UMEÅ UNIVERSITY

Digital Innovation Management

Investigating Digital Trace Data in Online Communities

Vasili Mankevich

Department of Informatics
Umeå 2019

Dissertation for PhD

ISBN: 978-91-7855-111-8
ISSN: 1401-4572 RR-19.03
Cover design: Johan Bodén
Electronic version available at: http://umu.diva-portal.org/
Printed by: Cityprint i Norr AB
Umeå, Sweden 2019

"What makes a perfect question? Ironically, the best questions are not questions that lead to answers, because answers are on their way to becoming cheap and plentiful. A good question is worth a million good answers ... Good questions are what humans are for"

- *Kevin Kelly, from "The Inevitable"*

Table of Contents

Abstract

Firms and individuals are interacting online on an unprecedented scale. These interactions may lead to new digital products, services and practices, all of which are manifestations of digital innovation. This process relies on records users leave on various digital platforms which carry information about their activities – digital trace data. The data is generated on a massive scale, yet is just data until it is confronted with meaning – its value remains latent. Digital trace data is agnostic about future use, it carries records of interactions with digital artifacts and is available to wide numbers of actors to reinterpret them as sources of innovation and value creation. Online communities where data is generated can be sources of innovation, but are also extremely vulnerable. Digital trace data are not finitely expendable but may be used and passed along to any other individuals, partners, customers, or suppliers. To remain competitive, firms increasingly need to manage dynamic interactions of online community members, confront digital trace data with meaning, and facilitate innovation that is decentralized and requires heterogeneous knowledge resources.

This dissertation explores how digital innovation can be leveraged in the context of online communities. It is based on four empirical investigations in the context of firms interacting with online communities that are rich with digital trace data. Collectively, these studies illustrate the potential utility of digital trace data generated by online communities for digital innovation, and suggest possible strategies for effective management of digital innovation for value creation. The dissertation contributes to both theoretically and empirically oriented discourses on the use of digital trace data. Specifically, it does so by providing propositions for dealing with digital trace data through platform design, community sociality, and narration for digital innovation.

Acknowledgements

I want to express gratitude to my supervisor professor Jonny Holmström. You had inspired me to pursue doctoral research and gave me freedom and support to develop my interests even when they started as side projects and vague ideas. I have learned so much from you and your many stories, and I cannot believe the impact the doctoral program had on me. It changed my whole thinking process, and for that, I sincerely thank you.

I would also like to thank my second supervisor professor Ola Henfridsson at Miami Business School for the guidance and advice during these years. I also thank Aron Lindberg at Stevens Research Institute for the year I have spent in Hoboken as a visiting scholar. It was critical for developing my methodological competence that grew to become an important part of my academic interest.

Research environment plays a vital role in every academic's development, and I consider myself very lucky in this regard. My colleagues at the Department of Informatics at Umeå University made these years fun and exciting. Through our discussions, I experienced academic dialogue and friction that is necessary for good research. Especially, I would like to thank Katrin Jonsson, Lars Öbrand, Daniel A. Skog for serving as panelists at my pre-seminar and providing constructive feedback when I needed it. Lars, I have learned quite a bit from our informal discussions, perhaps more than from many research papers I have read as a doctoral student. Daniel, thank you for trying to warn me of the challenges that lay ahead when I was starting this journey (I wish I listened to you more closely). I am grateful to Johan Sanberg for discussing my work at length and providing the most original and unexpected feedback during our conversations and seminars. Your contribution was very important when this thesis was taking form, and your support helped me to recognize the value in my work. I have also learned from you to take ownership of my research, which helped me tremendously in the development of this thesis. To Taline Jadaan, I am grateful for your support at every step of the way, for sharing my passion for methodology, for deep conversations on every possible topic, and for sharing some of the superstitions; to Daniel Nylén for the guidance as a new doctoral student, and for the fun time developing pedagogical material; to Henrik Wimelius for support and advice during the toughest stages of the doctoral program; to Fatemeh Moradi for sharing a positive attitude and great conversations we had over pizza in the evenings at the office; to Johan Boden for design of the cover for this dissertation; to the Informatics management and staff for continuous support; and to the doctoral student group at the Informatics for going on this tough journey together.

I would like to sincerely thank my co-author Sanja Tumbas of IESE Business School. It was a genuinely positive experience working with you, developing our ideas and iterating on the data. I thank the hosts and participants of the KIN Summer School 2015 (VU University Amsterdam), where I found inspiration for my work (and the first study in particular) and got to know the European Information Systems community. I am also grateful to MIT research school for research funding that enabled my doctoral research and introduced me to the network of Swedish researchers in management and information technology.

I thank my family for supporting me at every stage of my education and academic development. My mom, Liudmila, have always given me a wise advice that helped me in everything from studying to project management. I thank Olegs Tkacevs for showing me very early the nature of academic career and helping me to set on the path that ultimately led me to doctoral education.

Finally, and most importantly, I want to thank my wife, Brooke, for supporting me during these years. Your integrity continues to inspire me to be a better researcher, and your curiosity leads me to think more clearly about how my research affects the society outside academia. Laughing with you gave me the energy to start and to complete this thesis, for which I am forever grateful.

Umeå, August 2019

Vasili

Preface

This dissertation is based on studies described in the following appended papers.

Paper 1: Mankevich, V., Lindberg, A., & Holmström, J. The Acknowledgement Paradox in a User Community Context: Evidence from a Business Intelligence and Analytics User Community. *Manuscript*.

Paper 2: Mankevich, V., Lindberg, A., & Holmström, J. (2018). Better Safe Than Sorry? Investigating Big Data Evangelists and Their Value Propositions. Presented at *Academy of Management Global Proceedings, Surrey (2018),* 58.

Paper 3: Tumbas, S., Mankevich, V., & Holmström, J. "Stories within Stories": Digital Multiplicity for Entrepreneurial Storytelling. (Under review by an international journal)

Paper 4: Mankevich, V., Holmström, J., & McCarthy, I. P. (2019). Why Zlatan Ibrahimovic is Bigger Than Manchester United: Investigating Digital Traces in Co-branding Processes on Social Media Platforms**.** *Presented at the 52nd Hawaii International Conference on System Sciences.*

Introduction

The ongoing and pervasive digitalization of today's society enables online activities of firms and individuals (Brynjolfsson & McAfee, 2014; Faraj, Pachidi, & Sayegh, 2018; Yoo, 2010). Firms distribute their products and services, while individuals interact with them and each other. Interactions between firms and individuals online (recorded in digital trace data, i.e. traces users leave on various digital platforms that carry information about their activities) may collectively leverage valuable innovation for the firms (Di Gangi & Wasko, 2009; Shaikh & Levina, 2019). Achieving such value creation is not a trivial task, since the sheer scale of digital trace data generation makes it difficult to even collect (George, Osinga, Lavie, & Scott, 2016), let alone understand (Berente, Seidel, & Safadi, 2018). Further, innovation of the firms is not limited by the amount of information generated in their environment, but by their ability to process and appropriate it (Cohen & Levinthal, 1990; Fayard, Gkeredakis, & Levina, 2016; Teigland, Di Gangi, Flåten, Giovacchini, & Pastorino, 2014). Therefore, firms need strategies for handling (acquiring, filtering, processing, interpreting and exploiting) the enormous amount of available digital trace data and other kinds of available data (Günther, Rezazade Mehrizi, Huysman, & Feldberg, 2017). Clearly, all forms of digital trace data have a direct function, e.g., a post on social media communicates a bit of information to a wider audience, and a user asking for help in a forum engages in a discussion with a community of likeminded professionals. However, digital trace data also provide detailed records of online activities that may enable innovation and new practices (Aaltonen & Tempini, 2015). Firms need to proactively build strategies for leveraging these resources for their benefit.

Practitioners have been mostly optimistic about the opportunities digital trace data can bring about (e.g. Deloitte, 2013). United under an ambiguous banner of Big Data, industry experts have suggested that technological tools (i.e. business intelligence and analytics) will allow firms to abstract and distill innovation-enabling insights from the sea of data (Davenport, 2006). However, since the original hype, experts are somberly warning of high failure rates of analytics projects[1] and the limitations of data-centric decision-making philosophy (Ransbotham, Kiron, & Prentice, 2016). According to these experts there has been too much focus on data and technology, and lack of focus on data's meaning and its idiosyncrasies in an organizational setting (Kallinikos & Constantiou, 2015; Kane, Phillips, Copulsky, & Andrus, 2019; Shah, Horne, & Capella, 2012). Practitioners have also been deferring judgement to collectives,

[1] Throughout this thesis, analytics refers to acquisition and analysis of big datasets for business rather than purely academic purposes. In addition, the modern convention of referring to such datasets using data as a singular term (rather than as the plural of datum) is followed

establishing crowdsourcing systems in which inputs from interactions are collected and evaluated by firms and associated communities (Prpić, Shukla, Kietzmann, & McCarthy, 2015). While widely used for problem-solving, crowdsourcing has limited utility for value capture (Bloodgood, 2013) and is constrained by the information system employed (Lukyanenko, Parsons, Wiersma, & Maddah, 2019; Majchrzak & Malhotra, 2013). Its effectiveness relies on precise problem definition (Blohm, Leimeister, & Krcmar, 2013), which is not always possible when engaging in digital innovation.

A number of productive research streams have addressed how information outside a firm may be approached to enable innovation. For example, ways that firms source external knowledge for internal innovation processes have been described under the umbrella term Open Innovation (Chesbrough, 2003). Collectives of individuals that participate in knowledge exchange in virtual spaces – online communities (Faraj, Krogh, Monteiro, & Lakhani, 2016a) – are important for both generation of new ideas (Di Gangi & Wasko, 2009) and development (Dahlander, Frederiksen, & Rullani, 2008) throughout the innovation process. Online communities have been productively studied according to the specific role they play in the innovation processes. Examples include ideation communities (Bayus, 2013), open source communities involved in development of digital artifacts (Lindberg, Berente, Gaskin, & Lyytinen, 2016), and epistemic communities that generate use practices involving a firm's products or services (Johnson, Safadi, & Faraj, 2015).

Contemporary firms that develop digital products and services interact with their environment across these activities: the ideas are sourced, digital artifacts are collectively developed, and practices are reconfigured and remixed – all increasingly online. Often, these processes are not sequential but happen simultaneously, online with active participation of users, partners, and other firms. Since these activities overlap and involve distributed actors, studies of them cannot ignore their interactions and must treat them as an ensemble of practices. While there is an understanding that online communities are diverse and their capacities for innovation may differ substantially (van Osch & Avital, 2010), there is still little guidance for firms that attempt to manage this process. In order to overcome the challenges described above, I synthesize my empirical work to address the following research question:

How can digital innovation be leveraged in the context of online communities?

The dissertation is based on four empirical investigations focusing on aspects of the abovementioned processes in the context of online communities that are considered to have generated rich digital trace data. First, my colleagues and I (hereafter we) conducted two studies (reported in two papers) of the Business Intelligence & Analytics community, to investigate digital innovation in the online community and how value from analytics was communicated during the covered period by Tableau – the leading BI&A platform vendor (Gartner, 2018). Next, we investigated digital innovation and digital trace date in a crowdfunding campaign – Pebble Time – the most successful Kickstarter campaign so far. This focused on critical online interactions for digital entrepreneurs, in which information about their product(s), funding campaign(s), users' experiences, and industry contexts entwine into a narrative across a multitude of digital platforms. Finally, we studied online social media co-branding between Manchester United football (soccer) club and the player Zlatan Ibrahimović. The empirical case shows an intricate process in which information pertinent to a firm's core business was remixed and recombined outside the firm's boundaries. The phenomena analyzed in all three empirical settings involved use of naturally occurring data providing records of real interactions by the focal actors. To address the research question stated above as fully as possible, these records (and complementary data) were collected and analyzed by diverse methods, drawing on both qualitative and computational traditions, such as network analysis, sentiment and psychometric text analysis, and structural break analysis.

It is important to address the research question in distinct empirical contexts for three reasons. First, the explosive growth of digital trace data will continue and intensify (Gantz & Reinsel, 2012; Günther et al., 2017). Continuing digitalization will generate even more data, and complex knowledge economies that rely on interaction and data exchange will accelerate this process. As the digital products and services will continue to unbind from physical infrastructures (Yoo, Henfridsson, & Lyytinen, 2010), firms' capacity for digital innovation will increasingly depend on their ability to interact with distributed agents through digital trace data and ability to leverage them for digital innovation. Thus, acquisition and use of digital trace data will become increasingly crucial for economic development and entrepreneurship (Kling, Ackerman, & Allen, 1995).

Furthermore, online communities that generate digital trace data are becoming larger and increasingly complex, spanning multiple digital ecosystems. In addition, recent studies have highlighted their sensitivity to platform design (Ren, Kraut, & Kiesler, 2007), manipulation (Marwick & Lewis, 2017), and special-interest meddling (Weninger, Johnston, & Glenski, 2015). Once flourishing online collectives have perished with the companies that created them (Garcia, Mavrodiev, & Schweitzer, 2013), or dispersed and re-formed in new configurations on other platforms or in other forums. Thus, there are clear needs to understand the intricate processes that leverage digital innovation within them. This dissertation provides a background for conversations regarding the impact of sociality (Faraj et al., 2016a) on the activities and resilience of online communities in general, and more specifically their roles in digital innovation.

Finally, the answer to the research question will have implications for the major socio-technical force of the XXI century – the use by firms of artificial intelligence (Bostrom, 2017; Davenport, 2018). Digitalization and recent advances in computation and algorithms have made AI and machine learning (ML) accessible to the extent that many jobs are under threat from automation (Brynjolfsson & McAfee, 2014). AI is superior for tasks that humans traditionally do not handle well because they require constant consistency, adherence to safe practices, and accuracy, leading many authors to believe that widespread use of AI that complements human abilities is 'inevitable' (Kelly, 2017). However, it is still unclear how AI can be used in a business setting (Chui, Manyika, & Miremadi, 2018), partly because AI (at least "narrow" AI and ML) requires a symbolic system based on explicit decision-delineating information space – i.e. attaching meaning to data that is produced by applied algorithms (Bostrom, 2017). Answering the research question posed above should enable reflection about the limits of using digital trace data in in all kinds of automated processes (including AI/ML), and the opportunities it provides.

The dissertation consists a cover section and a collection of four research papers. In the next section I present the research background on digital innovation. In section three I reflect on the nature of digital trace data and look into the ways it was approached in the past. In section four I describe my research design, including case selection and investigative context, data collection and analysis. In section five I present paper summaries and then discuss research findings against the theory in section five. I conclude the dissertation with implications, limitations of the conducted research, which is then followed by the appended research papers.

2. Digital innovation

Innovation is a fundamental part of progress and economic growth – any firm that seeks to maintain profits must innovate (Schumpeter, 1934). In his works, Schumpeter described innovation as a fundamental driver of competitiveness and creative destruction that profoundly changes societal structures "incessantly destroying the old one, incessantly creating a new one" (Schumpeter 2003: 83). His ideas were rediscovered with the growth of knowledge economies and became even more relevant in the wake of The Great Recession of the late 2000s (Śledzik, 2013). Since so much of contemporary economy is intangible and globalized, the traditional barriers for competition (e.g. regulatory or geographical) are eroding, giving way to fierce global competition. As a result, firms are increasingly forced to seek new ways of innovation.

Innovation may be considered narrowly in performative terms of the value it creates for a firm, and sometimes it is described as a detailed Schumpeterian innovation process, i.e. invention, innovation, diffusion, and imitation (Burton-Jones, 2001). While perspectives vary across disciplines, a generalized approach to innovation is to see it as a function of both novelty and utility. This means that the resulting products or services are expected not only to be different from existing offerings, but also to generate value.

Innovation in computation, personal computing, and communication technology has allowed firms to digitize the ways they work, e.g., email and instant messaging replaced post, and accounting software replaced inventory books. This direct digitization (or computerization) of firms generated substantial amounts of research concerned, for example, with implementation of new technologies in firms (Klein & Sorra, 1996). However, these research efforts stumbled upon a 'productivity paradox' – a realization that straightforward digitization does not provide efficiency gains that would justify exorbitant spending on the new technologies (Brynjolfsson, 1993). For digital technology to generate value it needs to change firms' organizational processes. This challenge is pertinent to many general-purpose technologies. Take the case of electricity: Electric engines were initially built to mimic large steam engines, but the real value was created when new types of small and portable electric engines enabled redesign of the factory floor and organization of the work (Brynjolfsson & Hitt, 2000). Hence the emergence of research embracing both the unique nature of the digital technology and associated innovation in the organizational context.

Digital technology is malleable and can easily be repurposed: The software in a car's control panel may change features and purposes with an update, while the purpose of the car's transmission is locked by its physical characteristics. Digital products and services are free from such lock-in, so the focal concerns of digital innovation management are the key traits of digital technology (Nambisan, Lyytinen, Majchrzak, & Song, 2017), and structural boundaries for innovation outcomes that are major constraints in traditional industries largely do not apply (Yoo et al., 2010). For example, the most serious competition to car manufacturers is now coming from Waymo, which started as a robotaxi project of Google (McGee, 2019) – perhaps the most emblematic digital firm. This means that both incumbent and challenger firms need to examine their potential competitive landscapes more broadly to succeed. It also means that no one firm possesses all the competence required to innovate (Van de Ven, 2005). In this highly diverse environment, closed R&D processes are replaced by open processes (Chesbrough, 2003) that require both heterogeneous knowledge resources and participation in distributed collaboration structures (Lyytinen, Yoo, & Boland Jr, 2016). The enablement of innovation of products and services by digital technology is called digital innovation (Nambisan, Lyytinen, Majchrzak, & Song, 2017), and the research discourse concerning this phenomenon, and the processes involved, is the focus of the next section.

2.1 The digital innovation discourse

Research on digital innovation is a developing stream within the Information Systems (IS) discipline. Papers on the topic "Digital Innovation" only started to be published in prestigious Senior Scholars' Basket of Journals in 2000. Since then, the number of papers published on the topic in these journals has been increasing (Figure 1), as have numbers of relevant conference tracks, publications in other journals, and education curricula including it (Fichman, Santos, & Zheng, 2014). In addition, a network of European Digital Innovation Hubs has been initiated (Goetheer & Butter, 2017) and an Association for Information Systems Special Interest Group on Digital Innovation, Transformation and Entrepreneurship (SIGDITE) was established in 2019. Digital innovation research offers foundations to challenge past theories of innovation and provide opportunities for novel theorizing (Nambisan et al., 2017). Like many emerging and developing research streams, establishing a comprehensive overview might be challenging. A hindrance is that many digital innovation studies have different labels. For the most recent literature review on digital innovation, see Kohli and Melville (2019).

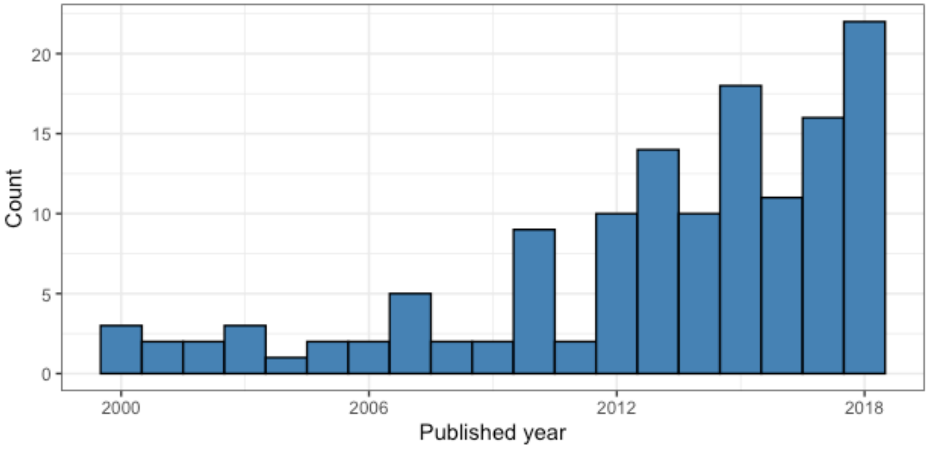

Figure 1: Numbers of papers on digital innovation published in the Senior Scholars' Basket of Journals between 2000 and 2018 according to a Web of Knowledge search

The IS discipline has been described as referential, i.e. referencing or drawing its intellectual foundations from other fields (Baskerville & Myers, 2002; Grover & Lyytinen, 2015; Keen, 1980; Lyytinen & Grover, 2017; Sidorova, Evangelopoulos, Valacich, & Ramakrishnan, 2008). Accordingly, the digital innovation discourse draws on theories developed within IS and in such fields as management (Porter, 1985), strategy (Teece, Pisano, & Shuen, 1997), social psychology (Weick, 1979), and economics (Schumpeter, 1934), as well as practitioner-oriented literature (Adner, 2006). While the referential approach was a debated issue in the past (Baskerville & Myers, 2002; Keen, 1980; King & Lyytinen, 2006), since then interaction with *adjacent* fields has been perceived as a critical aspect of the intellectual tradition of the IS discipline (Tiwana, 2019; Truex, Holmström, & Keil, 2006). However, a list of the 10 most commonly cited journals by digital innovation papers published in the Senior Scholars' Basket of Journals shows that digital innovation research has been particularly reliant on strategy discourse and more generally drawn on the management sciences (Figure 2). The special issue on digital innovation published in *Organization Science* in 2012 (Yoo, Boland, Lyytinen, & Majchrzak, 2012) exemplifies this notion.

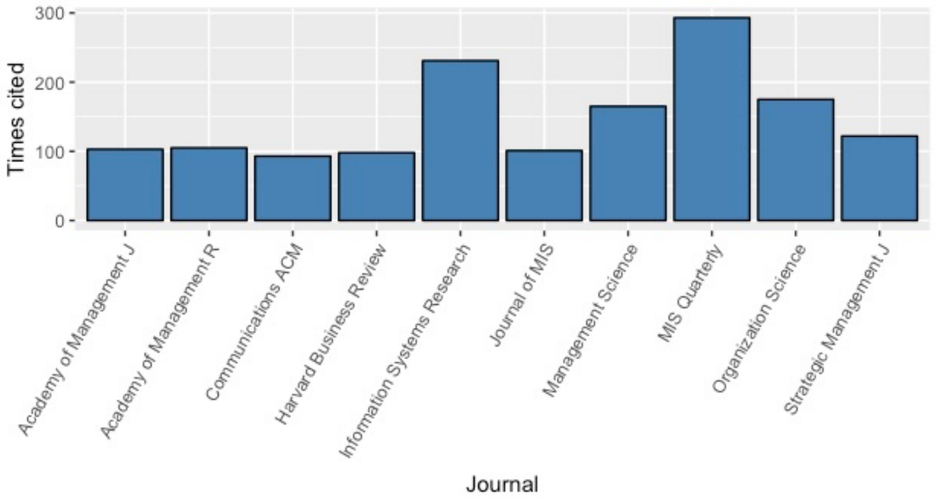

Figure 2: The ten most frequently cited journals in digital innovation research, illustrating the field's reliance on management science generally and strategy discourse particularly

To acquire a broad overview of the literature on digital innovation I first conducted a bibliometric analysis. Importantly, this was not done to conduct a detailed or systematic literature review, but rather to understand the emergent conversations among researchers in the field. I constructed a two-mode network of papers on the subject identified in a Web of Knowledge search and the authors they cited. Then I folded it into a one-mode network of digital innovation papers, in which links represented co-cited authors. To help identify clusters of the research papers, I used the algorithmic detection methods: Walktrap (Pons & Latapy, 2005), Spin-glass model (Traag & Bruggeman, 2009), and fast greedy modularity optimization (Clauset, Newman, & Moore, 2004). The results showed significant convergence of clusters by the applied methods (Figures 3 & 4).

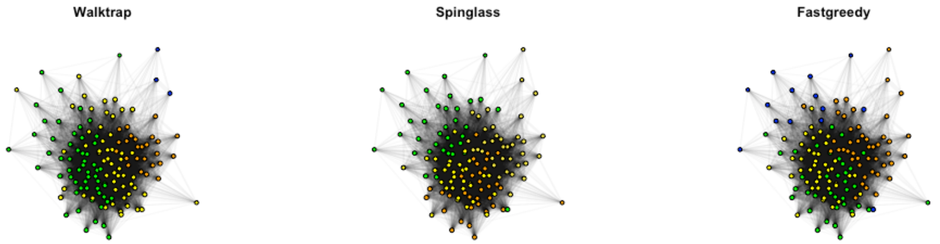

Figure 3: Bibliographic networks of digital innovation research publications generated by three indicated algorithms

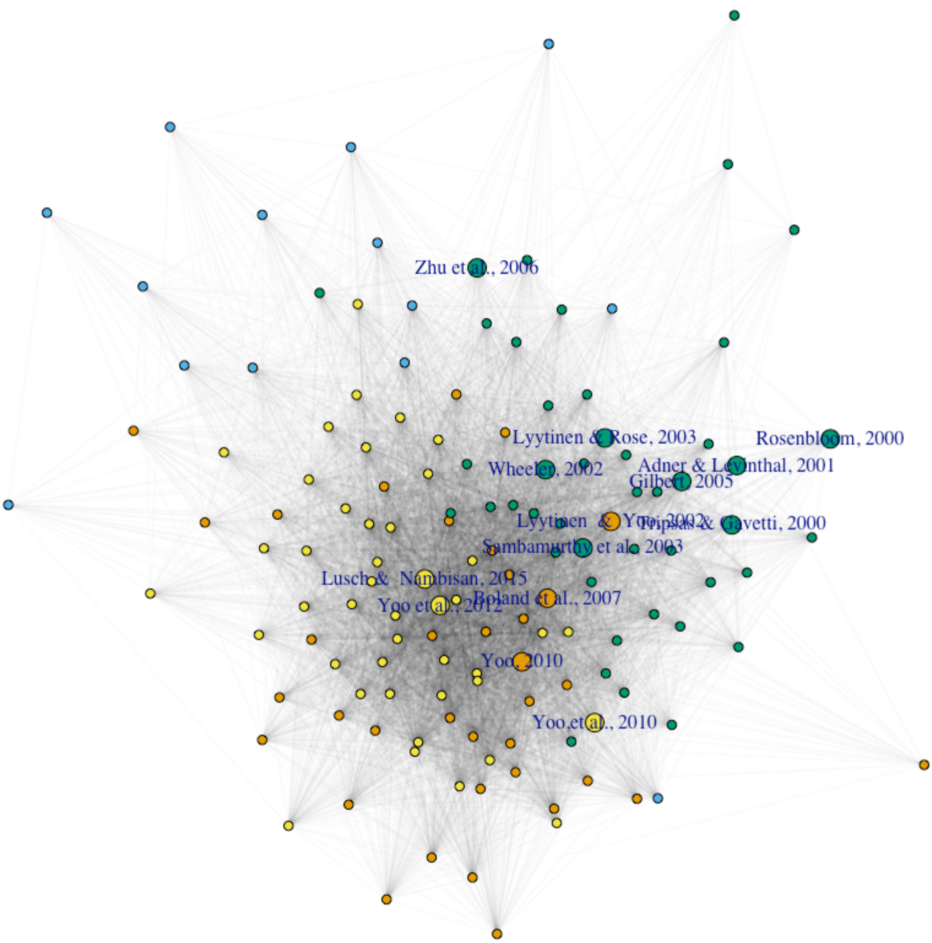

Figure 4: Bibliographic network of digital innovation research publications, with labels for top 10% of most frequently cited papers, clustering generated by the fast and greedy method

After the algorithmic clustering, I closely read papers in the clusters, compiled results obtained with the applied clustering methods, and identified four streams of particularly active digital innovation research in the last two decades, focused on digitalization, strategizing, ecosystems and infrastructures, and digital innovation practice. This list is not meant to be exhaustive, but rather represents ongoing conversations that are actively developing within the research community. See Table 1 for descriptions of the research streams, key references, and empirical examples.

Digital innovation studies are interconnected (see figure 4) and build on the unique properties of digital technology. Digital artifacts are characterized by programmability, addressability, communicability, memorability, sensibility, traceability, and associability (Yoo, 2010), liquification and open-endedness (Monteiro & Parmiggiani, 2019). Many of these properties have been known to the scientific community for some time. Some foundational characteristics of modern computation were laid by Charles Babbage, who developed the first mechanical computer ("difference engine") in the early XIX century (Swade, 2001). Or even earlier analog Antikythera mechanism of 1st or 2nd century AD provided rudimentary calculation capabilities (Freeth, Jones, Steele, & Bitsakis, 2008). The first digital computationally universal ("Turing complete") computer – ENIAC – was designed by 1945 (Goldstine & Goldstine, 1946). Communication networks that enable data transfer emerged with the telegraph in 1837 (Winston, 2002), and in 1974 internet protocol (IP) was developed (Cerf & Kahn, 1974). The long history of these technologies raises the question *Why does digital innovation need to be studied now, what is different*? Part of the answer is that the accessibility of personal computing, pervasive communicability, and overwhelming *digitalization* has created a complex technological landscape, in which technological novelties and apparently innocuous tools have become major, indispensable components of a spanning complex system (Brynjolfsson & McAfee, 2014). When complex systems emerge, they acquire properties beyond those of their separate components (Simon, 1996) or, colloquially, a difference in degree becomes a difference in kind! These fundamental changes render digital technology a new phenomenon, challenging existing innovation theories and requiring a new type of organizing (Nambisan et al., 2017).

Ensuing changes in digital innovation theory necessitate new forms of *strategizing*, representing an early research stream in digital innovation discourse. Researchers established that while previous IS studies accumulated significant knowledge regarding the systems' design, this might be inadequate for understanding digital innovation (Sambamurthy & Zmud, 2000). In a seminal study, Sambamurthy, Bharadwaj, and Grover (2003) described the strategizing process related to digital technology as the development of associated capabilities, digital options, agility, and alertness to increase firms' performance. Development of capabilities has been a central focus, with varying emphasis, for example on IT capabilities (Tripsas & Gavetti, 2000), digital capabilities (Sandberg, Mathiassen, & Napier, 2014), and improvisational capabilities (Pavlou & El Sawy, 2010). Importantly, capability building is increasingly an outward-directed process: firms seek and appropriate external capabilities situated in their extended ecosystems (Selander, Henfridsson, & Svahn, 2013). Empirically, such studies are often focused on digital innovation within large firms, such as camera manufacturers (Benner & Tripsas, 2012), the dairy industry (Sandberg et al., 2014), news (Gilbert, 2005) and media companies (Nylén & Holmström, 2015).

There has been more focus on interactions beyond firms' boundaries in the *ecosystems and infrastructures* discourse. Perhaps the most influential concept has been a layered model architecture (Yoo et al., 2010), the perspective that commands separation of "layers": content, service, network, and device. Such separation is propagated by the distinct characteristics of digital technology – reprogrammability, malleability, and self-reference (Kallinikos, 2006). Layered modular architecture was preceded by a modular architecture – nested rearrangeable components with fixed meaning. This provided adaptability and efficiencies at scale (Simon, 1996), but was often constrained by physical properties. Transition to layered modular architecture enabled development of digital infrastructures – "shared, unbounded, heterogeneous, open, and evolving sociotechnical systems comprising an installed base of diverse information technology capabilities" – and associated theorizing (Tilson, Lyytinen, & Sørensen, 2010: 1). One of the fundamental drivers of digital infrastructure's evolution is digital innovation (Henfridsson & Bygstad, 2013).

Digital innovation relies on digital artifacts that are imbued with implicit knowledge that can be hard to interpret (Leonardi & Bailey, 2008). Practices involved in knowledge creation and interaction between the artifacts and individuals are central concerns in DI research. Such interactions are often large-scale phenomena (such as many in online communities) and their analysis or exploitation requires careful considerations of network effects, participant diversity, and platform governance (Boudreau, 2012). For example, network effects are major drivers of the quality and diversity of independent digital innovation, but large numbers of submissions tend to limit firms' ability to appropriate them (Piezunka & Dahlander, 2015). Besides facilitating explicit information transfer, digital technology also enables implicit knowledge flows through sociality in online communities (Faraj et al., 2016a). Moreover, digital innovation is enabled by the interactions that facilitate knowledge re-use (Kyriakou, Nickerson, & Sabnis, 2017) and remix (Stanko, 2016), exemplifying the need to carefully study the key mechanisms pertaining to digital innovation.

Table 1: Four identified major digital innovation discourses

Stream	Focus	Seminal work	Recent examples, Empirical studies	Relevant concepts
Digitalization [Digitality]	Studies of the particular properties of digital technology and their fundamental consequences for organizations	Lyytinen and Yoo (2002), Yoo (2010)	Barrett, Davidson, Prabhu, and Vargo (2015), Jonsson, Mathiassen, and Holmstrom (2018)	Enabling properties: programmability, communicability, etc. of digital technology
[Capabilities] Strategizing	Exploration of capability-building that enables digital innovation and achievement of strategic goals	Sambamurthy and Zmud (2000), Sambamurthy et al. (2003)	Nylén and Holmström (2015), Sandberg et al. (2014), Tripsas and Gavetti (2000)	Recombinant capabilities
Ecosystems & infrastructures	Studies of how digital innovation is impacted by complex socio-technical environments	Yoo et al. (2010)	Lyytinen et al. (2016), Yoo et al. (2010), Svahn, Mathiassen, and Lindgren (2017)	Anarchic innovation networks
Digital innovation practice	Exploration of mechanisms that digital innovation	Leonardi and Bailey (2008), Boudreau (2012)	Faraj et al. (2016a), Kyriakou et al. (2017)	Sociality in online communities

2.2 Digital innovation management

2.2.1 Problem-solution pairing

As illustrated in the preceding section, digital innovation is a complex and multifaceted phenomenon, so managing it poses serious challenges for firms. Thus, various aspects of digital innovation management have been studied including (*inter alia*) its initiation, development, implementation, exploitation, environmental settings, and outcomes. Initiation, which has received the least attention to date (Kohli & Melville, 2019) is described as the definition of a problem to be solved with digital innovation. It can also be seen as capitalizing on digital technology. However, it has been argued that research should focus on a wider process of problem-solution design pairing and analyze digital innovation management "as a sporadic, parallel, and heterogeneous generation, forking, merging, termination, and refinement of problem-solution design pairs" (Nambisan et al., 2017: 226). To satisfy a need, a problem must be formulated, which is expensive and sometimes impossible (i.e. 'wicked') unless the problem formulation step can be bypassed, by firms focusing on connecting needs with solutions, or need-solution landscapes (Von Hippel & Von Krogh, 2015).

The essential process in this regard is searching within parts of the environment that fit the context, which is easiest for agents for whom the need or solution is most relevant: "this contextualization of the activity imbues the person with mindfulness, intent, and selective attention relevant to the need-pair identified" (Von Hippel & Von Krogh, 2015: 215). Jeppesen and Lakhani (2010) have described the process as broadcast searching in a rich solution space. It is also been noted that the pursuit of digital innovation is more complex than previous forms of innovation, and has distinct search paths, one of which is distant search involving recombination (Lopez-Vega, Tell, & Vanhaverbeke, 2016) of existing information (March, 1991). Distant search refers to searching for information that lies beyond a firm's current competences, for example through crowdsourcing (Afuah & Tucci, 2012; Piezunka & Dahlander, 2014). Innovators pursue entrepreneurial opportunities (Shane, 2000). While in the past entrepreneurship was mainly seen as recombination of resources (Schumpeter), more recently attention has shifted to information searching as a key to entrepreneurship (Cooper, Folta, & Woo, 1995). In digital innovation, the ultimate constraint of this searching is not the number of technological recombinations, but the firms' ability to process and appropriate them – a property of recombinant growth (Piezunka & Dahlander, 2015; Weitzman, 1998). Firms need to search for ways to attribute recombined innovations to the environment, for example contextualizing them through narratives (Garud, Gehman, & Giuliani, 2014a).

2.2.2 Recombination

As noted above, recombination is an essential and purposeful activity in digital innovation, and thus critical for digital innovation management. Besides recombination in search, recombination also occurs in use and design are described in the literature. Not surprisingly, however, the terminology used varies substantially as recombination is a powerful concept that describes a major characteristic of technology and innovation (Arthur, 2009). In IS and related disciplines related phenomena are referred to as remixing (Lessig, 2008), reconfiguration (Rai & Tang, 2010), bricolage (Senyard, Baker, Steffens, & Davidsson, 2014), and configurations (Amit & Han, 2017). These terms are important and often connected to a particular research tradition, e.g. bricolage in entrepreneurship research (for general management perspective see a recent review here Savino, Messeni Petruzzelli, & Albino, 2017). For the same reason, this study relies on the concept of recombination (Henfridsson, Nandhakumar, Scarbrough, & Panourgias, 2018). Similarly, in this thesis the term recombination, due to its association with digital innovation research (e.g. Henfridsson et al., 2018).

Recombination in design and creation is tightly connected to the generativity of digital artifacts, which has been a common theme in studies of digital innovation. Broadly, across various literature streams, generativity is described as the "drive to revitalize or rejuvenate; the production of novel configurations and new possibilities; as well as an attempt to challenge the normative status quo" (van Osch & Avital, 2010: 3-4). More specifically, it has been described as uncoordinated groups of participants' capacity for emergent innovation, or any entity's capacity to generate, through adaptability and ease of use, new things that upon distribution are sources of further innovation (Zittrain, 2005). Digital technology transforms linear processes into a matrix (Kallinikos, 2012) that constantly generates new functions and forms, in contrast to the fixed processes of analog technologies. Matrix (or grid) structure allows distribution of agency among multitudes of actors and consequently leads to emergent patterns of innovation activity (Zittrain, 2005). In digital innovation research, generativity is often used as a measure of value creation, a sought-after property of IS (Hukal & Henfridsson, 2017).

Lastly, recent research shows how recombination occurs in use (Henfridsson et al., 2018) and reuse (Kyriakou et al., 2017). Digital resources are reassembled-in-use, establishing new value paths across value spaces (Henfridsson et al., 2018), as firms use assemblages of digital technologies, dynamically changing them to meet immediate needs. In addition, digital innovation also embraces the refinement of "known technological combinations to solve new problems" (Carnabuci & Operti, 2013: 1592), a process also called knowledge reuse. Such recombination follows divergent (from many to one) and convergent (from one to many) recombination patterns (Flath, Friesike, Wirth, & Thiesse, 2017). Recombination-in-use requires capability deepening, i.e. finding and refining technological innovations for new contexts (Carnabuci & Operti, 2013). Community interactions are significant mediators of recombinant reuse (Stanko, 2016), since meanings of the innovations are grounded in practice-oriented communities (Tuomi, 2006), as consistently shown generally for digital innovation (Fleming & Waguespack, 2007).

Despite these insights, there is still insufficient understanding of the main dynamics of recombination outside firms' boundaries (Savino et al., 2017) and need-solution pairing generally (Von Hippel & Von Krogh, 2015). Prior studies have extensively addressed such dynamics in the context of technological patents (e.g. Kaplan & Vakili, 2015). More recently, the ability to tap into online communities (particularly regarding 3D printing and opensource software development) has also triggered investigations of recombination within these interactive spaces (e.g. Flath et al., 2017; Kyriakou et al., 2017; Stanko, 2016). However, the current understanding of digital innovation management must be extended by examining contexts in which boundaries of interaction (and recombination) are ill-defined. While significant interaction between a firm and its surroundings might occur in, say, innovation contests the firm initiates (e.g. Füller, Hutter, Hautz, & Matzler, 2014), much of the interaction is multi-dimensional (Scolari, 2009) and involves heterogeneous types of data and technologies (Sambamurthy et al., 2003). Innovation agency is distributed and is a result of a collective action (Lyytinen et al., 2016).

2.2.3 Distributed innovation in online communities

Distributed innovation is a central trait of digital innovation (Yoo et al., 2012). Digital networks allow for unprecedented connectivity and information sharing, which has generated great enthusiasm in society for distributed communities capable of great innovations (Kelly, 1996). These distributed communities are characterized by self-organization, voluntary participation, and decentralized problem-solving (Lakhani & Panetta, 2007). Various kinds have been recognized (e.g. open source software development, collective intelligence, Q&A communities, and wiki communities). However, the distributed systems of individuals and firms are mostly similar organizationally, with networks facilitating cognitive and social translations (Lyytinen et al., 2016) that enable digital innovation. Open source software (OSS) development emerged as a type of distributed innovation involving the collective development of technological artifacts (Raymond, 1999) that made a significant impact on organization science (Hippel & Krogh, 2003). As one of the earliest and well-studied cases of distributed innovation, it shows how innovation can be facilitated by extremely open development processes – the open innovation paradigm (Chesbrough, 2003). Collective intelligence refers (in this context) to voluntary group efforts to solve a task posed by an organization, often involving complex problem-solving and use of information systems as mediators (Majchrzak & Malhotra, 2013). The locus of the innovation lies not within an individual's interaction with the problem, but in the interactions between members – i.e., community-based generativity (Nambisan et al., 2017; van Osch & Avital, 2010) and "creative abrasion" (Leonard-Barton, 1995). Promoting these processes, whereby participants solve problems through interaction and friction (as opposed to individual problem-solving) is an identified causal mechanism of distributed innovation. However, creating conducive conditions for them remains one of the greatest challenges associated with distributed innovation systems (Majchrzak & Malhotra, 2013). The relational aspect of innovation creation has also been highlighted in studies of the groups' structural properties (analysis of their networks).

Network structure can play a significant role in facilitating digital innovation and solution of complex problems (Grewal, Lilien, & Mallapragada, 2006; Singh, Tan, & Mookerjee, 2011). Gargiulo and Benassi (2000) found that managers with numerous structural gaps in their networks were more adaptive to organizational change, while managers with cohesive networks resisted such adaptation. In addition, individuals occupying central network positions reportedly have higher performance and leadership levels, and most measures of network centrality are positively related to performance in distributed innovation systems (Faraj, Jarvenpaa, & Majchrzak, 2011; Wasko & Faraj, 2005). Few studies, however, have addressed dynamics of different types of networks and their effects on innovation – e.g. contrasting positions in (and dynamics of) social and professional affiliation networks as predictors of performance. Impacts of communication on innovation have been considered for several decades in management research (e.g. Ebadi & Utterback, 1984), but they have only recently been addressed in distributed innovation systems research, and online communities in particular. It is now known that simple and concise language use is associated with higher performance in distributed communities (Johnson et al., 2015). While sociality has been studied before (Faraj, Kudaravalli, & Wasko, 2015), there is a need for explicit inquiry into effects of interaction between positions in different networks on digital innovation.

The dependence of distributed innovation on IS and digitalization is self-evident, but the affordances (functionality that allows action) of such systems involved initially received cursory attention in IS research generally (Orlikowski & Iacono, 2001), and distributed innovation research particularly (Majchrzak & Malhotra, 2013). Subsequent adoption of a systems perspective and analysis of affordances of the systems that enable (or hinder) innovation provided new insights (Zhao & Zhu, 2014). However, the pendulum may have swung too far, with IS studies overemphasizing the role of specific features of digital platforms and neglecting the social aspects of knowledge flows in online communities (Faraj et al., 2016a).

In light of the central role of sociality in online communities associated with a firm's products or services, what can the firm do to manage the process? Research has been quite productive in identifying the reasons individuals initially participate in online communities. Studies, often focused on OSS communities, have identified various motivations, ranging from pursuit of hobbies and altruism (Raymond, 1999) to personal gain (Lakhani & von Hippel, 2003). At first, many participants try to solve a specific problem, improve the service, and/or improve existing practice, then leave, while a small number remain and form the core of the community (Shah, 2006). These individuals are willing to participate when the value generated in the community is perceived as public good rather than only beneficial to the firm (Wasko & Faraj, 2000). Individual contributions also vary greatly, following a power law distribution – most users do not contribute at all, while very few generate disproportionally high contributions (Johnson, Faraj, & Kudaravalli, 2014).

Since online communities act as knowledge exchanges (see discussion on epistemic practice in online communities in Faraj et al., 2016a) their value grows with each new member – exhibiting a network effect (or demand side economy of scale). Thus, maximizing the value of the community depends on the firm's ability to maximize the number of interactions between the members (Van Alstyne, Parker, & Choudary, 2016). Past studies show that inducing interactions and effort in online communities is not a straightforward process. A common approach to incentivize contribution to knowledge exchange can be counterproductive in some groups of users (Zhao, Detlor, & Connelly, 2016). At best such efforts by the firm may be temporary (Goes, Guo, & Lin, 2016). Moreover, when these incentives do work, there is a period of *learning* and decreased performance when the community members struggle to implement what is valuable for the firm in their contributions (Riedl & Seidel, 2018).

There have been inquiries into the explicit possibilities to manage innovation processes in online communities. Conditions that significantly increase performance are task modularity and option value – these factors decrease task complexity while increasing potential gain (Baldwin & Clark, 2006; Lakhani & Panetta, 2007). Functional decomposition must be accompanied by the ability to freely recombine the modules for further innovation. Previous studies both suggest that it may be helpful for firms to pose well defined problems for communities to solve (Blohm et al., 2013), while more uncertain problems are more likely to generate valuable proposals (Boudreau, Lacetera, & Lakhani, 2011). Further, firms need to orchestrate community interactions by building common identity with the community, creating opportunities for innovation and opening firm boundaries (Parmentier & Mangematin, 2014).

To summarize, I suggest there are a number of gaps in previous research on digital innovation management in online communities. First, the emphasis on platform design, while important, has significant limitations and fails to address roles of sociality in knowledge exchange (Faraj et al., 2016a). Second, while users' motivations and ways to incentivize them have been explored, there is still a lack of understanding of the subtleties of how individuals participate and contribute knowledge to online communities (Bogers, Afuah, & Bastian, 2010).

Third, the assumption that online communities are most effective when problems are clearly defined (Blohm et al., 2013) does not hold for the context of digital innovation, since such problems are difficult or even impossible to define in advance (Nambisan et al., 2017; Von Hippel & Von Krogh, 2015). In addition, the capability of crowdsourcing models to capture value beyond simple problem-solving has been recently contested (Bloodgood, 2013). Fourth, the literature suggests that online communities need to be orchestrated (Parmentier & Mangematin, 2014), yet many communities *multihome*, i.e., they are distributed among multiple digital platforms (Mital & Sarkar, 2011) and not under direct control of the firm. Finally, digital innovation in online communities relies on digital trace data, which essentially represents abstract (and consequently malleable) ideas, and has the properties of digital artifacts (Hedman, Srinivasan, & Lindgren, 2013), i.e. programmability, associability etc. (Yoo, 2010). Few studies have addressed this issue, which I review in the next section on digital trace data.

3. Digital trace data

In this section I reflect on the nature of digital trace data – the pervasive phenomenon that has become a conduit for interaction between numerous participants of online communities. A serious discussion is in order to understand how firms can leverage digital innovation within such complex contexts. Below I describe the origins of information research in the IS field. I reflect on the properties of data and its relationship to information. I highlight ways in which data is used and re-used in the innovation process. I address optimism and pessimism regarding the abundance of data and, most importantly, I discuss the significance of data contextualization.

The IS discipline became known in the 1960s for investigating the interplay between organizations and computers (Hirschheim & Klein, 2012), a long tradition began by addressing the ways information is implicated in organizational processes (Keen, 1980) and the role data plays in this interaction. From early works on information modeling (Lyytinen, 1987) to more recent considerations of information taxonomies (McKinney & Yoos, 2010), IS researchers have acknowledged that the adopted views (e.g., perception of data as a *token* or as a *form of translated meaning*) affect how digital systems are designed and used. Moreover, various digital tools have been studied to address the relationships between firms, information and digital technology (Arnott & Pervan, 2014), for example, management information systems (Ackoff, 1967), decision support systems (Gorry & Scott Morton, 1971), executive information systems (Walls, Widmeyer, & El Sawy, 1992), enterprise planning systems (Willcocks & Sykes, 2000), knowledge management systems (Alavi & Leidner, 2001), business intelligence and analytics (Chen, Chiang, & Storey, 2012), and intelligent executive systems (Chi & Turban, 1995). Possibly due to the referential nature of the field (Keen, 1980) the approaches to data and information across these distinct streams have been seen as inconsistent (Mingers, 1995).

Digitalization enables generation of information in unprecedented ways and scales (Yoo et al., 2010). Information is codified and transferred across digital networks, between individuals and firms. Futurists have called data the most valuable resource, projecting its exponential growth and adjusting the trend upwards every year (Gantz & Reinsel, 2012; Varian, 2014). The growth of information is self-referential: not simply representing complexities of knowledge-intense economies in codified form, but further expanding in reference to itself (Kallinikos, 2006). Digital data has been referred to as an essential resource for economic growth by governments and kickstarted “data economy” initiatives (e.g. EU Data policy, published 2017).

Implicitly, value creation occurs when data is re-interpreted and contextualized by firms (Varun Grover & Davenport, 2001). Further, the process of re-interpretation is highly volatile, sustained by the information structures present in the firms. Yet it is also an inward-forming process that leads to changes in organizational structures (Daft & Lengel, 1986) and individual behaviors (Boland, 1987). Information is intersubjective, since its re-interpretation is subject to overlapping meaning structures between a firm and the source (Mingers, 1995). These overlaps may result, for example, from established consensus between industry partners or standards imposed by government institutions. It is a serious question if such consensus is attainable in the context of global interactivity and heterogeneity of data sources used by digital firms.

The meaning data inherently holds can be fixed by an artifact that sediments part of the mental structure into the design (Tuomi, 1999). For example, a reading from a compass is re-contextualized by mental structures inherited from understanding of navigation and cartography. However, it is not clear how far this holds for *digital* artifacts, which are regarded as highly malleable and recombinational resources by IS researchers (Kallinikos, Aaltonen, & Marton, 2013; Yoo et al., 2010). Digital resources are prone to constant change and transfiguration, embedded in highly dynamic ecosystems, repurposed and often emerge as a result of collective distributed activities (Constantiou & Kallinikos, 2015). The embeddedness of mental structures into artifacts is challenged within digital space, in which meaning is often negotiated by infrastructure participants (Tilson et al., 2010), and redefined in-use (Henfridsson et al., 2018). Digital resources, data specifically, are critical for functioning of the digital economy (Surblyte, 2016), cannot be depleted, depreciate rapidly, and constantly renewed (Levitin & Redman, 1998).

Digital trace data is overabundant, yet its potential value depreciates rapidly (Borgmann, 1999). Firms operating in areas characterized by rapid technological development, such as digital startups, need to act fast under high uncertainty and equivocality (McMullen & Shepherd, 2006), where even novel and relevant data might be of limited use (Daft & Lengel, 1986). Furthermore, the organizations mentioned above are becoming more elusive, and their boundaries are not well defined. Through collaborating with customers or coalescing into meta-organizations (Gawer, 2014), or transcending industry boundaries (Nylén, Holmström, & Lyytinen, 2014; Scott & Orlikowski, 2012), firms cease to be information-processing units, as described in early works of organizational scholars (e.g. Galbraith, 1974). Instead, multitude of actors interact with data traces, that traverse fluid boundaries, being recombined, remixed, re-interpreted, and ultimately re-used, by translating information into specific contexts. Often these translations are emergent outside firms' control, as they result from generativity with loci in diverse collectives of heterogeneous actors.

That said, data is not information: information is derived from data that is relevant, accurate and valued in the context it is used in (Tushman & Nadler, 1978). In this sense, contextualization of data is crucial, as it represents the process of translating data into information. Moreover, due to the self-referential constitution of information (Kallinikos, 2006), these translations are enabled (or generated) from data translations either already or becoming in-use. Data contextualization is not a task, but a recursive process of constant recombination and re-use that depends on firms' ability to process and assimilate information. Further, digital trace data is not simply a set of symbols encoded as a result of de-contextualized information (Tuomi, 1999), it is also a set of digital artifacts that are interactive, distributed, editable, and reprogrammable (Kallinikos et al., 2013), which makes digital innovation practicable (Aaltonen & Tempini, 2015)

Recognition of the importance and challenges of deriving information, and thus creating value, from increasing amounts of data is not new. Since digitization's inception, researchers have voiced concerns that "The explosive growth of information technology has not been accompanied by a commensurate improvement in the understanding of information" (Stamper, 1973: 1). IS researchers have addressed the relationship between data and information by articulating hierarchical (data-information-knowledge-wisdom) conceptualizations (Ackoff, 1989; Henry, 1974) – while other exemplary studies have refined the concept (Checkland & Holwell, 1988; Hirschheim, Klein, & Lyytinen, 1995; Langefors, 1973). Some authors argue that use of the information concept in the field is inconsistent (McKinney & Yoos, 2010), but it is typically defined in relation to data, and often conceptualized as "processed data" that is endowed with meaning (Kettinger & Li, 2010). Information is decontextualized through data modelling into a symbolic form by adhering to a defined semantic structure. Data modeling cannot be separated from the underlying meaning structures, which mostly remain "unarticulated" in the data (Tuomi, 1999).

In practice, a widespread approach to digital data traces has been through fixation of analytics. Using a combination of domain knowledge, technical experience, and management consultants' inputs, practitioners have embraced analytics solutions, but underestimated the contextual complexity associated with procedural use of data (Sugimoto, Ekbia, & Mattioli, 2016). This issue is sometimes described as data or quant delusion (Blyth, 2018). Because of its diverse origins, digital trace data is agnostic about the contexts it might be used in (Constantiou & Kallinikos, 2015). That is, the massive datasets available are rarely specifically created for specific firms, however strongly they may be associated with some of the sources. Often, they are generated with a “more is better” philosophy in mind, sometimes including data in its rawest format or using aggregate measures relevant to the data provider, rather than potential user. For example, digital ecosystems that generate massive amounts of data offer it to firms through boundary resources (i.e. APIs) and tend to orchestrate the process according to platform strategy (Ghazawneh & Henfridsson, 2013), which has nebulous relevance for firms outside the ecosystem. The overabundance of data, and its tangential relevance to firms’ contexts, can easily lead to organizational paralysis unless it is confronted with the notion of meaning (Mutch, 1997) through dialogue, contextualization and disentangling friction, without which meaning cannot be explicitly articulated in data (Boland, 1987). Firms need to engage in information actualization, i.e. search for meaning that accommodates spaces for possibilities provided by digital trace data (Aaltonen & Tempini, 2015).

The amount of digital trace data has been growing exponentially, but it is not easy to give a comprehensive account of its origin. Perhaps the main source of growth has been transactional data – information on digitally mediated transactions (Varian, 2010). Amazon’s purchase records comprise an example of such data, which when recombined with the firm’s hardware (e.g. Kindle Fire) and software (AWS) portfolio enhanced performance (Bharadwaj, El Sawy, Pavlou, & Venkatraman, 2013). Other sources of growing data have been emergent networks of machines equipped to collect and transfer it between each other and human operators, often referred to as the internet of things (Saarikko, Westergren, & Blomquist, 2017). For example, data from sensors on industrial machinery is used for remote diagnostics, but also may enable boundary-spanning processes (Jonsson, Holmström, & Lyytinen, 2009; Jonsson et al., 2018).

The emergence of digital trace data has also been presented in a negative light. Some studies report that increases in data may be detrimental to individuals’ (Shenk, 1997) and firms’ performance (Wessel, 2016), as noted in contexts of organizational ethics (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016), innovation (Piezunka & Dahlander, 2015), marketing (Mela & Moorman, 2018), and research (Lazer, Kennedy, King, & Vespignani, 2014). Negative aspects of information overload usually mentioned include imbalances between available information processing capacities or skills and amounts of information supplied (Edmunds & Morris, 2000; O'Reilly, 1980). Consideration of the acceleration in data emergence in recent literature has been somewhat speculative, hence it has been given multiple labels (Eppler & Mengis, 2004), such as data (information) explosion, data growth, data smog (Shenk, 1997), data deluge (Alavi & Leidner, 2001), and infoglut (Andrejevic, 2013). In this thesis, I do not regard emergence of digital traces as a negative phenomenon, but simply treat digital trace data as overabundant (i.e. excessive in quantity). Regardless of claims made by proponents of Data optimism (Kelly, 1996; Kelly, 2017) or Data pessimism (Kling, 1996; Morozov, 2011), firms need to rethink existing approaches to data management and analysis (Jacobs, 2009), IS (Goes, 2013), and management generally (George, Haas, & Pentland, 2014).

4. Research design

In the doctoral work underlying this thesis I set out to answer the following research question: *How can digital innovation be leveraged in the context of online communities?* This is not easy, since it involves tackling the role of an emergent phenomenon (digital trace data generated by online communities) in a broad and heterogenous process (digital innovation). To address this question, my colleagues and I relied on an emerging approach to research design: computational case study. In the following sections I first describe the approach, then introduce the research context and data collection, and conclude with a description of the data analyses involved in the work.

4.1 Computational qualitative case study

Methodologically, information systems research generally - and digital innovation research in particular – has cumulatively achieved significant advances based on case studies (Eisenhardt & Graebner, 2007; Walsham, 1995). The interpretivist qualitative tradition has played a key role in fostering rich theorizing in the discipline (Chen & Hirschheim, 2004). Recent studies have acknowledged the opportunities provided by digital traces and the ways they open up new methodological opportunities (Berente et al., 2018; Nambisan et al., 2017). Digitalization in business and everyday life has brought profound changes in society (Lazer et al., 2009), contributing to novel ways of organizing work, e.g., involvement of distributed actors in innovation processes, coordinated by technological platforms. A profound consequence is that everyday human activities can be digitally traced, stored and analyzed. While such digital trace data is a rich source of insights for researchers to conceptualize social phenomena, it poses formidable needs for them to develop novel ways to conduct research (George et al., 2016; Lazer et al., 2009). Emerging opportunities in data collection have created potential for a paradigm shift to a new methodological tradition, designated computational social science (Chang, Kauffman, & Kwon, 2014). With no well-defined procedures, as yet, this refers to a wide array of approaches that heavily involve use of computational techniques for studying social phenomena.

Previous attempts to utilize digital trace data in combination with computational algorithms to conceptualize social phenomena had limited success because much human behavior does not fit abstract models, so human intervention with more qualitative approaches is required (Kerschberg, 2015). Data collected and generated from various digital sources provides measurements of variables, but with little indication of more intricate mechanisms or processes that help to explain or shape those phenomena. To acquire in-depth understanding of social phenomena, we need to explore the underlying mechanisms that guide processes driving change, so conceptualization of events and mechanisms is of central theoretical importance (Abbott, 1992).

As a result, a new emergent type of research methodology has been proposed, based on computational analysis of digital trace data in conjunction with techniques rooted in the rich qualitative interpretivist tradition. The approach is referred to as computational case study (Lindberg, 2015) and combines using a combination of qualitative case study inquiry (Walsham, 1995) and computational techniques, for example, automated web scraping and text mining (Debortoli, Müller, Junglas, & vom Brocke, 2016), network analysis (Wasserman & Faust, 1994) and psychometric language analysis (Tausczik & Pennebaker, 2010). Such combinations allow the researcher to demonstrate how potential patterns emerge in discourse-driven and direct-problem narratives derived from digital trace data with aid of computational methods and enhanced through the aid of qualitative inquiry (see Lindberg, Berente, & Lyytinen, 2015).

Previous computational case study methodology provided a number of approaches that this work was based upon, including conceptualization of digital trace data and related artifacts using a consistent lexicon to allow "zooming in and out" of the narrative (Gaskin, Berente, Lyytinen, & Yoo, 2014). For example, a series of responses on a forum were not simply treated as text, but rather assigned consistent lexical categories to facilitate analysis regarding the content, types of interaction involved and the digital platform. This also allowed generation of an analytical language that provides scaffolding for understanding relationships between the events and an appropriate vehicle for abstraction from unstructured qualitative data. A major challenge during the process lay in operationalizing concepts through the digital trace data.

The difficulty in assigning meaning to digital trace data (a major theme of the thesis) was an important consideration in the empirical studies. For example, what does a discussion on a support forum mean: Is it part of the knowledge exchange or socialization between professionals? On a more granular level, is a discussion between developers a search for a new solution or re-contextualization of old solutions to new problems? The process is doubly problematic when studying organizational processes, such as digital innovation based on digital trace data. To address these challenges, a recursive process of theorizing and data collection (Levina & Vaast, 2015) was applied. The first step in efforts to understand the nature of the phenomenon, and how it changes over time, involved iterations between identifying concepts and their evolution, and critically relating emergent theory to the wider socio-technical environment.

This recursive and inductive process resembles grounded theory (Strauss, 1987) and not coincidentally, as grounded theory provides foundations for emergent theorizing in computational case studies (Berente et al., 2018). For a detailed description of the process, see Figure 5.

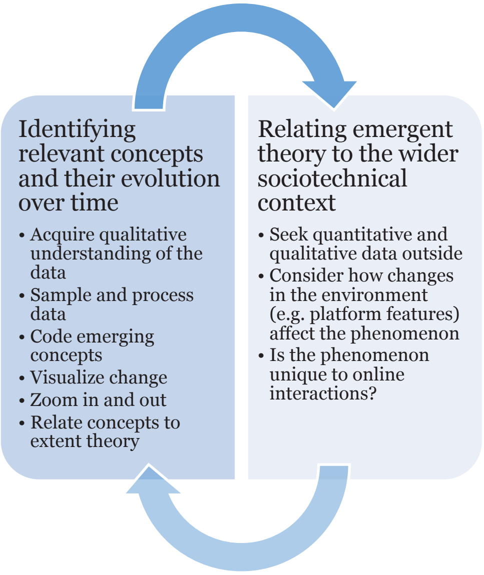

Figure 5: Recursive process of theorizing and data collection (adopted from Levina & Vaast, 2015)

In the work, while studying processes as described above, variance research methodology was also applied. There are fundamental differences in assumptions between the two traditions (Mohr, 1982): Variance research tends to concentrate on outcomes, while process studies often focus on changes leading to them. I take a position that the study of digital innovation requires attention to both elements: the journey *and* the destination, and applied an approach suggested to bridge differences between process and variance research traditions, called hybrid modeling (Ortiz de Guinea & Webster, 2017). In these efforts, a weakly blended approach was applied, combining selected elements from both traditions that seem most suitable for longitudinal studies (Ibid.). The approach also calls for extensive visual mapping – a technique commonly used in process research (Langley, 1999).

Moreover, a number of validity issues need to be considered when working with digital trace data (Howison, Wiggins, & Crowston, 2011), particularly in relation to construct validity: “the degree to which inferences can legitimately be made from the operationalizations in a study to the theoretical constructs on which those operationalizations are based" (Venkatesh, Brown, & Bala, 2013: 33). The validity issues related to digital trace data research are common, and the consensus is to justify one's interpretations through "example-based logical arguments and/or adduced empirical data" (Freelon, 2014: 69). In addition, once again, the methodological discourse focuses on the need to acquire an intimate qualitative understanding of the context in which digital trace data is produced (Geiger & Ribes; Howison et al., 2011; Levina & Vaast, 2015). In the next section I describe the research contexts of the empirical studies the thesis is based upon and the data collection procedures applied.

4.2 Data collection and research contexts

Data were collected in four studies (designated Studies 1-4 and reported in papers designated Papers 1-4, respectively) that focused on three distinct empirical contexts. Studies 1 and 2 both addressed phenomena associated with the business intelligence and analytics firm Tableau Software. Study 3 focused on Pebble Technology corporation and their Pebble Time crowdfunding campaign. Study 4 addressed a cobranding episode involving Manchester United football club and a player, Zlatan Ibrahimović. In the first two studies, a *typical* case selection strategy was applied (i.e., a case regarded as highly representative of a population was selected), while Studies 3 and 4 focused on *extremes* (strongly exemplifying studied phenomena) (Gerring, 2006; Seawright & Gerring, 2008). I elaborate on the selection strategy in the following sub-sections. In all four studies, digital trace data was the main source of empirical material. The empirical cases are overviewed in Table 2, and in the following sections I describe each research context and data collection process in detail.

Table 2: Overview of the empirical cases

Study No.	**1**	**2**	**3**	**4**
Case	Tableau Software		Pebble Technology Corporation	Manchester United FC, Ibrahimović
Selection strategy	Typical	Typical	Extreme	Extreme
Empirical setting	Business intelligence and analytics	Business intelligence and analytics	Entrepreneur, crowdfunding & social media	Social media & sports branding

	user community	software vendor		
Data	Digital traces	Digital traces	Digital traces	Digital traces
Source	Community platform	Corporate website and social media	Kickstarter platform, Pebble App store, Social media, GitHub, corporate blogs, news media	Social media
Data collection period	2016-2017	2017	2016-2018	2016
Investigated period	2008-2016	2016	2015-2016	July 2016

4.2.1 Tableau

As mentioned above, and reported in Papers 1 and 2, the first two studies focused on Tableau Software – a firm that develops and sells Business Analytics and Intelligence (BI&A) software. BI&A is a wide set of tools and techniques grounded in data mining, statistical analysis, and visualization (Chen et al., 2012). I consider it to be an excellent case to study digital innovation since it represents complex software designed for data analysis, which in itself is a generative tool that enables knowledge work[2]. BI&A is represented by growing number of platform developers that enable democratization of data analysis, by allowing business professionals (or anyone with sufficient resources and competence) to tap into extensive organizational data pools (Dinsmore, 2016). Tableau Software was selected as a typical case, i.e. one that "exemplifies a stable cross-case relationship" (Seawright & Gerring, 2008). The main goal in selecting a *typical* case is to find an instance that is representative of a general population. Tableau Software is representative of other BI&A vendors (e.g. Qlik, Microsoft Power BI) in the way it deals with its user communities. Similarly to

[2] I recognize the complexity of studying digital trace data in an online community that deals with data and data analysis, but this is exactly the type of self-referential property of digital technology that makes it a complex and interesting study object. I thank Lars Öbrand for noting that this is an intriguing case of analyzing analysts.

the other market participants, Tableau Software hosts a community platform that provides a space for its users to interact. These interactions focus on three areas: (1) knowledge exchange in relation to the product and service, (2) building industry-specific knowledge groups and geographically defined communities, and (3) improving the software by discussing existing problems, proposing new features, and debating potential feature roadmaps. Tableau Software has been highlighted by industry analysts as a successful example of utilization of a community as an asset driving customer experience and success (Gartner, 2018).

In Study 1, data were collected from the firm's community platform during the period 2016-2017. The data was collected using the Selenium engine (Harrison, 2016), stored and processed with R language and environment for statistical computing (R Core Team, 2015). Collected data covered all community activity during the period from 2008 till 2016. It included rich user profiles, problem submissions, solution suggestions, ideas for improvement, and a community digest (weekly overview of community events). As an additional source of rich qualitative data, several interviews were conducted with community participants and employees. The collected data is summarized in Tables 3 and 4.

Table 3: Data collected in Study 1

Data	**Description**	**Units collected**
User profiles	Profiles of online community users who had posted at least one message to the community.	28'147
Problem submissions	Problems submitted by the users. The data included descriptions of problems that were either solved or not solved.	67'800
Solution suggestions	Solutions to submitted problems. The author of the problem had an option to mark every proposed solution as accepted.	247'442
Ideas for improvement	Ideas submitted by users on ways to improve the software. All ideas were tagged depending of their acceptance	4'657

	level: active, archived, beta, by design, duplicate, in review, planned, released.	
Interviews	Interviews with active online community members and a Tableau Software employee.	4
Community Digest	Newsletter highlighting major news in the community, most popular posts, most active users.	98

Table 4: Data collected in Study 2

Data source	Units collected
Tableau white papers and blogposts	57

Study 2 also addressed phenomena in the same empirical context. The main difference is that Study 1 focused on interactions within the community, while Study 2 addressed Tableau's role as a provider of BI&A platform seeking to deal with data overabundance and extract value from it. The data collected included white papers and blogposts produced by Tableau during the study period.

4.2.2 Pebble crowdfunding

Study 3 focused on Pebble Technology corporation and its crowdfunding campaign for Pebble Time, selected as an extreme case, i.e., one that exemplifies "extreme or unusual values of [a phenomenon]" (Gerring, 2006: 89). Pebble Time is the most successful crowdfunding campaign to date. Pebble Technology was founded in 2012 by Eric Migicovsky, and relied on Kickstarter, the most popular crowdfunding platform, for funding. The "Pebble Time" campaign in 2015 attracted 20 million dollars funding from backers and is an exemplary case of successful use of digital trace data for digital innovation.

Pebble Time gathered the largest number of backers, more than 70,000, and raised more funds than any other Kickstarter campaign. There were 33 updates during the "Pebble Time" campaign, which generated around 4,000 comments, while the campaign itself generated more than 30,000 comments. Pebble Technology has relied on Kickstarter not only for raising funds but also for supporting other complementary products to their smartwatch. Kickstarter, like other crowdfunding platforms, was initially built without much emphasis on feedback mechanisms (Ingram et al., 2014). However, over time there has been increasing effort to incorporate features that support storytelling. Besides Kickstarter, Pebble Technology communicated its story across a number of platforms. Accordingly, as shown in Table 5, extensive data was collected from various digital platforms, including Facebook updates, Tweets, users' and developers' blogs updates, GitHub submissions, and trade news coverage.

Table 5: Data collected in Study 3

Data source	Units collected
Pebble campaign updates	33
Pebble update comments	33'350
Online sources related to mentioned digital platforms (Facebook updates, tweets)	1'580
Online sources related to all mentioned audiences (Facebook, twitter)	1'572
Other documents: Pebble blog updates, Pebble Developer blog updates, Pebble GitHub updates, Pebble subreddit, trade news coverage	150

4.2.3 Sports marketing

The case selected for Study 4 was the cobranding episode on social media between the football club Manchester United and player Zlatan Ibrahimović. The player was signed by the club, but instead of a regular announcement by the club's management to the press the transaction was announced on Instagram by the player. This is regarded as an extreme case, because the co-branding episode was considered an extreme break in protocol and "one of the worst kept secrets in football" (Gaughan, 2016). It illustrated the growing importance of social media for co-branding and paved the way for a new type of extravagant way of announcing player transfers (BBC, 2018).

The data was collected from social media accounts from Manchester United and Zlatan Ibrahimović pertaining to the co-branding episode. Nine posts were collected and, most importantly, responses from online communities the posts generated. In total, 65’555 comments were collected using automated scraping, including both written commentaries and a new type of non-verbal communication prevalent in social media – emojis used by the commentators. In addition, profiles of the commentators were collected, 57’000 in total. The collected data are summarized in Table 6.

Table 6: Data collected in Study 3

Data source	Units collected
Instagram updates	9
Comments	64’555
User profiles	57’000

4.3 Data analysis

All collected data was stored as raw files that were then analyzed with the use of Atlas.ti (Paper 2) and NVivo software (Paper 3) for qualitative analysis and with scripts written in R language for statistical computing (Papers 1 and 3). The data analyses in the four studies are described briefly below and in detail in the appended papers.

In the studies diverse methods were applied that enabled theorization from the digital traces of interactions in the focal online communities. In Studies 1 and 4 various computational language processing techniques were applied. Available algorithms can robustly identify a number of characteristics from text, such as sentiment and topic (Jockers, 2013, 2014). These algorithms, and associated methodology, have been successfully applied in social science research, organizational research (Hannigan et al., 2019), and specifically in analyses of phenomena in online communities (Johnson et al., 2015). Moreover, some of the approaches (such as AFINN sentiment analysis) have been optimized for short communication forms prevalently used in online communities and microblogging (as applied in Papers 1 and 4, respectively), resulting in highly accurate sentiment scoring (Nielsen, 2011). Further, in Study 1 a computerized method for psychometric linguistic assessment (Pennebaker, Francis, & Booth, 2001; Tausczik & Pennebaker, 2010) was used. This allowed scoring and comparison of text in relation to certain psychometric categories, such as *analytical*, *clout*, and *emotional*. Linguistic analysis is a promising element of IS methodology that already has been used for powerful theorizing. For an excellent example of the use of linguistic analysis in IS research, in a study of interlaced knowledge in development of a complex system (ATLAS in CERN) see Tuertscher, Garud, and Kumaraswamy (2014).

To assess how certain events bring change, in Study 1 structural break analysis, also known as Chow tests (Chow, 1960), was used. This allowed analysis of timeseries data and illustration of how an external force (user recognition by the firm) affected users' behavior. This type of econometric analysis has been previously used successfully in IS research in relation to large datasets from online communities (Nan & Lu, 2014).

Social network research methodology (Marsden, 1990; Wasserman & Faust, 1994) was also used for analyzing interactions between participants in online communities and studying network structures of knowledge areas (e.g. Tuertscher et al., 2014). The network paradigm has been productively and increasingly used in organizational research (Borgatti & Foster, 2003), to increase organizations' and individuals' awareness of networks, to aid analysis related to IS use, and to study change associated with technological platforms (Oinas-Kukkonen, Lyytinen, & Yoo, 2010). Network analysis has been used extensively in online community research (e.g. Mislove, Marcon, Gummadi, Druschel, & Bhattacharjee, 2007; Wasko & Faraj, 2005), but validity issues associated with network measures persist (Howison et al., 2011). To address those issues we followed methodological procedures that combine algorithmic network research with extensive qualitative work (Whelan, Teigland, Vaast, & Butler, 2016).

Qualitative analysis was conducted in all four studies (see section 4.1), but in Studies 2 and 3 it was methodologically explicit. It involved an iterative process, starting with open coding that resulted in themes emerging from the data, which were often very descriptive. The broad themes and related literature provided initial scaffolding for developing a general understanding of the investigated phenomena, thereby providing sensitization in the research process (Klein & Myers, 1999). Relevant categories were related to the literature (value of IT and entrepreneurial storytelling in Studies 2 and 3, respectively). Coding was done with the aid of Atlas.ti (Paper 2) and NVivo qualitative analysis software. For specific details regarding data collection and data analysis, please see the cited papers.

5. Summaries of research papers

In this chapter, I summarize the appended papers. The title of each chapter is presented, followed by keywords, a brief background, an overview of the research design, findings and main contributions. An overview is presented in Table 7.

Table 7: Overview of research papers

Paper No.	**1**	**2**	**3**	**4**
Empirical setting	Business intelligence and analytics community	Business intelligence and analytics vendor	Entrepreneur, crowdfunding & social media	Social media & sports marketing
Area of concern	User community innovation	Value of IT	Digital entrepreneurship	Co-branding
Theoretical framework	Knowledge exchange	Competing values	Digital storytelling	Online communities
Methodology	Computational case study	Case study	Computational case study	Computational case study
Data	Digital traces	Digital traces	Digital traces	Digital traces
Research question	What are the relationships between user acknowledgement, knowledge exchange and product innovation in an online user community?	How do big data evangelists construct value propositions related to big data analysis products and services?	How are entrepreneurial stories narrated in the digital age?	What are the challenges and opportunities associated with co-branding between corporate and individual brands on social media platforms?

5.1 Paper 1

Mankevich, V., Lindberg, A., & Holmström, J. **The Acknowledgement Paradox in a User Community Context: Evidence from a Business Intelligence and Analytics User Community**. *Manuscript*.

Keywords: digital innovation, online communities, recombination, knowledge exchange, knowledge creation

The first manuscript presents an investigation of knowledge exchange in a user community context. Online user communities be assets for software companies, as they may provide a space for sharing best practices, supporting new users, solving problems, and crowdsourcing product improvement ideas. While we know that user communities can become hotspots for innovation, it is poorly understood how interventions by the company can affect user leadership, and facilitate or hinder innovation outcomes in a user community context. The reported study focused on a business intelligence and analytics user community, in a forum where professionals solve problems regarding use of the software, socialize, and try to improve their work practices. We studied how participants in the online community created and exchanged knowledge, and how results of these processes were appropriated by the firm supporting the community – Tableau Software.

We describe the acknowledgement paradox – a decrease in product innovation performance if the user is publicly or internally acknowledged by the company. We illustrate how such acknowledgement leads to detrimental product innovation performance and manifests in changes in linguistic indicators of cognitive processes. Public recognition by the company disrupts sociality, which is the critical precursor for user community knowledge creation (Faraj, von Krogh, Monteiro, & Lakhani, 2016b). It triggers shifts in roles – public recognition requires maintenance of status, hence users may shift priorities from demanding innovation activities to sustaining the user community (e.g. attending to new users, maintenance of knowledge repositories, moderation, and evangelism). Associated changes in status also demonstrates an accumulation of online capital (Levina & Arriaga, 2014), as manifested in changes in user influence. Shifts in power structure may result in reductions in justification and contestation, which are required for interlaced knowledge and subsequently complex innovation (Tuertscher et al., 2014). Finally, public and internal recognition are forms of extrinsic motivation that have been shown to be short-lived at best, and in some cases detrimental to user innovation (Goes et al., 2016).

These findings reveal the complex social dynamics that firms must deal with when leveraging innovation in online communities. Interactions manifested through digital trace data rely on platform affordances, hence recombination is greatly affected by the platform design. The study also demonstrated how digital trace data of various forms was used by the firm to appropriate user innovation. By scanning user activity around the feature improvement submissions the firm was able to identify the most viable and critical inputs from the community. Results of the study have implications for research on knowledge flows in user communities and crowdsourcing innovation management.

5.2 Paper 2

Mankevich, V., Lindberg, A., & Holmström, J. (2018). **Better Safe Than Sorry? Investigating Big Data Evangelists and Their Value Propositions**. *Academy of Management Global Proceedings, Surrey (2018),* 58.

Keywords: business analytics, value of IT, technological evangelism, competing values framework, big data

In the second study we investigated value propositions offered by data evangelists – zealous advocates for using digital trace data for business purposes. The objective was to improve understanding of the phenomenon of digitalization that results in overwhelming data generation, from the perspective of the actors who are extensively involved with it – the solution providers. We studied value propositions by the leading provider of data analytics solutions – Tableau Software (Gartner, 2018), as expressed by its white papers and blog posts. Understanding how the value of their product was communicated was important for articulating opportunities and challenges of practitioners who were still learning how to innovate with data and analytics (Ransbotham et al., 2016). In particular, we applied a Competing Values Framework (Quinn & Rohrbaugh, 1983) – a theoretical framework used to assess the subject of the study against fine-grained value categories. In doing so, we addressed the call for diverse IT value research (Kohli & Grover, 2008) and revealed which types of value are communicated in the data evangelism space.

The results were puzzling. One of the most prominent big data evangelists emphasized the “safest” (operational efficiency) and most common type of value driven by data analysis (Philip Chen & Zhang, 2014). In contrast, contributors to the academic discourse (e.g. George et al., 2016; Günther et al., 2017; Leonardi & Contractor, 2018) tend to describe much more exciting and optimistic visions of the future, including capabilities for flexibility, experimentation, collaboration and organizational cohesion. This divergence shows the complexity of working with analytics – solution providers are often the most enthusiastic evangelists, but they also have a product to sell, so their "gospel" must be connected to specific features and hence show a more grounded vision of the potentially generated value. The results reflect immaturity (and potential) of the field to realize opportunities that digitalization is bringing about, particularly in relation to external flexibility value categories. The findings highlight limitations of the software solutions for dealing with digital trace data. We suggest that practitioners are missing the most significant potential for data – to assist organizations in developing information capabilities (Kohli & Grover, 2008) related to flexibility, experimentation, ideation, and dynamic problem-solution pairing (Nambisan et al., 2017). These capabilities are needed to deal with the overwhelming digitalization of today and the future (Yoo et al., 2012).

5.3 Paper 3

Tumbas, S., Mankevich, V., & Holmström, J. **"Stories within Stories": Digital Multiplicity for Entrepreneurial Storytelling.** (Under review by an international journal)

Keywords: digital entrepreneurship, entrepreneurial storytelling, digital storytelling, digital platform, online community, narrative

New firms rely on online communities to raise resources, establish legitimacy, and access partners and customers. In some cases, online communities play a central, critical role in this process. Notably, when handling crowdfunded products enabled by digital innovation, entrepreneurs rely on digital platforms to tell stories to associated communities in efforts to facilitate generation, dissemination and realization of their business ideas. Entrepreneurial storytelling is a constantly revised process with conflicting goals that are continuously reconciled. Increasing digitalization permeates all stages of new ventures' evolution, including the most critical phases such as inception and growth.

A key feature of digital entrepreneurial processes is oscillation between various digital platforms, between distinct online communities, making use of digital trace data. To explore how "less bounded" entrepreneurial processes unfold for new ventures seeking resources to survive in digital ecosystems, we analyzed entrepreneurial narratives of "Pebble Time": the highest earning Kickstarter campaign to date. Using various digital trace data sources, including from the Kickstarter platform, we followed activities of the firm over the course of the campaign. We also drew on complementary bodies of literature that were not previously coherently linked, covering entrepreneurial storytelling, digital entrepreneurship and digital storytelling.

Based on our findings, we identified three core processes of entrepreneurial storytelling that new ventures need to balance in the digital age. First, the crossing of borders of single digital platforms and transition to multi-platform engagement (via social media, microblogging or even open source platforms) which generates and recombines digital trace data from various sources. Second, the engagement of various online communities (both existing and emergent) in the storytelling process. Third, the blurring of borders between narratives and mobilization for action. We propose the concept of digital multiplicity to capture the processes and suggest that all three manifest through mechanisms of digital action – push, pull, and nudge. Digital actions facilitate generation and reuse of digital trace data in service of the entrepreneurial narratives, part of the balancing act new ventures must engage in when constructing a successful story in the digital age.

5.4 Paper 4

Mankevich, V., Holmström, J., & McCarthy, I. P. (2019). **Why Zlatan Ibrahimovic is Bigger Than Manchester United: Investigating Digital Traces in Co-branding Processes on Social Media Platforms.** *Paper presented at the 52nd Hawaii International Conference on System Sciences.*

Keywords: computational methods, digital traces, social media management, social media, co-branding, social media management in big data era, digital and social media

In the fourth paper we present a study examining the co-branding activity in online communities. We investigated co-branding on a social media platform – Instagram – involving the football club Manchester United and football player Zlatan Ibrahimović. Such use of digital trace data (on social media) for creation of business value is of interest, especially since in the investigated case the activity involved the most valuable athletic brand and a famous sports personality, yet still the activity was considered a fiasco. While there were many ways to interpret the failure to deliver on the promise of co-branding, we highlight failure of the participating actors (the club and player) to capitalize on the emergent activities from the online communities – the recombinant potential was never fulfilled. We demonstrated how the football club failed to capitalize on co-branding activity as measured through consolidating the audience, generating consistent emotional response, and creating a coherent message. In terms of digital innovation management, the case demonstrates how the lack of active positioning by firms can hinder materialization of value from the activity.

The study also illustrates how digital trace data can be used in research. In the study only one type of digital trace data was used – comments regarding Instagram posts. However, it is rich in meta information (time, authorship, sequence) and content, as Instagram is a textual and non-textual/symbolic communication medium. Consequently, a number of analytical methods were applied, including sentiment and emotional tone analysis, as well as assessment of intersection of the audience and illustrated non-verbal communication used by social media users. Even in the absence of rich contextual information, approaching digital trace data from various angles (sentiment, emotional tone, emoji analysis) provides a basis for corroboration (i.e. finding consistent evidence) for theoretical propositions. Finally, this paper contributes to social media management research by illustrating the difficulties associated with co-branding between personal and corporate brands as well as asynchronous communication. Further, our use of digital trace data and computational analysis illustrates how access to social media can illuminate research activities and provide insights about branding communication.

5.5 Other work

Besides the studies described in the appended papers, I have participated in a number of research projects during my doctoral work. For thematic and space limitation reasons they are not included in this work. However, they played an important role in the journey that is a doctoral education. Results of one of these projects were published under the title "The Hero With a Hundred Faces: The Role Of Coopetition in Innovation Network Evolution" (Mankevich, Skog, Wimelius, & Holmström, 2015). The project allowed me to explore the collaborative and competitive relationships firms must engage in to prosper and innovate in the digital age. The next project, reported in "Gateways to Digital Entrepreneurship: Investigating the Organizing Logics for Digital Startups" (Mankevich & Holmström, 2016) explored how digital artifacts firms deal with affect entrepreneurs and their supporting ecosystems – business developers and incubators. Finally, in a project entitled "Teaching Data Science the Interdisciplinary Way: Learning Cycles and Diverse Skillsets" (Mankevich & Sandberg, 2018) I explored how firms' new circumstances are generating needs for a new type of skillset – data science and change management – and together with my co-author reflected on how academia can provide them to future professionals.

- Mankevich, V., Skog, D., Wimelius, H., & Holmström, J. (2015). The hero with a hundred faces: the role of coopetition in innovation network evolution. Presented at the Strategic Management Society (SMS) Annual International Conference, The Interplay of Competition and Cooperation: Workshop and Discussion Forum.

- Mankevich, V., & Holmström, J. (2016). Gateways to Digital Entrepreneurship: Investigating the Organizing Logics for Digital Startups. In Academy of Management Proceedings (Vol. 2016, No. 1, p. 13995). Briarcliff Manor, NY 10510: Academy of Management.

- Mankevich, V., & Sandberg, J. (2018). Teaching Data Science the Interdisciplinary Way: Learning Cycles and Diverse Skillsets. Presented at The 6th Swedish Workshop on Data Science.

6. Discussion

Firms and online communities interact, and these interactions may lead to innovation, but not necessarily. Innovation is an essential element of firms' survival (Schumpeter, 1934), so they need to manage the process in great detail. However, interactions are complex: they take diverse forms, such as conversations on social media and problem-solving submissions in professional forums, and the results may vary from meaningless noise to critical insights for product or process development. Firms that want to manage the process need to find a way to interact, facilitate, and capitalize from the interactions in online communities.

Research gives some indications of ways to innovate with online communities. The discourse on open innovation suggests a new paradigm of organizational innovation that spans firms' boundaries, is open to external inputs (for example through licensing and acquisition) and ready for spinning out and divestment (Chesbrough, 2003). The insights from research on open source software development give indications of the variability of individuals' motivations to participate (Hertel, Niedner, & Herrmann, 2003), and that if firms want to draw value from communities they require utilization and governance mechanisms (Bogers & West, 2012; Reischauer & Mair, 2018). Such research has also provided significant insight into knowledge creation within communities themselves (e.g. Lindberg et al., 2016), prior to appropriation by the firm. Literature on innovation contests provides guidance for organizing digital innovation management to address predefined problems that require solutions from a wide range of potential solvers (Majchrzak & Malhotra, 2013). Finally, crowdsourcing literature (diverse as it is) shows the particularities of digital innovation management concerned with large numbers of entities, who have diverse motivations and generate inputs that require significant filtering (Piezunka & Dahlander, 2014).

While past studies have illuminated processes of digital innovation in distributed contexts (Lyytinen et al., 2016), there are still a number of challenges. Management of digital innovation is difficult: innovation is unbounded, innovation agency is not predefined, and the process and outcomes are complex and difficult to demarcate (Nambisan et al., 2017). Moreover, while digital innovation research has extensively investigated digital innovation in large firms (Svahn et al., 2017), distributed contexts such as online communities present a new set of challenges. They are elusive: they exist but are difficult to define (Baym, 2007), and valuable yet vulnerable (Barrett, Oborn, & Orlikowski, 2016; Garcia et al., 2013). Online communities are known to consist of heterogeneous crowds of users in pursuit of individual goals (Faraj et al., 2011).

We can observe their interactions through the traces they leave – digital trace data (Hedman et al., 2013) – which is generated in great amounts but is very difficult to interpret (Howison et al., 2011). Firms seeking to manage digital innovation in online communities need to understand digital trace data in a recondite process of meaning-making (Boland & Tenkasi, 1995). The difficulty is exacerbated by the fact that firms and the communities are often not situated in one place online – like a digital platform or a website – instead they spread across multitudes of "digital homes" (further on multihoming see Mital & Sarkar, 2011). It is not clear how firms could proactively facilitate digital innovation processes when users are situated on different platforms, motivated by distinct factors, have various levels of expertise, and the ultimate innovation goals are poorly defined (if at all). To tackle these challenges, in my doctoral work I addressed, with colleagues, the following research question:

How can digital innovation be leveraged in the context of online communities?

The four appended empirical papers address the question from different perspectives. Broadly, the thesis contemplates firms' purposeful recombinant reflection upon digital trace data as part of their digital innovation management practice. In the following sections, I discuss the three main contributions of the work. First, and most importantly I discuss the ways digital innovation is leveraged in online communities. Further, I reflect upon the nature of digital trace data and the difficulties associated with its appropriation for digital innovation management. Finally, I present the work's methodological contribution by reflecting on computational social science and computational qualitative case studies. The last section addresses limitations of the presented work and directions future research.

6.1 Leveraging digital innovation

In the introduction to this dissertation I suggested that firms that strive to promote digital innovation in online communities face a unique set of challenges. Digital innovation is complex, distributed, and unbounded (Nambisan et al., 2017). It relates to problems that are poorly defined (if at all) and constantly change with the changing socio-technical landscape. This raises questions about the optimal ways to harness online communities that are fluid, hard to define, heterogeneous and distributed among many digital platforms to tackle such problems, for which existing forms of interaction with online communities seem ill prepared. Take crowdsourcing – sourcing explicitly articulated tasks to a "crowd" of users (Howe, 2006). While clearly valuable for user feedback and input, the approach is not sufficient in itself for digital innovation and has been recently contested (Bloodgood, 2013). Firms need to embrace dynamic need-solution pairing, i.e. processes in which needs are dynamically approximated with possible solutions without explicit problem formulation (Von Hippel & Von Krogh, 2015). Overall, problem formulation might be completely unattainable in this context. A firm's goal in such cases is to provide opportunities for as much as possible generation (Zittrain, 2005), recombination (Henfridsson et al., 2018), and remixing (Stanko, 2016) of interactions (and resulting digital trace data), thereby enabling more effective pairing. Firms need to strive to maximize the volume of relevant community interactions (Van Alstyne et al., 2016), since these online communities are powered by demand side economies of scale (Nambisan & Zahra, 2016), i.e., the value of a community sharing a digital platform grows with each new participant (Singh et al., 2011). Hence the new mission of the management process is to increase the volume and quality of interactions between the members.

Firms can enable such generation, recombination, and remixing in online communities in a number of ways. They might provide opportunities for divergent (from many to one) and convergent (from one to many) recombination patterns (Flath et al., 2017) or apply internal capabilities to create and reuse technological combinations (Carnabuci & Operti, 2013). However, recombination outside firms' boundaries is still poorly understood (Savino et al., 2017). Paper 4 describes a case where potential for recombination was not fulfilled. I suggest that at least three factors should be considered to promote relevant recombination for digital innovation in online communities. The first is feature design, which a number of studies have linked to aspects of recombination, notably how platform features can incentivize interaction through sharing and gamifying (Morschheuser, Hamari, Koivisto, & Maedche, 2017), and creating social hierarchies (Goes et al., 2016). While these affordances create value for the online communities, their capacity to facilitate digital innovation might be limited. Dedicated features can act as important cognitive and social translations in innovation form when the nature of knowledge is largely homogenous, such as in dedicated R&D projects and decentralized open-source communities (Lyytinen et al., 2016). Paper 1 partly demonstrates elements of this innovation form. However, for cases where the knowledge resources needed are heterogeneous, platform design cannot accommodate all the necessary options. This is one of the reasons why cognitive and social translations in radical innovation circumventing standard feature design are facilitated by multiple platforms (Mital & Sarkar, 2011), through duplicate systems (Wimelius, 2011), recombined in-use (Henfridsson et al., 2018), and driven by workarounds (Petrides, McClelland, & Nodine, 2004). Paper 3 illustrates the point, by showing that instead of sticking to one digital platform and its features, Pebble Technology "followed" the online community making use of features that facilitated recombination of digital trace data via narratives. I suggest that affordances of the platforms are not sufficient to guide actions of users in pursuit of digital innovation.

The social dynamics of the online community must also be considered to promote recombination effectively. Digital innovation enabled by recombination of digital trace data shifts the center of influence from the firm to the interactive emergent collaborative structures (Von Hippel, 2005). Such structures necessitate careful balancing between the diverse goals of heterogeneous actors: individual users, the firm, partners etc. (Shah, 2006). These structures are complex: they take different forms, such as conversations on social media or problem-solving submissions on professional forums, and the results may vary from meaningless noise to critical insights for product or process development (Von Hippel, 2001). Firms that want to manage the process need to find a way to facilitate and capitalize from the interactions in these structures. Researchers (e.g. special issue introduction Faraj et al., 2016a) have explicitly emphasized the importance of sociality in online communities and its impacts on knowledge flows. Paper 1 illustrates the delicacy of the balance in the community and how even a slight intervention can affect recombination processes on a digital platform and its consequences for digital innovation outcomes. The findings support previous studies on the role of firms' incentive hierarchies (or public recognition of users' efforts) in facilitating innovation (Goes et al., 2016; Riedl & Seidel, 2018).

Lastly, I suggest that one of the most significant ways for a firm to promote recombination in online communities is to craft a narrative that enables digital innovation. Paper 3 demonstrates how a firm can enable digital innovation by weaving a narrative that builds on digital trace data scattered among multiple digital platforms. Much of the interaction between firms and online communities is unstructured, the process is sporadic, ongoing, and non-linear. Paper 3 describes a notion of multiplicity, the storytelling process that firms can employ to interact across platforms, using various media by exploiting an assemblage of digital actions. The actions (nudge, push, and pull) allow firms to confront meaning of digital trace data, to trigger recombination and further remixing. In doing so, narratives that are essential for a sense of purpose, direction, and legitimacy of the firm (Garud, Schildt, & Lant, 2014b) are further created in the process of immediate and ongoing storytelling. In contrast to the platform affordance perspective, this accommodates the fact that the firm is navigating a heterogeneous knowledge landscape that is unlikely to fit the constraints of a single platform or feature assemblage. This sharply contrasts with an innovation network orchestration perspective (Dhanaraj & Parkhe, 2006; Parmentier & Mangematin, 2014), since firms adopting narrative development strategies in online communities eschew a commanding role, which is often unfeasible in structures of heterogeneous knowledge (thus the name anarchic network form by Lyytinen et al., 2016). This form of digital innovation is disruptive, radical, and expected to become even more prevalent in the future.

To summarize, I have described three factors that should be considered in digital innovation management in online communities: platform design, sociality dynamics, and narrative development through multiplicity of digital actions. This should enable firms to create opportunities for recombination, remixing, and enhancing the generativity of digital trace data and further opportunities for digital innovation.

6.2 Digital trace data and the problem with metrics

In this section I argue that digital trace data is a new phenomenon that has profound implications on digital innovation management. I discuss the practices of deriving meaning from digital trace data and reflect on the risks of its metrification and quantification.

I set out (with my colleagues) to explore how digital innovation management can be leveraged in the context of online communities. Very soon, however, there was a realization that this discussion is impossible without paying close attention to digital trace data. In a sense, digital innovation manifests in online communities through digital trace data: it is one of the few conduits (or even only one) through which firms interact with online communities. Moreover, accumulating digital trace data may in itself be the goal of a firms' digital innovation management, as in open source software projects. Digital innovation management in the context considered here intimately interacts with digital trace data.

The concerns regarding overabundance of information in organizations have been voiced before (Stamper, 1973) and are at the core of the IS discipline. But what can be said about digital trace data? Its nature is paradoxical. Digital trace data is unexpendable, it can be reused again and again, and its marginal cost approaches zero (Shapiro, Carl, & Varian, 1998). However, it is also perishable – its value depreciates quickly (Borgmann, 1999). Digital trace data is unique – it is tagged and contains meta-information that sets it apart from other data, like geo-location, authorship information, timestamps, and platform-specific information (Hedman et al., 2013). At the same time, it grows in reference to itself (Kallinikos, 2006), i.e., it self-referentially builds upon itself (as in re-tweets and Wikipedia edits). For this reason, digital trace data, besides carrying direct information about an online activity (e.g., a submission in a crowdsourcing forum), makes innovation practicable (Aaltonen & Tempini, 2014). However, challenges faced in digital innovation management (e.g., high uncertainty and equivocality) cannot be met by simply acquiring more data (Daft & Lengel, 1986)!

Paper 3 illustrates how digital trace data is used in the process of digital innovation management. First, it is used as source, trigger, and clout in the digital innovation process: firms act upon users' submissions, but also guide innovation processes with their own actions, which in turn are recorded and the records are left online. There is a recombinant cycle between digital trace data left by all process participants (e.g. firms and users) that can have unintended consequences (as also shown in Paper 3), such as improving innovation outcomes or disrupting sociality needed for it (Faraj et al., 2016b). Firms confront digital trace data with meaning, but the process is complex and highly contextualized (Boland & Tenkasi, 1995). A key question is how can such contextualization or actualization (Aaltonen & Tempini, 2014) be realized in the high uncertainty situations of online communities? Deliberation, matching, and remixing activities, such as those initiated by Pebble Technology and co-created with the users, are essential. By developing and trying out competing visions (narratives) against the online community vision, firms seek actualization of digital trace data for digital innovation. Firms are capable of developing, and testing, many interpretations simultaneously, until the most viable of the competing visions emerge, while some remain parallel, and are further developed by the firms in the form of multiple distinct narratives feeding into various identities (Herzenstein, Sonenshein, & Dholakia, 2011). This process of digital multiplicity, paralleling and cross-validating digital narratives can be set to project viability of the venture (Garud et al., 2014b), but also to manage the process of digital innovation in distributed contexts relying on heterogeneous digital trace data (Paper 3).

An alternative approach firms apply is to use various configurations of digital trace data to triangulate it in efforts to obtain meaningful insights. For example, firms might use combinations of likes, comments, votes, and uploaded examples to assess the value of submissions (Paper 1). The important point is that digital trace data is not a viable metric for seeking meaning in itself (Mela & Moorman, 2018), particularly regarding digital innovation. While various platforms seek to standardize measurements of online communities' activities (Van Alstyne et al., 2016), such straightforward measurements fall far short of communicating the rich details present in (and required from) these interactive environments. Granted, certain metrics based on digital trace data have merit, such as parameters of platform engagement, match quality, or interaction failure, signaling the health of the online community (Ibid.). Wide arrays of affordances might help complex information to flow (even tacit knowledge - Faraj et al., 2016b), but as yet even top firms fall short in attempts to exploit these built-in platform affordances (Paper 4). While digital trace data can provide productive representations of individual users' activities and identities, using it to evaluate firms' actions and intentions might be challenging as digital platforms are poorly adapted for communicating complex identities of corporate entities. The challenge is salient in cases when there are few digital traces and the firm has little opportunity to contextualize them for a wide community (Paper 4).

Because of the abundance of digital trace data, firms and researchers are tempted to employ metric mentality when considering it. This brings the discussion to digital metrics – standardized systems of measurement based on digital trace data. Analytics literature has suggested metrics will improve innovation outcomes (Davenport, 2006), and can be tools for digital innovation management and research (Agarwal & Dhar, 2014). However, it is used as a catch-all term for rallying companies to use quantitative approaches, collecting data and hoping that innovation will emerge. Staggering numbers of analytics projects fail to deliver on this promise (Ransbotham et al., 2016). There are two challenges with using digital trace data for metrics, and more broadly analytics. First, metrification of rich digital trace data is very difficult. For example, how can a meaningful metric be derived from a message left on a support forum, or a rich user submission to improve software be abstracted? Much of the context will be lost in most abstracted versions of such traces. The second problem with metrification for digital innovation is that it involves transitioning to a data-driven innovation process, which might require significant organizational changes (Dremel, Herterich, Wulf, Waizmann, & Brenner, 2017). So far, organizations have widely defaulted to using analytics for the "safest" projects, such as reporting and accountability (Paper 2), attending to data with "stable" meanings (see semantic closure in Aaltonen & Tempini, 2015). Creation of value through use of data (including digital trace data) requires firms to move beyond the status quo of analytics: accountability and control. It requires firms to embrace experimentation (Pavlou & El Sawy, 2010) and a level of openness that will challenge many processes and organizational identities, but ultimately open data access to distributed structures (external actors and artifacts) required for digital innovation (Lyytinen et al., 2016). In the spirit of open innovation, firms have already opened up their innovation processes (e.g. for user contributions about their products). Now the industry needs to do the same with the wealth of data – open it for recombination and contextualization in the wider "generative matrix" (Kallinikos, 2012).

Firms and researchers need to have a nimbler approach to understanding digital trace data. Obsession with quantitative data is problematic (Wessel, 2016), especially if a firm sees quantities of digital trace data as straightforward metrics. Quantification of social phenomena is a pervasive trend (Mau, 2019), present in almost all social spheres, *inter alia* education (West, 2012), healthcare (Lin, Chen, Brown, Li, & Yang, 2017), and HR settings (Leonardi &

Contractor, 2018). However, obsession with numbers without seeking deep meaning is partially to blame for The Great Recession of the late 2000s as numerous financiers fell into the quant delusion (Blyth, 2018). Data generally, and digital trace data particularly, can leverage digital innovation management, but one must always consider Goodhart's law – as soon as the metric becomes a priority target, it ceases to be a good measurement and becomes void of its original meaning (Freeman & Soete, 2009; Goodhart, 1975).

6.3 Studying digital trace data

IS research generally and digital innovation research specifically, has made significant advances using case study research (Eisenhardt & Graebner, 2007; Walsham, 1995). The interpretivist qualitative tradition has provided important foundations for rich theorizing in the discipline (Chen & Hirschheim, 2004), but the rampant digitization of society obliges us to look beyond tried and tested methodological traditions. Everyday human activities can now be digitally traced, stored and analyzed. Digital trace data provides astonishing opportunities for research, but requires serious development of research methods that are adjusted to the particulars of this new type of data (George et al., 2016). There is a need to integrate approaches that facilitate analysis of large volumes of digital trace data, while retaining a serious focus on understanding the processes involved in their generation (i.e. underlying activities). Moreover, as it transforms confronting logics of innovation (being unbounded, with undefined agency, and processes meeting outcomes) digital innovation requires new methods of scientific research (Nambisan et al., 2017). In part, this requirement has been addressed by a new methodological tradition labeled computational social science (Lazer et al., 2009), which involves computational data processing and algorithmic pattern recognition and analysis.

While computational social science is a methodological advance that is pushing the discipline further (Chang et al., 2014), there have been a number of challenges. These include the difficulty of analyzing complex and messy social phenomena, and tendency of researchers to oversimplify (naturalize) these complex relationships, thereby curtailing the search for meaning (Törnberg & Törnberg, 2018). Ensuring construct validity (Howison et al., 2011), i.e., checking that digital trace data represents constructs that researchers claim (or simply hope) it represents (Venkatesh et al., 2013) poses further problems. Computational studies must confront the meaning of digital trace data, and seek conceptualization of the events and mechanisms it records (Abbott, 1992), which requires intimate qualitative understanding of the context(s) in which data is generated (Levina & Vaast, 2015).

The work underlying this thesis demonstrates how the challenges outlined above can be addressed. Computational processing of digital trace data (Lazer et al., 2009) must be combined with deep qualitative inquiry (Walsham, 1995). Moreover, it demonstrates how computational approaches can be combined to provide a triangulation procedure, by combining automated web scraping and text mining (Debortoli et al., 2016), network analysis (Wasserman & Faust, 1994), psychometric language analysis (Tausczik & Pennebaker, 2010), and structural breaks analysis (Chow, 1960). Perhaps one of the most challenging and important processes in this regard is creation of a consistent lexicon to allow "zooming in and out" of the narrative (Gaskin et al., 2014) and testing of relationships between constructs. For example, in the study reported in Paper 1, various submissions were assigned constructs such as product innovation and knowledge exchange, based on qualitative analysis of the Business Analytics forum. Qualitative inquiry is critical for establishment of such constructs.

Interaction with the context is crucial for theorizing from digital trace data. Researchers need to work recursively, oscillating between identifying relevant concepts and their evolution over time, and relating emergent theory to the wider socio-technical context (Levina & Vaast, 2015), as illustrated in Paper 3. Because of the data-driven open-ended theorizing, this recursive and inductive process resembles grounded theory (Strauss, 1987). Such similarity is not coincidental, since grounded theory method is a foundation of emergent theorizing (Berente et al., 2018). An important distinction is the use of both researchers' and machine capabilities for analysis (or pattern recognition) to draw the most reasonable inferences from available digital trace data. This *strategy* or *approach* is congruent with abductive inquiry of the pragmatist philosophical tradition (Lindberg, Forthcoming).

Further, studies dealing with large volumes of digital trace data while focusing on a process are likely to also apply variance research methodology. There are fundamental differences in assumptions between the two traditions (Mohr, 1982), which have occupied much space in methodological discussion related to the discipline (e.g. Sabherwal & Robey, 1995). While variance research tends to concentrate on outcomes, process studies often focus on the changes involved. As previously mentioned, I hold that the study of digital innovation requires attention to both elements: the journey *and* the destination. Researchers can apply hybrid modeling (Ortiz de Guinea & Webster, 2017), combining a few elements from variance and process analysis, to bridge the differences between the research traditions. To communicate the combination, visual mapping can be used (as illustrated in Paper 3): a technique commonly used in process research (Langley, 1999).

To summarize, I pose a necessity to integrate various approaches when probing the nature, roles and utility of digital trace data that facilitate analysis of both large volumes of data and the processes involved in their generation. I suggest that computational social science needs to incorporate much more qualitative research methodology to enable rich understanding of the socio-technical contexts of the studied phenomena. Recursive process theorizing and human-machine pattern recognition are promising approaches in this regard (Levina & Vaast, 2015; Lindberg, Forthcoming). Further, I demonstrate various ways to combine several computational approaches for analysis and triangulation. To my knowledge, the studies reported in the appended papers are some of the first to combine structural break analysis with psychometric text analysis in IS research. These contributions extend a much needed methodological toolkit for digital innovation management research (Nambisan et al., 2017).

6.4 Limitations and further research

The work this thesis is based upon, and the thesis itself, have a number of limitations imposed by the choices made in relation to theoretical framing, empirical settings, and methodology. To address the research question, I have reflected (in conjunction with colleagues) in-depth on problem-solution paring and recombinant innovation in online communities in efforts to understand digital innovation. I present the logic applied in reaching the decisions in previous sections, but acknowledge the associated limitations. As suggested by previous researchers, digital innovation (in addition to recombination and distributed innovation) is deeply rooted in platformization (Yoo et al., 2012). Due to inevitable limitations in scope, platformization has not been explicitly discussed in relation to digital trace data, which presents an opportunity for future research. Empirical elements of the work focused on contexts considered rich in digital trace data – online communities interacting on dedicated platforms and in social media. Future research will benefit greatly from studying digital innovation in contexts of online communities from the internal perspective of a firm, particularly how organizations combine meaning-making with computational methods when dealing with digital trace data in their organizational processes. Methodologically, the work has limitations in its reliance on available records of interactions by focal actors. Future research can generate important insights using interpretative case studies (Walsham, 1995) obtained through working directly with investigated firms. Also, due to the blended approaches applied, the computational potential for analyzing data was realized to a limited extent. Future research should continue the emerging trend of applying rigorous algorithmic methods (Hannigan et al., 2019; Lindberg, Forthcoming). I hope that reflections presented in this thesis will inspire future researchers to consider carefully the limitations of these methods, while seriously addressing digital trace data.

References

Aaltonen, A., & Tempini, N. 2014. Everything counts in large amounts: a critical realist case study on data-based production. ***Journal of Information Technology***, 29(1): 97-110.

Aaltonen, A., & Tempini, N. 2015. Everything counts in large amounts: a critical realist case study on data-based production. In L. P. Willcocks, C. Sauer, & M. C. Lacity (Eds.), ***Formulating Research Methods for Information Systems: Volume 2***: 325-357. London: Palgrave Macmillan UK.

Abbott, A. 1992. From Causes to Events - Notes on Narrative Positivism. ***Sociological Methods & Research***, 20(4): 428-455.

Ackoff, R. L. 1967. Management misinformation systems. ***Management Science***, 14(4): B-147-B-156.

Ackoff, R. L. 1989. From data to wisdom. ***Journal of applied systems analysis***, 16(1): 3-9.

Adner, R. 2006. Match your innovation strategy to your innovation ecosystem. ***Harvard Business Review***, 84(4): 98.

Afuah, A., & Tucci, C. L. 2012. Crowdsourcing As a Solution to Distant Search. ***Academy of Management Review***, 37(3): 355-375.

Agarwal, R., & Dhar, V. 2014. Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. ***Information Systems Research***, 25(3): 443-448.

Alavi, M., & Leidner, D. E. 2001. Knowledge management and knowledge management systems: Conceptual foundations and research issues. ***MIS quarterly***: 107-136.

Amit, R., & Han, X. 2017. Value creation through novel resource configurations in a digitally enabled world. ***Strategic Entrepreneurship Journal***, 11(3): 228-242.

Andrejevic, M. 2013. ***Infoglut: How too much information is changing the way we think and know***: Routledge.

Arnott, D., & Pervan, G. 2014. A critical analysis of decision support systems research revisited: the rise of design science. ***Journal of Information Technology***, 29(4): 269-293.

Arthur, W. B. 2009. ***The nature of technology: What it is and how it evolves***: Simon and Schuster.

Baldwin, C. Y., & Clark, K. B. 2006. The architecture of participation: Does code architecture mitigate free riding in the open source development model? ***Management Science***, 52(7): 1116-1127.

Barrett, M., Davidson, E., Prabhu, J., & Vargo, S. L. 2015. Service Innovation in the Digital Age: Key Contributions and Future Directions. ***MIS Quarterly***, 39(1): 135-154.

Barrett, M., Oborn, E., & Orlikowski, W. 2016. Creating Value in Online Communities: The Sociomaterial Configuring of Strategy, Platform, and Stakeholder Engagement. ***Information Systems Research***, 27(4): 704-723.

Baskerville, R. L., & Myers, M. D. 2002. Information systems as a reference discipline. ***MIS Quarterly***, 26(1): 1-14.

Baym, N. K. 2007. The new shape of online community: The example of Swedish independent music fandom. ***First Monday***, 12(8).

Bayus, B. L. 2013. Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community. ***Management Science***, 59(1): 226-244.

BBC. 2018. Football transfer news: How announcements got extravagant, ***BBC***.

Benner, M. J., & Tripsas, M. 2012. The influence of prior industry affiliation on framing in nascent industries: The evolution of digital cameras. ***Strategic Management Journal***, 33(3): 277-302.

Berente, N., Seidel, S., & Safadi, H. 2018. Research Commentary—Data-Driven Computationally Intensive Theory Development. ***Information Systems Research***, 30(1): 50-64.

Bharadwaj, A., El Sawy, O., Pavlou, P., & Venkatraman, N. 2013. Digital business strategy: toward a next generation of insights. ***MIS Quarterly***, 37(2): 471-482.

Blohm, I., Leimeister, J. M., & Krcmar, H. 2013. Crowdsourcing: how to benefit from (too) many great ideas. ***MIS Quarterly Executive***, 12(4): 199-211.

Bloodgood, J. 2013. Crowdsourcing: Useful for Problem Solving, But What About Value Capture? ***Academy of Management Review***, 38(3): 455-457.

Blyth, S. 2018. Big Data and Machine Learning Won't Save Us from Another Financial Crisis. ***Harvard Business Review Digital Articles***: 7-11.

Bogers, M., Afuah, A., & Bastian, B. 2010. Users as Innovators: A Review, Critique, and Future Research Directions. ***Journal of Management***, 36(4): 857-875.

Bogers, M., & West, J. 2012. Managing Distributed Innovation: Strategic Utilization of Open and User Innovation. ***Creativity and Innovation Management***, 21(1): 61-75.

Boland, R. J. 1987. ***The in-formation of information systems***. Paper presented at the Critical issues in information systems research.

Boland, R. J., & Tenkasi, R. V. 1995. Perspective Making and Perspective Taking in Communities of Knowing. ***Organization Science***, 6(4): 350-372.

Borgatti, S. P., & Foster, P. C. 2003. The Network Paradigm in Organizational Research: A Review and Typology. ***Journal of Management***, 29(6): 991-1013.

Borgmann, A. 1999. ***Holding on to reality : the nature of information at the turn of the millennium***. Chicago: University of Chicago Press.

Bostrom, N. 2017. ***Superintelligence***: Dunod.

Boudreau, K. J. 2012. Let a thousand flowers bloom? An early look at large numbers of software app developers and patterns of innovation. ***Organization Science***, 23(5): 1409-1427.

Boudreau, K. J., Lacetera, N., & Lakhani, K. R. 2011. Incentives and problem uncertainty in innovation contests: An empirical analysis. ***Management Science***, 57(5): 843-863.

Brynjolfsson, E. 1993. The productivity paradox of information technology. ***Communications of the ACM***, 36(12): 66-77.

Brynjolfsson, E., & Hitt, L. M. 2000. Beyond computation: Information technology, organizational transformation and business performance. ***Journal of Economic perspectives***, 14(4): 23-48.

Brynjolfsson, E., & McAfee, A. 2014. ***The second machine age: Work, progress, and prosperity in a time of brilliant technologies***: WW Norton & Company.

Burton-Jones, A. 2001. Knowledge capitalism: Business, work, and learning in the new economy. ***OUP Catalogue***.

Carnabuci, G., & Operti, E. 2013. Where do firms' recombinant capabilities come from? Intraorganizational networks, knowledge, and firms' ability to innovate through technological recombination. ***Strategic Management Journal***, 34(13): 1591-1613.

Cerf, V., & Kahn, R. 1974. A protocol for packet network intercommunication. ***IEEE Transactions on communications***, 22(5): 637-648.

Chang, R. M., Kauffman, R. J., & Kwon, Y. 2014. Understanding the paradigm shift to computational social science in the presence of big data. ***Decision Support Systems***, 63: 67-80.

Checkland, P., & Holwell, S. 1988. ***Information, Systems and Information Systems: Making Sense of the Field***. New York, NY.: John Wiley & Sons.

Chen, H., Chiang, R. H., & Storey, V. C. 2012. Business Intelligence and Analytics: From Big Data to Big Impact. ***MIS Quarterly***, 36(4): 1165-1188.

Chen, W., & Hirschheim, R. 2004. A paradigmatic and methodological examination of information systems research from 1991 to 2001. ***Information Systems Journal***, 14(3): 197-235.

Chesbrough, H. W. 2003. ***Open innovation: The new imperative for creating and profiting from technology***: Harvard Business Press.

Chi, R. T., & Turban, E. 1995. Distributed intelligent executive information systems. ***Decision Support Systems***, 14(2): 117-130.

Chow, G. C. 1960. Tests of equality between sets of coefficients in two linear regressions. ***Econometrica: Journal of the Econometric Society***: 591-605.

Chui, M., Manyika, J., & Miremadi, M. 2018. What AI can and can't do (yet) for your business. ***McKinsey Quarterly***(1): 96-108.

Clauset, A., Newman, M. E. J., & Moore, C. 2004. Finding community structure in very large networks. ***Physical review E***, 70(6): 066111.

Cohen, W. M., & Levinthal, D. A. 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. ***Administrative Science Quarterly***, 35(1): 128-152.

Constantiou, I. D., & Kallinikos, J. 2015. New games, new rules: Big data and the changing context of strategy. ***Journal of Information Technology***, 30(1): 44-57.

Cooper, A. C., Folta, T. B., & Woo, C. 1995. Entrepreneurial information search. ***Journal of business venturing***, 10(2): 107-120.

Daft, R. L., & Lengel, R. H. 1986. Organizational information requirements, media richness and structural design. ***Management Science***, 32(5): 554-571.

Dahlander, L., Frederiksen, L., & Rullani, F. 2008. Online communities and open innovation. ***J Industry innovation***, 15(2): 115-123.

Davenport, T. H. 2006. Competing on analytics. ***Harvard Business Review***, 84(4).

Davenport, T. H. 2018. ***The AI advantage: How to put the artificial intelligence revolution to work***: MIT Press.

Debortoli, S., Müller, O., Junglas, I. A., & vom Brocke, J. 2016. Text mining for information systems researchers: an annotated topic modeling tutorial. ***CAIS***, 39: 7.

Deloitte. 2013. The Analytics Advantage: We're just getting started: Deloitte.

Dhanaraj, C., & Parkhe, A. 2006. Orchestrating innovation networks. ***Academy of Management Review***, 31(3): 659-669.

Di Gangi, P. M., & Wasko, M. 2009. Steal my idea! Organizational adoption of user innovations from a user innovation community: A case study of Dell IdeaStorm. ***Decision Support Systems***, 48(1): 303-312.

Dinsmore, T. W. 2016. Self-Service Analytics, ***Disruptive Analytics: Charting Your Strategy for Next-Generation Business Analytics***: 199-230. Berkeley, CA: Apress.

Dremel, C., Herterich, M. M., Wulf, J., Waizmann, J. C., & Brenner, W. 2017. How AUDI AG established big data analytics in its digital transformation. ***MIS Quarterly Executive***, 16(2): 81-100.

Ebadi, Y. M., & Utterback, J. M. 1984. The Effects of Communication on Technological Innovation. ***Management Science***, 30(5): 572-585.

Edmunds, A., & Morris, A. 2000. The problem of information overload in business organisations: a review of the literature. ***International Journal of Information Management***, 20(1): 17-28.

Eisenhardt, K. M., & Graebner, M. E. 2007. Theory building from cases: Opportunities and challenges. ***Academy of Management Journal***, 50(1): 25-32.

Eppler, M. J., & Mengis, J. 2004. The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. ***The information society***, 20(5): 325-344.

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. 2011. Knowledge Collaboration in Online Communities. ***Organization Science***, 22(5): 1224-1239.

Faraj, S., Krogh, G. v., Monteiro, E., & Lakhani, K. R. 2016a. Special Section Introduction—Online Community as Space for Knowledge Flows. ***Information Systems Research***, 27(4): 668-684.

Faraj, S., Kudaravalli, S., & Wasko, M. 2015. Leading collaboration in online communities. ***MIS Quarterly***, 39(2): 393-412.

Faraj, S., Pachidi, S., & Sayegh, K. 2018. Working and organizing in the age of the learning algorithm. ***Information and Organization***, 28(1): 62-70.

Faraj, S., von Krogh, G., Monteiro, E., & Lakhani, K. R. 2016b. Special Section Introduction Online Community as Space for Knowledge Flows. ***Information Systems Research***, 27(4): 668-684.

Fayard, A.-L., Gkeredakis, E., & Levina, N. 2016. Framing Innovation Opportunities While Staying Committed to an Organizational Epistemic Stance. ***Information Systems Research***, 27(2): 302-323.

Fichman, R. G., Santos, B. L. D., & Zheng, Z. 2014. Digital innovation as a fundamental and powerful concept in the information systems curriculum. ***MIS Quarterly***, 38(2): 329-354.

Flath, C. M., Friesike, S., Wirth, M., & Thiesse, F. 2017. Copy, transform, combine: Exploring the remix as a form of innovation. ***Journal of Information Technology***, 32(4): 306-325.

Fleming, L., & Waguespack, D. M. 2007. Brokerage, Boundary Spanning, and Leadership in Open Innovation Communities. ***Organization Science***, 18(2): 165-180.

Freelon, D. 2014. On the interpretation of digital trace data in communication and social computing research. ***Journal of Broadcasting & Electronic Media***, 58(1): 59-75.

Freeman, C., & Soete, L. 2009. Developing science, technology and innovation indicators: What we can learn from the past. ***Research Policy***, 38(4): 583-589.

Freeth, T., Jones, A., Steele, J. M., & Bitsakis, Y. 2008. Calendars with Olympiad display and eclipse prediction on the Antikythera Mechanism. ***Nature***, 454(7204): 614.

Füller, J., Hutter, K., Hautz, J., & Matzler, K. 2014. User roles and contributions in innovation-contest communities. ***Journal of Management Information Systems***, 31(1): 273-308.

Galbraith, J. R. 1974. Organization Design: An Information Processing View. ***Interfaces***, 4(3): 28-36.

Gantz, J., & Reinsel, D. 2012. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east - United States. ***IDC iView: IDC Analyze the future***, 2007(2012): 1-16.

Garcia, D., Mavrodiev, P., & Schweitzer, F. 2013. ***Social resilience in online communities: The autopsy of friendster***. Paper presented at the Proceedings of the first ACM conference on Online social networks.

Gargiulo, M., & Benassi, M. 2000. Trapped in Your Own Net? Network Cohesion, Structural Holes, and the Adaptation of Social Capital. ***Organization Science***, 11(2): 183-196.

Gartner. 2018. Magic Quadrant for Analytics and Business Intelligence Platforms.

Garud, R., Gehman, J., & Giuliani, A. P. 2014a. Contextualizing entrepreneurial innovation: A narrative perspective. ***Research Policy***, 43(7): 1177-1188.

Garud, R., Schildt, H. A., & Lant, T. K. 2014b. Entrepreneurial Storytelling, Future Expectations, and the Paradox of Legitimacy. ***Organization Science***, 25(5): 1479-1492.

Gaskin, J., Berente, N., Lyytinen, K., & Yoo, Y. 2014. Toward Generalizable Sociomaterial Inquiry: A Computational Approach for Zooming In and Out of Sociomaterial Routines. ***MIS Quarterly***, 38(3): 849-871.

Gaughan, J. 2016. Zlatan Ibrahimovic announces Manchester United transfer on Instagram as Swedish superstar says: 'I am coming', ***Mail Online***.

Gawer, A. 2014. Bridging differing perspectives on technological platforms: Toward an integrative framework. ***Research Policy***, 43(7): 1239-1249.

Geiger, R. S., & Ribes, D. ***Trace ethnography: Following coordination through documentary practices***.

George, G., Haas, M. R., & Pentland, A. 2014. Big Data and Management. ***Academy of Management Journal***, 57(2): 321-326.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. 2016. Big data and data science methods for management research, ***Academy of Management Journal***, Vol. 59: 1493-1507: Academy of Management.

Gerring, J. 2006. ***Case study research: principles and practices***: Cambridge University Press.

Ghazawneh, A., & Henfridsson, O. 2013. Balancing platform control and external contribution in third-party development: the boundary resources model. ***Information Systems Journal***, 23(2): 173-192.

Gilbert, C. G. 2005. Unbundling the structure of inertia: Resource versus routine rigidity. ***Academy of Management Journal***, 48(5): 741-763.

Goes, P. B. 2013. Big Data in IS Research. ***MIS Quarterly***, 38(1): iii-viii.

Goes, P. B., Guo, C., & Lin, M. 2016. Do Incentive Hierarchies Induce User Effort? Evidence from an Online Knowledge Exchange. ***Information Systems Research***, 27(3): 497-516.

Goetheer, J. D., & Butter, M. 2017. Digital Innovation Hubs Catalogue.

Goldstine, H. H., & Goldstine, A. 1946. The electronic numerical integrator and computer (eniac). ***Mathematical Tables and Other Aids to Computation***, 2(15): 97-110.

Goodhart, C. A. E. 1975. Monetary Relationships: A View from Threadneedle Street in Papers in Monetary Economics. ***Sydney: Reserve Bank of Australia***.

Gorry, G. A., & Scott Morton, M. S. 1971. A framework for management information systems. ***Working paper***.

Grewal, R., Lilien, G. L., & Mallapragada, G. 2006. Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems. ***Management Science***, 52(7): 1043-1056.

Grover, V., & Lyytinen, K. 2015. New State of Play in Information Systems Research: The Push to the Edges. ***MIS Quarterly***, 39(2): 271-296.

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. 2017. Debating big data: A literature review on realizing value from big data. ***The Journal of Strategic Information Systems***, 26(3): 191-209.

Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S., & Jennings, P. D. 2019. Topic Modeling in Management Research: Rendering New Theory from Textual Data. ***Academy of Management Annals***, 13(2): 586-632.

Harrison, J. 2016. RSelenium: R Bindings for Selenium WebDriver. R package version 1.4.0.

Hedman, J., Srinivasan, N., & Lindgren, R. 2013. Digital traces of information systems: Sociomateriality made researchable, ***Thirty Fourth International Conference on Information Systems***. Milan.

Henfridsson, O., & Bygstad, B. 2013. The Generative Mechanisms of Digital Infrastructure Evolution. ***MIS Quarterly***, 37(3): 907-+.

Henfridsson, O., Nandhakumar, J., Scarbrough, H., & Panourgias, N. 2018. Recombination in the open-ended value landscape of digital innovation. ***Information and Organization***, 28(2): 89-100.

Henry, N. L. 1974. Knowledge management: a new concern for public administration. ***Public Administration Review***: 189-196.

Hertel, G., Niedner, S., & Herrmann, S. 2003. Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel. ***Research policy***, 32(7): 1159-1177.

Herzenstein, M., Sonenshein, S., & Dholakia, U. M. 2011. Tell me a good story and I may lend you money: The role of narratives in peer-to-peer lending decisions. ***Journal of Marketing Research***, 48(SPL): S138-S149.

Hippel, E. V., & Krogh, G. V. 2003. Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science. ***Organization Science***, 14(2): 209-223.

Hirschheim, R., & Klein, H. K. 2012. A Glorious and Not-So-Short History of the Information Systems Field. ***Journal of the Association for Information Systems***, 13(4): 188-235.

Hirschheim, R., Klein, H. K., & Lyytinen, K. 1995. ***Information systems development and data modeling: conceptual and philosophical foundations***: Cambridge University Press.

Howe, J. 2006. The rise of crowdsourcing. ***Wired magazine***, 14(6): 1-4.

Howison, J., Wiggins, A., & Crowston, K. 2011. Validity Issues in the Use of Social Network Analysis with Digital Trace Data. ***Journal of the Association for Information Systems***, 12(12): 767-797.

Hukal, P., & Henfridsson, O. 2017. Digital innovation–a definition and integrated perspective, ***The Routledge Companion to Management Information Systems***: 360-369: Routledge.

Jacobs, A. 2009. The Pathologies of Big Data. ***Communications of the ACM***, 52(8): 36-44.

Jeppesen, L. B., & Lakhani, K. R. 2010. Marginality and Problem-Solving Effectiveness in Broadcast Search. ***Organization Science***, 21(5): 1016-1033.

Jockers, M. 2013. ***Macroanalysis: Digital Methods and Literary History***: University of Illinois Press.

Jockers, M. 2014. ***Text analysis with R for students of literature***. New York: Springer.

Johnson, S. L., Faraj, S., & Kudaravalli, S. 2014. EMERGENCE OF POWER LAWS IN ONLINE COMMUNITIES: THE ROLE OF SOCIAL MECHANISMS AND PREFERENTIAL ATTACHMENT. ***MIS Quarterly***, 38(3): 795-A713.

Johnson, S. L., Safadi, H., & Faraj, S. 2015. The Emergence of Online Community Leadership. ***Information Systems Research***, 26(1): 165-187.

Jonsson, K., Holmström, J., & Lyytinen, K. 2009. Turn to the material: Remote diagnostics systems and new forms of boundary-spanning. ***Information and Organization***, 19(4): 233-252.

Jonsson, K., Mathiassen, L., & Holmstrom, J. 2018. Representation and mediation in digitalized work: evidence from maintenance of mining machinery. ***Journal of Information Technology***, 33(3): 216-232.

Kallinikos, J. 2006. Information out of information: on the self-referential dynamics of information growth. ***Information Technology & People***, 19(1): 98-115.

Kallinikos, J. 2012. Form, function, and matter: Crossing the border of materiality. ***Materiality and organizing: Social interaction in a technological world***: 67-87.

Kallinikos, J., Aaltonen, A., & Marton, A. 2013. The Ambivalent Ontology of Digital Artifacts. ***MIS Quarterly***, 37(2).

Kallinikos, J., & Constantiou, I. D. 2015. Big data revisited: a rejoinder. ***Journal of Information Technology***, 30(1): 70-74.

Kane, G. C., Phillips, A. N., Copulsky, J. R., & Andrus, G. R. 2019. ***The Technology Fallacy: How People are the Real Key to Digital Transformation***: MIT Press.

Kaplan, S., & Vakili, K. 2015. The double-edged sword of recombination in breakthrough innovation. ***Strategic Management Journal***, 36(10): 1435-1457.

Keen, P. G. 1980. MIS research: reference disciplines and a cumulative tradition, ***Proceedings of the First International Conference on Information Systems***: 9-18.

Kelly, K. 1996. The electronic hive: embrace it. In R. Kling (Ed.), ***Computerization and Controversy***, Vol. 2: 75-78. San Diego: Academic Press.

Kelly, K. 2017. ***The inevitable: understanding the 12 technological forces that will shape our future***: Penguin.

Kerschberg, B. 2015. How Xerox Uses Analytics, Big Data and Ethnography To Help Government Solve, ***Forbes***.

Kettinger, W. J., & Li, Y. 2010. The infological equation extended: towards conceptual clarity in the relationship between data, information and knowledge. ***European Journal of Information Systems***, 19(4): 409-421.

King, J. L., & Lyytinen, K. (Eds.). 2006. ***Information systems : the state of the field***: John Wiley & Sons Ltd.

Klein, K. J., & Sorra, J. S. 1996. The challenge of innovation implementation. ***Academy of Management Review***, 21(4): 1055-1080.

Kling, R. 1996. Hopes and Horrors: Technological Utopianism and Anti-Utopianism in Narratives of Computerization. In R. Kling (Ed.), ***Computerization and Controversy***, 2 ed.: Academic Press.

Kling, R., Ackerman, M. S., & Allen, J. P. 1995. ***Information entrepreneurialism, information technologies, and the continuing vulnerability of privacy***: Academic Press, Inc.

Kohli, R., & Grover, V. 2008. Business value of IT: An essay on expanding research directions to keep up with the times. ***Journal of the Association for Information Systems***, 9(1): 23-39.

Kohli, R., & Melville, N. P. 2019. Digital innovation: A review and synthesis. ***Information Systems Journal***, 29(1): 200-223.

Kyriakou, H., Nickerson, J. W., & Sabnis, G. 2017. Knowledge Reuse for Customization: Metamodels in an Open Design Community for 3D Printing. ***MIS Quarterly***, 41(1): 315-322.

Lakhani, K. R., & Panetta, J. A. 2007. The Principles of Distributed Innovation. ***Innovations: Technology, Governance, Globalization***, 2(3): 97-112.

Lakhani, K. R., & von Hippel, E. 2003. How open source software works: "free" user-to-user assistance. ***Research Policy***, 32(6): 923-943.

Langefors, B. 1973. Theoretical analysis of information systems.

Langley, A. 1999. Strategies for theorizing from process data. ***Academy of Management Review***, 24(4): 691-710.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. 2014. The Parable of Google Flu: Traps in Big Data Analysis. ***Science***, 343(6176): 1203-1205.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. 2009. Social science. Computational social science. ***Science***, 323(5915): 721-723.

Leonard-Barton, D. 1995. Wellsprings of knowledge: Building and sustaining the sources of innovation. ***University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.***

Leonardi, P., & Contractor, N. 2018. Better PEOPLE Analytics Measure Who THEY KNOW, Not Just Who THEY ARE. ***Harvard Business Review***, 96(6): 70-81.

Leonardi, P. M., & Bailey, D. E. 2008. Transformational technologies and the creation of new work practices: Making implicit knowledge explicit in task-based offshoring. ***MIS Quarterly***, 32(2): 411-436.

Lessig, L. 2008. ***Remix: Making art and commerce thrive in the hybrid economy***: Penguin.

Levina, N., & Arriaga, M. 2014. Distinction and status production on user-generated content platforms: Using Bourdieu's theory of cultural production to understand social dynamics in online fields. ***Information Systems Research***, 25(3): 468-488.

Levina, N., & Vaast, E. 2015. Leveraging archival data from online communities for grounded process theorizing. In K. D. Elsbach, & R. M. Kramer

(Eds.), ***Handbook of Qualitative Organizational Research: Innovative Pathways and Methods***: 215-224.

Levitin, A. V., & Redman, T. C. 1998. Data as a Resource: Properties, Implications, and Prescriptions. ***Sloan Management Review***, 40(1): 89-101.

Lin, Y.-K., Chen, H., Brown, R. A., Li, S.-H., & Yang, H.-J. 2017. Healthcare predictive analytics for risk profiling in chronic care: A Bayesian multitask learning approach. ***MIS Quarterly***, 41(2).

Lindberg, A. 2015. The Origin, Evolution, and Variation of Routine Structures in Open Source Software Development: Three Mixed Computational-Qualitative Studies.

Lindberg, A. Forthcoming. **Developing Theory through integrating Human & Machine Pattern Recognition**. ***Journal of the Association for Information Systems***.

Lindberg, A., Berente, N., Gaskin, J., & Lyytinen, K. 2016. Coordinating Interdependencies in Online Communities: A Study of an Open Source Software Project. ***Information Systems Research***, 27(4): 751-772.

Lindberg, A., Berente, N., & Lyytinen, K. 2015. Towards an open source software development life cycle: A study of routine heterogeneity and discourse across multiple releases, ***75th Annual Meeting of the Academy of Management***.

Lopez-Vega, H., Tell, F., & Vanhaverbeke, W. 2016. Where and how to search? Search paths in open innovation. ***Research Policy***, 45(1): 125-136.

Lukyanenko, R., Parsons, J., Wiersma, Y. F., & Maddah, M. 2019. Expecting the unexpected: effects of data collection design choices on the quality of crowdsourced user-generated content. ***MIS Quarterly***, 43(2): 623-647.

Lyytinen, K. 1987. Two views of information modeling. ***Information & Management***, 12(1): 9-19.

Lyytinen, K., & Grover, V. 2017. Management Misinformation Systems: A Time to Revisit? ***Journal of the Association for Information Systems***, 18(3): 2.

Lyytinen, K., & Yoo, Y. 2002. Research commentary: The next wave of Nomadic computing. ***Information Systems Research***, 13(4): 377-388.

Lyytinen, K., Yoo, Y., & Boland Jr, R. J. 2016. Digital product innovation within four classes of innovation networks. ***Information Systems Journal***, 26(1): 47-75.

Majchrzak, A., & Malhotra, A. 2013. Towards an information systems perspective and research agenda on crowdsourcing for innovation. ***The Journal of Strategic Information Systems***, 22(4): 257-268.

Mankevich, V., & Holmström, J. 2016. Gateways to Digital Entrepreneurship: Investigating the Organizing Logics for Digital Startups. ***Academy of Management Proceedings***, 2016(1).

Mankevich, V., & Sandberg, J. 2018. ***Teaching Data Science the Interdisciplinary Way: Learning Cycles and Diverse Skillsets***. Paper presented at the The 6th Swedish Workshop on Data Science.

Mankevich, V., Skog, D., Wimelius, H., & Holmström, J. 2015. ***The hero with a hundred faces: the role of coopetition in innovation network evolution***.

March, J. G. 1991. Exploration and exploitation in organizational learning. ***Organization science***, 2(1): 71-87.

Marsden, P. V. 1990. Network Data and Measurement. ***Annual Review of Sociology***, 16: 435-463.

Marwick, A., & Lewis, R. 2017. Media manipulation and disinformation online. ***New York: Data Society Research Institute***.

Mau, S. 2019. ***The Metric Society: On the Quantification of the Social***: John Wiley & Sons.

McGee, P. 2019. Robotaxis: can automakers catch up with Google in driverless cars?, ***Financial Times***. Frankfurt.

McKinney, E. H., & Yoos, C. J. 2010. Information About Information: A Taxonomy of Views. ***MIS Quarterly***, 34(2): 329-344.

McMullen, J. S., & Shepherd, D. A. 2006. Entrepreneurial Action And The Role Of Uncertainty In The Theory Of The Entrepreneur. 31(1): 132-152.

Mela, C. F., & Moorman, C. 2018. Why Marketing Analytics Hasn't Lived Up to Its Promise. ***Harvard Business Review Digital Articles***: 2-6.

Mingers, J. C. 1995. Information and meaning: foundations for an intersubjective account. ***Information Systems Journal***, 5(4): 285-306.

Mislove, A., Marcon, M., Gummadi, K. R., Druschel, P., & Bhattacharjee, B. 2007. Measurement and Analysis of Online Social Networks. ***Imc'07: Proceedings of the 2007 Acm Sigcomm Internet Measurement Conference***: 29-42.

Mital, M., & Sarkar, S. 2011. Multihoming behavior of users in social networking web sites: a theoretical model. ***Information Technology & People***, 24(4): 378-392.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. 2016. The ethics of algorithms: Mapping the debate. ***Big Data Society***, 3(2): 2053951716679679.

Mohr, L. B. 1982. ***Explaining organizational behavior***: Jossey-Bass.

Monteiro, E., & Parmiggiani, E. 2019. Synthetic Knowing: The Politics of the Internet of Things. Ithaca: Cornell University Library, arXiv.org.

Morozov, E. 2011. ***The net delusion: How not to liberate the world***: Penguin UK.

Morschheuser, B., Hamari, J., Koivisto, J., & Maedche, A. 2017. Gamified crowdsourcing: Conceptualization, literature review, and future agenda. ***International Journal of Human-Computer Studies***, 106: 26-43.

Mutch, A. 1997. Information literacy: An exploration. ***International Journal of Information Management***, 17(5): 377-386.

Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. 2017. Digital Innovation Management: Reinventing Innovation Management Research in a Digital World. ***MIS Quarterly***, 41(1): 223-238.

Nambisan, S., & Zahra, S. A. 2016. The role of demand-side narratives in opportunity formation and enactment. ***Journal of Business Venturing Insights***, 5: 70-75.

Nan, N., & Lu, Y. 2014. Harnessing the power of self-organization in an online community during organizational crisis. ***MIS Quarterly***, 38(4): 1135-1158.

Nielsen, F. A. 2011. ***A new ANEW: Evaluation of a word list for sentiment analysis in microblogs***. Paper presented at the Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts':Big things come in small packages 718 in CEUR Workshop Proceedings.

Nylén, D., & Holmström, J. 2015. Digital innovation strategy: A framework for diagnosing and improving digital product and service innovation. ***Business Horizons***, 58(1): 57-67.

Nylén, D., Holmström, J., & Lyytinen, K. 2014. Oscillating Between Four Orders of Design: The Case of Digital Magazines. ***Design Issues***, 30(3): 53-68.

O'Reilly, I. C. A. 1980. Individuals and Information Overload in Organizations: Is More Necessarily Better? , 23(4): 684-696.

Oinas-Kukkonen, H., Lyytinen, K., & Yoo, Y. 2010. Social Networks and Information Systems: Ongoing and Future Research Streams. ***Journal of the Association for Information Systems***, 11(2): 61-68.

Orlikowski, W. J., & Iacono, C. S. 2001. Research commentary: Desperately seeking the "IT" in IT research - A call to theorizing the IT artifact. ***Information Systems Research***, 12(2): 121-134.

Ortiz de Guinea, A., & Webster, J. 2017. Combining variance and process in information systems research: Hybrid approaches. ***Information and Organization***, 27(3): 144-162.

Parmentier, G., & Mangematin, V. 2014. Orchestrating innovation with user communities in the creative industries. ***Technological Forecasting and Social Change***, 83(1): 40-53.

Pavlou, P. A., & El Sawy, O. A. 2010. The “Third Hand”: IT-Enabled Competitive Advantage in Turbulence Through Improvisational Capabilities. ***Information Systems Research***, 21(3): 443-471.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. 2001. Linguistic inquiry and word count: LIWC 2001. ***Mahway: Lawrence Erlbaum Associates***, 71(2001): 2001.

Petrides, L. A., McClelland, S. I., & Nodine, T. R. 2004. Costs and benefits of the workaround: inventive solution or costly alternative. ***International Journal of Educational Management***, 18(2): 100-108.

Philip Chen, C. L., & Zhang, C.-Y. 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. ***Information Sciences***.

Piezunka, H., & Dahlander, L. 2014. Distant Search, Narrow Attention: How Crowding Alters Organizations' Filtering of User Suggestions. ***Academy of Management Journal***, 1: 12513.

Piezunka, H., & Dahlander, L. 2015. Distant Search, Narrow Attention: How Crowding Alters Organizations' Filtering of Suggestions in Crowdsourcing. ***Academy of Management Journal***, 58(3): 856-880.

Pons, P., & Latapy, M. 2005. ***Computing communities in large networks using random walks***.

Porter, M. E. 1985. Competitive advantage: Creating and sustaining competitive advantage: New York: Free Press.

Prpić, J., Shukla, P. P., Kietzmann, J. H., & McCarthy, I. P. 2015. How to work a crowd: Developing crowd capital through crowdsourcing. ***Business Horizons***, 58(1): 77-85.

Quinn, R. E., & Rohrbaugh, J. 1983. A Spatial Model of Effectiveness Criteria - Towards a Competing Values Approach to Organizational Analysis. ***Management Science***, 29(3): 363-377.

R Core Team, R. F. f. S. C. 2015. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Rai, A., & Tang, X. 2010. Leveraging IT capabilities and competitive process capabilities for the management of interorganizational relationship portfolios. ***Information Systems Research***, 21(3): 516-542.

Ransbotham, S., Kiron, D., & Prentice, P. K. 2016. Beyond the hype: the hard work behind analytics success. ***MIT Sloan Management Review***, 57(3).

Raymond, E. 1999. The cathedral and the bazaar. ***Knowledge, Technology & Policy***.

Reischauer, G., & Mair, J. 2018. How Organizations Strategically Govern Online Communities: Lessons from the Sharing Economy. ***Academy of Management Discoveries***, 4(3): 220-247.

Ren, Y., Kraut, R., & Kiesler, S. 2007. Applying Common Identity and Bond Theory to Design of Online Communities. ***Organization Studies***, 28(3): 377-408.

Riedl, C., & Seidel, V. P. 2018. Learning from Mixed Signals in Online Innovation Communities. ***Organization Science***, 0(0): null.

Saarikko, T., Westergren, U. H., & Blomquist, T. 2017. The Internet of Things: Are you ready for what's coming? ***Business Horizons***, 60(5): 667-676.

Sabherwal, R., & Robey, D. 1995. Reconciling variance and process strategies for studying information system development. ***Information Systems Research***, 6(4): 303-327.

Sambamurthy, V., Bharadwaj, A., & Grover, V. 2003. Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. ***Mis Quarterly***, 27(2): 237-263.

Sambamurthy, V., & Zmud, R. W. 2000. Research commentary: The organizing logic for an enterprise's IT activities in the digital era - A prognosis of practice and a call for research. ***Information Systems Research***, 11(2): 105-114.

Sandberg, J., Mathiassen, L., & Napier, N. 2014. Digital Options Theory for IT Capability Investment. ***Journal of the Association for Information Systems***, 15(7): 422-453.

Savino, T., Messeni Petruzzelli, A., & Albino, V. 2017. Search and recombination process to innovate: a review of the empirical evidence and a research agenda. ***International Journal of Management Reviews***, 19(1): 54-75.

Schumpeter, J. A. 1934. ***The Theory of Economic Development***: Cambridge University Press.

Schumpeter , J. A. 2003. ***Capitalism, Socialism and Democracy***.

Scolari, C. A. 2009. Transmedia storytelling: Implicit consumers, narrative worlds, and branding in contemporary media production.

Scott, S. V., & Orlikowski, W. J. 2012. Reconfiguring relations of accountability: Materialization of social media in the travel sector. ***Accounting, organizations and society***, 37(1): 26-40.

Seawright, J., & Gerring, J. 2008. Case selection techniques in case study research: A menu of qualitative and quantitative options. ***Political research quarterly***, 61(2): 294-308.

Selander, L., Henfridsson, O., & Svahn, F. 2013. Capability search and redeem across digital ecosystems. ***Journal of Information Technology***, 28(3): 183-197.

Senyard, J., Baker, T., Steffens, P., & Davidsson, P. 2014. Bricolage as a path to innovativeness for resource-constrained new firms. ***Journal of Product Innovation Management***, 31(2): 211-230.

Shah, S., Horne, A., & Capella, J. 2012. Good Data Won't Guarantee Good Decisions. ***Harvard Business Review***, 90(4): 23-25.

Shah, S. K. 2006. Motivation, governance, and the viability of hybrid forms in open source software development. ***Management Science***, 52(7): 1000-1014.

Shaikh, M., & Levina, N. 2019. Selecting an open innovation community as an alliance partner: Looking for healthy communities and ecosystems. ***Research Policy***, 48(8): 103766.

Shane, S. 2000. Prior knowledge and the discovery of entrepreneurial opportunities. ***Organization Science***, 11(4): 448-469.

Shapiro, C., Carl, S., & Varian, H. R. 1998. ***Information rules: a strategic guide to the network economy***: Harvard Business Press.

Shenk, D. 1997. ***Data smog: Surviving the information glut***: HarperCollins Publishers.

Sidorova, A., Evangelopoulos, N., Valacich, J. S., & Ramakrishnan, T. 2008. Uncovering the intellectual core of the information systems discipline. ***MIS Quarterly***, 32(3): 467-482.

Simon, H. A. 1996. ***The sciences of the artificial*** (3rd ed.). Cambridge, Mass.: MIT Press.

Singh, P. V., Tan, Y., & Mookerjee, V. 2011. Network effects: The influence of structural social capital on open source project success. ***MIS Quarterly***, 35(4): 813-829.

Śledzik, K. 2013. Schumpeter's view on innovation and entrepreneurship. ***Management Trends in Theory and Practice,(ed.) Stefan Hittmar, Faculty of Management Science and Informatics, University of Zilina & Institute of Management by University of Zilina***.

Stamper, R. 1973. ***Information in business and administrative systems***: John Wiley & Sons, Inc.

Stanko, M. A. 2016. Toward a theory of remixing in online innovation communities. ***Information Systems Research***, 27(4): 773-791.

Strauss, A. L. 1987. ***Qualitative analysis for social scientists***: Cambridge University Press.

Sugimoto, C. R., Ekbia, H. R., & Mattioli, M. 2016. ***Big data is not a monolith***: MIT Press.

Surblyte, G. 2016. Data as a Digital Resource. ***Max Planck Institute for Innovation & Competition Research***, 16.

Svahn, F., Mathiassen, L., & Lindgren, R. 2017. Embracing Digital Innovation in Incumbent Firms: How Volvo Cars Managed Competing Concerns. ***MIS Quarterly***, 41(1): 239-253.

Swade, D. 2001. ***Difference engine: Charles Babbage and the quest to build the First Computer***: Viking Penguin.

Tausczik, Y. R., & Pennebaker, J. W. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. ***Journal of Language and Social Psychology***, 29(1): 24-54.

Teece, D. J., Pisano, G., & Shuen, A. 1997. Dynamic capabilities and strategic management. ***Strategic management journal***, 18(7): 509-533.

Teigland, R., Di Gangi, P. M., Flåten, B.-T., Giovacchini, E., & Pastorino, N. 2014. Balancing on a tightrope: Managing the boundaries of a firm-sponsored OSS community and its impact on innovation and absorptive capacity. ***Information and Organization***, 24(1): 25-47.

Tilson, D., Lyytinen, K., & Sørensen, C. 2010. Research Commentary—Digital Infrastructures: The Missing IS Research Agenda. ***Information Systems Research***, 21(4): 748-759.

Tiwana, A. 2019. The curse of "reference" disciplines. ***Journal of Information Technology***, 34(2): 190-192.

Törnberg, P., & Törnberg, A. 2018. The limits of computation: A philosophical critique of contemporary Big Data research. ***Big Data & Society***, 5(2): 2053951718811843.

Traag, V. A., & Bruggeman, J. 2009. Community detection in networks with positive and negative links. ***Physical Review E***, 80(3): 036115.

Tripsas, M., & Gavetti, G. 2000. Capabilities, cognition, and inertia: Evidence from digital imaging. ***Strategic management journal***, 21(10-11): 1147-1161.

Truex, D., Holmström, J., & Keil, M. 2006. Theorizing in information systems research: A reflexive analysis of the adaptation of theory in information systems research. ***Journal of the Association for Information Systems***, 7(12): 797-821.

Tuertscher, P., Garud, R., & Kumaraswamy, A. 2014. Justification and interlaced knowledge at ATLAS, CERN. ***Organization Science***, 25(6): 1579-1608.

Tuomi, I. 1999. Data Is More than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational

Memory. ***Journal of Management Information Systems***, 16(3): 103-117.

Tuomi, I. 2006. ***Networks of innovation***: Oxford University Press.

Tushman, M. L., & Nadler, D. A. 1978. Information processing as an integrating concept in organizational design. ***Academy of Management Review***, 3(3): 613-624.

Van Alstyne, M. W., Parker, G. G., & Choudary, S. P. 2016. Pipelines, platforms, and the new rules of strategy. ***Harvard Business Review***, 94(4): 54-62.

Van de Ven, A. H. 2005. Running in packs to develop knowledge-intensive technologies. ***MIS Quarterly***.

van Osch, W., & Avital, M. 2010. ***Generative Collectives***. Paper presented at the ICIS.

Varian, H. R. 2010. Computer mediated transactions. ***American Economic Review***, 100(2): 1-10.

Varian, H. R. 2014. Beyond Big Data. ***Business Economics***, 49(1): 27-31.

Varun Grover, & Davenport, T. H. 2001. General Perspectives on Knowledge Management: Fostering a Research Agenda. ***Journal of Management Information Systems***, 18(1): 5-21.

Venkatesh, V., Brown, S. A., & Bala, H. 2013. Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. ***MIS Quarterly***: 21-54.

Von Hippel, E. 2001. Innovation by User Communities: Learning from Open-Source Software. (cover story). ***MIT Sloan Management Review***, 42(4): 82-86.

Von Hippel, E. 2005. Democratizing innovation: The evolving phenomenon of user innovation. ***Journal für Betriebswirtschaft***, 55(1): 63-78.

Von Hippel, E., & Von Krogh, G. 2015. Crossroads—Identifying viable "need–solution pairs": Problem solving without problem formulation. ***Organization Science***, 27(1): 207-221.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. 1992. Building an Information System Design Theory for Vigilant EIS. ***Information Systems Research***, 3(1): 36-59.

Walsham, G. 1995. Interpretive Case-Studies in Is Research - Nature and Method. ***European Journal of Information Systems***, 4(2): 74-81.

Wasko, M. M., & Faraj, S. 2000. "It is what one does": why people participate and help others in electronic communities of practice. ***The Journal of Strategic Information Systems***, 9(2-3): 155-173.

Wasko, M. M., & Faraj, S. 2005. Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. ***MIS Quarterly***, 29(1): 35-57.

Wasserman, S., & Faust, K. 1994. ***Social network analysis: Methods and applications***: Cambridge university press.

Weick, K. E. 1979. ***The social psychology of organizing***.

Weitzman, M. L. 1998. Recombinant Growth*. ***The Quarterly Journal of Economics***, 113(2): 331-360.

Weninger, T., Johnston, T. J., & Glenski, M. 2015. Random Voting Effects in Social-Digital Spaces: A Case Study of Reddit Post Submissions, ***Proceedings of the 26th ACM Conference on Hypertext & Social Media***: 293-297. Guzelyurt, Northern Cyprus: ACM.

Wessel, M. 2016. You Don't Need Big Data -- You Need the Right Data. ***Harvard Business Review Digital Articles***: 2-5.

West, D. M. 2012. Big data for education: Data mining, data analytics, and web dashboards. ***Governance studies at Brookings***, 4(1).

Whelan, E., Teigland, R., Vaast, E., & Butler, B. 2016. Expanding the horizons of digital social networks: Mixing big trace datasets with qualitative approaches. ***Information and Organization***, 26(1): 1-12.

Willcocks, L. P., & Sykes, R. 2000. Enterprise resource planning: the role of the CIO and it function in ERP. ***Communications of the ACM***, 43(4): 32-38.

Wimelius, H. 2011. Duplicate systems: investigating unintended consequences of information technology in organizations.

Winston, B. 2002. ***Media, technology and society: A history: From the telegraph to the Internet***: Routledge.

Yoo, Y. 2010. Computing in everyday life: A call for research on experiential computing. ***MIS quarterly***: 213-231.

Yoo, Y., Boland, R. J., Lyytinen, K., & Majchrzak, A. 2012. Organizing for Innovation in the Digitized World. ***Organization Science***, 23(5): 1398-1408.

Yoo, Y., Henfridsson, O., & Lyytinen, K. 2010. Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. ***Information Systems Research***, 21(4): 724-735.

Zhao, L., Detlor, B., & Connelly, C. E. 2016. Sharing knowledge in social Q&A sites: the unintended consequences of extrinsic motivation. ***Journal of Management Information Systems***, 33(1): 70-100.

Zhao, Y., & Zhu, Q. 2014. Evaluation on crowdsourcing research: Current status and future direction. ***Information Systems Frontiers***, 16(3): 417-434.

Zittrain, J. L. 2005. The generative internet. ***Harv. L. Rev.***, 119: 1975.