Ghost in the Machine: Theorizing data knowledge in the Age of Intelligent Technologies

Panagiota Koukouvinou
Umeå University, panagiota.koukouvinou@umu.se

Gemza Ademaj
Lund University, gemza.ademaj@ics.lu.se

Saonee Sarker
Lund Universty, saonee@vt.edu

Jonny Holmström
Umeå University, jonny.holmstrom@umu.se

Ghost in the Machine: Theorizing data knowledge in the Age of Intelligent Technologies

Short Paper

Panagiota Koukouvinou
Umeå University
Mit-huset, Campustorget 5, 90187
Umeå
panagiota.koukouvinou@umu.se

Gemza Ademaj
Lund University
Ole Römers väg 6, 223 63 Lund
gemza.ademaj@ics.lu.se

Saonee Sarker
Lund University
Ole Römers väg 6, 223 63 Lund
saonee.sarker@ics.lu.se

Jonny Holmström
Umeå University
Mit-huset, Campustorget 5, 90187
Umeå
jonny.holmstrom@umu.se

Abstract

AI technologies have led to new ways of thinking about data, knowledge, and organizations. Despite the arguments that data speak for themselves, the era of datafication demands revisiting data and knowledge and reflecting on new ways of theorizing. Considering that working with data is important for most employees, there is a need to investigate how the knowing of data can be achieved. In this paper, we move beyond the factual view of data and the hierarchical view of data and knowledge, to introduce data knowledge as a new type of knowledge. We present a first step towards a theory of explanation of what is data knowledge in today´s organizations. To investigate this, we apply an etymological lens, and review systematically the IS literature. Our preliminary findings demonstrate unveiling data, balancing between intuition and data, acknowledging external and internal capabilities, and realizing data, as the four main concepts of data knowledge.

Keywords: Data, knowledge, theorizing, etymology, systematic review

Introduction

The recent development and maturation of AI technologies has not only led to unpredictable changes in organizational life but also in the way we think about theory and knowledge across disciplines seeking to theorize the effects of new AI technologies. Some have even gone so far as to predict that the progress of new AI technologies will doom theory to obsolescence (Steadman, 2013). Advancing a similar argument, Kitchin (2014) made the case that "the volume of data, accompanied by techniques that can reveal their inherent truth, enables data to speak for themselves free of theory."(p.3) Wired magazine published the provocatively titled piece “End of theory: Data deluge makes the scientific method obsolete,” in which it was argued that in this new era of ML, “science can advance even without coherent models, unified theories, or really any mechanistic explanation at all” (Anderson, 2008). In this short paper, we argue for the need and set the first step of a theory of explanation of data knowledge as a new approach of knowing and working with data in the digital era (Gregor, 2006).

Admittedly, the key result of the digital revolution is the way in which data has become an important part of the social, economic, and political spectrum (Kitchin, 2014; Zuboff, 2019) and a medium and resource for organizing that reshapes the process of knowing within organizations (Alaimo and Kallinikos, 2019; Alaimo and Kallinikos, 2022). Hence, understanding, knowing, and working with data has been perceived as a building block for cultivating a digital mindset that allows organizations to gain and sustain competitive advantage (Neeley and Leonardi, 2022). Although the ability to understand data to analyze and provide recommendations has been traditionally discussed in the context of data scientists (e.g., Stice-Lusvardi, 2023), it is becoming increasingly important for managers and employees as well (Brown, 2023). The need for cultivating data-related knowledge and educating employees about data has also been highlighted by Waardenburg et al. (2022). However, despite that the importance of data as a strategic organizational resource has been a focal concern within IS literature (see Constantiou and Kallinikos, 2015; Chen et.,2017; Aenen et al., 2022 for some examples); the surge of datafication brings new questions for both research and practice. These questions come to challenge the assumptions of “data in principle,” namely how data come into being (Jones, 2019), and “data in practice,” namely how data come to be used in organizations to become objects of knowing (Jones, 2019; Monteiro, 2022), but also the validity of traditional and dominant conceptualization of data and knowledge.

Currently, the ways in which data and knowledge have been framed within IS have been dominated by two misleading assumptions, the factual view of data and the hierarchical relationship between data and knowledge. First, in the factual view (see Alaimo and Kallinikos, 2022; Kallinikos et al., 2013; Monteiro, 2022 for exceptions), data are perceived as “raw facts that describe a particular phenomenon” (Haag and Cummings, 2013, p. 508), or as facts that provide a certain, accurate and faithful representation of reality (Jones, 2019; Monteiro, 2022). Second, the IS field predominantly follows the hierarchical approach of the data-information-knowledge-wisdom pyramid (Ackoff, 1989) to conceptualize and depict the relationship between data and knowledge. This approach has been challenged by Tuomi (1999), while more recent scholarly work highlights the importance of re-approaching data, knowledge, and organizations (see e.g., Alaimo and Kallinikos, 2022) or argues that data represents knowledge (Monteiro, 2022) calling for further exploration. However, current research still focuses on distinguishing ontological and epistemological differences between the two concepts depicted and explored separately in the literature (Tuomi,1999; Buckland,1991). Additionally, the importance of knowing data has been explored under the spectrum of data scientists (see Parmigianni et al., 2022). Echoing this discussion and considering how the organizational contexts and boundaries are shifting, making it also important for employees to work with data, in this work we introduce the term data knowledge as a specific type of knowledge. We posit that data knowledge differs from seemingly similar terms that appeared in literature, for instance, data literacy and data sensemaking (e.g., Lycett, 2013). Data sensemaking refers to the process of recognizing cues to form an understanding of complex and often uncertain data (Lycett, 2013; Lycett and Marshan, 2016). It is primarily concerned with the way people generate meanings, develop understandings, and create interpretations of data (Ibid). On the other hand, data literacy refers to the individual capability of dealing with data (Brown, 2023). Contrariwise, data knowledge goes deeper in further these terms to capture what is actually to know data, and thus encompasses these concepts. Following Cook and Brown (1999), we perceive what is possessed as knowledge whereas we approach knowing as an action and recognize their reciprocal relationship. To start the theorization process, we visit IS literature to investigate what the building blocks of data knowledge are and then we aim to provide a pragmatic framework for organizations to achieve data knowledge. To elucidate this, we formulate the following research questions: “What does constitute data knowledge and how is data knowledge enacted in organizational settings?”

To approach the research question, we build on the idea formulated by Grover and Lyytinen (2023). Following the suggestion of the authors, we argue that while it is critical to collect, organize, and provide access to digital trace data from various sources to analyze contemporary socio-technical phenomena, this should not come at the cost of a good abstract thinking and theory building effort (Ibid.). In relation to this, following Compeau et al. (2022)’s recommendation to re-assess the validity of constructs as context changes and technologies evolve, we stress the need for revisiting the concepts of data and knowledge to assess their validity and re-evaluate their relation. To understand the origins of data and knowledge, we first approach the concepts through etymological lens, inspired by Schwarz and Chin (2007). Working towards the described goal of identifying what constitutes data knowledge and responding to the first aspect of our research question, we systematically review the IS literature following Levy and Ellis (2006) ´s guidelines to understand how data and knowledge are intertwined. In this work, we present the preliminary findings of our systematic literature review (SLR) and the first step of understanding the concepts that construe data knowledge, namely: unveiling data, balancing between data and intuition, acknowledging internal and external capabilities, and realizing data. Unveiling the concepts that constitute data knowledge is a first step to understand the ghost the machine, the intangible elements, while going beyond the duality and hierarchical relationship of data and knowledge. As such, we aim to set the stage for contributing to IS field by reviewing dominant assumptions on data and knowledge, and identifying what the primary building blocks of data knowledge are. Our goal at the next stage is to develop a theory of explanation of data knowledge (Gregor, 2006) that synthesizes the new challenges and opportunities of organizations acquiring knowledge in the era of intelligent technologies. Then, we aim to provide an organized, pragmatic framework that presents how data knowledge can be turned into actionable processes in organizational settings for non-data scientists. Considering the emphasis that organizations put on data and the necessity for employees to work efficiently with data (Grover et al., 2018; Günther et al., 2022), we posit the importance to theorize on what is to know data and moving beyond the separation of data and knowledge.

Theoretical Background

Assumptions on Data: On the long-standing views and the state-of-the-art

Admittedly, within IS literature, to define data, some authors distinguish and juxtapose the concept with knowledge and information. The long-standing debate, which stresses a hierarchical classification, presents a conventional approach. Specifically, under this view, data are defined as symbols that represent the properties of objects and events and have not yet been interpreted (Spek and Spijkervet, 1997; Ackoff,1989). Data is perceived as "symbols that represent properties of objects and events and their environments" (Ackoff, 1989, p.3). This is a dominant view in many IS textbooks that define data as "raw facts that describe a particular phenomenon" (Haag and Cummings, 2013, p.50), a pre-existing process or quality (Milaksen and Monteiro, 2021). Subsequently, information consists of processed data, or according to Liew (2007), data are "unrefined and unfiltered information". A similar view on data is presented by Davenport and Prusak (1997), arguing that data are "simple observations about states of the world" (p. 9) that users perceive as objective and trustworthy (Anthony, 2021). This nuance of processing is directed at increasing the level of usefulness (Ackoff, 1989). Furthering the conceptual distinction between information and knowledge, Ackoff (1989) stresses that information focuses on descriptions while knowledge is conveyed by instructions, targeting questions that gear around exploratory 'how-to' questions. The pyramid view is challenged by Tuomi (1999), who sets knowledge first before information and data. Hence, knowledge needs to pre-exist for information to be formed and before data can be measured against information (see also Alavi and Leidner, 2001; Alaimo and Kallinikos, 2022).

Along the same line, Jones (2019) focuses on electronic medical records in hospitals and criticizes the current approaches. Specifically, the author refutes that data is referential (represents a world independent of themselves), objective (represents a world without interpretation), equal (same techniques and processes can be used to analyze all the data), and lastly that is foundational (it is based upon which our worldview is built). Contrariwise, the author calls for research on data in practice and data in principle. In alignment, raw data do not exist since the most elementary piece of "data" includes a pre-understanding, influence, or knowledge. More recent scholarly discussions depict data as objects of knowing (Monteiro, 2022), stressing the importance of understanding data as instruments of knowledge. In a similar vein, Alaimo and Kallinikos (2022) conceptualize data "as semiotic artifacts, instruments of knowing used to capture or represent, know, and act upon the world" (p.20). Jarvenpaa and Marcus (2020), discussing IS sourcing, elaborate on and differentiate three views of data: the commodity view of data, the processual view, and the relational view. The commodity view approaches data as a technical object and is perceived as easily editable, processable, and distributable. The processual view is in alignment with Jones (2019) and stresses the dynamic interdependencies between data come into being and into used, namely the ontological and epistemological stances. Lastly, the relational view, which is less investigated in IS literature, and it is more prominent in the work of Leonelli (2015). The author goes beyond the processual view and explicates that data are not only controlled by processes and current knowledge but also interactions among actors. Other scholars describe that data are highly constructed (Baskerville et al., 2020), continuously extended, recombined, updated (Ekbia, 2009), editable (Kallinikos et al., 2013), and contextually and historically situated (Parvinen et al., 2020).

Assumptions on Knowledge: On different approaches and state-of-the-art

Capturing the notion of knowledge is a tricky endeavor, as it is multifaceted and penetrates different disciplines (Nissen and Jennex, 2005). The notion of knowledge may be viewed through several perspectives (1) a state of mind, (2) an object, (3) a process, (4) a condition of having access to information, or (5) a capability (Avali and Leidner,2001). There are three predominant epistemological stances about knowledge: cognitivistic, connectionistic, and autopoietic (Joshi et al., 2007). First, the cognitivistic perspective approaches knowledge as a "fixed and representable entity (data) universally stored in computers, databases, archives, and manuals. According to this perspective, knowledge resembles data, and can be "unproblematically shared from one entity to another" (Venzin et al., 1998). Contrary to this view, the connectionistic perspective is seen to be contextual-dependent with "local differences between the rules and stocks of knowledge exist". Lastly, the autopoietic perspective implies that knowledge is history-dependent (Joshi et al., 2007). In the same discussion, knowledge has been categorized as embedded and affected by contextual factors (Carlile, 2002), situated (Monteiro, 2022), explicit (Alavi and Leidner, 2001), tacit (Novaka and Krong, 2009), and procedural (how things are done) (Marcus, 2001). Within organization studies, the notion of knowledge has often been conceptualized through the practice lens. Specifically, Orlikowski (2006) adopts the performativity lens, derived from the practice approach to knowledge, arguing that knowledge is not an external nor enduring substance. Instead, according to this view, knowledge is described as a dynamic social accomplishment. The practice view of knowledge explicates knowing as being emergent, always in the making, embodied and socially, culturally, historically embedded (Orlikowski,2006; Orlikowski, 2002) and material (Orlikowski,2006). By using the metaphor of "scaffolding knowledgeability", Orlikowski (2006) depicts knowledge as a balance and intersection between culture and material. In the same discussion, Carlile (2002), offering a pragmatic view, describes knowledge as localized, embedded, and invested in practice.

Method

Our research shows the need to theorize data knowledge motivated by the emerging IS literature (see Monteiro, 2022; Alaimo and Kallinikos, 2022). To do so, we followed a systematic literature review (SLR) in the AIS Senior Scholars' Basket of Journals using Scopus's advanced search, spanning the time 2000-2023. Intending to explore the data knowledge beyond the perspective of data scientists, we selected articles that focused on data work within different types of business organizations while trying to emphasize non-technical employees. We started the search with synonym words for data knowledge, which resulted in 71 articles. Then, abstracts and introductions were screened (Webster and Watson, 2002) to evaluate their relevance. After removing articles that did not emphasize organizational settings, we ended up with 34 articles. Later, a second screening was conducted to assess the relevant articles that a) do not only focus on data scientists, b) discuss the socio-technical process of data knowledge-making. After this, the remained results are 18 articles.

This research-in-progress follows, at this stage, an inductive approach (Shepherd and Sutcliffe,2011) (see Table 1, sub-concepts). We followed an iterative process of analyzing and categorizing our findings. Open coding was used to generate the initial data sub-concepts. Then to make a connection with knowledge and further categorize them, we draw from an etymological approach (Schwarz and Chin, 2007). Specifically, the etymological approach was used to explore the roots of the word knowledge and served as a basis to come up with main themes for theorizing data knowledge. Thus, to provide a definition of data knowledge, we first explore the definitions of data and knowledge in the literature, inspired by the etymological approach (Schwarz and Chin, 2007). Following Schwarz and Chin (2007) etymological methodology, we started by using the derivative definition methodology, where the "definition of the word (is found) by reference to its derivation; in English, usually by its derivation from Anglo Saxon, from Latin, and from Greek" (Borsodi 1967, p. 25). The chosen methodological direction is one of the most common etymological approaches and demands the researchers to determine the origin of the data and knowledge and utilize the original language to derive the current definitions (Schwarz and Chin, 2007). To discover the root of the word “knowledge”, we used Barnhart (1988) dictionary of etymology, where we discovered that the word “knowledge” was first recorded before year 1400 and means "capacity for knowing, understanding; familiarity;" "fact or condition of knowing, awareness of a fact;" or "news, notice, information; learning; organized body of facts or teachings" [1]. The word “knowledge” originates from Old English ‘cnawan’ which means ‘to know” (https://www.etymonline.com/search?q=knowledge). Following the etymological approach and drawing from an Old English Dictionary (Borden and Arthur, 1982), we understood that the word ‘cnawan’ had multidimensional meanings: *to recognize, to know, to perceive, to acknowledge and to declare*. These two dictionaries provided us with the multidimensional meaning of the word knowledge and its original root. In combination with the sub-concepts, four main themes derived. The value of the etymological approach “as first and foremost the explanation of (one) word by means of investigating its connection with (another) word” (Saussure, 1986, p. 259) assists us in understanding the relation of the two concepts that will later allow us to pinpoint the components and develop a definition of data knowledge. We remained open-minded to exploring IS materials that could unveil assumptions, relations, and nuances of data and knowledge (Smythe and Spence, 2012).

Based upon the results derived from two dictionaries, we explored that the definition of this verb is multidimensional, consisting of five dimensions: *to recognize, to know, to perceive, to acknowledge and to declare*. To further understand the meaning of each dimension, we used the Merriam-Webster dictionary (https://www.merriam-webster.com/) from where we excluded the verb ‘to recognize’ as we identify the similarities with the verb ‘perceive”. The dictionary meaning of etymological dimensions served us to categorize the sub-concepts of data, which were then further coded. The combination of etymological meaning with the sub-concepts of data led to the main concepts of data knowledge: *Unveiling data, balancing between intuition and data, Acknowledging External and Internal capabilities, and Realizing Data*.

Synthesizing preliminary results

Our findings demonstrate that there is a conceptual interrelation between the concepts. The concepts present what constitutes data knowledge, whereas the sub-concepts demonstrate the activities that lead to concepts. In this section, we briefly elaborate on the findings.

Dimension of the Verb	Dictionary Meaning	Elaborating the relations	Sub-concepts	Main concepts	Relevant Articles
To Know	-“To be acquainted or familiar with.” -“To have experience with.” -“To have understanding of.”	Employees´aacts of getting familiar with data and their expertise. Getting to know the data.	Data capturing Data curation Contextualing Data discovery Data sourcing Data use Identification Re-organize and map	Unveiling Data	Boldosova (2019);Grover et al.,(2018);Mikalsen and Monteiro(2021); Parmiggianni et al., (2022);Sternkopf and Mueller,(2018);Saghafi et al.,(2022); Tamm et al., (2022).
To perceive	-“The ability to see, hear, or become aware of something through the senses: cognition; understanding.”	Cognite outputs of both data and the surrounding environment.	Data-centric knowing Managing data cognitive of data Interpreting data outputs Contextual enablers Data understanding Pro-activeness	Balancing between data and cognition	Aaltonen et al., (2021); Grover et al. (2018); Mikalsen and Monteiro(2021); Sternkopf and Mueller,(2018); Tamm et al., (2022).
To acknowlegde	-“Accept or admit the existance or truth of.” -“To disclose knowledge of or agreement with.” -“To notice.”	Acknowledge occuring in multiple levels including: the organizational capabilities (e.g., readiness, resources, and negotiations)leading to agreements on data management and collaboration. It refers to acknowledging the competive pressure, and the data proccessed.	Organizational capabilities Individual absorptive capacity Negotiation Cross-department Collaboration Technological capability Environmental dynamism Reproposed data Re-used data	Acknowledging external and internal capabilities	Aaltonen et al., (2021); Aaen et al., (2022); Boldosova (2019);Chen et al., (2015);Mikalef and Krogstie (2020); Shao et al., (2022); Tamm et al., (2022).
To declare	-“To make known formally, officially or explicitly.”	All aspects of data communication and understanding.	Data sharing Deliberate storytelling Multi-perspective storytelling (Situated) data interpretion Flexibility in data representation	Realizing Data	Aaen et al., (2022); Almklov et al., (2014); Boldosova (2019); Mikalsen and Monteiro(2021); Sternkopf and Mueller,(2018);Park and Mithas (2020); Tamm et al., (2022);Someh et al., (2023)

Table 1. Visual Representation of the etymological meaning, and conceptualization of each category and its elements

Unveiling data

To be familiar with something is to know of it. In the context of data knowledge, this refers to being familiar with the data. Data is perceived as signals that represent ideas or objects (Sternkopf and Mueller, 2018), and working with data is a continuous process that involves two main practices: acquiring the correct data and preparing the data for use (Parmiggianni et al., 2022). It also requires initiating data training for those with little technical expertise (Boldosova, 2019) and collaboration between data analysts and non-technical professionals (Tamm et al., 2022). Additionally, generating value from data through methods such as transparency, access, discovery, experimentation, prediction, optimization, crowdsourcing, and proactive adaptation (Grover et al., 2018) is essential. When dealing with new data types, practices include testing different scenarios or hypotheses, and making sense of the data through connecting pieces, reorganizing, and mapping (Saghafi et al.et al, 2022). Kallinikos and Tempini (2014) ´s study in a healthcare organization describes data actions that include coding, merging and splitting data as practices that turn data into medical facts. Mikalsen and Monteiro (2018), expand this discourse by referring to data handling in knowledge infrastructures, including, adding the practices of modeling and disseminating data. Regarding the practice of dissemination, the role of data sourcing, as the process through which the meta-data travels with the data and the meta-data, is crucial for mitigating contextual and interpretive use of data (Jarvenpaa and Markus, 2020). Also, the authors highlight data sourcing including use and recombination.

Balancing between data and intuition

A second dimension of knowledge involves the ability to perceive something through the senses, such as sight, or hearing, or become aware of something through the senses: cognition; understanding. As Zuboff (1988) argues, “I see, I touch, I smell, I hear and therefore, I know” (p.62). This concept outlines the cognitive processes of data in relation to the surrounding environment within organizations. There, individuals strive to balance data and cognition through managing data cognitive outputs, interpreting data outputs, and being aware of contextual enablers (Aaltonen et al., 2021). This is important because data serves as a medium for sensemaking and creating knowledge, and it changes depending on the way it is managed (Aaltonen et al., 2021). The aim is to achieve "data-centric knowing" through practices such as accumulating, reframing, and prospecting (Mikalsen and Monteiro, 2021). Additionally, balancing analytical capabilities with the area of expertise is important for minimizing the risk that employees lose the ability to analyse with creativity over time (Tamm et al., 2022).

Acknowledging internal and external capabilities

The third dimension of knowledge in relation to data that we identify, is ‘to acknowledge’. This verb bears the meaning of accepting or admitting the existence or truth of something or disclosing knowledge of or agreement with it. For instance, Shao et al. (2022) viewed analytical capabilities as individual, dynamic and adaptive skills which depend on technology, environment, and individual characteristics. Therefore, this requires acknowledging capabilities at multiple levels: within the organization (Boldosova, 2019; Chen et al., 2021), for instance between cross-department level (Boldosova, 2019; Tamm et al., 2022), at the individual level (Shao et al., 2022), and outside the organization (Aaen et al., 2022). Organizational capabilities include readiness, organizational commitment (Boldosova, 2019), and the ability to unlock the value of data (Chen et al., 2021). On the other hand, cross-departmental capabilities include collaboration between data experts and non-technical experts. Data experts inform about analytical capabilities needed whereas non-technical experts inform on organizational processes (Tamm et al., 2022). Finally, the individual level is about an employees' ability to notice the value of data (Shao et al., 2022).

Realizing Data

The fourth dimension of knowledge involves sharing data with one another in an explicit way. Data sharing, contrary to data sourcing, indicates the direct or indirect act of reciprocal exchange. In this case, the meta-data is shared through tacit or intact knowledge among the shares. This process usually occurs among communities of practice (Jarvenpaa and Markus, 2020). Data sharing happens through practices of storytelling and interpretations. Storytelling conveys different forms of understanding and at the same time may influence others in the organization to achieve specific goals e.g., implementing a new technology (Boldosova, 2019). Thus, it conveys information about technology, the organization, its members, and customers (Boldosova, 2019). In addition, storytelling happens through interpretations of a situation when trying to understand a phenomenon which are accumulated by previous knowledge (Almklov et al., 2014; Someh et al., 2023). Finally, flexibility in interpretations allows for easily accommodating new data and insights into analysis (Saghafi et al., 2022).

Conclusion and Outlook to Further Research

This research-in-progress is the first step for a theory of explanation of data knowledge (Gregor, 2006). Koestler (1967) wrote of 'the Ghost in the machine' as an allusion to Descartes and the spirit within a physical body. We build on this idea to scrutinize today's AI systems by indicating that we go beyond the duality and the dominant hierarchical relationship between data and knowledge by developing a theory that bridge them together. Specifically, we proposed the construct 'data knowledge' as an alternative term that goes beyond the procedural, embedded, and situated categorization of knowledge to capture specifically *what is to know data*. Thus, we refer to the process and the outcome of the act of knowing (Cook and Brown, 1999). We draw from the etymological lens of data knowledge to understand the ontological stances. The discourse on data and knowledge stands strong, but still, their intersection and relation remain ambiguous, grounded in a separation. The etymological approach allows us to map the "what is data" and "what is to know". Through the SLR, we present some preliminary findings of what data knowledge is within IS literature. As a next step, we will first refine the presented findings by exploring additional IS but also management literature. Adding the management nuance is essential for conceptualizing further the notion of knowledge. We will then proceed to theorize. After its completion, the research will provide a theory of explaining (Gregor, 2006) and an actionable, organized pragmatic framework on how data knowledge unfolds in organizational settings while going beyond data professionals.

References

Aaen, J., Nielsen, J.A. and Carugati, A., 2022. The dark side of data ecosystems: A longitudinal study of the DAMD project. *European Journal of Information Systems*, *31*(3), pp. 288-312.

Aaltonen, A., Alaimo, C. and Kallinikos, J., 2021. The making of data commodities: Data analytics as an embedded process. *Journal of Management Information Systems*, *38*(2), pp. 401-429.

Ackoff, R. L. (1999) Ackoff's Best. New York: John Wiley & Sons, pp. 170 – 172.

Alaimo, C., & Kallinikos, J. (2022). Organizations decentered: Data objects, technology and knowledge. *Organization Science*, *33*(1), pp. 19-37.

Alaimo, C., Kallinikos, J., & Aaltonen, A. (2020). Data and value. In *Handbook of Digital Innovation*. Edward Elgar Publishing.

Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. MIS Quarterly, 25(1), pp. 107-136.

Anderson, C. 2008. The end of theory: The data deluge makes the scientific method obsolete. Retrieved from https://www.wired.com/2008/06/pb-theory/

Almklov, P.G., Østerlie, T. and Haavik, T.K., 2014. Situated with infrastructures: Interactivity and entanglement in sensor data interpretation. *Journal of the Association for Information Systems*, *15*(5), p.2.

Anthony, C. (2021). When knowledge work and analytical technologies collide: The practices and consequences of black boxing algorithmic technologies. *Administrative Science Quarterly*, *66*(4), pp. 1173-1212. https://doi.org/10.1177/00018392211016755.

Barnhart, R.K., 1988. *The Barnhart dictionary of etymology*. New York: HW Wilson Company.

Baskerville, R. L., Myers, M. D., & Yoo, Y. (2019). Digital first: The ontological reversal and new challenges for IS research.

Brown, S. (2023). Data literacy for leaders. MITSloan.https://mitsloan.mit.edu/ideas-made-to-matter/data-literacy-leaders(Accessed 29/04/2023).

Boldosova, V., 2019. Deliberate storytelling in big data analytics adoption. *Information Systems Journal*, *29*(6),pp. 1126-1152.

Borden, A.R., 1982. *A comprehensive Old-English dictionary*. University Press of America.

Chen, D.Q., Preston, D.S. and Swink, M., 2015. How the use of big data analytics affects value creation in supply chain management. *Journal of management information systems*, *32*(4), pp. 4-39.

Chen, D.Q., Preston, D.S. and Swink, M., 2021. How Big Data Analytics Affects Supply Chain Decision-Making: An Empirical Analysis. *Journal of the Association for Information Systems*, *22*(5), pp. 1224-1244.

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

Compeau, D., Correia, J., & Bennett Thatcher, J. (2022). WHEN CONSTRUCTS BECOME OBSOLETE: A SYSTEMATIC APPROACH TO EVALUATING AND UPDATING CONSTRUCTS FOR INFORMATION SYSTEMS RESEARCH. *MIS Quarterly*, *46*(2).

Cook, S. D., & Brown, J. S. (1999). Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. *Organization science*, *10*(4), pp. 381-400.

Grover, V., & Lyytinen, K. (2023). Not only data: The necessity of abstract thinking to make sense of digital phenomena. *Journal of Information Technology*, *38*(1), pp. 74-78.

Ekbia, H. R. (2009). Digital artifacts as quasi-objects: Qualification, mediation, and materiality. *Journal of the American Society for Information Science and Technology*, *60*(12), pp. 2554-2566.

Grover, V., Chiang, R.H., Liang, T.P. and Zhang, D., 2018. Creating strategic business value from big data analytics: A research framework. *Journal of management information systems*, *35*(2), pp. 388-423.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., Deken, F., & Feldberg, F. (2022). Resourcing with data: Unpacking the process of creating data-driven value propositions. *The Journal of Strategic Information Systems*, *31*(4), 101744.

Haag, S., Cummings, M., 2013. Management Information Systems for the Information Age, ninth ed. McGraw-Hill Irwin, New York, NY.

Jarvenpaa, S. L., & Markus, M. L. (2020). Data sourcing and data partnerships: Opportunities for IS sourcing research. In *Information Systems Outsourcing* (pp. 61-79). Springer, Cham.

Joshi, K.D., Sarker, S. and Sarker, S., 2007. Knowledge transfer within information systems development teams: Examining the role of knowledge source attributes. Decision Support Systems, 43(2), pp. 322-335.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & society*, *1*(1), 2053951714528481.

Koestler, A. (1967). *The ghost in the machine*. New York: Macmillan.

Levy, Y., & J. Ellis, T. (2006). A systems approach to conduct an effective literature review in support of Information Systems research. *Informing Science*, 9, pp. 181–212.

Leonelli, S. (2014). What difference does quantity make? On the epistemology of Big Data in biology. *Big data & society*, *1*(1), 2053951714534395.

Liew, A. (2007). Understanding data, information, knowledge and their inter-relationships. *Journal of Knowledge Management Practice,* 8(2), pp. 1-16.

Lycett, M. (2013). 'Datafication': making sense of (big) data in a complex world. *European Journal of Information Systems*, *22*(4), pp. 381-386.

Lycett, M. and Marshan, A. (2016). Capturing sensemaking pattern during data analysis: A conceptual framework. *25th International Conference on Information Systems Development*, pp. 106-116

Mikalef, P. and Krogstie, J., 2020. Examining the interplay between big data analytics and contextual factors in driving process innovation capabilities. *European Journal of Information Systems*, *29*(3), pp. 260-287.

Mikalsen, M., & Monteiro, E. (2021). Acting with Inherently Uncertain Data: Practices of Data-Centric Knowing. *Journal of the Association for Information Systems*, *22*(6), pp. 1715-1735.

Mikalsen, M., & Monteiro, E. (2018). Data handling in knowledge infrastructures: A case study from oil exploration. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1-16.Monteiro, E. (2022). *Digital Oil: Machineries of Knowing*. MIT Press.

Neeley, T., & Leonardi, P. (2022). Developing a Digital Mindset. *Harvard Business Review*, *100*(5-6), pp. 50-55.

Nonaka I, von Krogh G (2009). Tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory. Organ. Sci. 20(3):635–652.

Orlikowski, W. J. (2006). Material knowing: the scaffolding of human knowledgeability. *European Journal of Information Systems*, *15*(5), pp. 460-466.

Orlikowski, W. J. (2002). Knowing in practice: Enacting a collective capability in distributed organizing. *Organization science*, *13*(3), pp. 249-273.

Park, Y. and Mithas, S., 2020. Organized Complexity of Digital Business Strategy: A Configurational Perspective. *Mis Quarterly*, *44*(1).

Parmiggiani, E., Østerlie, T., & Almklov, P. G. (2022). In the Backrooms of Data Science. *Journal of the Association for Information Systems*, *23*(1), pp. 139-164.

Parvinen, P., Pöyry, E., Gustafsson, R., Laitila, M., & Rossi, M. (2020). Advancing Data Monetization and the Creation of Data-based Business Models. *Communications of the Association for Information Systems, 47(1), 2.*

Saghafi, A., Wand, Y. and Parsons, J., 2022. Skipping class: improving human-driven data exploration and querying through instances. *European Journal of Information Systems*, *31*(4), pp. 463-491.

Saussure, Ferdinand de. (1986) Course in General Linguistics. Translated by Roy Harris. Chicago: Open Court Press.

Schwarz, A., & Chin, W. (2007). Looking forward: Toward an understanding of the nature and definition of IT acceptance. *Journal of the association for information systems*, *8*(4), 4.

Shao, Z., Benitez, J., Zhang, J., Zheng, H. and Ajamieh, A., 2022. Antecedents and performance outcomes of employees' data analytics skills: an adaptation structuration theory-based empirical investigation. *European Journal of Information Systems*, pp. 1-20.

Shepherd, D. A., & Sutcliffe, K. M. (2011). Inductive top-down theorizing: A source of new theories of organization. *Academy of Management Review*, *36*(2), pp. 361-380.

Someh, I., Wixom, B., Davern, M., & Shanks, G. (2023). Configuring Relationships between Analytics and Business Domain Groups for Knowledge Integration. *Journal of the Association for Information Systems*, *24*(2), pp. 592-618.

Spek, R.v.d., and Spijkervet, A. Knowledge Management: Dealing Intelligently with Knowledge.Utrecht: Kenniscentrum CIBIT, 1997.

Sternkopf, H., & Mueller, R. M. (2018, January). Doing good with data: Development of a maturity model for data literacy in non-governmental organizations. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Steadman, I. 2013, January 25. Big data and the death of the theorist. Wired UK. Retrieved from https://www. wired.co.uk/article/big-data-end-of-theory

Stice-Lusvardi, R., Hinds, P. J., & Valentine, M. (2023). Legitimating Illegitimate Practices: How Data Analysts Compromised Their Standards to Promote Quantification. *Organization Science*, 1-21.https://doi.org/10.1287/orsc.2022.165.

Tamm, T., Hallikainen, P. and Tim, Y., 2022. Creative analytics: Towards data-inspired creative decisions. *Information Systems Journal*.

Tuomi, I. (1999, January). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers* (pp. 12-pp). IEEE.

M. Venzin, G. von Krogh, J. Roos, 1998, Future research into knowledge management, in: G.v. Krogh, J. Roos, D. Kleine (Eds.), Knowing In Firms: Understanding, Managing, and Measuring Knowledge, Sage Publications, Thousand Oaks, CA, pp. 26-66.

Waardenburg, L., Huysman, M., & Agterberg, M. (2021). *Managing AI wisely: From development to organizational change in practice*. Edward Elgar Publishing.

Webster, J. & Watson, R. T. (2002).Analyzing the Past to Prepare for the Future: Writing a Literature Review, MIS quarterly, vol. no. xiii-xxiii.