

Big Data Value Chain: Making Sense of the Challenges

Milla Wirén, Matti Mäntymäki, A. Najmul Islam

▶ **To cite this version:**

Milla Wirén, Matti Mäntymäki, A. Najmul Islam. Big Data Value Chain: Making Sense of the Challenges. 18th Conference on e-Business, e-Services and e-Society (I3E), Sep 2019, Trondheim, Norway. pp.125-137, 10.1007/978-3-030-29374-1_11 . hal-02510149

HAL Id: hal-02510149
https://inria.hal.science/hal-02510149

Submitted on 17 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Big Data Value Chain: Making Sense of the Challenges

Milla Wirén[1], Matti Mäntymäki[1], and A.K.M. Najmul Islam[2]

[1] Turku School of Economics,
[2] Department of Future Technologies
University of Turku, FI-20014, Finland

milla.wiren@utu.fi
matti.mantymaki@utu.fi
najmul.islam@utu.fi

Abstract. In this paper, we highlight how the value of data accumulates through the stages in a value chain. We introduce a Big data value chain where the value adding stages are decoupled from the technological requirements of data processing. We argue that through viewing the stages of value accumulation, it is possible to identify such challenges in dealing with Big Data that cannot be mitigated through technological developments. Our proposed Big Data value chain consists of eight stages that we subsequently cluster into three main phases, namely sourcing, warehousing and analyzing. In scrutinizing these three phases we suggest that the technologically immitigable challenges in sourcing relate to the veracity of data, and the challenges in warehousing concern ownership and power distribution. Finally, in the phase of analyzing the problems are manifold, including the black boxed nature of the algorithms, the problematics of standards of desirability, and the mandatory trade-offs. Our discursive article contributes to the literature discussing the value, utility and implications of Big Data.

Keywords: Big Data, value chain, Big Data value chain, Big Data Challenges, data strategy,

1 Introduction

According to a popular, yet highly controversial saying, data – especially the Big Data – is the new oil. However, while the analogue has its merits in terms of the value potential of Big Data for contemporary businesses, a deeper scrutiny of the analogue reveals certain discrepancies that can be used to explore the value chain and challenges of Big Data. For example, unlike oil, Big Data streams from seemingly unlimited sources, and as such, is to quite an extent continuously renewable. Second, unlike raw oil, raw data has no such consistent constitution, which would always yield value when refined [1) – a big portion of the raw data is merely useless. And thirdly, raw data has not emerged as a result of evolutionary processes guided by the immutable laws of nature, but is a creation of intentional and unintentional human agency, guided by the exactly same haphazardness that accompanies all human activities.

These simple insights lead towards the focal discussion of this article. We map out the value chain of Big Data and identify the key challenges associated with each stage of the value chain. We specifically focus on challenges that are particularly difficult to overcome with solely technological means. For example, the increasing sophistication of data processing algorithms notwithstanding, datafication of the entities from the physical, and particularly from the subjective realms is prone to various errors and inaccuracies. Addressing these challenges essentially requires human judgement along the process of reaping the benefits of Big Data.

Far from being an unanimously defined concept [2], there is however an understanding of what types of contents the label 'Big Data' contains. The constitution of Big Data includes not only the traditional type of alphanumerical and relatively homogenous precategorized data found in institutional databases, but also the transsemiotic (images, sounds, scents, movements, digital action tokens, temperature, humidity to name a few) and highly heterogeneous data that is not categorized prior to its harvesting [3–7]. This latter type of data is sourced through sensor technology, from the intentional and unintentional interactions between humans and machines, from the surveillance systems, and from the automated digital transaction traces [2,8–14]. As a result, of the increasing prowess of sourcing the data (i.e. datafication) and the equally increasing and cheapening computational capacity, the accumulation of data is extremely rapid, also resulting in the continuous change in its constitution. In short, Big Data is a constantly growing and changing nebulous and amorphous mass.

However, while the mass of existing Big Data is impossible to delineate, viewing the phenomenon from the perspective of its anticipated utility reveals a value chain spanning from datafication to potential business value creation. In this article, we identify eight stages of the value chain, discussed in more detail later, but named here as datafication, digitizing, connectivity, storage, categorizing, patterning, cross-analysis and personalization. Indeed, there are other proposals for the Big Data value chain [15], however, most of the existing value chain proposals delineate diverse technological stages required in rendering data useful. In this paper, we delineate the stages according to the increase in the value of the data, meaning that some of the stages include several technologies, and some technologies span more than one stage. In short, we decouple the technological requirements of processing data from the value adding activities in refining data.

The research question of this paper is twofold: i) what types of challenges exist in different stages of the Big Data value chain, and ii) which of these challenges are particularly difficult to mitigate without human-based intelligence and judgement? In order to explore these questions, the remainder of the article flows through first delving the eight stages and clustering them into three main phases to identify the relevant accompanying challenges, towards the conclusion listing the contributions, limitations and future research possibilities.

2 Big data value chain

The process of obtaining insights from Big Data can be divided into three main stages, namely sourcing, warehousing, and analyzing data [16–18]. These stages have been rearranged and complemented, for example resulting in the stages of data acquisition, data analysis, data curation, data storage, and data usage [15]. These typologies are delineated from the perspective of technology, through clustering the stages along the technological requirements, a useful approach when the focus is on the side of technological developments needed for realizing the potential value of Big Data.

However, if we shift the focus beyond technology and zoom in to the actual value adding processes, the technology-driven boundaries do not match the boundaries between the stages of value add. Therefore, we propose another conceptualization of the Big Data value chain, where each of the stages differs from its neighbors in terms of its value accumulation potential. Our approach further divides the established three main stages and consists of the stages of datafication, digitizing, connectivity, storage, categorizing, patterning, cross-analysis and personalization, introduced next.

Datafication. The emergence of the phenomenon of Big Data is underpinned by the developments in technologies that enable datafying different types of entities. For example, the developments in the sensor technology have enabled producing data about movements, humidity, location, sounds, composition or smell to name a few [1]. On the other hand, the diffusion of digital devices and the accompanying increase in human-computer interaction is making it possible to deduce and produce data about the subjective preferences of the individuals, based on the traces of these interactions [2,19–22]. These developments in the technologies enabling myriad forms of datafication are the core source of Big Data, and the first fundamental building block of the Big Data value chain.

Digitizing. As data is created from entities of a wide variety of ontological natures, capturing the data through analog technologies resulted in various data types, each requiring their own processing technologies. When all data is digitized, made into binary digits of zeroes and ones, in theory any machine capable of processing bits could process the data [23,24] though in practice this is not yet the case. In terms of Big Data, the major value creating step is the homogenization of the data from the diverse sources, because it creates the foundation for cross-analyzing and -referencing data sets originating from very diverse sources.

Connectivity. Even if we had homogenous data from a variety of phenomena, without the capability to connect that data, each individual parcel of data would be relatively useless. However, with the emergence of the TCP/IP protocol and internet mandating how the data should be packaged, addressed, transmitted, routed and received, and the developments in the communications technologies, the uniform data from a diversity of sources can be transmitted somewhere to be pooled and accessed. The technologies enabling this are multiple and continuously developing. However, the value add transcends the technologies: while there are imperfections in the complex communications technologies and even in the very design of the internet [25], the idea of connectivity is a major value add by itself.

Storage. Even though digital data exists only in the form of zeroes and ones, it is not without a physical representation – quite the contrary, as storing the masses of data require hardware that enable accessing and processing the pooled homogenous data. The developments in the computational power and the cheapening of the storage capacity is critical for this value adding stage. However, the value itself emerges from the existence of these pools of data, which enables processing together data from the variety of sources. The technological solutions of data centers, data warehouses and data lakes are complex, however as with the stage of connectivity, the value add emerges from the mere possibility of having the data mass stored in pools fed and accessed from diverse points of entry.[15]

Categorizing. Unlike data in the traditional data sources, one of the defining features of Big Data is its automated and autonomous accumulation, which in other words means that the data is not categorized on the way in [7]. Instead, any sensemaking of the vast data masses must begin – or at least be guided by – designing mechanisms and principles based on which the data can be categorized. This value adding stage of categorizing is the stage where the end use of the data needs to be accounted for, because of the generative nature of data [26]: not only can the data be categorized in many ways, but the same data may yield different utility depending on the context of its use [27]. This is the stage where the algorithms are essential. Due to the volume, variety, and velocity of data, human computational capabilities are insufficient for processing. Therefore, algorithmic processing capabilities are needed. [3,10,28–30]

Patterning. The importance of the algorithms increases towards the end of the value chain. At the stage of patterning, the task is to identify patterns from the categorized masses of data. The patterns constitute the first stage in the value chain possessing to identify business value potential. Due to the volume and variety of the data, it is possible to identify patterns that may be invisible in smaller (in scope or quantity) data sets [31]. As an example, the customer behavior from CRM systems can be patterned to better understand the behavior of certain customer group.

Cross-analysis. Even more valuable than identifying novel patterns, is the ability to cross-analyze diverse patterns to seek correlations – for example through cross-analysing the data patterns of customer behavior against the patterns from marketing campaigns. Most of the current data use cases are grounded on this stage of data utility [13], which excels in creating generalized knowledge about a wide variety of phenomena. For example, through cross-analyzing the data from traffic accidents and driver demography, it is possible to find correlations between the age and gender of the drivers and accidents. This value adding stage also enables increasingly efficient customer segmentation for example in social media marketing. If certain preferences and demographic features are correlated, an offering can be marketed to the exact demography.

Personalization. The most value potential of Big Data is embedded in the final stage, namely personalization, which means that the data can be used for behavioral predictions [32,33]. This value adding capability is built on first having such cross-analyzed patterns that reveal correlations, and then analyzing the behavioral datafied history of an individual against those correlations [2,20]. Continuing the example in the previous stage, here the increase in value emerges from the possibility of harvesting the infor-

mation of the driving behavior of an individual, and cross-analyzing that personal history with such generalized driving behavior patterns that correlate with an increase in accidents. Also, in terms of targeted marketing, at this stage it is possible to deduce the preferences on the level of the individual, based on the traces left in human-device interactions, and to personalize the offerings accordingly [19].

3 Challenges in the big data value chain

As our focus is on identifying such challenges that are particularly difficult to overcome through technological developments alone, we will not offer a comprehensive view on the current state-of-the-art in any of the underlying technologies. In other words, the boundary between what can and cannot be solved through technological developments is blurred and bound to a specific point of time.

The next subchapters cluster the aforementioned value chain into three phases familiar in other data value chain approaches, the first of which we refer to as sourcing to encompass datafication and digitizing, to be followed with warehousing consisting of connectivity and storage, and finally analyzing, covering the stages from categorization and patterning to cross-analyzing and personalization. Table 1 summarizes the discussion.

Stage in the big data value chain	Cluster	Challenges	
Datafication	Sourcing	Veracity	
Digitizing			
Connectivity	Warehousing	Ownership, power distribution	
Storage			
Categorizing	Analyzing	Black boxes, standards of desirability	
Patterning			
Cross-analysis		Trade-offs	Privacy vs personalization
			Convenience vs independence
Personalization			Collective safety vs individual freedom
			Data security vs machine learning optimization
			Ease of outcomes vs validity of process

Table 1: Challenges in the Big Data value chain

3.1 Challenges in sourcing: veracity

The validity of the end results of any data refining processes is dependent on the validity of the source data, and unlike with oil, the quality of the raw data varies vastly [1]. First, in digitizing existing data sets, the imperfections, biases, unintended gaps and intentional results of human curation of data also end up in the mass of Big Data [34]. In

other words, such data that originates from times preceding the digital age, the acts of datafying and storing that data required a lot of work, which means that only a small part of all relevant data ended up in a form that yields itself to digitizing.

In a more limited scale, this applies also to such data that does exist in the traditional databases. Due to the costs embedded in storing data with pre-digital means, the databases hold pre-prioritized, pre-categorized data that someone has chosen at some stage to store. This means that such queries that require accounting for historical data can never be quite accurate, because most of what has been gone without a retrospectively datafiable trace, and the rest is already curated and thus subjected to human biases and heuristics [9]. This problem is referred to as veracity [35], which means coping with the biases, doubts, imprecision, fabrications, messiness and misplaced evidence in the data. As a result, the aim of measuring veracity is to evaluate the accuracy of data and its potential use for analysis [36].

However, the older databases are only one source of data, and the issue of veracity is not limited to it. Another source, the interactions between humans and digital devices [for example in using the mobile phones, browsing the internet or engaging in social media) is a vast torrent of data. Only Facebook generates more than 500 terabytes of data per day. The sheer volume and variability of such data presents its own problems in terms of technological requirements; however, the veracity of that data is even more problematic. As illustrated by Sivarajah, Irani, & Weerakkody [37] data from human-to-human online social interaction is essentially heterogenous and unclear in nature. Furthermore, malicious tools or codes can be used to continuously click on the performance-based ads creating fake data. There are bots that create traces mimicking human behavior. In addition, individuals vary in the level of truthfulness of their traceable activities. Part of the problems of veracity stem from the intentional human actions, which can be biased, misleading, and overall random. Furthermore, the samples of population participating in online interactions is skewed – not to even mention the geographical discrepancies emerging from the varying levels of technological penetration around the globe.[1]

In turn, the data from the sensor technologies, surveillance systems and digital transaction traces does not suffer from similar biases such as intentional human actions. However, these sources have their own veracity issues. As the processes are automated, and not a priori prioritized, a portion of that data is irrelevant and useless. This challenge relates to cleaning data, i.e. extracting useful data from a collected pool of unstructured data. Proponents of Big Data analytics highlight that particularly developing more efficient and sophisticated approaches to mine and clean data can significantly contribute to the potential impact and ultimately value that can be created through utilizing big data [3]. At the same time, however, developing the tools and methods to extract meaningful data is considered an ongoing challenge [27].

The problem of veracity is partially mitigable through technological developments. With the development of data cleaning and mining technologies, it becomes easier to filter the vast data masses to extract the valuable nuggets. However, the inaccuracies resulting from the imperfect older data sources and the haphazardness of human action are fundamentally immitigable.

3.2 Challenges in warehousing: ownership and power

The primary problems in connectivity and storage relate to technologies enabling the transmission, storing and accessing the vast data masses. However, not all problems are even in these stages solvable through technological progress. Scrutinizing data security highlights the issue, as it is only partially a technological question.

As argued by Krishnamurthy and Desouza [38], companies and organizations are facing challenges in managing privacy issues, thus hindering organizations in moving forward in their efforts towards leveraging big data. For example, smart cities, where collected data from sensors about people's activities can be accessed by various governmental and non-governmental actors [39]. Furthermore, the distributed nature of big data leads to specific challenges in terms of intrusion [40] and thus may lead to challenges to various threats such as attacks [41] and malware [42].

However, underpinning these data security capabilities is the question of data ownership. As Zuboff (2015) notes, one of the predominant features in digitalization is the lack of the possibility to not to opt-in as a data source – as the everyday life of an individual is embedded in the invisible digital infrastructures [23,43,44] the individuals have little control over the data exhaust being created about them.

The boundaries of ownership and control rights are a serious problem that surpasses the technological problems of ensuring the data security of specific data sets or applications. The issue of ownership is ultimately a question of data driven power distribution: the agents possessing not only the data sourcing capabilities but also the data sharing resources, are harnessing power to not only create business value but to have also socio-political influence. [45]

3.3 Challenges in analyzing: black boxes, standards of desirability and tradeoffs

There are two major aspects to algorithms that pave the way for discussing the challenges, here dubbed as standards of desirability and black boxes. Firstly, the algorithms cannot come up with priorities or questions by themselves, but humans are needed to provide them, and secondly, as we need the algorithms because the human computational capacity cannot deal with the vast masses of digital data, the human computational capacity cannot follow the algorithmic processes dealing with those masses of data.

To begin with the problem of standards of desirability [46,47], any question or task given to the algorithm to be processed must be underpinned by a set of goals that are to be reached. However, at any given moment of designing a goal there is no way of knowing whether that goal is still relevant or preferable at the time of reaching that goal: both the environmental circumstances and the internal preferences can have undergone changes. Especially, considering the generative nature of digital data [26] and the digital affordances [48], the data itself does not mandate a specific use or specific questions; instead the utility of the versatile data is dependent of the contextual fit and quality of the questions guiding the analyzing processes at any given time. However, as we know from both history and human focused research, we humans are far from

infallible – the quality of questions, or the relevance of them is never guaranteed [49–53], which means that this challenge is immitigable through technological advances.

Secondly, as we cannot follow the algorithmic computational processes, we cannot detect if there are errors in the processes [34,54] – ultimately the processes are black boxes. Furthermore, the black boxed nature of the outcomes does not end at the revelation of the outcomes of algorithmic processes: the outcomes of patterning and cross-analysis reveal correlations, and due to the sheer volume of the data masses there are high possibilities to find significant correlations between any pair of variables. This means that the identified correlations can be mere noise, unless supported by theoretical mechanisms (1]: unlike Anderson [55], McAfee et al. [9] claim, the scientists are not rendered obsolete, however the focal usefulness of scholars shifts from hunting correlations to understanding the underpinning theoretical causalities. The black boxes of algorithms reveal the black boxes of correlations, leaving it to humans to assess the relevance and validity of the outcomes.

Thirdly, the algorithmic analysis of Big Data creates such opportunities that require scrutinizing the accompanying tradeoffs. Newell and Marabelli [28] identify three, and next we introduce five. The tradeoffs next discussed are privacy-personalization, convenience-independence, collective security-individual freedom, data security-machine learning optimization, and ease of outcomes-validity of process.

Privacy vs personalization. Reaping the benefits from the ultimate stage of the Big data value chain through providing personalized offerings means that the agent making the offerings has to have access to personal data – in other words, has to breach the privacy of the targeted customer to an extent [56,57]. The ethical valence of this tradeoff needs to be considered contextually, meaning that there are both cases where the loss of privacy is easily offset by the benefits resulting from accessing the personalized service, and cases where the value of the personalized offering does not justify the breach to privacy [13,58–60].

Convenience vs independence. The more convenient it is to rely on a specific technology, for example the navigation devices and systems in cars and vessels, the more dependent on that technology one typically becomes. In a corporate use of Big Data and analytics, the widespread utilization of game analytics has even led into a situation where not using the technologies in decision-making is being referred to as "flying blind" and the use of analytics is considered as a necessity by game developers [61]. Taken together, there is an evident tradeoff between convenience and independence that requires acknowledging.

Collective safety vs individual freedom. Newell and Marabelli [28] share a case where Facebook was accused for not reacting to a threatening post by a person who carried out the threat and shot an individual. This example highlights this tradeoff amply: it is possible with the help of Big Data and behavioural prediction to anticipate threats to collective security, however reacting on those threats a priori of incidents limits the freedom of the individual being detained before having committed anything.

Data security vs machine learning optimization. The different policies of EU and China in terms of access to data in developing artificial intelligence highlight this issue nicely. In Europe the priority is to protect the privacy and data security, which means that there is less data available for developing machine learning [62]. This results on

the one hand the improved rights of the European citizens and on the other hand slower progress in developing artificial intelligence. In turn, China is investing heavily in developing artificial intelligence, for example facial recognition technologies through utilizing all available data from the ubiquitous mobile applications that billions of Chinese people use daily [63–65], resulting in less individual level data privacy, but competitiveness in the AI race.

Ease of outcomes vs validity of process. Traditionally, the value of accounting information has resided in the transparent and accessible processes through which the financial information has been gathered and processed. The credibility of the ensuing financial figures has been built on the validity of these processes. However, with the increasing use of the automated accounting systems and algorithms, the outcomes are achieved faster, however through the black boxes of algorithms – the validity of the process of creating the end results is no longer visible [54]. This is again a choice for the humans: when and why does the swift outcome have more value, and when and why is it mandatory to be able to observe the processes?

4 Conclusion

This study was set out to explore i) what types of challenges exist in different stages of the Big Data value chain, and ii) which of these challenges are particularly difficult to mitigate without human-based intelligence?

By addressing this research question: our paper adds on the literature on the roles and interplay between algorithmic and human-based intelligence in reaping the benefits and business value from Big Data [66] with two specific contributions. First, we present a Big Data value chain decoupled from the technological underpinnings and grounded on the stages of value accumulation, and secondly, we highlight a set of such challenges in utilizing Big Data that cannot be mitigated through technological developments.

We advance the understanding of the value and utility of data by putting forward a Big Data value chain that consists of eight stages in three clustered phases: sourcing (datafication and digitizing), warehousing (connectivity and storage) and analyzing (categorizing, patterning, cross-analyzing and personalization). In the first cluster, the immitigable challenges reflect the problems of veracity whereas in the second one, the problems relate to ownership and power distribution. Finally, in the third cluster, the issues include the black-boxed nature of the algorithms and implications thereof, the need for standards of desirability, and five tradeoffs that require acknowledging.

By elaborating on these tradeoffs related to the Big Data value chain, our study adds on the prior discussions related to data and privacy [2,56,57,59,60], strategic utility of data [3,4,10,12] reliability of data [1,9] and data ownership [45] .

Like any other piece of research, this study suffers from a number of limitations that in turn call for additional research. First, due to its conceptual nature, empirical scrutiny of our Big Data value chain is a self-evident area for future research. Second, since the value chain can manifest itself differently across different contexts and under different contingencies, future research focusing on contextual aspects of Big Data value chain would be highly insightful [67]. Third, since the stages of the Big Data value chain

often consist activities undertaken various actors, it is relevant to consider how does trust manifest itself among these actors [68], what kind of business ecosystems and networks emerge for the utilization of big data, and how different actors strategize their utilization of big data [69]

5 References

1.J. Condliffe, China Turns Big Data into Big Brother.; Retrieved 2017, from https://www.technologyreview.com/s/602987/china-turns-big-data-into-big-brother/

2.M. A. Vasarhelyi, A. Kogan and B. M. Tuttle, Big Data in accounting: An overview, Accounting Horizons **29** (2015) 381–396.

3.Y. Wang and M. Kosinski, Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. (2017).

4.H. R. Varian, Computer mediated transactions, American Economic Review **100** (2010) 1–10.

5.T. Davenport and J. Harris, *Competing on Analytics: Updated, with a New Introduction: The New Science of Winning*. Harvard Business Press.

6.P. Barnaghi, A. Sheth and C. Henson, From data to actionable knowledge: big data challenges in the web of things, IEEE Intelligent Systems (2013) 6–11.

7.J. L. Casti, *X-events: the collapse of everything*. Harper Collins.

8.J. G. March, Bounded rationality, ambiguity, and the engineering of choice, The Bell Journal of Economics (1978) 587–608.

9.A. McAfee, E. Brynjolfsson, T. H. Davenport, D. J. Patil and D. Barton, Big data: the management revolution, Harvard business review **90** (2012) 60–68.

10.D. J. Teece, Research directions for knowledge management, California management review **40** (1998) 289–292.

11.S. Zuboff, Big other: surveillance capitalism and the prospects of an information civilization, Journal of Information Technology **30** (2015) 75–89.

12.J. D. Thompson, *Organizations in Action: Social Science Bases of Administrative Theory*. Transaction Publishers.

13.M. Kosinski, D. Stillwell and T. Graepel, Private traits and attributes are predictable from digital records of human behavior, Proceedings of the National Academy of Sciences of the United States of America **110** (2013) 5802–5805.; DOI:10.1073/pnas.1218772110 [doi]

14.U. Sivarajah, Z. Irani and V. Weerakkody, Evaluating the use and impact of Web 2.0 technologies in local government, Government Information Quarterly **32** (2015) 473–487.

15.B. Macaes, Europe's AI delusion. Politico.; Retrieved 2018, from https://www.politico.eu/article/opinion-europes-ai-delusion/

16.X. Yi, F. Liu, J. Liu and H. Jin, Building a network highway for big data: architecture and challenges, IEEE Network **28** (2014) 5–13.

17.T. Peterson, Facebook will target ads to people based on store visits, offline purchases, calls to businesses. Marketing land.; Retrieved 2018, from https://marketingland.com/facebook-will-target-ads-people-based-store-visits-offline-purchases-calls-businesses-224668

18.J. Van Dijck, Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology, Surveillance & Society **12** (2014) 197–208.

19.R. Krishnamurthy and K. C. Desouza, Big data analytics: the case of the social security administration, Information Polity **19** (2014) 165–178.

20.D. Weinberger, Optimization over Explanation. Berkman Klein Center.; Retrieved 2018, from https://medium.com/berkman-klein-center/optimization-over-explanation-41ecb135763d

21.A. A. Cárdenas, P. K. Manadhata and S. P. Rajan, Big data analytics for security, IEEE Security & Privacy **11** (2013) 74–76.

22.J. Kallinikos and I. D. Constantiou, Big data revisited: a rejoinder, Journal of Information Technology **30** (2015) 70–74.

23.D. Lyon, *Surveillance after Snowden*. John Wiley & Sons.

24.S. Akter, S. F. Wamba, A. Gunasekaran, R. Dubey and S. J. Childe, How to improve firm performance using big data analytics capability and business strategy alignment?, International Journal of Production Economics **182** (2016) 113–131.

25.A. Labrinidis and H. V. Jagadish, Challenges and opportunities with big data, Proceedings of the VLDB Endowment **5** (2012) 2032–2033.

26.R. Abbas, K. Michael and M. G. Michael, The regulatory considerations and ethical dilemmas of location-based services (LBS) A literature review, Information Technology & People **27** (2014) 2–20.

27.J. Sutanto, E. Palme, C.-H. Tan and C. W. Phang, Addressing the personalization-privacy paradox: an empirical assessment from a field experiment on smartphone users, Mis Quarterly (2013) 1141–1164.

28.S. Newell and M. Marabelli, Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification,' The Journal of Strategic Information Systems **24** (2015) 3–14.

29.T. Dinev, P. Hart and M. R. Mullen, Internet privacy concerns and beliefs about government surveillance–An empirical investigation, The Journal of Strategic Information Systems **17** (2008) 214–233.

30.H. A. Simon, Administrative behavior; a study of decision-making processes in administrative organization. (1947).

31.X. Wu, X. Zhu, G.-Q. Wu and W. Ding, Data mining with big data, IEEE Transactions on Knowledge and Data Engineering **26** (2014) 97–107.

32.T. Shah, F. Rabhi and P. Ray, Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions, Cluster Computing **18** (2015) 351–367.

33.H. J. Smith, T. Dinev and H. Xu, Information privacy research: an interdisciplinary review, MIS quarterly **35** (2011) 989–1016.

34.M. Mäntymäki, S. Hyrynsalmi and A. Koskenvoima, How do small and medium-sized game companies use analytics? An attention-based view of game analytics, Information Systems Frontiers (In press).

35.J. Martines, Forget Big Data-Little Data is Making Learning Personal
, WIRED.; Retrieved 2019, from https://www.wired.com/2016/11/forget-big-data-little-data-making-learning-personal/

36.R. V. Zicari, Big data: Challenges and opportunities, Big data computing **564** (2014).

37.Y. Yoo, Computing in everyday life: A call for research on experiential computing, MIS quarterly (2010) 213–231.

38.H. Chen, R. H. Chiang and V. C. Storey, Business intelligence and analytics: from big data to big impact, MIS quarterly (2012) 1165–1188.

39.N. Brunsson, The irrational organization (1985).

40.P. Quattrone, Management accounting goes digital: Will the move make it wiser?, Management Accounting Research **31** (2016) 118–122.

41.D. Tilson, K. Lyytinen and C. Sorensen, Desperately seeking the infrastructure in IS research: Conceptualization of" digital convergence" as co-evolution of social and technical infrastructures (pp. 1–10). Presented at the System Sciences (HICSS), 2010 43rd Hawaii International Conference on, IEEE.

42.Z. Huang, Y. Lei and S. Shen, China's personal credit reporting system in the internet finance era: challenges and opportunities, China Economic Journal **9** (2016) 288–303.

43.D. Boyd and K. Crawford, Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon, Information, communication & society **15** (2012) 662–679.

44.T. H. Davenport, P. Barth and R. Bean, How big data is different, MIT Sloan Management Review **54** (2012) 43.

45.D. Kahneman, *Thinking, fast and slow*. Macmillan.

46.A. E. Marwick and D. Boyd, Networked privacy: How teenagers negotiate context in social media, New Media & Society **16** (2014) 1051–1067.

47.S.-H. Liao, C.-C. Wu, D.-C. Hu and K.-A. Tsui, Relationships between knowledge acquisition, absorptive capacity and innovation capability: an empirical study on Taiwan's financial and manufacturing industries, Journal of Information Science **36** (2010) 19–35.

48.M. Mäntymäki, *Does e-government trust in e-commerce when investigating trust? A review of trust literature in e-commerce and e-government domains*, Towards Sustainable Society on Ubiquitous Networks (pp. 253–264). Springer.

49.C.Anderson, The end of theory: The data deluge makes the scientific method obsolete, Wired magazine **16** (2008) 16–07.

50.M. Boncheck, Little Data Makes Big Data More Powerful. Harvard Business Review.; Retrieved 2018, from https://hbr.org/2013/05/little-data-makes-big-data-mor

51.J. Kallinikos, A. Aaltonen and A. Marton, The Ambivalent Ontology of Digital Artifacts., Mis Quarterly **37** (2013).

52.M. Wirén and M. Mäntymäki, Strategic Positioning in Big Data Utilization: Towards a Conceptual Framework (pp. 117–128). Presented at the Conference on e-Business, e-Services and e-Society, Springer.

53.J. H. Abawajy, A. Kelarev and M. Chowdhury, Large iterative multitier ensemble classifiers for security of big data, IEEE Transactions on Emerging Topics in Computing **2** (2014) 352–363.

54.I. D. Constantiou and J. Kallinikos, New games, new rules: big data and the changing context of strategy, Journal of Information Technology **30** (2015) 44–57.

55.Z. Aldama, China's big brother: how artificial intelligence is catching criminals and advancing health care. Post Magazine.; Retrieved 2017, from http://www.scmp.com/magazines/post-magazine/long-reads/article/2123415/doctor-border-guard-policeman-artificial

56.E. Curry, *The big data value chain: definitions, concepts, and theoretical approaches*, New horizons for a data-driven economy (pp. 29–37). Springer, Cham.

57.W. A. Günther, M. H. R. Mehrizi, M. Huysman and F. Feldberg, Debating big data: A literature review on realizing value from big data, The Journal of Strategic Information Systems **26** (2017) 191–209.

58.S. Kaisler, F. Armour, J. A. Espinosa and W. Money, Big data: Issues and challenges moving forward (pp. 995–1004). Presented at the 2013 46th Hawaii International Conference on System Sciences, IEEE.

59.J. L. Zittrain, The generative internet, Harvard law review (2006) 1974–2040.

60.D. Weinberger, *Everything is miscellaneous: The power of the new digital disorder*. Macmillan.

61.J. G. March, Rationality, foolishness, and adaptive intelligence, Strategic Management Journal **27** (2006) 201–214.

62.N. F. Awad and M. S. Krishnan, The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization, MIS quarterly (2006) 13–28.

63.E. Autio, S. Nambisan, L. D. Thomas and M. Wright, Digital affordances, spatial affordances, and the genesis of entrepreneurial ecosystems, Strategic Entrepreneurship Journal **12** (2018) 72–95.

64.K. Lyytinen and Y. Yoo, Ubiquitous computing, Communications of the ACM **45** (2002) 63–96.

65.E. Barbierato, M. Gribaudo and M. Iacono, Performance evaluation of NoSQL big-data applications using multi-formalism models, Future Generation Computer Systems **37** (2014) 345–353.

66.D. Tilson, K. Lyytinen and C. Sørensen, Research commentary—Digital infrastructures: The missing IS research agenda, Information systems research **21** (2010) 748–759.

67. P. Mikalef, I. O. Pappas, J. Krogstie and M. Giannakos (2018). Big data analytics capabilities: a systematic literature review and research agenda. Information Systems and e-Business Management, **16**(3), (2018), 547-578.

68..J. G. March, Theories of choice and making decisions, Society **20** (1982) 29–39.

69. .C. Cadwallar and E. Graham-Harrison, How Cambridge Analytica turned Facebook 'likes' into a lucrative political tool. The Guardian.; Retrieved 2018, from https://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm?CMP=share_btn_tw