A Review Paper on the Application of Big Data by Banking Institutions and Related Ethical Issues and Responses

Victor Chang[1][a], Lina Xiao[2], Qianwen Xu[2] and Mitra Arami[3]
[1]*School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, U.K.*
[2]*IBSS, Xi'an Jiaotong-Liverpool University, Suzhou, China*
[3]*PARDIS Ltd, London, U.K.*

Keywords: Big Data, IoT, Banks, Ethical Issues of Big Data, Suggested Solutions.

Abstract: Nowadays, Big Data and the Internet of Things (IoT) are one of the most popular topics. This review paper demonstrates an overview of the application of Big Data and IoT in banking institutions. In the beginning, a brief definition of Big data and IoT is provided and the integration of technologies by banks is illustrated. Then, this paper explains the potential sources where banks could generate Big Data. Next, the major works that banking institutions use Big Data are listed. In the final two parts, some acute ethical concerns are raised and appropriate solutions are suggested for the banking industry and other organizations.

1 INTRODUCTION

Due to the financial crisis and its impacts between 2007 and 2009, the public has realized the importance of corporate to disclose financial and non-financial data regularly and adequately. Central banks and regulators should take responsibility to make objective evaluations and implement strict monitoring. This practice indicates for many parties consisting of firms, central banks, regulators, and so on, they have to deal with massive and complicated data in a limited time. The era of Big Data enables a sharp increase in different forms of data such as client data, trade figures, health data, management data, etc. From the picture below, Big Data has been widely used in all walks of life, such as marketing, HR, healthcare, supply chain, agriculture, finance, and so on (Invested Development, 2015; Marr, 2015; Stackowiak et al., 2015). Therefore, it is essential for businesses to know how to develop Big Data to improve their business operations, processes, communications and opportunities.

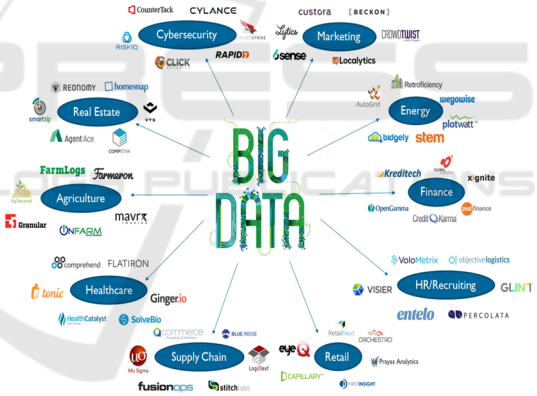

Figure 1: The era of Big Data employed.

It is a universal truth that Big Data is deemed as a competitive advantage and organizations could benefit a lot from these advanced technologies. For example, for the meteorological department and agricultural sector, Big Data can be utilized for disaster prediction. For firms, they may make use of Big Data to understand consumer behavior and increase their sales revenue. For banking institutions, they harness the ability of Big Data to avoid risk and better serve their customers. It is worth noting that Big Data itself is neutral with no harm. However, the negative impacts of Big Data may come from the

[a] https://orcid.org/0000-0002-8012-5852

Chang, V., Xiao, L., Xu, Q. and Arami, M.
A Review Paper on the Application of Big Data by Banking Institutions and Related Ethical Issues and Responses.
DOI: 10.5220/0009427701150121
In *Proceedings of the 2nd International Conference on Finance, Economics, Management and IT Business (FEMIB 2020)*, pages 115-121
ISBN: 978-989-758-422-0

intentions, methods and people who use it. Overall, this paper discusses the role of Big Data and IoT in banking industries and highlights some ethical issues later. Several responses from different perspectives are also put forward within the context to address those concerns.

2 BIG DATA AND IOT

In recent years, firms, governments and other sectors have been deployed Big Data partly due to the inherent limitation of traditional analytical tools. That is, traditional data processing methods are inadequate to process pretty large and complex data sets. Grable and Lyons (2018) stated that "when analyzed computationally, Big Data can provide more precise insights into hidden patterns, trends, and associations, especially in the context of human decision making" (p.17). In the early 2000s, Doug Laney (2001) defined Big Data, including three concepts: volume, variety, velocity. Based on his original work, other concepts have been complemented: veracity and value (Grable and Lyons, 2018). Bholat (2015) claimed that even in the banks, data are generated of high volume, high velocity and diverse in form.

Generally, IoT is considered as the 'third industrial revolution'. The term was explained as "a conceptual framework, which involves embedding connectivity and intelligence across a wide range of devices over the cloud" (Dutta and Ghosh, 2018, p. 1). Saxena and Ali Said Mansour Al-Tamimi (2017) argued that it enables better monitoring and interaction. The following picture vividly depicts a relationship between Big Data and IoT in terms of the business model maturity index. Briefly, IoT produces large volumes and complex, diversified data and Big Data assistants IoT. Additionally, Big Data cannot do without cloud computing. Therefore, IoT, Big Data and cloud computing are the basis of each other.

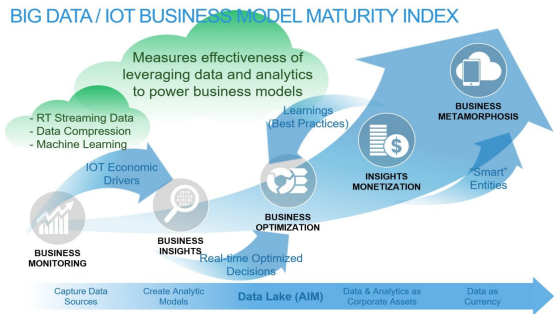

Figure 2: Big Data/IoT Business Model Maturity Index.

According to Saxena and Al-Tamimi (2017), the banking industry may combine Big Data and IoT technologies altogether to establish a more robust framework. IoT technologies have promoted the innovation of financial service, improve interaction with clients, help optimize the structure of organizations and even assist in designing better business models. Therefore, IoT could be tapped by banks to improve their customer relationship management. In case of theft or misplace of customers' plastic cards, IoT technologies come in handy to find lost objects. Big Data can be generated by IoT technologies in their own turn and data sharing could be facilitated by IoT. Overall, banks' efficiency can be significantly improved and customer trust can be built through the application of Big Data analytics and IoT technologies.

3 POTENTIAL SOURCES OF BIG DATA FOR BANKS

Figure 3: Three potential sources of Big Data for banks.

Big Data can be generated by banks in many ways. As shown in the picture, offline and online channels and social banking activities are the main sources of Big Data for banks. Generally, offline banking mode is the most common method to manage customer relationships, such as customers visiting the bank physically. Except for the offline channel, banks could also transact with customers via online modes such as internet banking, telephone banking, ATM, WAP-banking and other means (Cheng et al., 2006). Besides, customer relationships may be forged through banking activities on social networking media like Wechat or Weibo, as Ghazinoory et al. (2016) said that "social banking" allows customers to be more involved and their behavioral features can be identified through insights from social data. For instance, a Wechat Subscription called "ShanRong", the e-commerce platform of CCB, releases the news of the latest and popular promotion for users. Data analysts investigate user behavior characteristics

based on massive data and finally, three outcomes can be achieved. Firstly, operators could make timely optimization decisions by evaluating the attractiveness of functions and marketing activities introduced by ShanRong subscription to users. Secondly, it can implement real-time monitoring, active forewarning for online fraud, and timely detection of inhuman ticket brushing and intrusion. Thirdly, formulate personalized online marketing activities and push individual notifications through the precise clustering of customers. Thus, these social banking activities are again a potent source of Big Data, in the sense that banks are appraised of the needs and requirements as well as feedback and grievances regarding various issues.

4 MAJOR AREAS BANKING INSTITUTIONS UTILIZE BIG DATA

Banks utilize Big Data mainly to intensify their risk management frameworks to become more transparent and auditable (Srivastava and Gopalkrishnan, 2015). According to The Economist (2012), high-performance hardware and software are used by financial services to investigate complicated patterns of fraud within unstructured data. Nunan and Domenico (2013) argued that the application of Big Data "has enabled the cost-effective provision of financial services in areas that would previously have been regarded as too risky to be sustainable". For instance, before a statement is issued, credit card issuers usually apply Big Data analytics to detect a cardholder's fraudulent behavior. Moreover, Srivastava and Gopalkrishnan (2015) demonstrated that data analysis is useful to identify and assess finance crime management solution rules by detecting the correlation between financial crimes and transactions' attributes in advance. The followings are specific ways.

MIS/ Regulatory reporting	Disclosure reporting
Real time keyboard conversation tracking	Anti-money laundering

Figure 4: Ways to manage risk by utilizing Big Data.

In addition, by using Big Data, banks are able to improve their marketing strategies. The data helps them to profile and categorize their customers and then to learn what the customer's need and predict their behavior. In this way, it becomes easier for banks to identify their potential customers or provide better service for their existing customers (Hassani, Huang & Silva, 2018).

5 SOME ETHICAL ISSUES ARISING FROM BIG DATA

Due to the immense commercial and social value that Big Data may bring, today, people consider Big Data as a competitive advantage for financial institutions and other sectors. However, a range of ethical concerns may arise as a matter of time when using and dealing with Big Data technologies. More specifically, as Bratu (2018) said that Big Data analytics covers a series of processes, including data acquisition, storage, distribution, evaluation, and information implementation. Hence, the concerns relate not only to the gathering, retention of data and its security, but also with its analysis and interpretation by data scientists, and with the commercial trade of personal data. This paper is going to investigate, depict, and assess several ethical concerns, and then relevant suggestions may be proposed later.

5.1 Privacy and Consent

Bratu (2018) argued that breaching and invading privacy and information suppliers' approval are the most commonplace ethical topics at a challenging level. The right to privacy attracts much attention and debate from the public regarding the usage of Big Data. Privacy is quite vital to be protected as a human right because it is tremendously beneficial to individuals and society as a whole (Nersessian, 2018). When one's data is to be harvested and stored, notification should be explicitly conveyed and the informed consent should be obtained. Fuller (2017) found that in practice, most service providers offer information, generally in a written text form, to users to obtain their consent. Then, a form needs signing, or a box needs to be ticked by users to confirm that they understand and accept all the terms represented to them.

In practice, however, Wilbanks (2014) pointed out that this is a process intentionally limiting the liabilities of the parties who obtain users' information rather than sincerely notify the data subjects. Consequently, the ability of users whose data are gathered tends to be imperceptibly minimized. To conclude, there are three main practical difficulties in managing privacy and acquire informed consent (Solove, 2013). Firstly, for those who are not aware

of the importance of privacy, they tend to ignore relevant policies. Secondly, some people acknowledge the significance and read the terms but do not understand them. Thirdly, although some people understand the privacy policies literally, they may make unwise choices due to insufficient background knowledge.

5.2 Security Problem

Since the banking industry has gradually achieved the network, banks have always been the coveted objects by hackers as they are capital-intensive areas. Landwehr (2014) emphasized that customers' data should be kept complete and be protected from accidental or intentional threats. Accordingly, he proposed strategies to handle such threats through prevention, detection, reply and recovery. There is no doubt that it is data handlers' responsibility to prevent the data from exposure, misuse, or cyber-attack. However, establishing a robust system is equally vital because not every disaster may be foreseeable. These systems can assist in detecting attacks and responding to them timely and appropriately. Besides, other suitable new safeguards should be in place if necessary.

Nevertheless, Fuller (2017) mentioned that whatever means are taken, hacked computers and leakage of confidential data have frequently been happening, which indicates that breaches of security will always be a tricky problem. Recently, HSBC was exposed that hackers attacked its customer accounts between 4 October and 14 October in 2018 (Davis, 2018). Consequently, around 1% of American customer personal information was leaked out. Besides, in May 2016, managers and employees of Agricultural and Commercial Bank and China CITIC Bank illegally sold customers' personal credit reports, involving 2.57 million personal information items.

5.3 Commercial Usage of Customer Data

Big Data may possess commercial usefulness in various ways. For example, Fuller (2017) claimed that a website or a store card could intelligently record information about a user's previous purchase. With the footprint left by consumers, firms may carry out precision marketing, target promoting and persuade consumers to purchase further. Likewise, banks also make commercial use of Big Data to recommend appropriate wealth management products to specific customer groups. On the one hand, many people acknowledge that the commercial usage of data helps them to access the products they might want and like in a more convenient way. However, on the other hand, as people's data generate a huge financial gain, there is a tendency for firms to sell personal data for profit, even illegally. It is necessary and urgent to have some effective regulations around us in case personal data is used for evil and illegal purposes. Nevertheless, Kitchin (2014) pointed out that currently, few particular laws and codes have been formulated to regulate data brokers and their power has not been limited to a reasonable level yet. Thus, regard to this field, effective and timely actions should be taken, such as imposing relevant ethical or legal restrictions upon data brokers before this problem is more acute and severe.

5.4 Unfairness

Big Data has been widely used by financial institutions such as banks, trust companies in assessing people's suitability for loans. As O'Neil (2016) investigated that this may aggravate the existing social inequalities. For instance, for an individual who has a poor credit record and lives in a poor neighborhood, it is probable for institutions to judge that person as an inappropriate candidate for a loan application or charge him or her higher premiums of insurance. This practice may deprive the opportunities of the poor and reinforces their existing poverty. Therefore, it is unreasonable and inequitable for banks to use Big Data to categorize and sort people as this may further deepen social unfairness to a new level.

It is worth investigating the Big Data's role in accessing financial services in China. Compared with developed countries, China's financial services penetration rate is lower (Kshetri, 2016). Kshetri (2016) also highlighted that, especially for low-income families and small and midsize enterprises (SMEs), the low penetration problem is more acute and severe than for high-income households and large firms. Klein and Cukier (2009) claimed that SMEs in China contribute to 70% of GDP, but only 20% of financial resources are available to them. What is more, around 89% of Chinese SMEs face barriers in meeting banks' requirements when applying for loans (Jing, 2014). Two main reasons are explainable for this situation. Firstly, borrowing loans to poor people and micro-enterprises will bring traditional banks higher transaction costs and inconvenience of processes, so banks are usually reluctant to serve these borrowers. Secondly, information opacity is another reason to interpret why the weak and small

businesses have difficulty in reaching financial services and other favorable policies.

Kshetri (2014) claimed that information opacity is partly since Chinese credit rating agencies could not provide adequate and complete information on SMEs' creditworthiness. This practice may perhaps be because certified audited financial statements are not compulsory for those micro firms. However, it has become crucial for banks to minimize their lending risks based on reliable and transparent credit information. Kunt and Maksimovic (2004) found that as foreign banks grasp less credit information of Chinese people and SMEs than banks in China, they are more cautious even reject those "opaque" borrowers' loan service applications. Big Data may help increase chances for Chinese low-income families and micro-enterprises to obtain financial services via evaluation and analysis of potential borrowers' creditworthiness and reduction of transaction costs. Nevertheless, the unfairness cannot be eliminated in a short time.

5.5 Data Cleaning, Analysis and Presentation of Results

Except for the ethical concerns discussed above, another issue also captures people's attention that is the data clean, analysis and presentation of the outcomes by data analysts. As Fuller (2017) argued that data expert s' one role is to be taleteller. That is, before presenting persuasive final results, data scientists need to carry on a series of data processing, including data cleaning. In people's minds, Big Data is objective and just "is", but the data processing may embed various biases and human factors' interference. First, each data does not exist without ground. It could be obtained in different ways such as data acquisition, data sharing even cyber-attack illegally. Biases may be introduced in these ways at the same time. Second, data cleaning is usually implemented before detailed analysis conducted. Data experts have discourse power to determine the approach to handle the missing data, such as imputing missing variables, transforming variables and removing outliers. However, data cleaning and analysis processes are rarely documented. The experts' choices on the approaches may influence the analysis, interpretation and presentation in the next steps. The influence is long-term and it is not limited to the one-step but the whole progress (Borgman, 2015). Third, the specific tools adopted in the data analysis process may incorporate preconceptions. The algorithms employed to carry out the analysis of the data are given specific values and are embodied in specific scientific methods (Kitchin, 2016). Fourth, aside from the tools and algorithms used to analyze the data, biases may also result from data scientists themselves. In order to take responsibility for the interpretation result and try to minimize the biases, the analysts have to identify the influence of their analyzing experience and opinions in their interpretation (Boyd and Crawford, 2012). Last but not least, apart from the biases resulting from the analysis and interpretation, the presentation of final results from analysis deserves attention as well. In the presentation, analysts commonly use tables, graphs, charts, diagrams, etc. to realize data visualization. However, some conscious or unconscious biases may result from these visualizations, which might encourage readers to see and understand results in particular ways. According to Gitelman and Jackson (2013), using different visualization methods in the presentation may produce different effects. Additionally, in that the decision-makers will make decisions on the presentation of the analysis results, the presenters need to present them properly.

6 PROPOSED SOLUTIONS

The ethical issues discussed above are somewhat complicated and overlapping. Each of them should be treated seriously by taking appropriate actions. Broadly, solutions are proposed from three perspectives: technical, legal and ethical.

6.1 Technical Solutions

Regard to some ethical issues raised before, like security problems, technological development and innovation may assist in addressing. That is, continuously update high-performance hardware and software systems that store and process Big Data. For instance, in terms of frequent cyber-attacks of banks, as advised by Fuller (2017), the security holes can partly be handled through continuous oversight and strengthening infrastructure construction and development. Overall, integration of experienced data handlers and a robust framework can mitigate the security problems to a great extent. As data science is an advanced domain with rapid changes, so relevant techniques to improve security will change and develop correspondingly. Similarly, for those who deliberately damage security will transform means and find new coping ways as well. Ultimately, at least, notifications of the risks from technical limitations, failures and security breaches should be clearly conveyed to people who may be influenced.

Besides, acquiring interested parties' consent to those risks is necessary.

6.2 Legal Solutions

Many legal and regulatory methods are formulated to deal with plenty of matters. Nevertheless, there are still some challenges that should not be underestimated. Among them, the most three acute issues are explained below. Firstly, as Big Data, its analysis and its commodification change so rapidly and frequently, it is urgent to develop corresponding legislations in time (Fuller, 2017). If legislations lag behind current practices, not only these legal and regulatory means would be old-fashioned, but also bad people will make unbridled attacks and individual rights and interests cannot be guaranteed promptly. Secondly, although relevant legislation already exists in some circumstances, in reality, they are not feasible or/and maybe routinely overlooked. In such cases, laws must be adapted to make sure they are meaningful, believed and enforceable. Thirdly, making applicable and global legislation is crucial. Actually, each country enacts laws and regulations based on its conditions and differentiation may exist among different nations. However, there is no border restricting data, so it is necessary to coordinate and cooperate between countries to produce some agreed international standards. These significant undertakings may consume much cost, time and effort, but they are undoubtedly crucial and deserve these valuable resources.

6.3 Ethical Solutions

The financial crisis reflected the absence of professional ethics and low moral standards. After that, banking supervisors and politicians have been aware that the corporate governance system and financial institutions' ethical culture play a vital role in the occurrence and development of the crisis. That is, poor performance in corporate governance and unethical behavior of management and employees are partly responsible for the financial crisis. Consequently, the global financial stability and social welfare will be influenced negatively. For instance, Enron's managers colluded with Arthur Andersen to manipulate accounts, which indicated the deficit of professional ethics. The incident caused both Enron and Andersen to suffer from a terrible loss of reputation (Markham, 2015). Thus, for financial institutions, including the banking industry, ethical status is rather significant. In order to improve the overall level of financial practitioners' ethics, first and foremost, board characteristics count. Baselga-Pascual et al. (2018) demonstrated that within an organization, the board of directors (BOD) is considered as the most vital internal governance mechanism because formulating and monitoring the ethical culture of the whole organization is BOD's responsibility. Baselga-Pascual et al. (2018) investigated a positive relationship between ethical reputation and board characteristics. More generally, the larger board size, more diversified gender, and CEO duality may contribute to more effective monitoring and oversight. However, if board members are too busy to have regular meetings, poor monitoring and a low ethical reputation may be formed. Therefore, it is crucial for financial institutions to make a balance in the structure of BOD. Additionally, BOD themselves are required to conduct ethically and establish appropriate and feasible ethics codes for employees. Last but not least, regular evaluation of the company's ethical atmosphere and strict monitoring of employees' behavior ought to be in place.

Besides, data scientists and employees have been advised to undertake an oath of practicing specific ethical codes, showing they fully understand those standards and are willing to take responsibility and accept the oversight from others. However, in order to ensure the effectiveness, two requirements should be satisfied. One thing is that all data scientists must comply with the oath and everyone is equal before the regulations. The other is that severe punishment measures like penalties would be acted on any data analysts who breach of it.

7 CONCLUSION

To sum up, in this review paper, we give a brief introduction to Big Data and IoT. Then we focus on the usage of Big Data and IoT in banking institutions through analyzing the potential sources where banks could acquire Big Data and the major areas that banks adopt data. From these two parts, we find that the banking industry frequently generates Big Data in three ways. They are offline channels, online channel and social banking activities. Generally, the banking industry makes use of Big Data to reinforce risk management frameworks to reduce financial risks. In addition, Big Data can also be utilized to develop more accurate marketing strategies, reduce transaction and to operate costs and provide better service to consumers. Next, we list some ethical issues related to the application of Big Data such as privacy and consent, security problems, commercial

usage, unfairness and data cleaning, presentation process. Finally, several responses are provided from technical, legal and ethical aspects. On the whole, Big Data is a strategic resource if it is used legally and adequately by organizations. We expect Big Data could benefit people and the whole society to the greatest extent.

FUNDING

We are grateful to VC Research to support this research, with grant number VCR 0000026.

REFERENCES

Baselga-Pascual, L., Trujillo-Ponce, A., Vähämaa, E., & Vähämaa, S. (2018). Ethical Reputation of Financial Institutions: Do Board Characteristics Matter?. *Journal of Business Ethics*, 148(3), 489–510. DOI: 10.1007/s10551-015-2949-x.

Bholat, D. (2015). Big data and central banks. *Bank of England Quarterly Bulletin*, 55(1), 86–93.

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT press.

Boyd, D. and Crawford, K. (2012). Critical Questions for Big Data. Information, *Communication and Society* 15: 662–75.

Bratu, S., (2018). The Ethics of Algorithmic Sociality, Big Data Analytics, and Data-Driven Research Patterns.*Review of Contemporary Philosophy*, 17, 100–106. DOI: 10.22381/RCP1720187.

Cheng, T. C. E. , Lam, D. Y. C. , & Yeung, A. C. L. . (2006). Adoption of internet banking: an empirical study in hong kong. *Decision Support Systems,* 42(3), 1558-1572.

Davis, J. (2018) HSBC Bank Discloses Security Incident. News Report, Security today. Available on: https://securitytoday.com/articles/2018/11/08/hsbc-bank-discloses-security-incident.aspx (assessed on 27 March 2020).

Dutta, N. S. and Ghosh, M., (2018). IoT Driving Social Innovation. *Telecom Business Review*, 11(1), 1–5.

Fuller, M. (2017). Big Data, Ethics and Religion: New Questions from a New Science. *Religions*, 8(5), 1–11. DOI: 10.3390/rel8050088.

Gitelman, L., & Jackson, V. (2013). Introduction. In "Raw Data" is an Oxymoron. Edited by Lisa Gitelman. *Cambridge: MIT Press*, 1–14.

Grable, J. E. and Lyons, A. C. (2018). An Introduction to Big Data. *Journal of Financial Service Professionals*, 72(5), 17–20.

Hassani, H., Huang, X., & Silva, E., 2018. Digitalisation and big data mining in banking. *Big Data and Cognitive Computing,* 2(3), 18

Invested Development, 2015. 'Big Data and the Internet of Things', Available at: http://investeddevelopment.com/2015/05/big-data-and-the-internet-of-things-weekly-review-54-58/ (Accessed: 28 November 2018).

Jing, M. (2014). Alibaba, leaders team up for SME financing. China Daily. Available on: http://www.chinadaily.com.cn/business/2014-07/23/content_17901230.htm (assessed on 31 March 2020).

Kitchin, & Rob. (2016). Thinking critically about and researching algorithms. *Information, Communication & Society*, 1-16.

Kshetri, N. (2016). Big data's role in expanding access to financial services in China. *International Journal of Information Management*, 36, pp. 297–308. DOI: 10.1016/j.ijinfomgt.2015.11.014.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, 6(70), 1.

Markham, J. W. (2015). *A financial history of modern US corporate scandals: From Enron to reform*. Routledge.

Marr, B. (2015). *Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance*. John Wiley & Sons.

Nersessian, D. (2018). The law and ethics of big data analytics: A new role for international human rights in the search for global standards. *Business Horizons*, 61(6), 845–854. DOI: 10.1016/j.bushor.2018.07.006.

Nunan, D. and Di Domenico, M. (2013). Market research and the ethics of big data. *International Journal of Market Research,* 55(4), 2–13.

Saxena, S. and Ali Said Mansour Al-Tamimi, T. (2017). Big data and Internet of Things (IoT) technologies in Omani banks: a case study. *Foresight*, 19(4), p. 409.

Solove, Daniel J. (2013). Privacy management and the consent dilemma. *Harvard Law Review* 126: 1880–903.

Srivastava, U. and Gopalkrishnan, S. (2015). Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks. *Procedia Computer Science,* 50, 643–652. DOI: 10.1016/j.procs.2015.04.098.

Stackowiak, R., Licht, A., Mantha, V., & Nagode, L. (2015). *Big Data and the Internet of Things: enterprise information architecture for a new age*. Apress.