Determining Potential Failures and Challenges in Data Driven Endeavors: A Real World Case Study Analysis

Daniel Staegemann, Matthias Volk, Tuan Vu, Sascha Bosse, Robert Häusler, Abdulrahman Nahhas, Matthias Pohl and Klaus Turowski
Magdeburg Research and Competence Cluster Very Large Business Applications,
Faculty of Computer Science, Otto-von-Guericke University Magdeburg, Magdeburg, Germany

Keywords: Data Driven, Big Data, Case Study, Failures, Analysis, Categorization.

Abstract: The utilization of data in general and big data in particular offers large opportunities, but is at the same time accompanied by a huge number of potential causes for failure. To avoid those pitfalls when realizing such undertakings, at the beginning, it is necessary to develop an in-depth understanding of those causes. This contribution analyses twelve real world case studies, from the big data and related domains, which were facing issues. The causes for the experienced problems were extracted and thereupon categorized, facilitating the understanding of practitioners and researchers that are engaged in the big data domain. Furthermore, potential avenues for future research are highlighted.

1 INTRODUCTION

With the growing amount (Yin and Kaynak 2015; Dobre and Xhafa 2014) and complexity (Yang et al. 2017) of data produced by humanity and the increasing expectations regarding its utilization (Jin et al. 2015), traditional technologies and approaches are often overstrained. For this reason, big data projects are becoming a promising solution for those challenges. While the term big data itself has no single, universally utilized explanation (Hartmann et al. 2016), the definition provided by the National Institute of Standards and Technology (NIST) is widely accepted. According to that definition, big data "consists of extensive datasets primarily in the characteristics of volume, velocity, variety, and/or variability that require a scalable architecture for efficient storage, manipulation, and analysis" (NIST 2019). The application areas of those technologies are multifarious. Examples comprise, but are not limited to, the construction industry (Bilal et al. 2016), procurement (Staegemann et al. 2019c), tourism (Gajdošík 2019), urban transportation management (Fiore et al. 2019), civil protection (Wu and Cui 2018) and weather data analysis (Onal et al. 2017). However, even though the potentials of big data are manifold and high (Müller et al. 2018; Bughin 2016; Maroufkhani et al. 2019; Alharthi et al. 2017), the same applies for the challenges and risks (Alharthi et al. 2017; Philip Chen and Zhang 2014; Staegemann et al. 2019b; Wenzel and van Quaquebeke 2018; Staegemann et al. 2019a; Volk et al. 2019). Hence, to ensure the best results, it is necessary to have a deep understanding of those challenges, to avoid potential pitfalls, when implementing projects. For this reason, the following research question will be explored in the course of this work:

What are major factors contributing to failure in big data endeavors?

To answer that research question, an analysis of twelve real world cases from the big data or conjoined domains, with negative outcomes, is conducted. The focus goes beyond pure big data applications, since failures are not as commonly publicized as success stories, which limits the available resources. To provide value, and offer comprehensive insights into the domain, it however appears reasonable, to cover a variety of settings and objectives. Furthermore, big data projects are a subset of data driven endeavors (Günther et al. 2017). Therefore, it appears reasonable, to slightly extend the scope to also include those. Hence, the work is structured as follows. After providing an introduction, the second section describes the respective cases. Afterward, the analysis and its findings are presented. Finally, a conclusion is drawn, which also includes the contemplation of limitations and future perspectives.

Staegemann, D., Volk, M., Vu, T., Bosse, S., Häusler, R., Nahhas, A., Pohl, M. and Turowski, K.
Determining Potential Failures and Challenges in Data Driven Endeavors: A Real World Case Study Analysis.
DOI: 10.5220/0009792504530460
In *Proceedings of the 5th International Conference on Internet of Things, Big Data and Security (IoTBDS 2020)*, pages 453-460
ISBN: 978-989-758-426-8

2 CASE STUDIES

To increase the success rate of future big data analytics projects, it is at first necessary to understand the common factors leading to failures in big data endeavors. For this purpose, the analysis of examples from the past can provide valuable insights. Therefore, in the following, twelve real world cases will be analysed under this aspect. Due to the wide range of possible data driven projects, the cases were chosen from heterogeneous contexts, regarding application area, intent, occurred issues, and financial background, facilitating a better insight into the topic as a whole (Khan and van Wynsberghe 2008).

2.1 Kmart

The retail company Kmart found itself in a price competition with Walmart in the mid-to-late 1990s. However, in contrast to its competitor, it neglected technical advancements and therefore had no functioning "just-in-time" inventory management. As a result, customers were often faced with out of stock items, resulting in poor shopping experiences (24/7 Wall St. 2012; Turner 2003). Furthermore, Kmart did also not decide on a true identity and instead tried to appeal to everyone. This approach, however, makes it hard to define clear goals, which are a prerequisite for a purposeful analysis and the formulation of strategies (Marr 2017, 21 f.). In contrast, the competitors, Walmart and Target, defined their own niche (Leinwand and Mainardi 2010). While Kmart eventually at least acknowledged the need for technical support in planning and invested in a business intelligence platform in 2003 (BusinessWire 2003), the damage had already been done. As a result, a major decline in business success occurred, which led to a merger with the department store chain Sears in 2005 (24/7 Wall St. 2012; Egan 2015).

2.2 UK National Health Service

In 2002, the UK National Health Service launched a project to achieve a top-down digitization in the healthcare system in England, implementing new technologies and IT systems. The idea was to incorporate the plethora of different applications and databases to create one integrated solution. While the initial budget and duration were estimated to be £6.2 billion and ten years, the lackluster inclusion of health care end-users, professionals and facilities, project management issues, technical difficulties, and changing specifications resulted in major problems. In September 2011 it was dismantled after spending about £9.8 billion without achieving the desired results (Justinia 2017; Syal 2013; Hefford 2011).

2.3 Solid Gold Bomb T-shirt Company

To increase the variety of offered T-shirts, the startup wrote a software that autonomously created designs by combining the phrase "Keep Calm and", which was coined by the British as "Keep Calm and Carry On" during the second world war, with another verb and pronoun that were randomly taken from a prepared list. Subsequently, those shirts were automatically offered on Amazon, without further review by the employees. Since the list with the words that were to be used was not carefully curated and checked in advance, this procedure resulted in slogans like "Keep Calm and Rape Her" or "Keep Calm and Hit Her", leading to a public outrage and the shutdown of the shop's Amazon account, which was also its main distribution channel, resulting in a severe reduction of sales for the company (Pagliery 2013; McVeigh 2013).

2.4 Target

New parenthood is one of the occasions, where retailers have the highest chance of changing customer's shopping behavior. Therefore, in 2002, Target started developing a pregnancy-prediction system to attract potential soon-to-be parents by sending out corresponding coupons before the competition even knew about the pregnancy. While sales significantly increased, the specific coupons also resulted in disclosing pregnancies to third persons. For instance, a pregnant teenager's father was informed of the pregnancy due to the received ads, which were promoting baby-related items. This kind of unsolicited notification not only violates the privacy of the expectant mother, but can potentially also result in negative consequences for her. As a result, Target started mixing customer-specific promotions with randomly selected offers, reducing the assignability of the advertisements with life circumstances while still presenting a selection of highly relevant items (Duhigg 2012; Albert 2015).

2.5 Pinterest

In contrast to Target's contentwise correct but intrusive analysis of its customer's life circumstances, in 2014 the social media company Pinterest falsely congratulated a share of its community on getting married. While weddings and the corresponding accessories constitute a significant part of the

website's presented content, many of the recipients were not planning to get married or at least in a relationship and in some cases, there was not even a clear connection between the consumed contents and marriage at all. As a result, some of the falsely addressed users complained on twitter, attracting attention by further media. Later on, Pinterest apologized, stating that they wanted to reach people who are interested in wedding-related content and the wording, suggesting the imminence of an actual wedding, was a mistake (Kosoff 2014; Roy 2014).

2.6 Google Flu Trends

Google Flu Trends (GFT) was a system that analysed the frequency of flu-related search terms input into google to predict flu outbreaks. To do so, about 50 million search terms were mapped to 1152 data points (Ginsberg et al. 2009) which were identified as flu indicators by the U.S. Centers for Disease Control and Prevention. While there might be a connection between those searches and actual cases of the flu, the system completely missed the H1N1 flu and often overestimated flu rates by high margins. Furthermore, GFT lacked the transparency for outstanding persons to properly evaluate the results, turning it into a black box and preventing the submission of concepts for possible improvements to the algorithms. In 2013, GFT's public website was shut down. However, the concept and the data itself are still being used. For instance, researchers continue working on models and concepts to predict flue occurrences ahead of time (Pappas 2014; Lazer et al. 2014; Comstock 2015; Pervaiz et al. 2012).

2.7 Ferrari

In 2010 Fernando Alonso lost the almost certain Formula 1 championship during the last race of the season. To win the title by himself, independent from his competitors' performance, a fourth place would have sufficed. As it is common, he was using a race strategy that was chosen by his chief race strategist. The process was supported by a decision support system (DSS). By design, the DSS offered only two options to pick. According to his team's regulation, the strategist was required to choose one of those, which implies disregarding all other possibilities. However, in contrast to the other races of the season, the DSS could not build upon previous experiences with the course, because it was its first appearance in the circuit. Furthermore, as another limitation of the system, insights from the ongoing race were not included. In theory, the strategist could have ignored that rule, ordered a different strategy and potentially secured the title. Yet, he would have taken a personal risk by violating the policy. In case of a failure and with a prevailing blame-culture in the team, he would have been in a hard-to-justify position. As a result, he abided the rules and even though he chose the better of the two proposed options, it was inappropriate for the status of the race and the particularities of the course, resulting in a seventh place, losing the almost certain title (Aversa et al. 2018).

2.8 OfficeMax

In 2014, the office supply company OfficeMax sent an advertisement mail to one of their off-and-on customers. Even though, that procedure itself is not noteworthy, the address printed on the envelope not only included the name of the recipient, but also, instead of the name of his business, the addition that his daughter had died in a car crash. While the information was correct, and his daughter, along with her boyfriend, actually died in a car crash in the previous year, the occurrence was disturbing for the addressee and traumatizing for his wife. Subsequently, the question arose, how and why the company even had the information. Furthermore, this highly delicate information was falsely labelled, being the only reason for its revelation. In the following, instead of apologizing, the contacted representatives doubted the customer's claims, leading to the media being notified of the story. While the company later on apologized and referred to a (not disclosed) data broker as the culprit, stating that they never ordered information that exceeded the addresses itself and also announced they would implement additional filters to flag inappropriate information, the damage to the affected family was already done (Hill 2014; Pearce 2014).

2.9 US Election Prediction 2016

With the 2016 US presidential election campaign approaching its end, the final vote between Donald Trump and Hillary Clinton, many media outlets tried to predict its outcome. For example, the renowned statistician Nate Silver prognosticated a 71.4 percent (Silver 2016) chance of Clinton becoming the next president. In doing so, he was even rather conservative compared to other forecasts. The New York Times estimated Clinton's chances at 85 percent, Reuters gave her 90 percent, the Huffington Post 98 percent and the Princeton Election Consortium even 99 percent. While the timeframe to be forecasted was only a few days and there were

plenty of data from polls as well as sophisticated algorithms for the analysis, the results differed severely from each other and even more from the real outcome. While it is hard to precisely determine the cause, several factors were mentioned in the review. Those comprise mistakes in the sampling of the questioned population, dishonest answers due to societal pressure, the dynamic and often emotional nature of elections, the sometimes irrational actions of humans. But also the fact, that the public opinion was used as the only indicator of voting behavior instead of utilizing additional sources like the number and facial expressions of the attendance at rallies or social media limited the significance of the obtained results (Stone 2017).

2.10 Orca

Another example related to presidential elections, which took place four years earlier and potentially even influenced the outcome, is the mobile data analytics platform Orca. As his opponent Barack Obama (Scherer 2012), republican candidate Mitt Romney intended to harness the power of big data for the steering and coordination of his 2012 campaign. The idea was to analyse what was happening at polling stations and subsequently use that knowledge to direct the efforts of campaign volunteers towards potential Romney voters in contested states who had not yet voted. While this general strategy is not new and has been applied for many years, this time it was supposed to be digitized, therefore allowing for a real time analysis in contrast to the previously used physical lists. However, the system itself, as well as the corresponding communication, were flawed. The designated users were not properly trained, the according materials were not adequately distributed and many of the volunteers could not even log into the system, which also crashed repeatedly. This, in turn, resulted in a huge waste of not only money for the development of Orca, but also the manpower of the volunteers who otherwise could have had a positive impact on the outcome of the election, and frustrated many of the most devoted supporters (Casaretto 2012; Terkel 2012; Marcus 2012).

2.11 Facebook

With targeted advertisement being a major source of income and over two billion users (Statista 2019), the automated management of the ads constitutes a highly important operation for Facebook. However, in 2017 it was discovered that the automatically created categories for the definition of the target demographic, which were based on the information gathered on users, not only allowed explicitly addressing “teachers” or “nurses”, but also a group of about 2300 “Jew haters”. While this category, along with others, has been subsequently deleted, a list of around 5000 manually checked terms was curated and a general increase in human oversight of its ad targeting was promised by Facebook, this has not been the first controversy regarding this part of the company’s business (Elder 2014; Angwin and Parris 2016). Over the years, there were repeatedly issues accompanying the automation and the options for self-service and customization of the ads and their distribution, leading to violations of the law, a loss of trust and damage caused to the image. Yet, since this automation is unavoidable, considering the magnitude of the task, Facebook is challenged to improve the underlying concepts and mechanisms (Angwin et al. 2017; Dua 2017).

2.12 Tay

In 2016, Microsoft launched an AI chatbot named Tay, which was supposed to interact with English speaking people on Twitter, emulating the chatting behavior of a teenage girl. The idea was to research conversational understanding and at the same time create interactions with a young target group. A similar experiment had already been conducted in China, where “XiaoIce” was running successfully. However, this time, the endeavor did not proceed as expected. After less than 24 hours and about 100.000 tweets, Tay had to be shut down, having turned into a genocide promoting, anti-semitic racist. While many of its insulting tweets were originated in a “repeat after me” functionality and just copied messages that were written by users, others were an effect of the AI learning from the obtained inputs and various other online sources. This resulted in statements like “bush did 9/11 and Hitler would have done a better job than the monkey we have now. donald trump is the only hope we’ve got.” (Kleeman 2016). Besides shutting Tay down, Microsoft also deleted the inappropriate tweets and issued an apology, referring the scientific and unpredictable nature of the experiment, but also submitting an oversight on their part, underestimating the potentially disruptive nature of the internet, respectively a part of its users. This incident not only showed how an internet-based-learning AI could be corrupted, even by comparatively few people, and therefore, how important a sufficient supervision is. It also exemplified the enormous influence of cultural differences in the usage of technologies and the necessity of taking those into account when creating new concepts or implementing common ones

(Ohlheiser 2016; Vincent 2016; Steiner 2016; Hunt 2016).

3 ANALYSIS

This work mainly aims to deepen the understanding regarding the practical implementation of big data in enterprises by analyzing and categorizing the mistakes in the presented cases. To increase the objectivity in the extraction of those mistakes, three of the authors have independently analyzed them. Afterward the results were conjointly preprocessed. This included, for example, the merging of synonyms and the aggregation of similar causes. In cases where divergences occurred, those were collectively discussed to finally create a joint list of identified failures (Orwin and Vevea 2009). The final list comprises the factors *Bad Data Quality*, *Ensuring Data Security*, *Problematic Data Integration*, *Lack Of Technological Understanding*, *Intransparent Analysis*, *Relying Too Much On The System*, *Insufficient Quality Assurance*, *Unrealistic Budgeting*, *Wrong Data Interpretation*, *Insufficient End-User Engagement*, *Lack Of Skills*, *Privacy Concerns*, *Lack Of Vision*, *Bad Project Management*, *Bad Company Culture* and *Lack Of Strategy*. Subsequently, using the same approach as before, the identified failures were classified following the categories presented in (Alharthi et al. 2017). Those are *Technology*, *Human*, and *Organization*. While in the first category, the factors are caused by technical issues in a broader sense, and human causes can be attributed to the actions and properties of individuals, organizational issues originate from relations, customs, structures and hierarchies within those organizations. This also makes them probably the hardest to overcome, since isolated measures are unlikely to resolve the issue and extensive interventions or restructurations have to be performed. Furthermore, to take account of the complexity of the regarded subject and the inherent connections, also combinations of the three proposed main categories were regarded. Those can involve two or even all three of them, generating four more classes, for a total of seven general fault types. The results of the analysis, also constituting the answers to the research question, are depicted in Figure 1, showing the processed list of causes for failure as well as their mapping to the categories. While the applied procedure severely differs from the approach followed in (Weibl and Hess 2018), the results show large similarities, suggesting their validity.

4 CONCLUSION

Big data does not only offer opportunities, but is also highly susceptible to failure. In the publication at hand, twelve real world cases, whose data driven

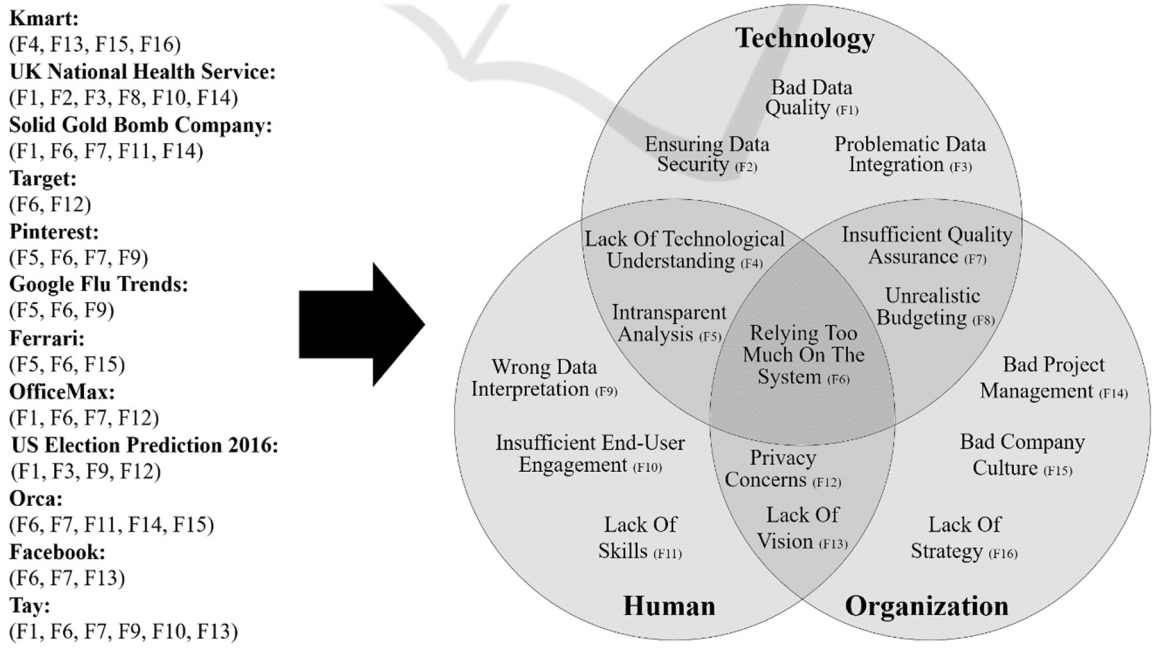

Figure 1: Factors contributing to the failure of data driven endeavors.

approaches failed or had drawbacks of varied severity, were analyzed and the determined issues categorized, providing insights to practitioners and researchers. In the future, the number of analyzed cases shall be further increased and the results will be fused with insights from existing studies as well as expert interviews, to enhance the significance of the findings. As a subsequent step, it is also intended to use the enhanced data basis to develop concrete solutions on how to avoid the identified pitfalls.

REFERENCES

24/7 Wall St. (2012): The Worst Business Decisions of All Time. Available online at https://247wallst.com/special-report/2012/10/17/the-worst-business-decisions-of-all-time/, updated on 10/17/2012, checked on 12/18/2019.

Albert, Kendra (2015): But What Did the Daughter Think? Available online at https://medium.com/@Kendra_Serra/but-what-did-the-daughter-think-8d9233789b4f, updated on 8/31/2015, checked on 12/18/2019.

Alharthi, Abdulkhaliq; Krotov, Vlad; Bowman, Michael (2017): Addressing barriers to big data. In *Business Horizons* 60 (3), pp. 285–292.

Angwin, Julia; Parris, Terry (2016): Facebook Lets Advertisers Exclude Users by Race. Available online at https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race, updated on 10/28/2016, checked on 1/3/2020.

Angwin, Julia; Varner, Madeleine; Tobin, Ariana (2017): Facebook let advertisers target 'Jew Haters'. Available online at https://www.businessinsider.com/facebook-let-advertisers-reach-jew-haters-through-ad-buying-tool-2017-9?r=DE&IR=T, updated on 9/14/2017, checked on 1/3/2020.

Aversa, Paolo; Cabantous, Laure; Haefliger, Stefan (2018): When decision support systems fail: Insights for strategic information systems from Formula 1. In *The Journal of Strategic Information Systems* 27 (3), pp. 221–236.

Bilal, Muhammad; Oyedele, Lukumon O.; Qadir, Junaid; Munir, Kamran; Ajayi, Saheed O.; Akinade, Olugbenga O. et al. (2016): Big Data in the construction industry: A review of present status, opportunities, and future trends. In *Advanced Engineering Informatics* 30 (3), pp. 500–521.

Bughin, Jacques (2016): Big data, Big bang? In *Journal of Big Data* 3 (1).

BusinessWire (2003): Kmart Selects Business Objects as Its Business Intelligence Platform. Available online at https://www.businesswire.com/news/home/20030610005293/en/Kmart-Selects-Business-Objects-Business-Intelligence-Platform, updated on 6/10/2003, checked on 12/18/2019.

Casaretto, John (2012): Romney's Project Orca – a Big Data Fail. Available online at https://siliconangle.com/2012/11/12/romneys-project-orca-a-big-data-fail/, updated on 11/12/2012, checked on 1/1/2020.

Comstock, Jonah (2015): Google Flu Trends website shuts down; will send data to Boston Children's, Columbia, CDC. Available online at https://www.mobihealthnews.com/46248/google-flu-trends-website-shuts-down-will-send-data-to-boston-childrens-columbia-cdc, updated on 8/21/2015, checked on 12/18/2019.

Dobre, Ciprian; Xhafa, Fatos (2014): Intelligent services for Big Data science. In *Future Generation Computer Systems* 37, pp. 267–281.

Dua, Tanya (2017): Facebook promises more human oversight of its ad targeting, as COO Sheryl Sandberg says recent anti-Semitic mishap is a 'fail on our part'. Available online at https://www.businessinsider.com/facebook-sheryl-sandberg-introduces-manual-reviews-of-ad-targeting-after-propublica-article-reveals-anti-semitic-categories-2017-9?r=DE&IR=T, updated on 9/20/2017, checked on 1/3/2020.

Duhigg, Charles (2012): How Companies Learn Your Secrets. Available online at https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?ref=magazine, updated on 2/16/2012, checked on 12/18/2019.

Egan, Matt (2015): Kmart's sales have fallen off a gigantic cliff. Available online at https://money.cnn.com/2015/06/08/investing/kmart-sales-decline-sears-eddie-lampert/, updated on 6/9/2015, checked on 12/18/2019.

Elder, Jeff (2014): Nude Webcams and Diet Drugs: the Facebook Ads Teens Aren't Supposed to See. Available online at https://www.wsj.com/articles/facebook-ads-teens-werent-supposed-to-see-1393541465, updated on 2/27/2014, checked on 1/3/2020.

Fiore, Sandro; Elia, Donatello; Pires, Carlos Eduardo; Mestre, Demetrio Gomes; Cappiello, Cinzia; Vitali, Monica et al. (2019): An Integrated Big and Fast Data Analytics Platform for Smart Urban Transportation Management. In *IEEE Access* 7, pp. 117652–117677.

Gajdošík, Tomáš (2019): Big Data Analytics in Smart Tourism Destinations. A New Tool for Destination Management Organizations? In Vicky Katsoni, Marival Segarra-Oña (Eds.)*Smart Tourism as a Driver for Culture and Sustainability, vol.* 77. Cham: Springer International Publishing (Springer Proceedings in Business and Economics), pp. 15–33.

Ginsberg, Jeremy; Mohebbi, Matthew H.; Patel, Rajan S.; Brammer, Lynnette; Smolinski, Mark S.; Brilliant, Larry (2009): Detecting influenza epidemics using search engine query data. In *Nature* 457, 1012 - 1014.

Günther, Wendy Arianne; Rezazade Mehrizi, Mohammad Hosein.; Huysman, Marleen; Feldberg, Frans (2017): Debating big data: A literature review on realizing value from big data. In *The Journal of Strategic Information Systems* 26 (3), pp. 191–209.

Hartmann, Philipp Max; Zaki, Mohamed; Feldmann, Niels; Neely, Andy (2016): Capturing value from big data – a taxonomy of data-driven business models used by start-

up firms. In *International Journal of Operations & Production Management* 36 (10), pp. 1382–1406.

Hefford, Rhys (2011): Why the NHS National Programme for IT didn't work. Available online at https://www.cio.co.uk/it-strategy/why-nhs-national-programme-for-it-didnt-work-3431723/, updated on 12/2/2011, checked on 12/18/2019.

Hill, Kashmir (2014): OfficeMax Blames Data Broker For 'Daughter Killed in Car Crash' Letter. Available online at https://www.forbes.com/sites/kashmirhill/2014/01/22/officemax-blames-data-broker-for-daughter-killed-in-car-crash-letter, updated on 1/22/2014, checked on 1/2/2020.

Hunt, Elle (2016): Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. Available online at https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter, updated on 3/24/2016, checked on 1/4/2020.

Jin, Xiaolong; Wah, Benjamin W.; Cheng, Xueqi; Wang, Yuanzhuo (2015): Significance and Challenges of Big Data Research. In *Big Data Research* 2 (2), pp. 59–64.

Justinia, Taghreed (2017): The UK's National Programme for IT: Why was it dismantled? In *Health services management research* 30 (1), pp. 2–9.

Khan, Samia; van Wynsberghe, Robert (2008): Cultivating the Under-Mined. Cross-Case Analysis as Knowledge Mobilization. In *Forum: Qualitative Social Research* 9 (1).

Kleeman, Sophie (2016): Here Are the Microsoft Twitter Bot's Craziest Racist Rants. Available online at https://gizmodo.com/here-are-the-microsoft-twitter-bot-s-craziest-racist-ra-1766820160, updated on 3/24/2016, checked on 1/4/2020.

Kosoff, Maya (2014): Pinterest Accidentally Sent Emails To Single Women Congratulating Them On Getting Married. Available online at https://www.businessinsider.com/pinterest-accidental-marriage-emails-2014-9?r=DE&IR=T, updated on 9/4/2014, checked on 1/1/2020.

Lazer, David; Kennedy, Ryan; King, Gary; Vespignani, Alessandro (2014): Big data. The parable of Google Flu: traps in big data analysis. In *Science (New York, N.Y.)* 343 (6176), pp. 1203–1205.

Leinwand, Paul; Mainardi, Cesare (2010): Why Can't Kmart Be Successful While Target and Walmart Thrive? Available online at https://hbr.org/2010/12/why-cant-kmart-be-successful-w, updated on 12/15/2010, checked on 12/18/2019.

Marcus, Stephanie (2012): Mitt Romney's Project ORCA Failure: Broken ORCA App Cost Him Thousands Of Votes. Available online at https://www.huffpost.com/entry/mitt-romney-project-orca-broken-app-cost-thousands-votes_n_2109986, updated on 11/10/2012, checked on 1/1/2020.

Maroufkhani, Parisa; Wagner, Ralf; Wan Ismail, Wan Khairuzzaman; Baroto, Mas Bambang; Nourani, Mohammad (2019): Big Data Analytics and Firm Performance: A Systematic Review. In *Information* 10 (7), p. 226.

Marr, Bernard (2017): Data strategy. How to profit from a world of big data, analytics and the internet of things. New York: Kogan Page Ltd. Available online at http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=1494509.

McVeigh, Tracy (2013): Amazon acts to halt sales of 'Keep Calm and Rape' T-shirts. Available online at https://www.theguardian.com/technology/2013/mar/02/amazon-withdraws-rape-slogan-shirt, updated on 3/2/2013, checked on 12/18/2019.

Müller, Oliver; Fay, Maria; Vom Brocke, Jan (2018): The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. In *Journal of Management Information Systems* 35 (2), pp. 488–509.

NIST (2019): NIST Big Data Interoperability Framework: Volume 1, Definitions, Version 3. Gaithersburg, MD.

Ohlheiser, Abby (2016): Trolls turned Tay, Microsoft's fun millennial AI bot, into a genocidal maniac. Available online at https://www.washingtonpost.com/news/the-intersect/wp/2016/03/24/the-internet-turned-tay-microsofts-fun-millennial-ai-bot-into-a-genocidal-maniac/, updated on 3/25/2016, checked on 1/4/2020.

Onal, Aras Can; Berat Sezer, Omer; Ozbayoglu, Murat; Dogdu, Erdogan (2017): Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning. In Jian-Yun Nie, Zoran Obradovic, Toyotaro Suzumura, Rumi Ghosh, Raghunath Nambiar, Chonggang Wang (Eds.)*Proceedings of the 2017 IEEE International Conference on Big Data.* Boston, MA, 11.12.2017-14.12.2017. Piscataway, NJ: IEEE, pp. 2037–2046.

Orwin, Robert G.; Vevea, Jack L. (2009): Evaluating Coding Decisions. In Harris M. Cooper, Larry V. Hedges, Jeff C. Valentine (Eds.)*The handbook of research synthesis and meta-analysis.* 2nd ed. New York: Russell Sage Foundation, pp. 177–206.

Pagliery, Jose (2013): Man behind 'Carry On' T-shirts says company is 'dead'. Available online at https://money.cnn.com/2013/03/05/smallbusiness/keep-calm-and-carry-on/, updated on 3/5/2013, checked on 12/18/2019.

Pappas, Stephanie (2014): Data Fail! How Google Flu Trends Fell Way Short. Available online at https://www.livescience.com/44089-google-flu-trends-problems.html, updated on 3/13/2014, checked on 12/18/2019.

Pearce, Matt (2014): OfficeMax executive apologizes over 'daughter killed' mailer. Available online at https://www.latimes.com/nation/la-na-officemax-mess-20140121-story.html, updated on 1/20/2014, checked on 1/2/2020.

Pervaiz, Fahad; Pervaiz, Mansoor; Abdur Rehman, Nabeel; Saif, Umar (2012): FluBreaks: early epidemic detection from Google flu trends. In *Journal of medical Internet research* 14 (5), e125.

Philip Chen, C. L.; Zhang, Chun-Yang (2014): Data-intensive applications, challenges, techniques and

technologies: A survey on Big Data. In *Information Sciences* 275, pp. 314–347.

Roy, Jessica (2014): Pinterest Accidentally Congratulates Single Women on Getting Married. Available online at http://nymag.com/intelligencer/2014/09/pinterest-congratulates-single-women-on-marriage.html?, updated on 9/4/2014, checked on 1/1/2020.

Scherer, Michael (2012): Inside the Secret World of the Data Crunchers Who Helped Obama Win. Available online at http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/, updated on 11/7/2012, checked on 1/1/2020.

Silver, Nate (2016): Final Election Update: There's A Wide Range Of Outcomes, And Most Of Them Come Up Clinton. Available online at https://fivethirtyeight.com/features/final-election-update-theres-a-wide-range-of-outcomes-and-most-of-them-come-up-clinton/, updated on 11/8/2016, checked on 12/31/2019.

Staegemann, Daniel; Hintsch, Johannes; Turowski, Klaus (2019a): Testing in Big Data: An Architecture Pattern for a Development Environment for Innovative, Integrated and Robust Applications. In *Proceedings of the WI2019*, pp. 279–284.

Staegemann, Daniel; Volk, Matthias; Jamous, Naoum; Turowski, Klaus (2019b): Understanding Issues in Big Data Applications – A Multidimensional Endeavor. In *25th Americas Conference on Information Systems, AMCIS 2019, Cancun, Q.R, Mexico, August 15-17, 2019:* Association for Information Systems.

Staegemann, Daniel; Volk, Matthias; Turowski, Klaus (2019c): Mobile Procurement Management. In Tobias Kollmann (Ed.)*Handbuch Digitale Wirtschaft, vol.* 9. Wiesbaden: Springer Fachmedien Wiesbaden (Springer Reference Wirtschaft), pp. 1–15.

Statista (2019): Number of monthly active Facebook users worldwide as of 3rd quarter 2019. Available online at https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/, updated on 11/19/2019, checked on 1/3/2020.

Steiner, Anna (2016): Was Microsoft durch „Tay" gelernt hat. Available online at https://www.faz.net/aktuell/wirtschaft/netzwirtschaft/was-microsoft-mit-dem-bot-tay-von-der-netzgemeinde-gelernt-hat-14146188.html, updated on 3/26/2016, checked on 1/4/2020.

Stone, Adam (2017): When Big Data Gets It Wrong. Available online at https://www.govtech.com/data/When-Big-Data-Gets-It-Wrong.html, updated on March 2017, checked on 12/31/2019.

Syal, Rajeev (2013): Abandoned NHS IT system has cost £10bn so far. Available online at https://www.theguardian.com/society/2013/sep/18/nhs-records-system-10bn, updated on 9/18/2013, checked on 12/18/2019.

Terkel, Amanda (2012): Project ORCA: Mitt Romney Campaign Plans Massive, State-Of-The-Art Poll Monitoring Effort. Available online at https://www.huffpost.com/entry/project-orca-mitt-romney_n_2052861, updated on 12/6/2017, checked on 1/1/2020.

Turner, Marcia Layton (2003): Kmart's ten deadly sins. How rogue managers ruined an American icon. New York, Chichester: Wiley.

Vincent, James (2016): Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. Available online at https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist, updated on 3/24/2016, checked on 1/4/2020.

Volk, Matthias; Staegemann, Daniel; Pohl, Matthias; Turowski, Klaus (2019): Challenging Big Data Engineering: Positioning of Current and Future Development. In *Proceedings of the 4th International Conference on Internet of Things, Big Data and Security.* 4th International Conference on Internet of Things, Big Data and Security. Heraklion, Crete, Greece, 02.05.2019 - 04.05.2019: SCITEPRESS - Science and Technology Publications, pp. 351–358.

Weibl, Johannes; Hess, Thomas (2018): Success or Failure of Big Data: Insights of Managerial Challenges from a Technology Assimilation Perspective. In *Proceedings of the MKWI 2018.* Multikonferenz Wirtschaftsinformatik 2018. Lüneburg, Germany, 06.03.2018-09.03.2018, pp. 47–58.

Wenzel, Ramon; van Quaquebeke, Niels (2018): The Double-Edged Sword of Big Data in Organizational and Management Research. In *Organizational Research Methods* 21 (3), pp. 548–591.

Wu, Desheng; Cui, Yiwen (2018): Disaster early warning and damage assessment analysis using social media data and geo-location information. In *Decision Support Systems* 111, pp. 48–59.

Yang, Chaowei; Huang, Qunying; Li, Zhenlong; Liu, Kai; Hu, Fei (2017): Big Data and cloud computing: innovation opportunities and challenges. In *International Journal of Digital Earth* 10 (1), pp. 13–53.

Yin, Shen; Kaynak, Okyay (2015): Big Data for Modern Industry: Challenges and Trends [Point of View]. In *Proceedings of the IEEE* 103 (2), pp. 143–146.