 100 million records being stolen. (Computerweekly.com, 2014). Showing that even with security tools, a system is only as vulnerable as the staff working within it.

Trust

Who can you trust?

As described under the security section trust is critical to any organisation, and to have as many as 37% (Symantec, 2013) of all breaches occurring from a “human factor” is disconcerting, whether these breaches are deliberate or an innocent mistake the percentage is none the less high.

As a company it is imperative that an employee is trustworthy, a company cannot spend all its time monitoring its employees all day and night. An employee however should know what is and what isn’t expected of them to control and to prevent breaches, again this can be done with training and with
policies to ensure that they know where their responsibilities begin and end.

Inherently it is naïve to simply rely upon trust, a previously mentioned example of where trust existed but failed was the target breach of 2014 (Computerweekly.com,2014) where employees made the incorrect decision to ignore a security warning. This is an example of where the company to employee wall of trust is broken (Anash, 2011).

V. ETHICS

A. The cost of big data

The cost of big data can be counted not only the monetary costs from the moment of breach through to when the breach is fully stopped. The cost is much greater, its damage can be far reaching. Trust and reputation can be damaged severely by a breach (infosecisland.com, 2011).

B. Obligation

A company is morally, ethically and in the European Union legally obliged to ensure that any data held by that company is stored in a secure manner and is not kept, and in following with the European Data Protection Directive (EU Directives, 1995) any and all data should be kept only under the following principles:

- Notice – All data should only be taken given notice to the owners knowledge
- Purpose – That the data should be only be used for a given purpose, and any other purpose must obtain explicit permission from the owner to alter or add to that purpose
- Access – Data should be accessible to the data owner to make corrections
- Disclosure - Data owners should be knowledgeable of who has access to this data
- Consent – Data owner must provide consent for any data to be used
- Security – Collated data should be secure from any unauthorised purpose of that data
- Accountability – The holders of the data can be held accountable for all of these principles

These principles written out by OECD (Oecd.org, 1980) are the guiding light for the European directive governing the data protection and also the principles which all other European Data protections acts are created on. These were incorporated into what is now the UK data protection act 1998 (Gov.uk, 1998) However these principles form the ethics of the directive and therefore should be, for the sake of a company to stand from a strong ethical stand point should be followed. Within the UK and every EU state data protection is enforceable up to the value of £500,000 (itgovernance.com, 2013) and therefore is not light penalty for any SME. The penalty issues by the UK information commission (ICO) is one of the many fines that would be imposed on any company found guilty of breaching the data protection act of 1998 in the UK (Gov.uk, 1998).

VI. DATA GOVERNANCE

A. Data governance?

Data governance can be defined as the “the formal orchestration of people, process, and technology to enable an organization to leverage data as an enterprise asset and mitigate risk” (datagovernance.com, 2010). Data management within data governance is how data is managed can be the difference between a breach and the data being secured, if the said data is encrypted then a breach can be more easily brushed away as an unfortunate incident, with a far lower public affairs impact and therefore a lower overall cost to the company, and significantly reducing the regulatory fine received for a data breach. It is also important to note that it is not only governments which have the power to levy fines but also non-governmental regulatory bodies such as best practice bodies too. (Zurich, 2012)

B. How can data management mitigate the issue of data breaches

Data management can be used to mitigate the risk of data breaches by ensuring that as few people have direct authorisation to access the said data, in minimising the number of people and ultimately machines that can access the data in question it will reduce its footprint in the network and ultimately the risk of a data breach (oracle.com, 2011).

VII. RECOMMENDATIONS

A. Remove Old Data

Ensure that all data that has been used or is no longer necessary is forensically, this will prevent any ethical breaches in holding onto data with no purpose and will reduce the number of records that could possibly be breached. Which is a part of the UK Data protection act (gov.uk, 1998), however it is regularly not followed.

B. Ensure that there is a strict internet policy that employees sign and adhere too

This is a simple aspect of training on responsibilities to the data owners and what is expected by the company. This will allow employees to be more comfortable in how they handle the data and will put the company in better stead for preventing data breaches and if one arises to react to a data breach.

C. Ensure that there is the lowest possible surface area for the data

Only allow authorised access where it be application based or human based access, ensure that the data is kept on a tight leash and that the only people and applications that access the data are the ones which the data owner has given permission for that purpose.

D. Regularly review the battlefield

Regularly review the possible attack vectors of the company, find the weaknesses first, whether they be human, or system based weaknesses. If this can be adhered to it will reduce the number of breaches dramatically as you will stay on
top of any weaknesses or flaws in your security system. There is no one solution to all of the security issues.

E. Never underestimate attackers

This is to never be negligent on the security of the company, never assume total safety no matter when that may be. Always assume there is a hole in the system, always look for it.

F. Monitor external breaches, ensure they aren’t repeated

Watch other companies, see when they are breached, what the attack vector was, how would you deal with that, how could you combat that attack, are you also vulnerable to that type of attack, how can you prevent it.

G. Have a plan

Always have a plan to enact if a breach is found, to mitigate the breach and to prevent further data loss. This must be for many different possible eventualities. If there’s a plan in place then if a breach occurs there’s no time wasted strategizing on how to initially react, therefore reducing the impact. There must also always be money set aside in case of a breach. Preferable more than enough money; set aside, it is better to be over prepared than under prepared.

VIII. CONCLUSION

This paper has explored the issues with big data security, in the contemporary small-medium enterprise environment. It has also explored commonly overlooked factors and costs ensued within the event of a data breach. The risks which come when handling big data have been addressed with recommendations on how to improve a business’s approach to the event of a data breach. With focus on how to plan against a data breach.

IX. REFERENCES


Abstract—SMEs (Small-Medium Enterprises) have become a bigger target for cyber-crimes, security breaches have increased rapidly over the recent years and their information is at greater risks. The purpose of this article is to identify security threats within SMEs and inform them of policies and security features that are available to improve an organisation’s security.

Index Terms—SME, Security, Threats, Prevention

I. INTRODUCTION

An organisation’s information has constantly been at risk, a workplace is never fully secure and will always have vulnerabilities and there are several different types of attacks that will exploit them. Despite being a SME (Small-Medium Enterprise) it is important to keep your data safe and secure as security breaches have increased over recent years especially in SMEs which are now experiencing a drastic increase in attacks (PWC, 2013) (See Fig. 1). With such a high rise in attacks it is important that changes are made to improve the security within an organisation.

Security guidance has been issued in 2012 by the UK government to show organisation’s what needs to be done in order to protect themselves from security threats. However, many small businesses still believe that a simple firewall and anti-virus software is enough to provide effective security for their network and therefore have not followed the guidance that the government has provided whereas 30% of large organisation’s have (PWC, 2013). SMEs are lacking in adequate security, it could possibly be that they think being a small business makes them less likely to be noticed and not worth the time and effort for cyber criminals but as a result attacks have increased (MacInnes, 2013). It is clear that security is overlooked in SMEs which is most likely the reason why cyber criminals have been targeting them as accessing an unsecure network would be much easier than secure one (Ashford, 2014). With unauthorised intrusions and malicious attacks on the rise it is vital that some kind of security is implemented within a workplace no matter how large or small the organisation to help protect against the threats of cyber criminals and breaches (Ashford, 2014).

This article will discuss how important security is to SMEs and the risks they take by not investing in security features and also discuss a few types of popular attacks that cyber criminals use.

II. IMPORTANCE OF SECURITY

An organisation relies on its information, system and hardware that support it and therefore having a compromised system can result in devastating consequences (GFI, n.d.). Security of information must be arranged in order to guarantee that an organisation’s intellectual property is protected against the many threats especially as an increasing number of businesses store their sensitive information on their IT systems (Anon., n.d.). Several malicious incidents that SMEs experience are because of a lack of internal and external security and this could potentially bring an organisation to its knees.

Security is a necessity as being a victim to the many possible malicious threats can have many consequences, compromised systems can result in a loss of business as some attacks have the potential to shut down the operations of a workplace especially if the business depends on the IT infrastructure (Dimopoulos, et al., n.d.). The long term consequences can result in customers leaving as a company that is constantly breached puts their private information at risk. However, even though the consequences are high, SMEs are still not equipped to deal with cybercrime as well as large organisation’s as you can see in (Fig. 2) and a study polled 2000 SMEs where 58% of respondents have said that management do not think that cyber-attacks pose a significant risk (PWC, 2013). Another problem that organisation’s face when it comes to security is that it can be very costly especially to a SME as they have restricted budgets, 42% said their budget is not adequate for achieving effective security against cyber-criminal (Ashford, 2014).

- 63% of small businesses were attacked by an unauthorised outsider in the last year (up from 41% a year ago)
- 23% of small businesses were hit by denial-of-service attacks in the last year (up from 15% a year ago)
- 15% of small businesses detected that outsiders had successfully penetrated their network in the last year (up from 7% a year ago)
- 9% of small businesses know that outsiders have stolen their intellectual property or confidential data in the last year (up from 4% a year ago)
III. ADDRESSING SECURITY THREATS

SMEs information are vital assets and therefore it is important that they are protected. According to Trend Micro, 3.5 malicious threats are created every second (Trend Micro, 2012). However, SMEs are still failing to protect their information from the many risks that are out there. SMEs are at risk of losing data, employee productivity, profit and reputation with the increasing number of cyber-attacks (Trend Micro, 2012). The damages can be permanent especially in a competitive business environment as having a breached system can result in loss of business. So what kind of threats do SMEs face and what can be done?

A. Malware

Malware (abbreviation of malicious software) are very common threats that many organisation’s will be hit with (GFI, n.d.). They are malicious script or code with the intention to exploit system vulnerabilities in order to infect, disrupt and damage a computer system (Bullguard, n.d.). Examples of different types of malware include Trojans, worms, viruses, adware and spyware. According to a survey carried out by PWC, 78% of SMEs have had malicious security incidents in the last year (PWC, 2013).

Malicious software can become very frustrating and potentially cause a great a great amount of damage to an SME and can be challenging to deal with without the appropriate security software (Distler, 2007). A well designed anti-virus protection software is one of the main technological defence that every organisation should have as it is capable of scanning, detecting and removing malicious files that may have accessed the network. Anti-virus software can also warn against suspicious websites that may contain malicious links to websites used for phishing (Kaspersky, n.d.). Anti-virus software plays a big role when it comes to protecting a system but a single software/hardware is not enough and should not be the only line of defence used for malicious attacks. They will not be able to detect every single malicious threat especially if it has not been updated to recognise new malicious signatures. It is also possible for malware to deactivate an anti-virus, therefore it is best to have more than just an anti-virus software (GFI, n.d.).

B. Distributed Denial-of-Service Attack

SMEs are presenting themselves as targets for distributed denial of service attacks as the number of DoS attacks have increased over recent years and are becoming more of a threat towards SMEs. Surveys have shown that DoS attacks have increased to 23% which is a rise 15% from 2012 (PWC, 2013). DDoS attacks are where a hacker would compromise multiple computers by infecting them with a Trojan, once in control of multiple systems the attacker would bombard the targets network infrastructure with continuous requests in an attempt to exhaust the resources available to a network so it is no longer accessible by the genuine user (Arbor Networks, n.d.). A DoS attack is capable of causing catastrophic damage to an organisation, according to a survey, 37% of companies that were a victim of DoS attacks reported that the attack had lasted over 24 hours and resulted in severe financial loss (Daniel, 2013). Despite the threat, there are many ways to defend and protect against a DDoS attack. As DDoS attacks are endless requests, the one way to protect against it is to identify the source and block it. Having an intrusion detection system (IDS) to identify the source by distinguishing legitimate requests from the false request can then allow you to block the source with a firewall. However, this is easier said than done (Carabott, 2012). As the number of DDoS attacks are growing it is vital that SMEs implement protection against it as this threat is one of the most damaging cyber-attacks that a hacker can orchestrate (Kwan, 2014).

C. Employee Negligence

The vulnerabilities in an organisation is not necessarily just from computers and software being exploited. Employees can
be a company’s greatest asset but also their greatest weakness as they can also put a workplace at risk as they tend to be very careless (Trend Micro, 2012). Many malicious attacks require the victim to execute it for the attack to spread and affect a network. Staff tend to have a habit of opening and downloading attachments on emails from unknown sources that could possibly contain malicious software and access malicious links they have been sent, doing this causes the malicious file to be executed (GFI, n.d.). An employee’s actions can greatly affect computer security and employees do not realise how dangerous their actions can be. According to research conducted by Ponemon Institute, employee negligence is the main cause of data breaches (Symantec, 2012) (see Fig. 3).

Simply telling staff to not open or access unknown attachments and links on websites is not enough, educating staff so they develop good security habits should be a main priority for SMEs when it comes to security for their system (GFI, n.d.). Educating staff about the severity of potential threats and techniques to detect and avoid them as well as understanding the nature and risk of malware can benefit an organisation as it is a simple and effective approach, it is also more cost efficient for an organisation which is helpful as SME’s do not have a sufficient budget for security software and hardware (Anon., 2006)

These are just a few examples of the rising threats that SMEs face. There are several more attacks that are just as harmful to an organisation and require different methods to protect against.

IV. POLICIES AND PROCEDURES

While having technological security is very important for addressing security threats and protecting big data it is also important to have policies and procedures to manage the human factor as this will also aid in the protection of an organisation’s IT system infrastructure. Policies and procedures are the foundations for guaranteeing a secure IT workplace. An IT network may not be very secure without policies even if there are security tools implemented as staff will most likely still be careless (GFI, n.d.). With policies and procedures in a workplace employees will have rules to abide by so they do not do anything that may compromises the systems. It is important that staff comply with the rules as it will benefit in the security and safety of an organisation (Chi, 2011).

V. LEGAL RISKS

Malware and cyber criminals pose a great threat and can cause internal and external issues for SMEs but there are also legal risks that come with security incidents and breaches. The Data Protection Act 1998 (DPA) is a law in the United Kingdom that defines the way personal information of individuals are to be handled in order to protect the persons information from being misused or abused (Anon., 2014). The law applies to any person or business that processes personal data, they have to abide by the law and secure their customers data or they may be taken to court or fined (Wright, 2011).

VI. BIG DATA

Big data is the term used to define the exponential growth and availability of data that is both structured and unstructured (SAS, n.d.). Big data holds valuable patterns and information that were previously concealed because of the amount of work needed to extract the information (Dumbill, 2012). Large organisation’s such as Google have been able to acquire the valuable information from their data for some time, but at a very high cost. SMEs however do not really have the funds for this but technological advancements over the years have allowed big data to become feasible as cost effective methods have developed to allow the volume, velocity and variability of large amounts of data to be tamed (Dumbill, 2012). This has allowed SMEs and less well-resourced organisation’s to be able to process big data.

There are five defining properties of big data.

1) Volume

Volume refers to the amount of data that is generated (Marr, 2014). The volume of data has increased dramatically over time, there are many factors that add to the increase in data volume such as e-mails, photos, videos, and sensor data. Businesses often have to deal with a huge amount of data (H, 2013). Big data technology allows the large volume of data to be stored (Marr, 2014).

2) Velocity

The velocity refers to the speed of data processing (H, 2013), with more data comes increased speed, number and frequency of transactions (Beulke, 2011).

3) Variety

Variety refers to the many different types of formats that data comes in. The data can be structured and fit into relational databases neatly such as financial data or unstructured which cannot be put into databases so easily (SAS, n.d.). Unstructured data includes photos and videos. 80% of data worldwide is now unstructured. Big data technology allows both structured and unstructured to be brought together (H, 2013).

4) Variability

With a large volume, velocity and variety of data there will be bad data. The quality of data will not always be 100% accurate (Beulke, 2011). Data does not always need to be flawless but it is important that it is as close to perfect as possible to gain relevant insight from the analysis (H, 2013).

5) Value

Value is considered the most important feature of big data. While being able to access big data is good, unless the data can be turned into something of value it can be useless. The cost of implementing new systems to handle and store big data can be a lot and organisation’s require a return on investment (H, 2013).

VII. BIG DATA AND CYBER CRIMINALS

If businesses are using the latest technology then so are cyber-criminals. Cyber-criminals are also using the principles of big data to improve and develop a range of tools to help
SMEs are underestimating the growing complexity of technology and are failing to see the scale of the potential risks. They are vulnerable to several different types of attacks that can affect their system and they are becoming more advanced and sophisticated as technology evolves. Fortunately however, there are also several different IT security software, hardware and policies and procedures that are available to help protect and possibly reduce the damages that malicious threats cause. It is essential that SMEs realise that they are no longer going unnoticed, changes need to be made to their security and policies and procedures for the organisation’s security to become more effective especially now that criminals are focusing their attacks on smaller organisation’s. As new threats are constantly being created and weaknesses being found in software it will be impossible to completely prevent against every threat. However, effectively managing threats and vulnerabilities will reduce the chances of any incidents involving malware and cyber criminals breaching an organisation’s system and will therefore make a safer IT environment and benefit an organisation in the long run.

VIII. CONCLUSION

SMEs are underestimating the growing complexity of technology and are failing to see the scale of the potential risks. They are vulnerable to several different types of attacks that can affect their system and they are becoming more advanced and sophisticated as technology evolves. Fortunately however, there are also several different IT security software, hardware and policies and procedures that are available to help protect and possibly reduce the damages that malicious threats cause. It is essential that SMEs realise that they are no longer going unnoticed, changes need to be made to their security and employees need to comply with policies and procedures for the organisation’s security to become more effective especially now that criminals are focusing their attacks on smaller organisation’s. As new threats are constantly being created and weaknesses being found in software it will be impossible to completely prevent against every threat. However, effectively managing threats and vulnerabilities will reduce the chances of any incidents involving malware and cyber criminals breaching an organisation’s system and will therefore make a safer IT environment and benefit an organisation in the long run.

IX. REFERENCES


European Data Protection Reform
Implications on SME’s Big Data

Nathan Whittington
University of Derby
Derbyshire, UK
100163804@unimail.derby.ac.uk

Abstract— This paper goes through the changes that are contained in the new European data protection reforms. Three of the main areas of change have been reviewed in relation to organisational structure of oversight body, customer data and data breaches. There will be many changes that are required to meet the reforms for example the ability to identify all information relating to a customer and then have the ability to delete it.

Index Terms—European data protection law, Big data

I. INTRODUCTION

In the near future the way that businesses collect, store and use customer personal data will change dramatically with the introduction of a new European data protection law. This law will affect any organisation using data on European citizens, inside or outside of the EU. The new data protection law is currently in the process of being passed after it was recently agreed by a large majority vote in the European Parliament, it has now being passed onto the European Council. Within the law there is currently expectations made be small and medium size businesses (SMEs) that have under 250 staff that don’t have data processing as their main business focus. These exceptions include not requiring a data protection office or an impact assessment along with different sanctions. This paper will focus on the effects of the law on SME’s and so will take into account these exceptions.

During the time while this is being passed there will be changes made to the details of the law, this paper will focus on three areas of changes that will be introduced with the law. The first is a change in oversight and legislation, this will change the structure of the law, from 28 laws to “one single law and one single authority will apply to business based in the EU”. (European Commission, 2011) The next area to be covered is the change in the way that organisations handle their customer’s personal data. There will be an implementation of a right to be forgotten along with informing the customer about what data is collected and why before asking for their explicit consent for these processes. This is meant to help restore confidence in how their personal data is handled by businesses. (Curtis, 2014). The final area covered is data breach notification for businesses, which will require a notification to be sent to their data protection body and affected customers as soon as possible.

II. BIG DATA

The change in data protection law will have an impact on the way that big data is collected and processed by businesses. Big data is a term used for the collection of data in the modern age, being collected in larger volume, higher variety, quicker velocity and increased veracity. The volume of data collection is so large that it is estimated that 2.3 trillion Gb of data is created each day by 2020. (IBM, N.D.) This data is from multiple different sources for example from the customer account registration, their purchases or their social networking. There are many benefits in using big data for businesses for analytical purpose, with it estimated that citizen’s data was worth €315 billion in 2011 (European Commission, 2014b). While there will still be benefits with the introduction of the new law it will make the collection and storage of data that is classified as personal more complex. Personal data can be defined as “Data are personal data if they relate to an identified or at least identifiable person, the data subject.” (The Council of Europe & European Court of Human Rights, 2013). For example in the case of an e-commerce website it would affect the personal details like name and home address but would not affect purchase history unless the data had a link directly to the customer. This will allow for the company to retain details of stock sold without knowing who brought it if the customer asked to be deleted.

III. CHANGES TO THE LAW AND OVERSIGHT BODY

A. Changes to structure of laws

With the implementation of the new law there will be a change to a single set of rules and oversight body to reduce the fragmentation of 28 different sets of rules and authorities. The “fragmentation of rules between EU countries is a costly administrative burden that makes it harder for many companies, particularly small and medium-sized businesses (SMEs), to access new markets.” (European Commission, 2011) An example that was given of the existing rule is a chain of shops in multiple different EU countries that collected data in each with a home office in France would fall under French data
B. National Data Protection Authorities

With the new law and the change to a singular law a change in oversight body was required. Currently there is one per country that carry out their own investigations based upon their countries laws. This is intended to change so that there is an over sight body per country called a National Data Protection Authority (DPA). The DPA is responsible for ensuring that this new law is followed within its region. In addition to this change there is another oversight body that is there to monitor the DPAs and to ensure that all the rulings are fair and even across Europe, this is called the European Data Protection Board. “Under the draft Regulation DPAs would be responsible for regulating companies that have their “main establishment” in the country in which they conduct their regulatory activities. ‘Main establishment’ refers to the premises in which companies take their main decisions about personal data processing. If companies take those decisions outside of the EU a main establishment will be taken as any “place where the main processing activities in the context of the activities of an establishment of a controller in the Union take place”, according to the draft.” (Out-Law, 2013) This will mean that companies are only required to interact with their main DPA rather than all the data protection organisations in which they fall under as present. The only exception to this rule currently is that a customer or data subject has the right to lodge a complaint with a supervisory authority in any state they consider that the processing of personal data relating to them does not comply with the regulation. (European Commission, 2014)

C. European Data Protection Board

To ensure that responses are done fairly and evenly across the whole of Europe the European Data Protection Board has the right to express their opinion and challenge the response of the DPA. The way this is done is that the DPA communicates their proposed measures they intend to take following a regulatory investigation of a data protection incident to the board. The board will then review the measure and the evidence supporting this from the DPA and express their option to the DPA. While the board doesn’t over rule the DPA it is the board’s responsibility to ensure that the response is fair and correct. In the case of if the incident is across multiple countries or not in the territory of the companies home DPA the board is responsible to ensure that the DPAs ruling is correct and takes into account any other effected DPA. The main DPA is the only oversight body that will investigate the incident and make the judgement.

D. The sanctions of the law

The sanctions from not reacting to the new data protection act and breaching its requirements will become singular and much larger. Rather than each country affected by a breach of their data protection act having to investigate and then punish a company, a single authority will to do this. While this amount has yet to be finalised it was previously expected to be up to €1 million or up to 2% of company’s global annual turnover whichever is higher. (Clarke, 2012) This has recently been increased to 5% of global turn over or €100 million when it was passed through the European Parliament. (Stevens & Bolton LLP, 2014) To contrast an example that happened to google in 2012 when they changed their privacy policy, Spain and France fined them for €150 000 and €900 00 respectively. With the implementation of the new law this fine could be as high as €731 million. (Reding, 2014) While such a high fine is unlikely for a SME as they will be more lenient on them especially for a first time or accidental breach. The change in structure can lead to large fines with less serious offenses currently expected to start at €250,000 or 0.5% of global turnover and go up to €500,000. This results in a requirement for businesses, especially SME’s to look at the new law and implement the changes that are require to ensure compliance with it. (Clarke, 2012)

IV. CUSTOMER DATA

A. Customer consent

The changes made to customer’s personal data revolve around the idea customers are responsible to customers for their personal data. The first change to the current way data is collected on people is that they firstly have to be informed on any and all reasons behind the data collection and that consent must be explicit rather than the customer staying silent and being assumed to be saying yes. Although explicitly consent does not mean that it has to be given in writing and that it can be done by clicking on an icon or ticking a box that is clearly defined for this purpose. (European Commission, 2012) While this is seen as an improvement by some people, by others it is seen as a complication. Some have said that the draft regulation proposed raises the bar for consent rather than simplifies it. With the requirement of explicit consent in all cases. This may lead to businesses having less confidence as to if they are compliant to this regulation. (Travers Smith, 2013) This change will have to be implemented into any data collection method to ensure that this consent was given unambiguously and that it was explicitly done or implied by the way that the customer acted. (The Council of Europe & European Court of Human Rights, 2013) In addition to this change the customer must be informed on any intended use of the collected data before consent can be given. The data can only be used for the purpose the customer has been informed of and given permission for. The collected data must then be removed from the system once the defined process has been completed.
B. Methods to meet changes in customer consent

To help adapt to the new changes to customer data collection a review should be carried out of the current data that is collected and why. This will provide a foundation of a new policy to see what data is required and what can be removed from existing databases along with collection tools. This will help reduce excess data as well as informing the customer of what data and why it is collected. This could be done by adding the basics onto the website or data collection tool and more detailed information in a privacy policy will help to gain their confidence and trust in providing the company with their data. (Henderson, 2014) This will allow the company to meet the new rules by informing the customer the reasons why the data is being collected and giving them the choice to allow for it to be collected. In addition to this option, an option could be given to customers to let them provide different levels of information and the benefits that this could provide. (Henderson, 2014) For example if it was an e-commerce website, the account creation could provide an option that allows for the retention of order details for example home address, phone number which would allow for faster ordering of goods and updates of offers.

C. Right to forget

The new law states companies will have to give their customers the ability to find out all the data the company stores on them as well as have the data removed through the ‘right to be forgotten’. (Ross, 2012) The new law in its current form provides the ability for customers to request all the personal data a company has on them that is stored both electronically and on paper. It is proposed that this request for data is done for free and completed within three months of the request. (European Parliament, N.A.) SME’s will be able to charge a fee if the requests are made frequently or require a lot of work to gather the data if it is wide ranging or a large quantity. (European Commission, 2014a) The right to be forgotten by companies will require that reasonable steps shall be taken to delete the data including technical measures in relation to data. This includes informing third parties processing the data that a request to erase any links to or replication of the personal data. (European Commission, 2014). This right to forget can only be used if there is no legitimate reason to keep the data. (European Commission, 2014a) The right to be forgotten will require companies to be able to provide this function for customers across all stored records that are controlled by the company. This will mean both active databases as well as any other stored locations like back-ups or any secondary databases used for business continuity. There is not currently a time frame as to how long companies will have to conduct this deletion but it is expected that it will have to be carried out swiftly.

D. Methods to meet right to forget

The two changes will require companies to have the ability to locate all of the personal data on a customer to inform them of it as well as for deletion if required. Considering the requirements both technologically and the policies it would be worth companies looking into it now. “Consider the practicality of implementing the 'right to be forgotten' now, especially in customer and marketing databases, as retro-fitting could be more expensive than early incorporation.” (Ross, 2012) The changes made to databases will likely be the biggest area affected, requiring a link across all storage to identify all the data on a particular person. This could be done with the use of a unique number across all the databases to tie all the records together through a single item of data rather than requiring multiple different items for example a customer ID. This could further be strengthen with encryption to the personal data and anonymys any data that does not require a personal link to them. This would allow for personal data to be secure and deleted swiftly while retaining none personal data for analysis. Any changes that businesses make to fulfil these requirements will cost money and time depending on the systems and policies currently in place, with retro-fitting systems likely to be even more costly.

V. DATA BREACHES

A. Data Breaches

The next big change is to data breaches, there will now be a requirement to inform the data protection authority along with effected customers of any breaches without undue delay. This will require a change in policy in most organisations with the requirement of notification so quickly compared to the current rules of not needing to notify. The idea behind implementing the rule is that “stolen records are most valuable to criminals immediately after they have been stolen”. (European Commission, 2012) Once people have been informed they can take measures to protect themselves from the criminals. Statistics show that countries which require quick notifications have fewer data breaches, strong rules in data protection encouraging companies to manage personal data more securely. (European Commission, 2012) This suggests that there is a big benefit to companies with quick notification and an increase of consumer trust with the knowledge that they will be informed of any breaches affecting them.

B. Methods to meet data breach requirements

For companies to comply with these rules they may need to completely rework their security policy with the way that they handle and investigate breaches. They will likely need to be able to identify the breaches and the details of it more rapidly than previously. To be able to inform the oversight body and the customers with as much detail as possible to obey the new law and limit reputational impact. Other methods that could be implemented to improve the performance of the company in case of breaches would include the introduction of encryption to hide the identity of the customer while allowing for data controllers to ensure they are dealing with the same person. (The Council of Europe & European Court of Human Rights, 2013) With the use of encryption a company would be able to inform effected customers that their data had been breached while ensuring the data is not readily usable. This will provide time for the company and customers to react to minimize the effect of the breach and minimise any future impact.
VI. CONCLUSION

A. Summary of the law

To summarise the three key areas of change highlighted from this report. There will be some dramatic changes made to the way that data protection is viewed and governed within the EU. The law has yet to be finalised with changes happening to aspects of the law while it has been passed through the EU parliament it has to go through the Council. There already has been some changes like the amount for fines more are expected before it is finalised. However the three areas of the law covered in this paper have remained generally the same with only some details changing. With the law having a focus on informing and giving customer’s control of their personal data, data breaches and the change in the structure of the data protection organisation and fines that can be used.

B. Summary of implementations

The change in organisational structure of the data protection bodies will allow for businesses to understand the rules better and know who they should contact. As long as the business covers the rules set out within the law and know who to contact this aspect should require the least amount of changes to react to the suggested rules. To react to the customer data planned changes it has been suggested that there is a system and polices in place to review the existing and planned systems. This will allow a business wide plan to be created to allow for implementation of the new features of customer information consolidation and deletion. This review will also help with the identification of required data collection and the purposes behind them to inform the customer. Other ways to meet this requirement mentioned was to anonymise any excess data to allow for permanent storage for data analysis without it affecting the initial customer. The changes required to meet the data breaches rule will force companies to review their policies in response to data breaches with a focus on trying to gather the details as quickly as possible to allow for an initial response to the appropriate parties. This may not change the way that some businesses work if they have already implemented such policies but in most it will require an overhaul of their responses.

VII. REFERENCES


Abstract—The concept of data analytics has come on leaps and bounds in recent years with the utilisation of Big Data technologies by a number of organisations. Big Data has opened the door to companies being able to recognise trends in their sizeable sets of information and, as a result, cater their services more efficiently. But is the hype around these technologies justified? Big Data analytics is still a rather new subject and so there are aspects of it that remain unclear. One major issue with Big Data is that of trust, stemming from a great deal of uncertainty within this new area. This paper aims to bring to attention the issues involved in Big Data, including problems of trust, the issues that are produced as a result of a lack of trust and what these considerations mean for a type of organisation that could benefit greatly from the idea, provided a sufficient understanding of the subject is gained. This type of organisation is an SME in the financial sector and with uncertain economic times, is it right to be moving to such an ambiguous technology?

Index Terms—Big Data, Economic Climate, Financial Sector, SMEs, Trust, Veracity.

I. INTRODUCTION

Big Data has been a topic of intense interest among businesses worldwide for many years now. The term Big Data refers to a collection of data that is so large in volume that the use of traditional database management software is rendered inadequate. Every day, technology users worldwide create approximately 2.5 quintillion bytes of data (Hogan, 2013) and organisations are now looking to tap into the opportunities of analysing massive data sets. But the usual database applications are simply no longer able to cope with the sheer size of these data requirements. Instead, more advanced analytical tools are utilised in order to generate links between the plethora of information. The flagging of correlations in the data by such tools can allow businesses to adapt the choices that they make in relation to the way that they operate.

But whilst Big Data has been given considerable attention for some time, it is only recently that the subject has become an increasing part of the typical modern business’s life. A surge in the amount of interest in Big Data naturally comes with an increase in the number of companies using it in their day-to-day activities. A recent forecast by IDC stated that the Big Data market was “expected to grow to $32.4 billion by 2017” (IDC, 2013), which is six times the predicted growth rate for the overall Information Technology market. But are users of this new technology aware of the risks of Big Data? Whilst there are numerous statistics thrown around the web that paint the topic in a good light, it is important for potential users to understand that Big Data might not always be the right solution for their issues in data analytics. The aim of this report is to highlight the risks involved in turning to Big Data and to show the reader that stepping back and refraining from jumping on the Big Data bandwagon is not necessarily a negative action, considering the technology’s relative infancy. In addition, this report looks to relate such risks to a specific type of organisation – Small and Medium Enterprises (SMEs) in the financial services sector.

SMEs tend to be on tighter budgets and have limited resources available to them in relation to the corporate giants, so an argument can be made that these companies need to be extra sure about the advantages and disadvantages of Big Data technologies before utilising them. Likewise, it is organisations in the financial sector that face similar considerations. Credit card and financial data are considered to be the two most private types of data for consumers (BCG, 2013) and it is therefore of the utmost importance that any shift in the way data are handled is met with a thorough understanding of the effects of the shift. As a result of all of this uncertainty, financial SMEs are the specific subject for this paper. One of the major issues within Big Data currently is that of a by-product of uncertainty – a lack of trust. How much are organisations, in particular financial SMEs, able to trust Big Data solutions to aid in their data analytics? Further complications arise when you consider the uncertainty about the global economic climate. Whilst most experts consider the Global Financial Crisis to largely be over, the effects are still being felt by key economic areas, such as the US and Europe. This is significant for financial firms as they are intrinsically linked to the economy. So with a number of issues still up in
the air about Big Data, should financial SMEs be making the move to the new technology at this moment in time?

II. BIG DATA AND THE ISSUE OF TRUST

A. The Vs of Big Data

Big Data is a complex issue that is still a new concept in the eyes of the average business. As a result, it helps in the process of understanding the idea to split Big Data into three or four major components. The components of Big Data are often referred to as the Vs (V words) of Big Data and it is widely considered that there are three standard Vs that are used to initially describe what Big Data is about:

- **Volume** – The obvious attribute of Big Data is that it is large in size. It is estimated that 40 zettabytes (43 trillion gigabytes) of total data will have been created by the year 2020, up from the official 2005 figure, which was 300 times smaller (IBM, 2013). Massive data sets render traditional database software to be almost useless, as such applications are unable to cope with the volume of information. Companies are therefore looking to Big Data tools, which are designed to deal with large data sets, to help enhance their analytical capabilities. But it is also important to understand that there are a number of sources of information from which numerous pieces of data can be gathered. It is the diversity of sources that links to the next V.

- **Variety** – Data can be presented in a variety of different formats, under various names and from a range of sources. On social networking sites, it is estimated that 30 billion pieces of information are shared on Facebook monthly, whilst 200 million active users on Twitter send 400 million tweets each day (IBM, 2013). And, as mentioned with volume, the types of technology that can produce vast amounts of data are rapidly growing. From smartphones to desktop PCs to wearable fitness gear, our lives are now dominated by data-producing tech. As a result, it is for this reason that variety is a major consideration of Big Data. Companies need to ensure that the scope of the variety of data that they measure is finely balanced to be within their needs but also not able to exceed resources. This can also be said for the previous V, volume, and is of particular significance to SMEs, whose resources will normally be more restricted than the larger corporations. But another point to consider with regards to appropriate scaling is the speed at which the data is being received and processed, i.e. data velocity.

- **Velocity** – Speed is the final key consideration in Big Data. In an ever-advancing technological world, the data of yesterday is increasingly being considered to be outdated and the movement of data is as close to real time as it has ever been. Organisations, especially SMEs, must be aware of the ‘need for speed’ in Big Data and act accordingly in making sure that their technologies are capable of sustaining the pace of the flow of data and the frequent updates required for such information. For example, the New York Stock Exchange, which captures a terabyte of information per trading session (IBM, 2013), relies on substantial sets of data being processed to precise figures and updated every few seconds. Because of this, the NYSE’s analytical tools need to be able to withstand high data velocity to effectively deal with the data received.

The three Vs explain the different aspects of the complicated subject that is Big Data, but there are other V words that are also often associated with the term. One such V is veracity, which in Big Data terms can be defined as the degree of uncertainty surrounding information. Veracity allows for a range of questions to be asked when put into this context. How accurate is the data? How effective are the analytical tools? Many calculations are based on approximations and can bear a degree of ambiguity, so whilst data analysis tools are powerful, they are not perfect.

B. The Trust Issue with Big Data for SMEs

Considering that Big Data is still a rather new concept, it is only natural that initial concerns related to veracity become apparent. Arguably the most notable concern for businesses is a product of low data veracity – distrust. If firms are unable to trust the data that they process, how can they get the information to effectively guide them in the everyday decisions that they make? As Bob Hayes writes in his article In Data We Trust, “the veracity of your data is about the accuracy and truthfulness of the data” (Hayes, 2013). Data can be considered trustworthy when errors and uncertainty are low. But it seems uncertainty about Big Data technologies is at a high level since statistics show a clear lack of faith in the idea. IBM’s John Easton states that “by 2015, 80% of all available data will be uncertain” (Easton, 2012). Another IBM source reports that one in three business leaders do not trust the information they use to make decisions, (IBM, 2013) and that represents a noticeable trust issue within Big Data. For SMEs, this should be an alarm bell. If one in three business leaders do not have faith in the data they make their decisions on, the move to utilising Big Data for a small or medium enterprise requires additional thought. Is it worth implementing a new, distrusted concept into a firm with relatively limited resources?

So trust is dependent on the factor of uncertainty, but it is worth noting that veracity links greatly to the three primary Vs associated with Big Data (Subramaniam, 2014). With regards to volume, if there is more data to analyse, then logic dictates that veracity will be reduced. Additionally, the more varied the information, the more likely it is that there will be less uncertainty. Furthermore, frequent updates of data (i.e. high velocity) will mean veracity will be lower than irregular revisions. On balance, SMEs are likely to have, in relation to larger organisations, a smaller pool of data available to them with less variety and updates to information over longer intervals. With all of these points comes the concern of trust in Big Data, and for SMEs it is of great importance that this is understood.
III. TRUST ISSUES FOR FINANCIAL SMEs AND THE UNCERTAIN ECONOMIC CLIMATE

A. Big Data Trust and Financial SMEs

The clear issue of trust in Big Data technologies paves the way for further complications with organisations. This is because trust is a highly influential factor within a business of any size and one of paramount importance for the success of a company (Solomon and Flores, 2001). A particular type of organisation that depends on trust to a great extent is a financial SME, which is generally perceived to have relatively limited resources, yet deals with data of crucial importance. In the BCG Global Consumer Sentiment study, it states that the privacy of personal data remains a top issue for 75% of consumers (BCG, 2013). As a result, it stands to reason that before any major decision is made with regards to the way that financial data are handled, all risks must be understood and the move must be deemed trustworthy. Big Data would certainly fall into the category of ‘major decision’. However, this is the problem with the concept - trust is not at a high level with Big Data, due to the uncertainty associated with the term.

An example of an issue of trust with Big Data is that of employing individuals that are capable of using and understanding the new technology’s tools. Unfortunately, this is an issue for many firms – Big Data has a skills crisis (Ranger, 2013). A study by Harvey Nash found that one in four CIOs experienced difficulty in locating people with the skills to utilise Big Data technologies (Harvey Nash, 2013). This is a major issue for all companies and especially SMEs, who tend to have limited resources in terms of their employment process in comparison with the corporate giants. With individuals that possess the necessary skills for Big Data analytics being so scarce, there is therefore a lack of trust in Big Data itself.

Another factor affecting the level of trust in Big Data is the high expectations that come with a concept that has widely been given rave reviews but has not been explained well to prospective subscribers to the idea. A survey by Forrsight highlighted a “hesitancy to dive into “big data” (Forrsight, 2013) for IT leaders, stemming from a plethora of questions about Big Data, as a result of a lack of information. For example, the Forrsight report on the survey highlights a scenario where representatives from a business reported that they were concerned about the company’s ability to internally operate the well-known Big Data analytical tool Hadoop (Hadoop, 2014). The company went on to state that the firm was turning to a partner organisation for support on the matter. This demonstrates a lack of confidence in and understanding of Big Data technologies for that firm and, as mentioned by Forrsight, “confidence comes from trust” (Forrsight, 2013). This highlights a serious problem with the current state of Big Data. If firms that currently deploy the technologies do not have confidence in them, the concept won’t come across as trustworthy. This again is an issue for financial SMEs and readdresses the point that investing in an idea that is surrounded by such uncertainty, lack of trust and depletion in confidence is not the right idea at this moment in time. And it turns out that a lack of understanding of Big Data is not specific to this example – a survey for IT professionals by SAS showed that 22% of respondents had (at best) a poor understanding of the technologies involved (SAS, 2013). In addition, the global economic climate accentuates this issue.

B. Utilising Big Data in Uncertain Economic Times

Economic uncertainty is a term that has been mentioned often since the start of the Global Financial Crisis in 2007-08. A Bank of England publication on the worldwide crisis shows that uncertainty is inversely related to GDP (Haddow et al., 2013). Consequently, as GDP has generally decreased in major economies, uncertainty about the future has risen in response. A high level of economic uncertainty is not a desirable condition for an SME or company of any size to operate under. A reduction in productivity is the most obvious effect for firms; in scenarios where there is an element of risk involved, the organisation more often than not selects the ‘safe’ option. For financial firms, this tends to be to not: proceed with a potential investment, enter a new market (Haddow et al., 2013) or post vacancies (Lazear and Spletzer, 2011). This is because potential actions, such as the ones mentioned, are normally based on forecasts, which become less easy to quantify as economic uncertainty grows. Economic uncertainty, which is still at high levels in 2014, is therefore a huge disadvantage to the prospect of financial SMEs considering the move to Big Data. It is a move with high risk and the statistics show that in such times firms prefer to ‘play it safe’ with their decisions.

IV. THE NEED FOR RESTRAINT – UNCERTAINTY SQUARED

It is clear that whilst Big Data has its advantages, the issues already discussed about this new concept are still providing question marks about the idea of an organisation implementing it, let alone a type of firm that must make certain that there are to be no complications for the adoption of an idea such as Big Data. SMEs in the financial services sector are a prime example of such an organisation. A high level of uncertainty surrounding Big Data coupled with consistently high rates of uncertainty within the economy make for a potent mix of distrust and lack of confidence – i.e. uncertainty multiplied by uncertainty, equaling uncertainty squared. So financial SMEs must ask themselves, are these the right conditions to be making a move to a technology that is still in the early stages of its evolution? Even the professional services organisation PwC recognises the infancy of the idea, stating that financial firms looking to make the move must know that “Big Data technologies are still evolving and require careful and ongoing needs assessments” (PwC, 2013). So, for financial SMEs, the best option right now is to take a step back and spend time understanding Big Data, its flaws as well as its benefits. And don’t be put off, there are a number of advantages for financial firms by utilising Big Data. On a general level, plus-points include improvements in operational efficiency, better customer service and more relevant marketing campaigns through the analysis of customer and employee data. For financial firms, Big Data can help to shorten processing times for transactions as well as aid in the tracking and flagging of suspicious behaviour that could be a sign of financial crime, such as money laundering or fraud (PwC, 2013). This point is particularly important as the 2014 Global Economic Crime
Survey showed that of all the major industries, the financial services sector was most at risk from economic crime, with 49% of respondents stating that they had been a victim of crime themselves (PwC, 2014). So it is clear that Big Data can help to improve a financial organisation, however, this paper is stressing the need for a pragmatic, thorough approach to the research into the idea before taking up this uncertain but potentially rewarding concept.

V. CONCLUSION

Big Data is a term that is firmly the zeitgeist of the modern world. Whilst it promises advantages in terms of steering financial SMEs towards becoming better businesses, it is apparent that some are naïve as to thinking that the concept will solve all of their issues. This is not the case and the lack of trust and high uncertainty with Big Data is evident. Although this is likely to become a frequently utilised technology in future years, the concept is not right for financial organisations that deal with such important data and SMEs that possess limited resources right now unless an extraordinarily thorough understanding of the topic has been grasped. An IBM study showed that 55% of respondents, who were from the financial industries, identified customer-centric objectives as the top priority for their firm (Turner et al., 2013). This gives a clear message that the customer comes first for these businesses and this is their personal data that is most valuable. With this being the case, why should financial firms put their customers’ data at risk? By implementing Big Data they would be doing so.

The same study also showed that almost three in four financial services firms have either developed a strategy for Big Data or implemented a pilot study on the concept (Turner et al., 2013). But have the firms that have begun the move to the technology considered all of the issues? In such uncertain times, they cannot afford not to.

VI. REFERENCES


<table>
<thead>
<tr>
<th>Author Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algahtani, Mohammed</td>
<td>6</td>
</tr>
<tr>
<td>Alkanan, Bader</td>
<td>11</td>
</tr>
<tr>
<td>Allen, Gareth James</td>
<td>15</td>
</tr>
<tr>
<td>Almarri, Abdulhadi</td>
<td>20</td>
</tr>
<tr>
<td>Bensbiet, Abdulmajeed</td>
<td>24</td>
</tr>
<tr>
<td>Cain, Michael</td>
<td>29</td>
</tr>
<tr>
<td>Castle, Daniel</td>
<td>34</td>
</tr>
<tr>
<td>Fairbrother, Rachel</td>
<td>38</td>
</tr>
<tr>
<td>Fowler, Connor</td>
<td>43</td>
</tr>
<tr>
<td>Fowler, Samuel</td>
<td>48</td>
</tr>
<tr>
<td>Fox, Jamie</td>
<td>52</td>
</tr>
<tr>
<td>Gamgoum, Youssef</td>
<td>57</td>
</tr>
<tr>
<td>Holding, Thomas</td>
<td>61</td>
</tr>
<tr>
<td>Johnson, Ben</td>
<td>66</td>
</tr>
<tr>
<td>Johnston-Aiken, Hannah</td>
<td>70</td>
</tr>
<tr>
<td>Kamoto, Yemurayi</td>
<td>74</td>
</tr>
<tr>
<td>Lovatt, James Peter</td>
<td>79</td>
</tr>
<tr>
<td>Malkan, Anish</td>
<td>84</td>
</tr>
<tr>
<td>Martin, Adam</td>
<td>89</td>
</tr>
<tr>
<td>McMahon, Joel</td>
<td>94</td>
</tr>
<tr>
<td>Moreton, Nathan</td>
<td>98</td>
</tr>
<tr>
<td>Morland, Sean</td>
<td>102</td>
</tr>
<tr>
<td>Nadeem, Mohammed Jawad</td>
<td>107</td>
</tr>
<tr>
<td>Nagi, Anwar</td>
<td>112</td>
</tr>
<tr>
<td>Nagi, Nazir</td>
<td>117</td>
</tr>
</tbody>
</table>
O’Sullivan, Patrick .................................................................122
Pimlott, Ben .................................................................128
Pocock, Andrew .............................................................132
Rankin, Mitchell .............................................................137
Raza, Ghulam Hussain ....................................................142
Rees, Elliot .................................................................146
Searson, Michael ...........................................................151
Sidhu, Pumandip ............................................................156
Silde, Alice .................................................................161
Sumner, Ben ...............................................................166
Tran, Lewis .................................................................171
Whittington, Nathan .....................................................176
Williams, Matthew .........................................................181