Proceedings Paper:

Yu, H. and Foster, J. orcid.org/0000-0002-9439-0884 (2017) Towards information governance of data value chains: Balancing the value and risks of data within a financial services company. In: Uden, L., Lu, W. and Ting, I.-H., (eds.) Knowledge Management in Organizations (KMO 2017). KMO: 12th International Conference on Knowledge Management in Organizations (KMO 2017), 21-24 Aug 2017, Beijing, China. Communications in Computer and Information Science, 731 . Springer Verlag , pp. 336-346. ISBN 978-3-319-62697-0

https://doi.org/10.1007/978-3-319-62698-7_28

Towards Information Governance of Data Value Chains: Balancing the Value and Risks of Data within a Financial Services Company

Haifangming Yu and Jonathan Foster

Abstract

Data is emerging as a key asset of value to organizations. Unlike the traditional concept of a business value chain, or an information value chain, the concept of a data value chain has less currency and is still under-researched. This article reports on the findings of a survey of employees of a financial services company who use a range of data to support their financial analyses, and investment decisions. The purpose of the survey was to test out the idea of the data value chain, as a useful abstract model for organizing the different discrete processes involved in data gathering, data analysis, and decision-making, to further identify issues, and suggest improvements. While data and its analysis is clearly a tool for supporting the delivery of financial services, there are also a number of risks to its value being realized, most prominently data quality, along with some reservations as to the relative advantages of data-driven over intuitive decision-making. The findings also raise further data and information governance concerns. If implemented such programs would aid in the realization of value from data within the organization, while also mitigating the risks of this value not being realized.

Key Words

Data value chain; competitive advantage; financial services; data governnace; information governance

---

Jonathan Foster
j.j.foster@sheffield.ac.uk

Information School, University of Sheffield
Regents Court, 211 Portobello Street
Sheffield S1 4DP, United Kingdom

1. Introduction

In an age of information and big data, the capture, processing analysis and use of data presents companies with both opportunities and challenges. Within this context, it can be argued that the implementation of a systematic data value chain becomes a prerequisite for companies to be able to achieve their business goals. From this perspective, the systematic capture, processing, analysis and use of data can be viewed as a programme and a technology for the gathering and processing of data from both internal organizational and external environments for, among other purposes, internal decision-making, analysis of customer intelligence, the monitoring of organizational performance, and organizational forecasting (Ofner et al. 2013; Vera-Baquero et al., 2015). The data value chain encompasses a sequence of processes including data capture, storage, distribution, analysis and use. By implementing a data value chain, an organization can take data-driven as well as intuitive decisions, and more generally increase the use that is made of the data that it captures. As well as consisting of a set of organizational processes, Kasim (2012) points out how data value chains can also encompass inter-organizational and networked processes. On the one hand, data driven decision-making presents the opportunity of increasing profits by aiding the targeting of returns from certain investments, and by aiding processes of risk analysis. On the other hand, as one process within a set of interdependent processes, data-driven decision-making also carries with it some of the potential risks associated with the other elements of the data value chain. Therefore, a better understanding of the data value chain, its operating standards and governance, can help to increase the value of data, while also mitigating against the potential risks.

The article reports the results of a survey of the employees of a Chinese branch of a financial services company. The aim of the research was to evaluate how the data value chain works within the company, and to identify what factors can influence the outcome of data-driven decision-making. More specifically, how does the data value chain benefit the company? What are employees' opinions of and attitudes towards the data value chain? What factors influence the data processing of data along the data value chain? Do companies currently have a preference for data-driven decision-making over intuitive decision-making? The structure of the paper is as follows. Section 2 provides a brief review of literature on the data value chain. Section 3 outlines the survey methodology, including sampling approach. Section 4 presents the findings. Section 5 discusses the results in light of previous literature. The conclusion provides recommendations for future research into the processes of the data value chain.

2. Data Value Chain

The concept and practice of the data value chain (DVC) is the focus of this research. The concept plays an important role in organizations that rely on the capture and processing of data from their external and internal environments, and its onward processing in a systematic way. Here we provide a number of current definitions of the data value chain and its elements. Miller et al., (2013) suggest that the data value chain can be decomposed into three parts: data discovery, data integration and data exploitation. Kasim (2012) approach is more granulated with the DVC including data collection, data management, data integration, data analysis, data simulation and data visualization. With trends towards the utilization of data for growth and well-being increasing, OECD (2015) has proposed a number of enabling factors including digitization, open data, a fast and open internet and the internet of things (IoT) in the area of data gathering; analytics (algorithms), cloud computing, and specialist data skills in the area of data analysis; and machine learning, automated decision making and simulations and other data experiments in decision-making. Along this data value chain, data are collected, stored and maintained, used, repaired, and finally destroyed.

In general, it is rare for a single model of the entire data value chain to be used in organizations (Ofner et al., 2013). Therefore, management of each individual element of the data value chain is considered a cornerstone of the effective working of an organization's data value chain (Gartner, 2008); with data value chain evaluation, data value chain analysis, and data value chain communication are considered important aspects of this management. Ofner (2013) holds the same opinion. The details on how the concept of a data value chain actually works in an organization is difficult to deduce from the literature. In other words, while Kasim (2012) and

Miller et al. (2013) allude for example to the structure of the data value chain, they do not mention how the capture and gathering of data is converted and realized into value that benefits the organizations. While positing the idea of ‘data as an asset’ Khatri and Brown (2010), and therefore the value of data as an entity and a resource, the authors point more to the accountability aspects of managing data, and the allocation of decision rights, rather than directing attention to the enabling aspects of converting data into value. Foster (2016) points to some of the initial models being used to realize value from data.

3. Methods

A survey of employees of a financial services company was conducted. The survey method was considered appropriate as a deductive method of inquiry for testing out the idea of the data value chain, as a useful abstract model for organizing the different discrete processes involving in data gathering, data analysis, and decision-making. The survey is organized into four parts. The initial part gathers demographic data on participants; the second section contains four questions related to data gathering the first process of the data value chain; the third section contains three questions related to data analysis, the second process of the data value chain; and the fourth section contains seven questions related to the final stage of the data value chain, decision-making. Participants in the survey were all members of an international financial services company based in Shenzhen. In order to avoid biasing a specific interest group, a randomized sample of 100 employees from the financial services company was compiled; with the expectation of 70 responses. The survey was posted to the online website *Wenjuanxing* and, in actuality, 82 responses were received.

4. Findings

Demographics. 82 employees of the Financial Services Company participated in the research. Figure 1 presents the distribution of the employees who responded, by age. As can be seen the overwhelming majority of those who are participating in data-driven decision-making are of a younger age, with 91.46% of the employees who are leveraging the data value chain being aged between 18 and 40;

and with only 8.54% of employees aged between 40 and 60 similarly doing so.

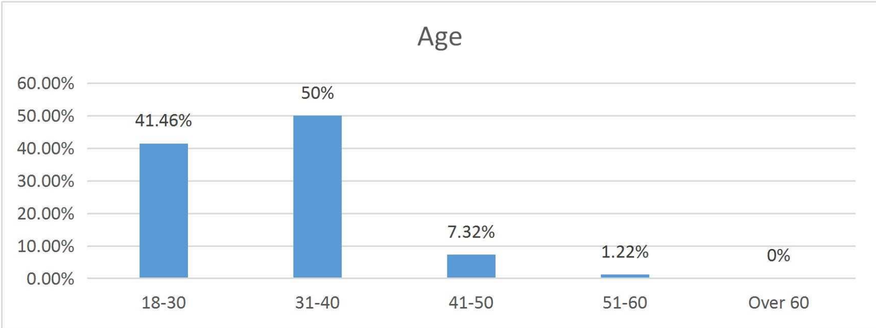

Fig. 1. Distribution of respondents by age

Figure 2 shows the distribution of employees by highest educational qualification held. The responses clearly demonstrate the educated nature of data users within the company with 63.41% of

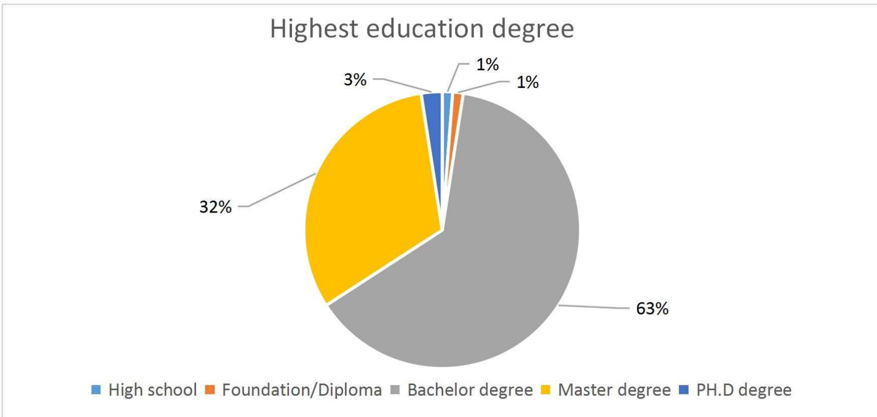

Fig. 2. Distribution of respondents by highest education degree

respondents holding a Bachelors degree, a further 31.71% a Masters degree, and 2.44% a PhD degree. In sum, 97.56% of the data users within the company are in possession of a Bachelors degree or above.

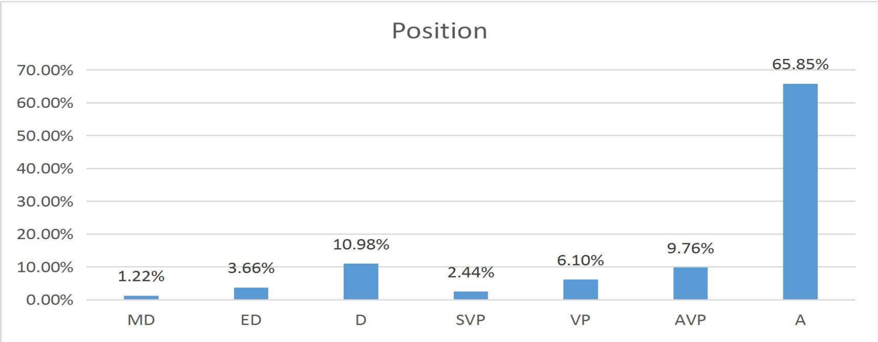

Fig. 3. Distribution of respondents by position

Figure 3 provides a distribution of respondents by job title. The general trend is that employees at a lower rank in the hierarchy will tend to be the greater users of data, with employees of a higher rank tending to be the lesser users of data.

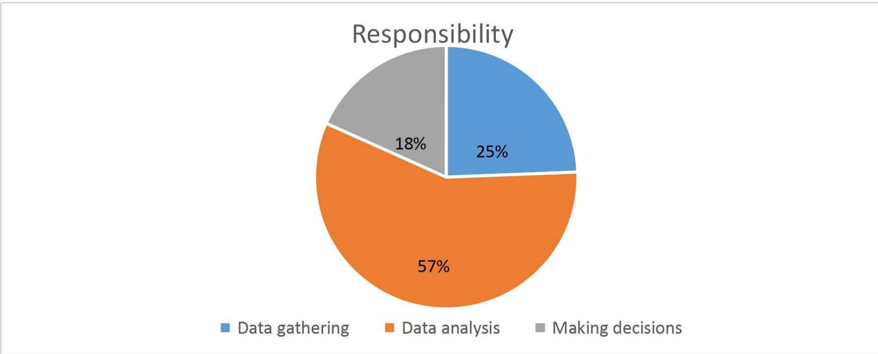

Fig. 4. Distribution of respondents by phase of data value chain

For example, Associate positions account for 65.85% of employees, and Managing Directors positions 1.22%. The exception to this is Director positions that account for 10.98% of data users. Figure 4 presents information on the allocation of responsibility for the different phases of the data value chain. In terms of distribution of responsibility the greatest number, and vast majority, of users are involved in the data analysis phase of the data value chain (57.32%), with 24.39% involved in data gathering, and the least number involved in data-driven decision-making (18.29%). Respondents were in clear agreement that data and its analysis was a useful tool in supporting their work.

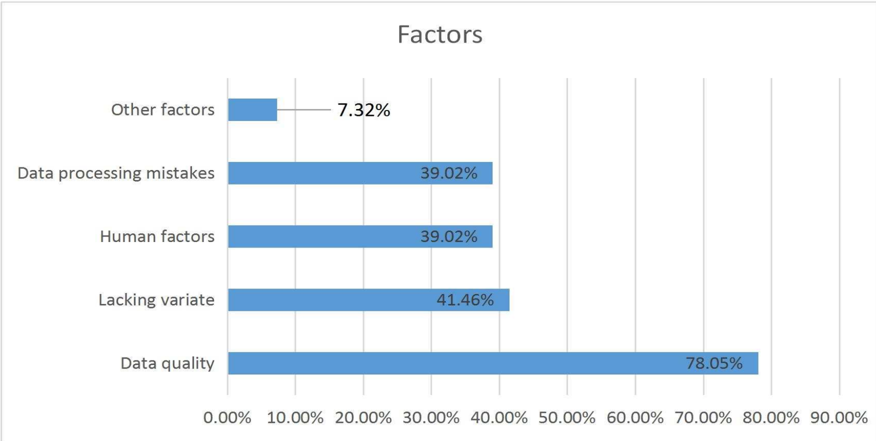

Fig. 5. Distribution of respondents by phase of data value chain

In response to the statement "I think big data analysis helps my work", 89.03% of the respondents were in either agreement or strong agreement with the statement; with only 10.98% of respondents either neutral about or disagreeing whether data analysis is helpful in supporting their work.

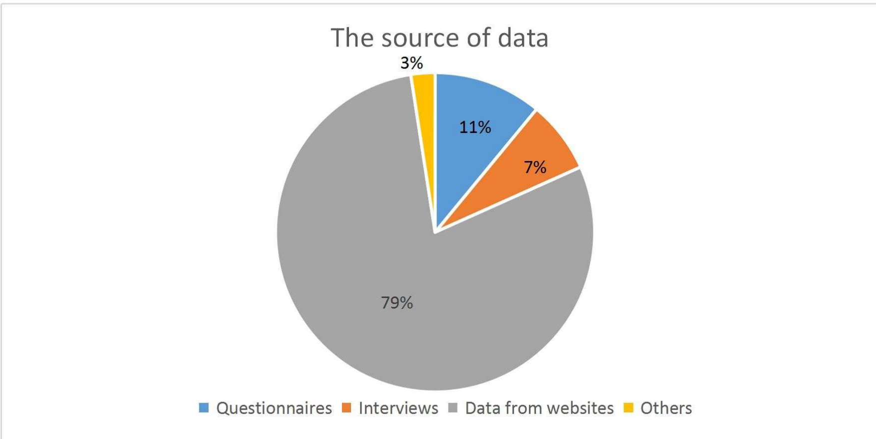

Fig. 6. Data by source type

Nevertheless respondents were also aware that there exist a number of risk factors (multiple choice) that could affect the accuracy of the data used. Figure 5 illustrates these problematic factors. The most frequented cited factor is data quality (78.05%), followed by a clutch of other factors: data lacking variation (41.46%), human factors (39.02%), and data processing errors (30.02%).

Data gathering phase: Figure 6 illusrates the sources from which employees gather their data. These are predominantly from the firm's website (79%), but also from quantitative (11%) and qualitative (7%) surveys, with only 3% coming from other sources e.g. professional data supply websites. The overwhelming majority of employees (91%) will then systematically store this data for future use. *Data analysis phase*: Figure 7 identifies the data analysis tool(s) that employees use to analyse data (multiple choice).

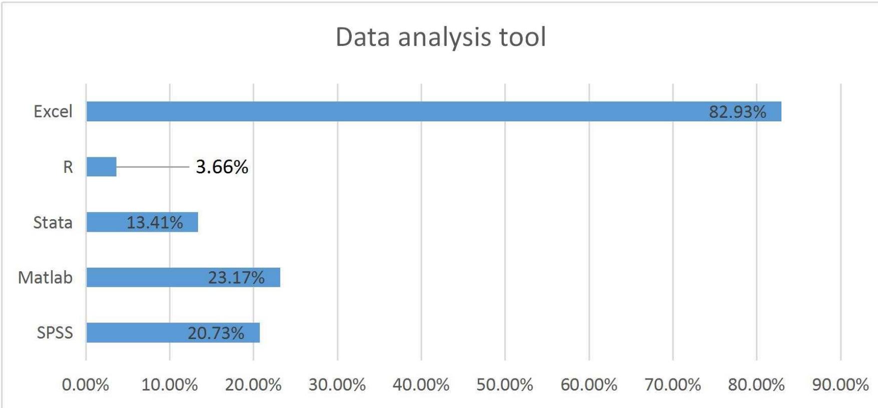

Fig. 7. Data analysis tools by frequency of use

The tools most frequently used for data analysis are: Excel (82.93%), Matlab (23.17%), SPSS (20.73%), Stata (13.41%), and R (3.66%). *Decision making phase*: Respondents were overwhelming favourable towards data-driven decision-making; with 87.8% of employees either favourable or very favourable towards the practice, with only 12.2% of employees exhibiting only an occasionally favourable or unfavourable attitude towards the practice. Figure 8 illustrates the perceived consistency between the results of data analysis and intuitive predictions as to financial performance. Here the figures are more circumspect. While 71.95% of employees perceive a consistency and therefore a useful complementarity between the two types of decision-making, nearly

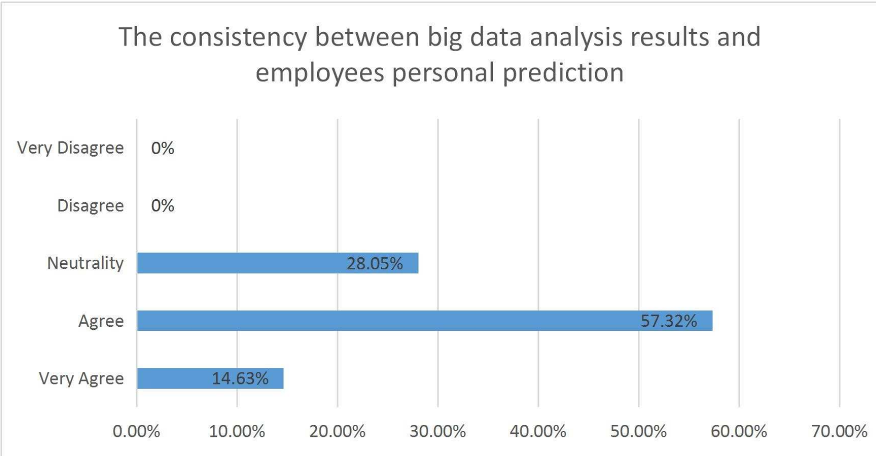

Fig. 8. Perceived consistency between big data analysis results and personal predictions

a third of employees (28.05%) were neutral on this issue. Figure 9 illustrates the number of people who will be typically involved in making decisions on the basis of the data analyzed.

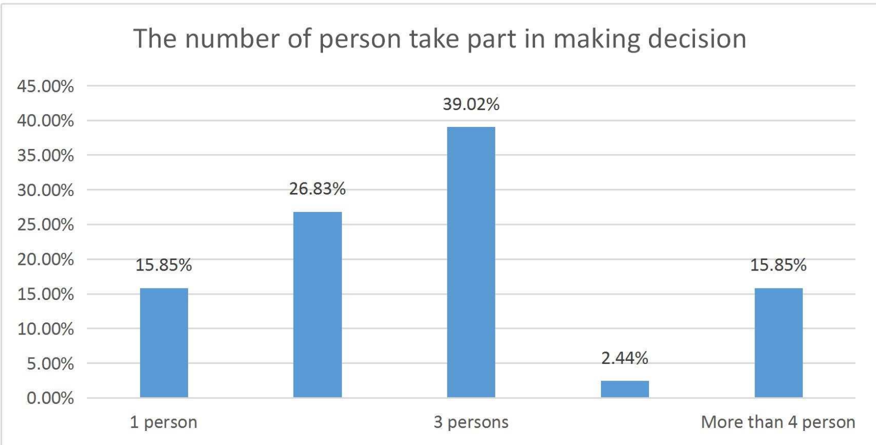

Fig. 8. Decision-making process by number of employees

The typical number of decision-makers being 3 employees (39.02%), although the number of decision-makers can range from 1 decision-maker to (15.85%) to more than 4 decision-makers (15.85%). What can be said is that the collaborative use of data, and available of data for joint interpretation and decision-making appears to be the norm; with 84.15% of decisions involving 2 or more people.

Two further bi-variate analyses of the data were conducted.

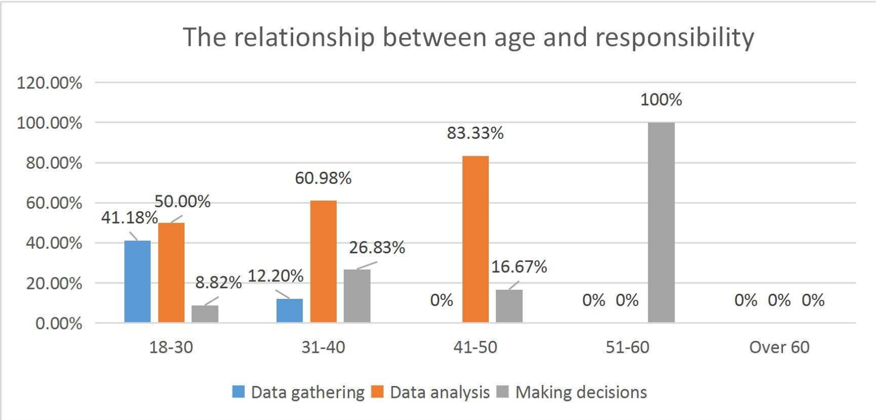

Fig. 9. Relationship between age and responsibility

Older and therefore typically more experienced employees tended to be responsible for the latter phases of the data value chain. Figure 9 illustrates the relationship between age and responsibility for the different phases of the data value chain. There is a tendency for the younger more in experienced employees to be involved in data gathering, for slightly older and middle-ranking employees

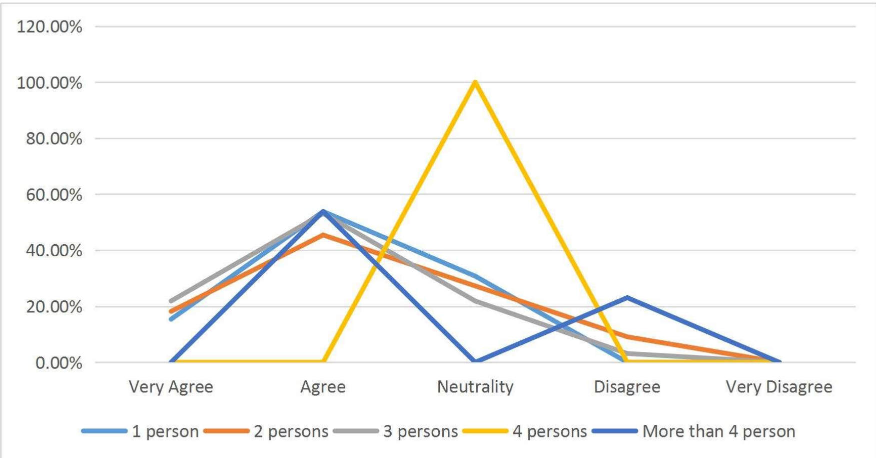

Fig. 10. Decision-making process by number of employees

tended to be involved in data analysis, while senior more experienced employees are 100% responsible for decision-making. After further testing there is some weak evidence that this association is statistically significant (p-value 0.0593). Given the importance attached to data quality as a barrier to accurate decision-making, this relationship is worth further investigation. The relationship between the number of employees participating in data-driven decision-making, and employee's attitudes towards the accuracy of the resulting decision was also tested. Although there is some evidence that in some situations, an increased number of employees can lead to an inaccurate decision (Figure 10), there is very little if any evidence that this is necessarily the case (p-value 0.089). A combination of a competitive financial institution, along with a low-trust culture may provide an explanation, when multiple people are employed.

5. Discussion

In this section we briefly discuss the findings in light of the previous literature. The value of data is largely determined by its accuracy (Liu et al., 2015; Jorge et al. (2015); while timeliness is also considered to be one of the key attributes of data quality (Blake & Mangianmeli, 2011; Pipino et al., 2002; Wand and Wang, 1996). Therefore, it is significant that the key risk factor, cited by 78.05% of respondents was data quality. It is clear that data quality is the key issue that can reduce the likelihood that value will be created from data. From a data resource perspective it will be important to assess the degree of quality attached to each stream of data. There is some weak evidence that younger, more experienced, employees are responsible for data gathering. If substantiated by further evidence, it will also be important to attend to this issue when allocating decision-making rights and accountability for the domain of data quality, and how this is a key criterion on evaluating the worth of data gathering (Khatri and Brown, 2010). Given good quality data, the value of data analysis for organizing and interpreting the data is not in question, with the majority of employees in this company also holding a positive view of its worth in supporting financial services. Nevertheless, the perceived gap for some between intuitive and evidence-based data-driven decision-making is worthy of further investigation.

5. Conclusion and Future Research

Data and its analysis are clearly assets valued by those who responded, as being a tool for supported the delivery of financial services. However, what appears to be the case is that the accumulation of this value is framed in a piecemeal fashion. What is required is much greater cumulative management and oversight of the data value chain; and treating the data value chain as a set of interlinked processes, rather than independent processes. In this regard previous work on the management (Liu et al., 2015; Jorge et. al., 2015;; Ofner et al., 2013; Gartner, 2008) and governance (Khatri and Brown, 2010) of the data value chain would appear to be a guide to future practice. Nevertheless Khatri and Brown (2010) tend to view data as an exercise in potential value and accountability; whereas the emerging field of information governance (e.g. Tallon, Ramirez and Short, 2011; Foster, 2016) tends to examine the factors that can shape the realization of the value of data, and its conversion into information. These include not only procedural practices associated with managing the data value chain, but also structural practices that assign decision-making and accountability rights; and relational practices that draw attention to the need for communication and co-ordination along the data value chain. In doing so the organization will enhance its capacity for enabling value to be realized from data e.g. allocating decision-making rights; while also mitigating the risks to that value being created e.g. data quality.

References

Foster, J. (2016). "Towards an understanding of data work in context: issues of economy, governance and ethics", Library Hi-Tech, 34 (2), 182-196.

Gartner. (2008). *Governance of Master Data Starts With the Master Data Life Cycle*. Gartner Research, Stamford, CT.

Jorge, M., Ismael, C., Bibiano, R. (2015). A Data Quality in Use model for Big Data. Future Generation Computer Systems, 1-8.

Kasim, H., Hung, T., & Li, X. (2012). Data Value Chain as a Service Framework: for Enabling Data Handling, Data Security and Data Analysis in the Cloud. International Conference on Parallel and Distributed Systems, 18, 804-809.

Khatri, V. and Brown, C.V. (2010). "Designing data governance", *Communications of the ACM* 53 (1), 148-152.

Ofner, M., Straub, K., Otto, B., & Oesterle, H. (2013). Management of the master data lifecycle: a framework for analysis. *Journal of Enterprise Information Management*, **26** (4), 472-491.

Pipino, L.L., Lee,Y.W., and Wang, R.Y. (2002). Data quality assessment. Communications of the ACM, **45** (4), 211-218.

Tallon, P.P., Ramirez, R.V., and Short, J.E. (2013). "The Information artifact in IT governance: Toward a theory of information governance", *Journal of Management Information Systems*, **30** (3), 145-181.

Vera-Baquero, A., Colomo-Palacios, R., Molloy, O. (2016). Real-time business activity monitoring and analysis of process performance on big-data domains. Telematics and Informatics, 33, 793-807.