Data as Strategic Resources

Gunther, W.A.

2019

document version
Publisher's PDF, also known as Version of record

CHAPTER 5

DISCUSSION

5.1 Summary of Findings and Contributions

In the preceding chapters, I presented the findings of three studies that, when taken together, provide complementary insights into *how organizations explore the opportunities of data as strategic resources*. In the following sections, I will first summarize the key findings, main contributions, boundary conditions and directions for future research as described in each of the studies separately. Afterwards, I will draw on the findings from these studies to address the main research question and discuss the implications for researchers and practitioners. I will end my discussion with a methodological reflection.

5.1.1 Debating Big Data

In Chapter 2, I set out to understand what tensions organizations face when they (try to) realize value from “big data”. To this end, I reviewed papers from (top) academic IS journals and conferences that discuss organizational changes, drivers, and actions related to big data value realization.

Key Findings: My review of the literature resulted in the identification of six debates in the literature regarding how organizations realize value from big data, at three levels of analysis. These debates are representative of the tensions that organizations face. At the work-practice level, actors who work with and analyze data may adopt inductive or deductive approaches. Additionally, organizations will have to decide how much they will rely on algorithmic or on human-based intelligence for developing insights from data. At the organizational level, organizations need to find ways to structure their data-related resources and capabilities, and may do so in a centralized or decentralized fashion. Organizations also need to decide whether they will leverage data to improve their existing business model, or will focus on innovating or even transforming their business model. At the supra-organizational level, organizations need to decide to what extent and how they will share their data with external stakeholders. They may also carefully consider the social risks of data and data analytics (e.g., ethics and privacy) or choose to ignore these potential implications.

I also asked: What are the potentially unique features of big data that influence how organizational actors work with big data, develop organizational models, and deal with stakeholder interests in practice? This helped me better understand why data may provide certain opportunities and challenges for realizing social and economic value. Based on the review of the literature, I identified two socio-technical features of big data that influence value realization: portability and interconnectivity. Whereas portability refers to the possibility to remotely access data from and transfer data to other contexts, interconnectivity refers to the possibility to synthesize data from various sources.

Contributions: The contributions of the study are threefold. First, the study contributes to the emerging stream of literature on data-driven value realization, by presenting an in-depth analysis of the recent discourse and explicating critical, yet unsolved debates. Thus, the study unpacks the different assumptions that are associated with “big data” as an emerging phenomenon. Second, by presenting two socio-technical features of “big data” that may influence this process, the study responds to the calls for more socio-technical characterizations of the phenomenon (Markus & Topi, 2015). Identifying these features helped me to better contextualize value realization with regard to big data use. As a final contribution, Chapter 2 highlights the need for understanding data-driven value realization as a *process* in which organizations continuously realign work-practices, organizational models, and stakeholder interests. This is in line with a stream of literature that acknowledges that technology-driven strategizing and value realization are complex, emergent, and dynamic processes that involve many social dimensions and increasingly reach beyond organizational boundaries (Galliers, 1995; Marabelli and Galliers, 2017; Merali et al., 2012). The study presents a number of propositions reflecting these interactions, and presents an integrated model to visualize the relation between the debates, the socio-technical features of big data, and interactions between the different levels.

Boundary Conditions and Future Research: The study calls for future empirical research that looks at when, if, and how the opposing sides of each debate are relevant for organizations aiming to realize value from big data. For instance, future research may consider under what conditions organizations are more likely to adopt centralized or decentralized approaches to structuring their data-related capabilities (e.g., Schüritz et al., 2017), or when actors are more likely to engage in inductive approaches to collecting and analyzing data. Future research is also needed to explore how the two socio-technical features—portability and interconnectivity—can be exploited in practice. For example, future research may look at when portability and interconnectivity can be facilitated at the work-practice level to allow for the emergence of new insights, and how organizations may effectively engage in partnerships to allow for data from different stakeholders to be exchanged and synthesized. Finally, the study proposes a number of opportunities for future research on the potential interactions between work practices, organizational models, and stakeholder interests (see Table 2.1 in Chapter 2).

In light of this dissertation, I would like to highlight that the papers reviewed predate the fall of 2016. Since then, a number of empirical studies have emerged that focus on the topic of data-driven value realization. For example, Côrte-Real et al. (2018) focus on the “antecedents” of big data analytics value; Lehrer et al. (2018) highlight the material agency and affordances of big data analytics-related technologies in the context of service innovation, and Müller and Vom Brocke (2018) critically assess whether big data analytics actually results in productivity improvements based on firms' performance data. Future research may effectively focus on how recent studies contribute to the debates that we found and examine how the field is evolving over time. In doing so, studies may also consider to what extent the term "big data" may still usefully capture the phenomenon.

5.1.2 Minding the Data

In Chapter 3, I conducted an in-depth case study of a European postal service organization that had been triggered to invest in data and data-related technologies, yet struggled to actually realize value from data. I adopted a strategizing perspective (Galliers, 2007; Marabelli & Galliers, 2017; Mintzberg & Waters, 1985) and perceived the process of data-driven strategy development and implementation as an emergent process in which data and strategic choices and actions co-evolve over time. My specific aim was to understand how data can shape strategic processes and outcomes.

Key Findings: I was able to identify four mechanisms through which data can shape strategic processes and outcomes, that is, by 1) opening up choices and actions beyond the historical purpose of the data, 2) imposing historical decisions on current choices and actions, 3) encouraging the inclusion of external actors in data-driven strategizing, and 4) hindering external actors to engage in data-driven strategizing. Reflecting on these mechanisms, I concluded that they essentially reflect two tensions. First, while data may allow actors to explore new strategic paths by opening up opportunities beyond the historical purpose of the data, they may also cause organizations to stay close to the historical purpose of the data. Even the same characteristics of data that contribute to opening up strategic possibilities, such as the level of detail and relationality of the data, may simultaneously impose historical decisions on current choices and actions and thereby limit organizations in their exploration. In essence, data are "historically situated", as past decisions may be inscribed in the data and influence current and future choices and actions. Second, we find that, because of the nature of the data, organizations may be encouraged to closely collaborate with external stakeholders. However, the data can also prevent stakeholders from becoming fully engaged in such collaborations. The findings suggests that data need to be "contextualized" (Zeng & Glaister, 2017) in the sense that organizations need to ensure different stakeholders can make sense of the data and see how they may be relevant to their specific contexts.

Contributions: The findings in Chapter 3 contribute to existing literature in several ways. First, the findings contribute to the emerging stream of literature that aims to understand how organizations may realize value from data, by highlighting the role of data and illustrating that data can actually shape strategic processes and outcomes. The study highlights that, rather than pushing data to the background, scholars need to treat data as dissimilar, contextualized, and dynamic "actors" in order to fully understand how organizations may effectively leverage data as strategic resources.

Second, the findings presented in Chapter 3 cast criticism on one major assumption that is associated with "big data", which is that data are "unbounded" and infinitely open to be explored with may emergent value propositions (e.g., Aaltonen & Tempini, 2014; Constantiou & Kallinikos, 2015). The findings show that in many ways, data may actually force organizations to stay close to the original purpose for which the data have been used. I explicated that even the same characteristics of data that allow organizations to explore new opportunities, may also force them to stay close to the data's original purpose.

Third, the findings suggest that data may also not be as "portable" as they are often described in the literature. Literature so far has assumed that data can be readily accessed and transferred from one context of application to be used in another context (see also Chapter 2). The findings from Chapter 3 illustrate, however, that even when data are portable from a technical perspective, characteristics such as ownership may prevent organizations from using them for different purposes. Similarly, the fact that data need to be effectively "contextualized" (Zeng & Glaister, 2017) suggests that data cannot be readily used in many different contexts without substantial effort.

As a final contribution, I stressed the importance of taking a process view for those aiming to understand how organizations may effectively leverage data as strategic resources. Not only are data dynamic resources of which the nature and role may change over time as actors work with them in practice, but they are also *historically situated*. Scholars should capture how past choices and actions are inscribed in the data, and how this historical situatedness shapes organizing processes.

Boundary Conditions and Future Research: Chapter 3 presents a number of directions for future research. Given the fact that data may impose historical decisions on current choices and actions—thereby trapping organizations into staying close to the historical purpose of the data—an interesting direction for future research would be to see how actors can potentially "break free" from the historical choices and actions that are inscribed in the data. Similarly, given the fact that data need to be "contextualized" in order for different stakeholders to be engaged, future research may examine how such contextualization of data can effectively be supported.

The findings in Chapter 3 are based on a single case study of an organization that offered a rather traditional service (i.e., the delivery of physical mail). Future research may focus on examining the role of data in contexts where the business strategy is inherently more "digital" (Bharadwaj et al., 2013). Future research may also effectively examine the role of data in the context of less-controlled markets, to see to what extent characteristics such as "ownership" and "personal sensitivity" influence data-driven strategizing in those contexts. Situations in which organizations more intensively collaborate with other stakeholders would also be especially interesting and may particularly lead to insights regarding how organizations can "contextualize" data across different contexts. Finally, future research may examine the role of data in strategizing processes in which organizations heavily rely on sophisticated algorithms and Artificial Intelligence.

5.1.3 Leadership in the Age of Data Analytics

In Chapter 4, I set out to explore what data analytics leadership entails. I collected job ads for senior managers with titles pertaining to "data analytics", "digital", and "information technology". I then applied a topic modeling algorithm on their contents to explore: 1) What are the data-related responsibilities that characterize data analytics leadership? 2) Which positions will be responsible for providing data analytics leadership? and 3) Which (combinations) of responsibilities are stressed as the most important for "data analytics leaders"? I hereby iterated between the patterns and the qualitative data in order to make sense of the patterns.

Key Findings: In Chapter 4, I described four data-related responsibilities of senior managers. Specifically, I found that senior managers may be held responsible for implementing a technological (big) data infrastructure ("data infrastructure"); ensuring that data can be accessed and used across the organization ("data control"); facilitating and initiating the creation of insights, products, and services from data through analytics ("applied analytics"), and considering the legal and ethical privacy implications of data analytics ("data privacy"). After identifying these data-related responsibilities, I explored how much they are highlighted in the job ads of different positions. The data suggest that job ads for "data analytics" positions often provide explicit information regarding data-related responsibilities, which implies that organizations are not just signalling these terms in the job titles (Nelson et al., 2014). Conversely, positions with titles pertaining to "information technology" and "digital" are much less likely to be responsible for providing data analytics leadership. Finally, I explored which (combinations of) responsibilities are highlighted as the most important for "data analytics leaders". I found that "applied analytics" is the most highlighted responsibility, which suggests that "data analytics leaders" in our sample are mostly responsible for the creation of insights, products, and services from data through analytics. Conversely, I found that "data privacy" is by far the least popular responsibility.

Contributions: The findings have a number of implications for how we perceive and study leadership in the age of data analytics. First, I discussed to what extent the data-related responsibilities (i.e., data infrastructure, data control, applied analytics, and data privacy) may be contemporary characteristics that are specific to leadership in the context of data analytics. For each of the responsibilities, I discussed how it relates to, and diverts from, responsibilities that have traditionally been considered part of IS leadership. I argued that new or complementary leadership theories may be needed that acknowledge the differences between data analytics leadership and more traditional IS leadership.

Second, by exploring which positions are typically expected to be responsible for providing data analytics leadership, the study helped lift some of the controversy surrounding this topic. The findings suggest that similar to when new IS positions at the level of senior management emerged as organizations began to see the real strategic opportunities in technology (Chun & Mooney, 2009), new senior data analytics positions may now be emerging because organizations take data analytics seriously at the level of strategy. The observation that new positions are being created also implies, however, that existing organizational structures and roles may not only “evolve” to facilitate the successful implementation of data-related initiatives (Sharma et al., 2014), but may be fundamentally reshaped through the introduction of new positions.

Third, the fact that “data privacy” is marginally discussed in job ads for “data analytics leaders” made me question to what extent these senior managers will actually be expected to “fully consider the context of their data through the lens of privacy and unique privacy interests” (Everson, 2017, p. 32) and evaluate how “datafication of our everyday lives and the associated algorithmic decision-making [..] will affect society” (Newell & Marabelli, 2015, p. 13). The finding seems somewhat alarming in light of recent debates regarding the numerous risks for consumers and society when organizations exploit data (Clarke, 2016; Newell & Marabelli, 2015; Markus, 2015).

Boundary Conditions and Future Research: I proposed a number of directions for future research on data analytics leadership. I suggested that future research is needed to assess to what extent data-related responsibilities also raise new and unique challenges for senior managers, and *how* they may effectively deal with these potentially unique challenges in such a way that they can indeed fulfil their important roles. This would require future studies to observe how data analytics leaders enact each of the data-related responsibilities that we identified in practice. Additionally, I called for future research that examines *how* organizational structures, processes, and relations change as organizations create new leadership positions; how different data analytics leadership roles may best be structured in organizations, and how organizations may prevent roles from overlapping. In light of the finding that “data analytics leaders” are not explicitly expected to think about the privacy implications of data analytics, I called for future research that focuses on how organizations ensure that responsibilities related to the social implications of data analytics are an integral part of “data analytics leadership”, especially when data privacy and related concerns are considered part of another domain.

Given that the findings provide a snapshot of the responsibilities of data analytics leaders and associated positions, future research may exploit the longitudinal nature of the data to examine how roles and responsibilities of data analytics leaders change over time. Moreover, whereas the study has been limited to the analysis of job ads collected from the U.S., future research may also explore patterns across different cultural contexts. Future research may also consider patterns across industries and different types of organizations, and may examine a complementary range of senior manager positions (e.g., Chief Marketing Officer) to see how these positions are involved in providing data analytics leadership.

	Chapter 2	Chapter 3	Chapter 4
Title	Debating big data: A literature review on realizing value from big data	Minding the data: An empirical analysis of how data shape data-driven strategizing	Senior leadership in the age of data analytics: Exploring organizational expectations
RQ	What tensions do organizations face in realizing value from data?	How do data shape the process of data-driven strategizing?	What are the expected responsibilities and positions of data analytics leaders?
Methods	Systematic literature review of articles from (top) academic journals and conferences on “big data” in organizational settings.	In-depth case study of a rich case that aimed to leverage data to create new data-driven products.	Analysis of job ads for senior managers with varying titles pertaining to “information technology”, “digital”, and “data analytics”.
Findings	- Six tensions that organizations face as they realize value from data, at different levels of analysis. - Two socio-technical features of “big data” that influence how organizations realize value from data in practice.	Data may shape strategizing by: - opening up choices and actions beyond the historical purpose of the data, - imposing historical decisions on current choices and actions, - encouraging the inclusion of external stakeholders in data-driven strategizing, - hindering stakeholders to become engaged in data-driven stratagizing.	- Four data-related responsibilities that characterize leadership in the age of data analytics. - Job ads for “data analytics” positions provide explicit information regarding data-related responsibilities, while “digital” and “information technology” positions are much less likely to be responsible for providing data analytics leadership. - Job ads highlight “applied analytics” as the most important responsibility for data analytics leaders, while “data privacy” is by far the least popular responsibility.
Contri-butions	- Provides an in-depth analysis of the recent discourse on “big data”. - Responds to calls for more socio-technical conceptualizations of “big data”. - Calls for a perspective that focuses on the continuous aligning between different levels of analysis and presents a number of propositions and an integrated model.	- Highlights the need for treating data as dissimilar, contextual, and “dynamic” actors. - Casts criticism on the assumption that data are inherently unbounded. - Casts criticism on the assumption that data are inherently portable. - Stresses the value of taking a process view when studying how organizations may effectively leverage data as strategic resources.	- Addresses whether data-related responsibilities may be contemporary responsibilities that uniquely characterize leadership in the age of data analytics. - Suggests that organizational structures may be reshaped through the introduction of new positions. - Highlights that privacy-related concerns may need to become an integral part of data analytics leadership.
Future research	Future research may: - look at when, if, and how the opposing sides of each debate are relevant for organizations. - consider how portability and interconnectivity are facilitated and exploited in practice. - study interactions between work-practices, organizational models, and stakeholder interests in the context of big data.	Future research may: - focus on how organizations can “break free” from historical choices and actions inscribed in the data. - examine how the contextualization of data across different contexts can be effectively supported. - examine non-similar cases to find other mechanisms, e.g., in contexts of digital organizations, less-regulated markets, collaborative environments and sophisticated algorithms.	Future research may: - observe how data analytics leaders actually enact their role in practice and face unique challenges in doing so. - observe how existing organizational structures change as new positions are introduced. - examine how ethical responsibilities can become an integral part of data analytics leadership. - examine changes over time; consider patterns across different cultural contexts, industries, and organization types, and include complementary senior manager positions.

Table 5.1. Summary of chapters.

5.2 Response to the Overall Research Question

In the introduction of this dissertation, I argued that there is much optimism and hope regarding the opportunities of data as strategic resources. I illustrated that these hopes are sometimes met with skepticism and that both scholars and practitioners have warned that the potentials of data as strategic resources may be overshadowed by hypes. My aim for this dissertation was to move beyond the hype and examine: *How do organizations explore the opportunities of data as strategic resources?*

The findings from this thesis strongly indicate that organizations *are* actively exploring the opportunities of data as strategic resources. In Chapter 3, I presented the story of an organization that seriously invested in developing data-driven products and saw data as strategic resources that could help them create value for business clients. In Chapter 4, I showed that organizations are creating new senior manager positions for people who can be held responsible for data and data analytics. I showed that these managers will not only be held responsible for data governance or data management, but also for ensuring the creation of insights, products, and services from data through analytics. These findings confirm that organizations are no longer looking at data as a by-product of IT processes, but as something that can be leveraged strategically and that requires dedicated attention at the tactical and even strategic levels.

Most of all, however, my findings suggest that the benefits that are often associated with data as strategic resources cannot readily be attained. Chapter 2 explicated that organizations face several tensions when exploring the opportunities of data as strategic resources, and organizations may differ quite substantially in how they approach these tensions. Here, I theorized that organizations may need to continuously re-align efforts at the work-practice, organizational, and supra-organizational levels in order to realize value from data. Similarly, Chapter 3 illustrated that organizations may encounter many (unforeseen) opportunities and challenges as they explore the opportunities of data as strategic resources, and that this process of exploration is a complex process in which the data and strategic choices and actions may continuously shape each other over time.

Based on my studies, I concur that while there is certainly some hype surrounding the topic—and while terms like “big data” or recently even “thick data” might die out—*the phenomenon* that data are becoming strategic resources is very real and may be “here to stay” (e.g., Abbasi et al., 2016; Madsen & Stenheim, 2016; Marr, 2016; Woerner & Wixom, 2015). Yet, scholars and practitioners alike should not be overly optimistic, as the expected benefits cannot be readily attained. The insights from this dissertation present learning opportunities both for scholars and practitioners who aim to understand how organizations may effectively leverage data as strategic resources. I will elaborate on these learning opportunities in the following sections.

5.2.1 Implications for Research on Data-Driven Value Realization

Based on the findings presented in this dissertation, I suggest three key learnings for scholars interested in studying how organizations can realize social and economic value from data.

The Need for Being Critical about the Nature and Role of Data

Scholars may learn from this dissertation that they need to be critical about the nature and role of "data" as strategic, digital resources. In different chapters of this dissertation, I showed that unpacking the characteristics of data may help us better understand how data can effectively be leveraged by organizations. For example, in Chapter 2, I took the opportunity to critically examine what characteristics of data lied at the heart of the debates that I identified. This allowed me to identify two socio-technical features of big data that may influence data-driven value realization: portability and interconnectivity. By critically reflecting on the nature of big data, I was able to move beyond technical characteristics such as volume, variety, and velocity, and observed that the phenomenon rests on the assumptions that data can readily be transferred and accessed from one context of application to be used in other contexts (e.g., Lycett, 2013; Tallon et al., 2013-14), and that new insights can be gained from the synthesis of data from various sources (e.g., Lycett, 2013; Malgonde & Bhattacherjee, 2014).

Similarly, in Chapter 3, I focused specifically on how data may shape the process of data-driven strategizing in an empirical setting. By foregrounding the data in the process of data-driven strategizing, I found that data may actually influence strategic choices and actions in several ways. I found that whereas characteristics like granularity and relationality may open up a range of opportunities for organizations, the same characteristics may also lead organizations into staying close to the original purpose of the data. Additionally, I found that different stakeholders may struggle to see how data are relevant to their context when data are too raw (or conversely, too abstracted), which suggested that data may not be as extremely portable as they are sometimes assumed to be. Thus, by unpacking the nature and role of data in Chapter 3, I was able to see how data may actually guide actors as they are making certain choices and performing certain actions.

In sum, while much of the literature on the topic has so far been pushing data to the background, I showed in both Chapters 2 and 3 that there is value in bringing data to the forefront and critically examining what it is about data that is influencing how organizations conduct their business, and how this happens. I thus concur with recent literature (Jones, 2018; Kitchin & McArdle, 2016; Tempini, 2017), which argues that to fully understand how organizations may leverage data as strategic resources, we need to pay more attention to the nature and role of data. As argued in Chapter 3, this requires scholars to examine what the different characteristics of data are (beyond e.g., volume, variety and velocity), how these data are being used in specific contexts, and how the data themselves and also the roles of data may change over time.

The Power of Adopting a Process Lens

Scholars may also learn from this dissertation that adopting a process lens can be especially helpful when studying how organizations (try to) realize value from data. Throughout this dissertation, I explicated several reasons for why this is the case. For example, in Chapter 2, I illustrated how important it is that organizational efforts at different levels of analysis (i.e., at the work-practice, organizational, and supra-organizational level) are aligned, and argued that achieving alignment between the different levels is not a sporadic effort but rather a continuous process of align*ing* (cf., Karpovsky & Galliers, 2015; Wilson et al., 2013). I argued that scholars aiming to understand how organizations realize value from data need to capture this process and see how organizations continuously realign data-driven work practices, organizational models, and stakeholder interests.

In the study presented in Chapter 3, I initially proposed two reasons for adopting a process lens when studying how organizations may leverage data as strategic resources. First, I argued that it is crucial to recognize that data are in practice very dynamic resources that change over time. Data can be extended, combined, deleted, abstracted, and continuously updated as actors work with data in practice and through other digital artifacts (Kallinikos et al., 2013; Yoo et al., 2010; Ekbia, 2009). Second, I acknowledged that even when data themselves do not change, actors may still continuously uncover new functions of the data as they interact with them in practice (Kallinikos et al., 2013), causing many insights, opportunities, and challenges to emerge *en route* (Aaltonen & Tempini, 2014; Constantiou & Kallinikos, 2015; Zeng & Glaister, 2017). By adopting a process lens in this study, I was also able observe that data are *historically situated*, that is, past choices and actions may become inscribed in the characteristics of the data that subsequently influence current and future choices and actions. This might be one important, complementary explanation for why organizations struggle to realize value from data. And, this finding highlights even more the need for adopting a process lens; scholars may only fully understand how data can be leveraged as strategic resources by also capturing the historical situatedness of data, which inevitably would require them to have knowledge about past choices and actions (Jones, 2018).

Thus, while much of the research on this topic so far has adopted a variance approach to understand which variables play a role when organizations aim to realize value from data (see also Chapter 2), this dissertation highlights that very relevant and complementary insights into how organizations may (fail to) leverage data as strategic resources can be gained by adopting a process lens. Adopting a process lens allows scholars to focus on “how and why things—people, organizations, strategies, environments [and data]—change, act and evolve over time” (Langley, 2007, p. 271), which is particularly helpful to capture the dynamic, interactive, and historically situated nature of data.

The Need to Care About the Social Implications

Scholars may also learn from this dissertation that there is a great deal of demand for studies that explicitly consider the social implications of data analytics. For example, in Chapter 2 of this dissertation, I showed that there is a debate in the IS literature regarding the extent to which different organizations should be allowed to pursue economic value, *at the expense* of individual and societal values. While the literature (e.g., Clarke, 2016; Newell & Marabelli, 2015; Zuboff, 2015) has presented examples of organizations that take extensive measures to ensure that they do not violate any legal or ethical norms, there are also examples of organizations that pursue data-driven goals while almost ignoring how this might harm individuals or society. So far, IS studies on big data value realization have provided little empirical evidence that illustrates how organizations can effectively deal with legal and ethical concerns in acceptable and innovative ways.

The case of LogiCo, presented in Chapter 3, serves as an example of an organization that actually greatly cared about the social implications of data use. Members of the organization expressed explicit concerns about the privacy of consumers. What this case shows, however, is that it might not even be clear if and how the use of particular data could negatively affect individuals and society. For example, LogiCo’s stakeholders experimented with “mail order data” to see if they could use these for targeted advertising. Mail order data are not by definition personally sensitive. These data *became* personally sensitive after finding out that by combining mail order data from different business clients, analysts could gain insights into the personal preferences of people living on addresses. This shows that even organizations with the best intentions may struggle to see how the use of data could impose risks for individuals and society, and that insights are needed regarding how organizations and other stakeholders may prepare for these kinds of events where data use could raise unintended consequences.

Finally, in Chapter 4, I highlighted even more the need for research that focuses on how organizations can organize themselves (and others) in order to deal with the (unforeseen) social implications of data analytics. In this study, I found that only very few of the job ads for “data analytics leaders” explicitly mention terms related to legal and ethical data privacy concerns. Although I admitted that this does not mean that organizations are not concerned about the social implications of data analytics at all, I did question to what extent this means that privacy is considered as a strategic dimension when organizations look for data analytics leaders. In this chapter, I explicitly called for research that looks at how the social implications of data can become an integral part of data analytics leadership.

5.2.2 Implications for Practice

The findings presented in this dissertation can also be helpful for practitioners aiming to leverage data as strategic resources. Below, I suggest three key learnings for managers and other organizational stakeholders.

A Cross-Level and Collaborative Effort

Managers should be aware that only hiring data scientists or implementing data-related technologies will not be enough to realize value from data. Instead, they may also have to redesign their company’s existing organizational models, and reconsider the different ways in which they interact with external stakeholders. For example, I illustrated in Chapter 2 that even when analysts at the work-practice level can gain relevant insights by analyzing the data, these insights may not lead to value when they are not accompanied with the creation of appropriate organizational models. Similarly, I argued that organizations may not be able to implement data-driven business models when they are unable to deal with stakeholder concerns. These examples illustrate that realizing value from data is in essence a *cross-level process*, which requires managers to carefully consider how they may facilitate data collection and analysis at the work-practice level, how they will design organizational models at the organizational level, and how they will deal with stakeholder interests at the supra-organizational level.

Because data-driven value realization is a cross-level process, it is also important that different organizational stakeholders collaborate to ensure that work-practices, organizational models, and stakeholder interests are aligned. This involves not only stakeholders who are directly involved in analyzing data (e.g., analysts and data scientists), but also senior managers who can be held responsible for developing organizational models. In Chapter 4, I showed that organizations are even creating new senior manager positions to be responsible for data analytics. In light of these findings, managers who are reading this dissertation should carefully reflect on their own role in data-driven value realization, and potentially assess the need for creating new positions in order to ensure alignment between the levels.

Practitioners should also be aware that collaboration efforts go beyond the boundaries of their organization. For example, in Chapter 2, I argued that organizational actors may have to consider the interests of several external stakeholders (e.g., users, consumers, competitor organizations) when they analyze data and develop organizational models. Additionally, in Chapter 3, I showed that—because of the nature of the data—organizations may be encouraged to actively involve external stakeholders such as data providers, business clients, and consumers in processes of developing and implementing data-driven strategies. Thus, practitioners who are interested in realizing value from data need to carefully consider to what extent they wish to engage with external stakeholders, how much they may actually rely on external stakeholders, and how they will deal with all the different stakeholder interests that could impact data-driven value realization.

A Data-Driven Journey of Exploration

Practitioners can also learn from the findings presented in this dissertation that the process of data-driven value realization is often *not* a linear process in which organizations simply collect data, analyze these data, and then engage in value-creating actions. Instead, practitioners may find themselves continuously revisiting the data, their choices, and their actions in a process of exploration. For example, I explained in Chapter 2 that when analysts or decision makers work with data in practice, unforeseen opportunities or challenges may emerge, requiring them to revisit their initial choices (Constantiou & Kallinikos, 2015). Similarly, in Chapter 3, we saw that managers in the case of LogiCo were surprised to find that mail order data allowed analysts to gain very personal insights about households, which not only opened up unforeseen opportunities for them, but also raised unforeseen privacy challenges. Strategy team members then had to reconsider using these data for the goal of offering targeted advertising and had to revisit the initial strategy.

The examples show that even when managers have a clear idea of how they wish to leverage data to attain economic or social value, they may at some point have to deviate from the original plans, either to exploit new opportunities or respond to unforeseen challenges. Similar to LogiCo, organizations may find themselves continuously readjusting their data and their data-driven strategies. This is not to say that there is no point in developing data-driven strategies or business cases. On the contrary, studies have argued that an initial strategy and vision is key to ensuring that value can be realized from data (e.g., Lavalle et al., 2011; Shollo & Galliers, 2015). However, the findings from this dissertation do imply that managers may need to offer organizational stakeholders the time to learn about the data and the different ways in which these data can enable or constrain strategies; that managers should frequently revisit their business cases and strategies, and that they should perhaps not purely evaluate data-driven initiatives on the basis of the original business case.

Mind Your Data

Last but certainly not the least, managers can learn from my dissertation that it is extremely important that they know *what data* they have and *how* these data may play a role in the development and implementation of data-driven strategies. Managers should be careful not to black-box the data that they are using; even when data are of high-volume, high-variety, and high-velocity, this does not automatically open up infinite possibilities. Managers should be extremely aware of what data are relevant in their context, thereby also moving beyond general quality characteristics such as accuracy and completeness. In Chapter 3, we explicated a number of additional characteristics such as granularity, relationality, and timeliness, that shaped data-driven strategizing in the case of LogiCo. Managers need to be open to see how *any* characteristic of the data may be relevant in *their* context. Consequently, this also requires managers at different levels to have affinity with data, including those residing at the level of senior management.

An important part of knowing what data you have also involves knowing where the data came from and how traditional decisions around the data may be inscribed in its characteristics (including the metadata). Just because you have access to data, does not mean that you can or should use it. We saw this in the case of LogiCo: the data that they collected from external sources were traditionally updated at a frequency and collected at a level of granularity that did not fit with the purpose for which LogiCo aimed to use the data. The circumstances under which data have been collected were also important: it turned out that because mail order data were uploaded by business clients, LogiCo did not actually own these data and therefore could not use them for any new purpose. Regardless of whether data come from outside the organization or inside organization, managers need to understand how the data “came to be” (Jones, 2018) in order to understand how they may leverage these data as strategic resources.

5.3 Methodological Reflection

In this dissertation, I relied on a range of different methods in order to answer my research questions. In Chapter 2, I performed a literature review; in Chapter

3, I conducted a longitudinal case study, and in Chapter 4, I applied a topic modeling algorithm to help me analyze the contents of job ads. In this section, I would like to briefly reflect on my experiences with the different methods adopted in the studies, and share some key learnings.

5.3.1 Identifying and Synthesizing Literature

In Chapter 2, I conducted an in-depth review of the IS literature on "big data", in order to identify tensions that organizations might face when they realize value from data. I aimed to move beyond a descriptive overview of the field and went in-depth to extract contradicting findings and opinions from relevant papers. This required me to put substantial effort into identifying relevant literature, and into effectively synthesizing papers in a way that would allow me to identify tensions.

Those who consider systematic literature reviews as a research approach have highlighted the need to be specific about how the researcher identified and selected "relevant" papers (e.g., what search queries were used, which databases were searched, and which inclusion and exclusion criteria were applicable) (Boell & Cecez-Kecmanovic, 2015; Jones & Gatrell, 2014; Webster & Watson, 2002). I experienced that the key to "systematically" doing this, is keeping a log of *all choices* that have been made over time. For example, I consistently kept track of all the specific papers (e.g., author, title, abstract, year, outlet) that resulted from my searches—including the ones that initially seemed "irrelevant". For all of these papers, I kept a record of why the particular paper was deemed "relevant" or "irrelevant", by specifying which inclusion and exclusion criteria it adhered to. Because of this, I was continuously able to reflect on earlier choices, reiterate if needed, and ensure that my approach for identifying and selecting relevant papers was "systematic" and consistent.

What proved to be a less straightforward task, however, was to effectively analyze and synthesize insights from the different papers, in order to be able to identify contradicting findings and opinions. I began by designing a review framework and using this as a guide for qualitatively coding each of the papers that I had found. This helped me to interpret the authors' views on "big data" in the context of organizations. However, I realized that coding individual papers does not allow you to easily see patterns *across* different papers; I needed a way to effectively synthesize papers. Eventually, I created one large table; each row in the table represented one of the papers, and each column represented an element from my review framework—similar to Webster and Watson's (2002) concept table. In the cells I summarized, for each paper, the different elements of the review framework using codes and quotes that resulted from my coding efforts. Having this table allowed me to more easily see how findings, assumptions, and theories described in the various papers differed. By continuously iterating between potential patterns and the different papers, and updating the table accordingly, I was eventually able to identify the six debates presented in Chapter 2.

In sum, I have learned that the key to systematically identifying and selecting relevant papers is to consistently keep a log of all the choices that have been made over time. I also learned that if scholars wish to move beyond general descriptions of a field and truly capture the theoretical foundations and assumptions that are associated with it (Webster & Watson, 2002), then the key is finding a way to effectively synthesize the relevant literature. Although one can try to do this "systematically", researchers will find that synthesizing literature and identifying patterns from their synthesis is actually a very iterative process in which you may continuously move back and forth between potential patterns and the different papers. Undoubtedly, however, such efforts will allow researchers to gain in-depth understanding of their field.

5.3.2 Understanding Organizational Contexts and Processes

In Chapter 3, I conducted an in-depth case study of an organization that was triggered to invest in data and data-related technologies by the promises of social and economic value, yet struggled to actually realize such value. My aim was to understand how data shaped the process of data-driven strategizing in this case. This required me to have a solid understanding of the organizational context, as well as how different elements changed *over time.*

To understand how data shaped the process of strategizing in the specific case, I needed to have knowledge about, at least, the organization's business model and strategy; organizational structure; dominant culture; running projects; relevant stakeholders; ecosystem partners, and even the regulatory environment. To gain access to such knowledge, I engaged in an "exploration" phase in which I collected numerous documents, attended meetings on a multitude of topics, and conducted explorative interviews. To effectively synthesize the data that I had collected in this phase, I adopted an approach inspired by Yin (1981), who argues that "data that address the same topic should be assembled together" (p. 60). Specifically, I created a "case report" in which I grouped raw excerpts from different interviews, documents, and meeting notes that were about similar topics under higher-level themes (e.g., "structure" and "culture"). At any time, I could go back to that document to immediately observe what had been said about a particular theme. This proved to be very helpful for me to become familiar with the organizational context.

The data that I collected after the exploration phase were more specifically focused on capturing the data-driven story—I was especially interested in capturing how data-driven strategic initiatives had emerged and evolved over time. I collected interview data, data from strategic documents, and meeting notes. I found that these data could be integrated and synthesized in a similar fashion as I had done during the exploration phase, that is, by integrating data elements into higher-level themes. This time, however, my focus was on capturing a *process,* which is why I put additional effort into chronologically ordering the themes. The resulting document functioned as a "case story" that specified how the organization had tried to create and appropriate value from data over time; who was involved and how; what the data looked like; what external and internal contextual factors and events happened, and what the (intermediate) outcomes of strategizing with data were. This allowed me to gain a detailed understanding of the process of data-driven strategizing—so detailed that interviewees were sometimes surprised when I helped them remember when something had happened.

In sum, I have learned that an "exploration" phase may be particularly useful for researchers to become familiar with the context in which phenomena of interest occur (Yin, 1981; Eisenhardt, 1989). As a way of making sense of the data collected during such a phase, researchers may consider integrating and combining data elements into a "case report", where similar data are combined under general themes (Yin, 1981). Over time, the themes may become more specific as the researcher becomes more focused on a particular research question. I learned that such a way of integrating data can also be very useful when adopting a process lens (Langley, 2007; Pettigrew, 1990; Pettigrew, 1992). The focus should then be on finding a way to chronologically order themes. Doing so would result in a type of "case story" in which raw data excerpts are put together when they are about similar issues that happened in the same time period. Researchers interested in studying organizing processes may also consider creating such a document, and may consider coding this document as an alternative to coding all the individual interviews, documents, and notes in isolation (c.f. Huang et al., 2014).

5.3.3 Relying on Algorithmic Intelligence

The approach that I adopted in Chapter 4 differs from the approaches adopted in the other chapters, as I largely relied on "algorithmic intelligence" to answer my research question. In this study, my aim was to understand what "data analytics leadership" entails. I used a tool to scrape 13191 job ads (including duplicates) for senior managers with varying titles. The study was very explorative; I was open to see which responsibilities may characterize data analytics leadership, and which positions may be responsible for providing data analytics leadership. I hereby largely relied on computing power and a topic modeling algorithm to analyze the job ads and arrive at relevant patterns.

As illustrated in Chapter 4, algorithms can be particularly useful to identify patterns, and provide a useful and fast alternative to processing large amounts of information. However, there are definitely also challenges and pitfalls. I noticed that, while there are quite some guidelines for how (not) to do a systematic literature review (e.g., Boell & Cecez-Kecmanovic, 2015; Jones & Gatrell, 2014; Webster & Watson, 2002) and how to conduct a case study (e.g., Yin, 1981; Eisenhardt, 1989), there is little guidance for IS scholars on what to look out for when relying on algorithmic intelligence to answer research questions. In this section, therefore, I would like to share three concrete suggestions for IS scholars who are interested in applying algorithms to help them answer their research questions.

Be Careful What You Feed the Algorithm

My first suggestion is to be careful what you feed the algorithm. Algorithms feed on data, and *humans* are typically the ones who feed them. Similar to my argument that managers should be aware of what data they are working with and how this may shape their organizing processes, so too do scholars need to critically evaluate what data they feed their algorithms and how this may influence the results. In fact, scholars can exercise quite a lot of agency in making choices regarding the data. I will illustrate this point using examples from my own experiences.

For the study described in Chapter 4, I collected data from a popular job website, Indeed.com. But before even collecting the data, I needed to ensure that I understood—as best I could—how Indeed collects the job ads (e.g., Indeed also scrapes external sources); what metadata they offered (e.g., company names, industries, level of experience); how accurate these metadata were (unfortunately, I found they were often not accurate and decided not to rely on them), and at what frequency new job ads were being added. I spent quite some time exploring the nature of the data that I was going to base my conclusions on. I then devised a search strategy to actually collect those data that would help me to answer my research question. In doing so, I had to make several additional choices, such as which version of the website to use (e.g., US-based, UK-based, China-based); how often I should scrape data, and what search queries to use (e.g., I found that a search for “senior manager” resulted in much noise).

Before I could apply Nonnegative Matrix Factorization to analyze the data that I had collected, I first had to devise ways to filter and clean them. This involved several steps in which I also had to make a range of different choices. For example, I needed to find a way to filter out job ads that did not actually comply to the inclusion criteria, and devise a way to decide when a job ad is a duplicate of another job ad. In the process of cleaning data, I had to decide which words I would include as “stopwords” (stopwords are generally defined as “trivial” words, but what makes a word trivial?); what to do about extremely frequent and infrequent words; how to treat punctuation and numbers (e.g., delete them or replace them by a whitespace?); whether to use stemming or lemmatization (or none of these?); what to do with abbreviations; when to treat something as a bigram, and what to do with these bigrams (e.g., replace them or add them as supplements to the text?).

Scholars aiming to rely on algorithmic intelligence may face similar questions as I did when they try to collect, select, and clean data for analysis. In making such choices, it is important that scholars realize that each choice could potentially affect the eventual outcomes. In fact, this is the key argument of scholars emphasizing that algorithms are not “unbiased” or “objective” (e.g., Jones, 2018; Newell & Marabelli, 2015); rather, algorithms are driven by humans and merely reflect the choices that have been made by humans. As scholars, it is our primary job to always carefully reflect on each of the choices that we make as we collect our data and prepare these data for analysis. And not only do we need to carefully reflect on our choices, but we also need to be transparent about them, such that anyone who interprets the results and relies on our work to inform their decisions can also become aware of the choices that were made along the way. Practically, this would mean that any work in which scholars relied on algorithmic intelligence should be accompanied by an extensive appendix in which scholars explicate the different choices they made in selecting and cleaning the data—see for an example Sidorova et al. (2008), and the Appendix for Chapter 4 in this dissertation.

Familiarize Yourself with the Algorithm(s)

My second suggestion is to familiarize yourself with the algorithm. Perhaps the most important decision that scholars need to make when relying on algorithmic intelligence, is *what* algorithm they will use. Evidently, not all algorithms are equally useful for the purpose at hand. When choosing an algorithm, it is important that scholars understand the assumptions that underlie the different algorithms.

Unfortunately, deciding which algorithm might be most applicable is not an easy task (unless perhaps when you are really into mathematics). Many of the papers in which algorithms are introduced are quite technical and focus on explaining the mathematical foundations of the algorithm, while it is not always clear how this relates to real-world assumptions, or when scholars are better off choosing one algorithm over the other. One may try different algorithms and then choose the one that gives the best results (e.g., Dai & Wu, 2017). However, scholars adopting this strategy need to always critically reflect on *why* some algorithms may produce “better” results than others, and also specify what they consider to be “better” results—this is a point that I will elaborate on in the next section. As this is perhaps not an ideal approach, my advice would also be to involve people who are familiar with the mathematical assumptions underlying the different algorithms, and engage in discussions with them to see how these algorithms may be applicable to your specific research question and context.

Once scholars have chosen an algorithm, many choices still remain to be made. Often, the implementation of the algorithm requires a number of parameters to be set. As described in Chapter 4, for example, we had to specify how many latent topics we expected to find in the dataset; what we considered to be a reasonable error margin; how many times we would allow the algorithm to iterate, and how the algorithm should initialize the procedure (e.g., just start at random or consider a more sophisticated initialization procedure). We found that even a seemingly trivial choice such as the choice of random seed (i.e., a number used to ensure that a model can be replicated, given that there is some degree of "randomness") can influence the results. Now, there are default values for each of these parameters, and there are even automated ways to find a combination of parameters that is optimal from a computational point of view. But as I will also highlight in the next section, what is "optimal" from a technical perspective is not always conceptually meaningful or particularly helpful. Rather, scholars may have to experiment with different parameters in order to find what works given their specific data, in their specific context, and given their specific research question. To illustrate—although I did not keep count—I may have run NMF on my data thousands of times. Each time I ran it, I also learned new things that I could subsequently use to optimize my methods for cleaning the data, and training a new model.

Mind the Unintelligent Nature of Algorithms

Finally, I would like to highlight that the results of algorithms like the one we used in Chapter 4 still need to be made sense of. In essence, there is no deeper meaning behind why algorithms produce certain results other than that it was the logical outcome of applying a set of mathematical steps (and often also incorporating some randomness). This implies that scholars may have to put substantial effort into making sure that the results can be interpreted, and also be *critical* when actually interpreting these results.

First, I would like to warn scholars that extra steps may be needed before they can get to the point where they can interpret the results produced by algorithms. At least in my case, interpreting the results produced by NMF was not a trivial task. In essence, NMF is a dimensionality reduction algorithm that results in two matrices (Berry et al., 2007). The products of these two matrices approximates the original input matrix, which in our case was a TFIDF-weighted document-word frequency matrix (Salton & Buckley, 1988). This is a quite technical explanation, and while there are many papers that explain the technicalities behind this algorithm, there is less guidance on how to interpret the matrices and assign meaning to them. The problem perhaps is that algorithms like these are often not designed to be interpreted; they are designed to automatically cluster computationally similar data. Thus, scholars may have to put substantial effort into making the results interpretable, while also critically reflecting on how this influences the conclusions that can be drawn from the results.

I would also like to highlight that because algorithms in their very essence represent a set of mathematical steps, they can behave in very "unintelligent" ways. For example, as I experimented with different data cleaning procedures and parameter selections, I was confronted with situations in which the algorithm that I used could not distinguish between two different uses of the same word. I found for instance that when "privacy" is mentioned in the disclaimer of a job ad, this job ad might score almost equally high on the topic "data privacy" as a job ad where the senior manager is actually expected to think about the social implications of data analytics. I could not rely on the algorithm to distinguish between these two different uses of the word. As this was undesirable given my research question, I eventually had to do more extensive pre-processing and remove disclaimers[40]. I also needed to continuously go back and forth between the patterns and the data, to see if the patterns actually made sense, and to see if my interpretation of the topics was also in line with what was described in the job ads.

Eventually, I needed to choose a trained model that would help me to further explore the data and gain insights into what data analytics leadership entails. Scholars have come up with a number of measures, such as "topic coherence" to guide analysts in their choice of a particular model (e.g., O'Callaghan, 2015). Related to my previous point, however, I would like to highlight here that models that computationally produce "coherent" results, may not always produce results that conceptually also make the most sense (Chang et al., 2009), or that are the most helpful given your particular topic of interest. For example, I identified several trained models that combined the topics "clinical trials" and "data management" (see Appendix, Table A.4.3). This might be interesting to a researcher who is interested in identifying leaders in healthcare. Conceptually, however, "data management" is not unequivocally related to "clinical trials"—I saw in the job descriptions that senior managers may be responsible for data management in very different settings as well. Given my research question, it made more sense for me to choose a model in which these topics were being separated. In the end, I thus relied very much on my own personal knowledge and judgements in order to decide whether a model produced topics that were

40 Nowadays, algorithms are being developed that are more sensitive to context, such as word2vec (Mikolov et al., 2015). Yet we still need to understand how these may be useful for understanding research questions like the ones presented in this dissertation, and how "intelligent" or "unintelligent" these may be in practice.

"conceptually coherent" and would help me to answer the research question.

The examples that I have provided in this section illustrate that a whole lot of "human intelligence" was needed in order for me to be able to rely on "algorithmic intelligence". Like me, scholars that rely on algorithmic intelligence may find themselves constantly going back to the data in an effort to make sense of the algorithmic results, thereby potentially changing the data or the algorithm. But this is not a bad thing; it would have been bad if we had not assessed our data or blindly trusted on default packages, methods, and parameters as we implemented the algorithm. Although algorithms are becoming increasingly sophisticated, it is our job as scholars to assume that algorithms are inherently "unintelligent" and be critical about what results they produce, and *why*. In doing so, we should also continuously and critically reflect on our own choices and interpretations that may have played a role.