De-Agentializing Data Practices: The Shifting Power of Metaphor in 1990s Discourses on Data Mining

Niels Kerssens

05.16.19

Introduction: data mining as extraction of natural resources

> Big Data represents a tremendous opportunity to drill down and tap into these critical insights. In fact, the powerful potential to mine and refine this vital, valuable resource points to a direct comparison to a similarly vital resource in the modern economy: crude oil.[1]

> Meanwhile, artificial-intelligence (AI) techniques such as machine learning extract more value from data.[2]

---

[1] Arvind Singh, "Is Big Data the New Black Gold?," *Wired* (2013).

[2] unknown, "The World's Most Valuable Resource Is No Longer Oil, but Data," *The Economist* (2017).

> The current generation of AI startups recognize this difference and apply machine learning models to extract value from the data they collect.[3]

> Like oil, for those who see Data's fundamental value and learn to extract and use it there will be huge rewards.[4]

How do metaphor-driven representations of data mining in corporate media discourse affect understandings of data and data practices? This is, in short, the focal question of this article. Current media discourses of "Big Data" are characterized by a high degree of metaphorization[5] and these metaphors play a key role in how we (mis)understand this ill-defined phenomenon. For example, media scholars Cornelius Puschman and Jean Burgess argue that associated metaphors of big data as *a force of nature to be controlled* and a form of *nourishment/fuel to be consumed* lead to a deceptive understanding of big data in terms of natural phenomena. These metaphors are misleading, they argue, as by evoking the image of big data as a force of nature their value is presented as a naturally given, while the fact that value and meaning are actually inscribed by analysis is obscured.[6] This article aims to further deconstruct the representational politics of big data metaphors, directing specific attention to changes in the discursive—specifically metaphorical—framing of data practices in the 1990s, comparing academic discourses with corporate media discourses.

The *data mining is the extraction of natural resources* metaphor first emerged as a dominant figure in the 1990s. Nonetheless, it continues to shape popular understandings of data practices due to its pervasive presence in today's corporate media discourses on Big Data and data analytics, as evidenced by the excerpts presented above. In current discourse, the metaphor evokes a concept of value production in big data as analogous to the extraction of oil, or other natural resources, from the earthly deposits that contain and conceal them. The metaphor contains the associated metaphor of *big data as natural resource,* more specifically 'the oil of the information economy,'[7] which is equally pervasive in today's corporate media discourses on big data—often employed to portray its financial importance. Spread wide in business-oriented data discourses, these associated metaphors play a key role in the conceptualization of data practices and data analytics, thus shaping experiences and understandings of big data in the cor-

---

[3] Jocelyn Goldfein and Ivy Nguyen, "Data Is Not the New Oil," *Techcrunch.com* (2018).

[4] Joris Toonders, "Data Is the New Oil of the Digital Economy," *Wired.com* (2014).

[5] Cornelius Puschmann and Jean Burgess, "Big Data, Big Questions| Metaphors of Big Data," *International Journal of Communication* 8 (2014).

[6] Puschmann and Burgess, "Big Data, Big Questions," 1699.

[7] Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (Boston: Mariner Books, 2014).

porate context. For that matter, these are metaphors that the corporate world of big data 'computes by', to borrow Marianne van den Boomen's[8]rephrasing of the well-known expression by metaphor scholars Lakoff and Johnson.[9]They provide a frame of meaning for corporate data practices, steering the attention of the business world to a perception of valuable information (e.g. business insights) as some kind of inherent property of data, just as oil already contains an economic value even when still locked between layers of rock deep within the earth. But however tempting this metaphor may be, data is not oil and data analytics is not a process of extracting it.

Although ubiquitous in current media discourse, the metaphor is far from new. It first emerged in the 1990s when business organizations in the United States and elsewhere started using database technologies and data mining algorithms to convert large amounts of digital data collected during the 1970s and 1980s into valuable business insights. These database technologies developed from work in the academic field of research that emerged under the name *Knowledge Discovery in Databases* (KDD) in the late 1980s and early 1990s. When American business organizations started using data mining systems, the American technology press reported extensively on the new phenomenon, despite its extremely complex technological nature—one that easily evaded apprehension of journalists. To make the seemingly incomprehensible nature of data mining understandable to a public of management executives, and to communicate its potential corporate benefits, journalists drew on the language associated with natural resource extraction. Data was discussed as the new gold, and its 'mining' as a process of natural resource extraction.

The metaphor of *data mining as the extraction of natural resources,* I argue, played a central role in the misrepresentation of data mining in 1990s corporate media discourse. It discursively shaped a now common yet false perception of value production in big data as natural, intermediary and unbiased process by cancelling out any concept of the agential role of technologies and humans contributing to the production of value and meaning from data. Such misrepresentations of big data and data mining still play an important role in the legitimation of these practices, supporting the unquestioned integration and implementation of data analytics in the worlds of business, science and public governance, especially in current times of abundant media hype arguing the financial benefits of 'Big Data'.[10]For that matter, such misleading accounts of the nature of data

---

[8]Marianne van den Boomen, *Transcoding the Digital: How Metaphors Matter in New Media* (Amsterdam: Institute of Network Cultures, 2014), 23.

[9]George Lakoff and Mark Johnson, *Metaphors We Live By* (Chicago: University of Chicago press, 1980).

[10]Nick Couldry and Jun Yu, "Deconstructing Datafication's Brave New World," *New Media & Soci-*

and its analyses, such as those that de-agentialize data analytics, pose a threat to the critical agency of all professionals—in business, academia, government—currently involved in data and algorithm-driven knowledge production. By attending to changes in the discursive—specifically metaphorical—framing of data practices in the 1990s, comparing academic KDD discourse with corporate media discourses, this article traces a significant shift in the meaning-making power of the data mining metaphor. In doing so, it aims to open up debate concerning more truthful representations and conceptualizations of data practices, which can form the starting point for a more thoughtful acceptance and implementation of big data in any domain—whether business, science, or government.

The analytical section starts with examining the role of the *data mining* metaphor as conceptualization device in the academic discourse of KDD. I will do so through the analysis of (metaphorical) discourse on data mining in academic books and articles that formed foundational publications for this emerging field of research into knowledge discovery. All of these sources were published between the late 1980s and early 1990s. To examine the role of metaphorical language in corporate media discourse, I use examples from American news items published in popular technology journals[11] that formed part of the trade press of the emerging data analytics industry in the 1990s, such as *Byte*, *Datamation*, *Infoworld*, and *Informationweek*.[12] All news items contain references to data mining and were published in the United States between 1992 and 1997.[13]

---

ety (2018); David Beer, "Envisioning the Power of Data Analytics," *Information, Communication & Society* 21, no. 3 (2018); Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (Sage, 2014).

[11] I chose to focus on popular technology magazines to represent corporate media discourse based on the assumption that these sources played the most important role in creating a media hype about data-mining, using metaphorical language. This is in contrast to other possible sources such as handbooks, training manuals, business education materials, technology reviews, and software manuals. Magazine advertisements and TV commercials about data mining and database technology did contribute to the media hype and also made use of metaphorics (e.g. IBM's 1996 television commercial that was part of its "Solutions for a Small Planet" advertising campaign). However, these do not show a different kind of metaphorical depiction than the news sources I studied—the main reason why advertising has not been included in the corpus.

[12] The initial choice was to focus on the popular technology magazines *Datamation* and *Byte* that were read by business personnel working in the entire field of business computing and data processing. Both magazines had large circulation numbers in the 1990s and were distributed internationally (including Europe), although published in the United States. Because an online search in English language publications pertaining to all geographical regions indexed by LexisNexis and the global business information database Dow Jones Factiva (which includes *Byte* and *Datamation*) returned news articles from other well circulated news publications as well, I chose to include them in the corpus. I queried both database's English language publications for "data-mining" and associated keywords including: "data processing"; "data warehouse"; "gold rush"; "OLAP"; and "knowledge discovery". The search retrieved ca. 90 relevant news articles.

[13] All articles analyzed in this research were printed in magazines and trade journals published in

These two analytical sections reflect the historical character of this research with which it attempts to meet recent calls urging researchers to start thinking historically, more specifically, genealogically, about big data[14] and deconstruct the assumed newness of the big data phenomenon by bridging the gap between (critical) data studies and the history of computing.[15] But first I will specify the analytical frame of this research in two interrelated sections. In the first I will discuss how and why I approach data and data mining as concepts that are discursively constructed. In the second section I discuss how I study metaphor as a conceptualization device that, to a greater or lesser extent, discursively shaped the meaning of data mining in the 1990s.

Treating data mining as a concept

Within existing literature about (big) data and algorithms there have been several moves towards criticizing notions of these phenomena as objective and unbiased. Some have stressed how 'raw' or' pure' data doesn't exist as the production of data is always preceded by an 'interpretative base' enabling data to be envisioned and imagined as data.[16] Others have emphasized the hidden biases in big data that lead to false senses of accuracy[17] and how biases in the process of data creation lead to bias in the database itself.[18] Yet critical media scholars have directed particular attention to algorithms, studying power relations deployed and expressed by computational processes.[19] Some have already started to investigate such powers by deconstructing the history of contemporary data practices, machine learning techniques (e.g. Bayes classifier; neural networks; vector machine; decision

---

the United States directed at an audience of business professionals. Although printed in the US, key publications such as *Byte* and *Datamation* were distributed internationally, which means it is safe to say they their discourses impacted industry professionals outside of the US context.

[14] Kitchin; David Beer, "How Should We Do the History of Big Data?," *Big Data & Society* 3, no. 1 (2016).

[15] Amelia Acker, "Toward a Hermeneutics of Data," *IEEE Annals of the History of Computing* 37, no. 3 (2015).

[16] Lisa Gitelman and Virginia Jackson, "Introduction," in *"Raw Data" Is an Oxymoron*, ed. Lisa Gitelman (Cambridge; London: The MIT Press, 2013).

[17] Daniel A McFarland and H Richard McFarland, "Big Data and the Danger of Being Precisely Inaccurate," *Big Data & Society* 2, no. 2 (2015).

[18] Julia Adams and Hannah Brückner, "Wikipedia, Sociology, and the Promise and Pitfalls of Big Data," ibid.

[19] David Beer, "The Social Power of Algorithms," *Information, Communication & Society* 20, no. 1 (2017); Tarleton Gillespie, "The Relevance of Algorithms," in *Media Technologies: Essays on Communication, Materiality, and Society*, ed. Tarleton Gillespie, Pablo J Boczkowski, and Kirsten A Foot (MIT Press, 2014).

trees; linear regression; k-means clustering) and operations (e.g. vectorization; optimization; probabilization; pattern recognition).[20]

Driven by either such historical or more empirical methodologies, media scholars increasingly debunk assumptions of objectivity that have historically been associated with algorithms, instead foregrounding their non-neutral, partial, biased, and subjective nature, and how this affects the knowledge they help produce, and the decision-making processes in which they are involved. For example, in his book *The Black Box Society* (2015) Frank Pasquale explicitly argues for the importance of looking beyond the data by placing research emphasis on the role algorithms play in the operationalization of data. As Pasquale points out, the guiding and organizing role of algorithms is central from the perspective that 'critical decisions are made not on the basis of the data per se, but on the basis of data analyzed algorithmically'.[21]

While in this current scholarly discussion of big data and algorithms concerns are primarily directed to the power that algorithmic technologies exert over our lives,[22] this study directs attention to the power of the 'discourse surrounding algorithms'.[23] This means I will direct analytical focus to the *concept* of data mining—and the metaphor as conceptualizing device—instead of data mining methods, techniques, and technologies themselves.[24] Recently, David Beer has argued for the importance of research studying how the conceptualization of big data[25] and algorithms takes place.[26] As Beer put it, '[w]e need to look beyond the algorithms themselves, to explore how the concept of the algorithm is also an important feature of their potential power'.[27] With this, Beer places research focus on understanding how 'notion[s] of the algorithm' take shape in 'the discourse surrounding algorithmic processes'.[28] He emphasizes in particular the importance of research into the way in which these discourses about algorithmic processes play a role in perpetuating and legitimizing particular rationalities and ideologies that 'promote certain values and forms of calculative objectivity',[29] or

---

[20] Adrian Mackenzie, *Machine Learners : Archaeology of a Data Practice* (Cambridge, MA: MIT Press, 2017); Bernhard Rieder, *Engines of Order* (Amsterdam: Amsterdam University Press, 2019 (To be published).

[21] Frank Pasquale, *Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2016), 21.

[22] John Danaher et al., "Algorithmic Governance: Developing a Research Agenda through the Power of Collective Intelligence," *Big Data & Society* 4, no. 2 (2017).

[23] Beer, "The Social Power of Algorithms," 11.

[24] Beer, "The Social Power of Algorithms," 11.

[25] David Beer, "How Should We Do the History of Big Data?," *Big Data & Society* 3, no. 1 (2016).

[26] Beer, "The Social Power of Algorithms."

[27] Beer, "The Social Power of Algorithms," 7.

[28] Beer, "The Social Power of Algorithms," 11.

[29] Beer, "The Social Power of Algorithms," 7.

what media scholar José van Dijck refers to as 'dataism'.[30]

In a similar way, media scholar Rob Kitchin emphasizes the role of corporate discourse in producing a business rationale of big data that legitimates its widespread adoption by promoting a narrative of managerial and financial bene-fits—more effective and efficient business management and leveraging idle data resources.[31] In doing so, such corporate discourses, as Couldry and Yu argue, neutralize data practices by treating data as 'as if they were "natural", part of "nature"' thus making them 'distinctively immune from critique'.[32] This research deconstructs the most powerful discursive mechanisms shaping corporate and public perceptions of data mining through a historical perspective that traces a discursive shift, more specifically, a shift in the meaning-making power of the data mining metaphor in the 1990s, when corporate media discourses started treating data mining as if analogous to natural resource extraction.

Studying metaphor as a discursive framing device in discourses on data mining

While this research follows Beer and treats data mining as a concept, it studies the degree to which metaphor framed its conceptualization. This study draws on some of the insights from metaphor theory and critical discourse analysis (CDA) to examine metaphors as representational instruments—as discursive framing devices driving the conceptualization of data mining in discourses about data practices of the 1990s. Here I follow new media scholar Marianne van den Boomen who argues that metaphors are fundamental to how we conceptualize and understand computational processes and practices in particular.[33] Because technically complex and highly abstract, understanding a computational phenomena such as data mining is facilitated by the conceptualizing power of metaphor to turn this unknown into a known. Metaphor, as van den Boomen states, is a 'meaning-making machine' that doesn't so much represent, but constitutes and transforms the technological objects and processes it aims to ren-

---

[30] José Van Dijck, "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology," *Surveillance & Society* 12, no. 2 (2014).

[31] Kitchin distinguishes an academic from a business rationale. According to Kitchin, the academic variant promotes big data in terms of advancing knowledge and developing a better understanding of the world. In the business rationale, on the other hand, big data's financial importance is emphasized.

[32] Couldry and Yu, "Deconstructing Datafication's Brave New World," 3.

[33] van den Boomen, *Transcoding the Digital*.

der intelligible.[34] Although other discursive framing devices exist,[35] metaphors appear as salient in corporate and popular discourse on big data and data mining, something that already started more than twenty years ago. For that reason, the study of these metaphors is important for understanding how, and by what kinds of politics, they transform the abstract and complex nature of computational processes into understandable practices that have meaning and value for the corporate world.

To study changes in the meaning-making power of the *data mining* metaphor in 1990s academic and corporate media discourses on data practices, the article draws upon the concept of 'discourse metaphor'.[36] Zinken, Hellsten en Nerlich define a discourse metaphor as 'a relatively stable metaphorical projection that functions as a key framing device within a particular discourse over a certain period of time'.[37] In the study of these metaphors particular attention is directed at media discourses and how metaphors frame a topic or debate in the service of particular rationales 'through the features and constraints they impose'.[38] Within critical discourse analysis attention is directed at how metaphors in media discourse are both strategically and subconsciously chosen by powerful actors including journalists, business people, and politicians to purposefully highlight certain aspects of a media phenomenon, whilst downplaying others. I will draw upon the concept of discourse metaphor to substantiate a change to the extent to which the *data mining as extraction of natural resources* metaphor shaped the meaning of data practices in corporate media discourse, compared to earlier KDD discourse.

To study such conceptualization in more detail, I draw on the theory of conceptual (or cognitive) metaphor, such as first extensively discussed by Lakoff and Johnson.[39] Work in the field of cognitive linguistics focuses on metaphorical conceptualization as a cognitive process, often through a detailed linguistics-based analyses of metaphor structure and operation. Such an approach is not the intention here. For this study the theory of conceptual metaphor provides an analytical lens for examining how metaphor conceptualizes data mining in academic and media discourses. Central to the theory is the assumption that 'the essence of metaphor is understanding and experiencing one kind of thing in terms of another'.[40] This other is often a more familiar concept that can serve as an explana-

---

[34] van den Boomen, *Transcoding the Digital*, 187.

[35] Zhongdang Pan and Gerald M. Kosicki, "Framing Analysis: An Approach to News Discourse," *Political Communication* 10, no. 1 (1993).

[36] Jörg Zinken, Iina Hellsten, and Brigitte Nerlich, "Discourse Metaphors," *Body, language and mind* 2 (2008).

[37] Zinken, Hellsten, Brigitte Nerlich, "Discourse Metaphors," 241.

[38] Puschmann and Burgess, "Big Data, Big Questions," 1696.

[39] Lakoff and Johnson, *Metaphors We Live By*.

[40] Lakoff and Johnson, *Metaphors We Live By*.

tory bridge for making sense of a more abstract concept. The familiar concept is known as the 'source domain', and metaphor is understood as the transference of meaning from this known domain to what is termed a 'target domain'. In discourse, the explanatory power of metaphors 'rests in the fact that the familiarity of the"known" [source] domain (e.g. the extraction of natural resources) can offer initial guidance to investigate and to plumb the ''unknown'' [target] domain" (e.g. data- and algorithm driven knowledge production).[41]

Yet, as Lakoff and Johnson explain, 'the very systematicity that allows us to comprehend one aspect of a concept in terms of another (…) will necessarily hide other aspects of the concept'.[42]Metaphors that highlight certain aspects of a concept, they argue, 'can keep us from focusing on other aspects of the concept that are inconsistent with that metaphor'. This also applies in particular to technological metaphors, as is the case in this article. Here transference does not occur between concepts, but between concepts associated with a (natural) source domain and a target domain that comprises 'a material-physical system that affects a state of affairs'—such as databases and data mining algorithms.[43]Technological metaphors are frequently criticized, because they have the potential to conceal certain aspects of technology, while 'making others appear natural'.[44]That is, the politics of such metaphors operate through these twin dimensions of naturalization (and ontologization) and obscuration, which encourage misleading interpretations of phenomena concerned.

Such politics of metaphor will, obviously, only have any effect on our understanding of the target domain when metaphor itself appears as a device driving conceptualization—that is, when it operates as a discourse metaphor. This study will demonstrate that it is only when corporate media discourse adopted the data mining metaphor from KDD, and repurposed it by connecting it with the domain of natural resources, that metaphor developed into a full-blown discourse metaphor. And it is only in possession of such meaning-making power that the politics of the mining metaphor—naturalizing and ontologizing—come to affect our perception of data and data mining practices, resulting in a de-agentialized and neutralized view of corporate data mining practices. Here de-agentialization refers to the role of metaphor in representing the production of meaning and value in data practices as 'brought about in […] ways, impermeable to human [and technological] agency' such as natural forces.[45]

---

[41]Rowan Wilken, "An Exploratory Comparative Analysis of the Use of Metaphors in Writing on the Internet and Mobile Phones," *Social Semiotics* 23, no. 5 (2013).

[42]Lakoff and Johnson, *Metaphors We Live By*, 13.

[43]van den Boomen, *Transcoding the Digital*, 51.

[44]Puschmann and Burgess, "Big Data, Big Questions," 1697.

[45]Theo van Leeuwen, "Representing Social Action," *Discourse & Society* 6, no. 1 (1995): 96.

Drawing on the theory of conceptual metaphor this article will study the *data mining is the extraction of natural resources* metaphor as a key framing device within a corporate discourse on data mining, and interpret the politics of such framing by examining the kind, and degree of, naturalization and obscuration involved.[46] Yet to be able to interpret this role of metaphor as a shift in the discursive framing of data practices, I will first study the role of the mining metaphor in academic KDD discourse.

Agentializing data-driven knowledge production: KDD and the mining metaphor

> Because computers have enabled humans to gather more data than we can digest, it is only natural to turn to computational techniques to help us unearth meaningful patterns and structures from the massive volumes of data. Hence, KDD is an attempt to address a problem that the digital information era made a fact of life for all of us: data overload.[47]

> The decade of the 1990s has brought a growing data glut problem to the worlds of science, business, and government. Our capabilities for collecting and storing data of all kinds have far outpaced our abilities to analyze, summarize, and extract "knowledge" from this data.[48]

Illustrated by the excerpts above, mining metaphors and associated terminology (e.g. extraction, unearthing), had a presence in the academic discourse of knowledge discovery in databases (KDD) in the early 1990s. Employing such metaphorical language, KDD's academic discourse prepared the much more extensive use of these metaphors in later corporate media discourse, which

---

[46] I started the analysis with an open-ended reading of all of the selected texts without very specific questions or hypotheses constraining the analysis. Because the size of the corpus was not too large, and the average length of the articles was not too long, I read all of the texts completely. In reading I did pay close attention to headlines, the first one or two paragraphs of all the articles constituting the corpus, what technological objects they discussed, and if and how they employed metaphor to constitute these objects discursively. I selected the quotes included in the article on the basis of the extent to which they aligned with the discourse metaphor *data mining is the extraction of natural resources*.

[47] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI magazine* 17, no. 3 (1996): 28.

[48] Usama M Fayyad et al., *Advances in Knowledge Discovery and Data Mining*, Advances in Knowledge Discovery and Data Mining (Menlo Park, CA: AAAI Press and MIT Press, 1996), xiii.

closely followed developments in the field. However, within KDD discourse, I argue, metaphorical language did not function as a key discursive framing device. Metaphors functioned mainly at the surface of discourse. They had little or no impact on how KDD researchers understood and theorized the process of knowledge discovery. KDD discourse, I argue, produced concepts of data and algorithm-driven knowledge production that *agentialized* these practices—meaningfully constructing knowledge discovery as a multi-agential practice of producing knowledge and value from data, involving a combination of algorithmic agencies (e.g. data summarization; classification; regression; clustering) and human agencies (e.g. data preparation, selection and cleaning; and the interpretation of mined patterns).

KDD emerged as an interdisciplinary field of research within the computer science community in the late 1980s and early 1990s. It brought together practitioners from a diverse array of fields including statistics, machine learning, artificial intelligence, information retrieval, and management information systems.[49] Demonstrated by the historical excerpts presented above, the field developed in response to a problem practitioners in the field articulated in terms of a 'data overload' and 'a growing gap between data generation and data understanding.'[50] In short, KDD researchers claimed that state-of-the-art tools and methods for generating meaning and value from data were increasingly inadequate to discover knowledge from exponentially growing amounts of data stored in corporate computer databases.

Since at least the mid-1970s, in close connection with the development and implementation of database management systems (DBMS) and the theory of the relational database model, American businesses developed an awareness, or mindset, of the potential for data to be turned into valuable business insights.[51] This formed a key incentive for systematically generating, collecting, and storing administrative and transactional data with the aim of generating valuable business information from it. In the late 1970s and 1980s analytical and other cognitive capacities of data analysts still drove the process of turning business data into valuable insights. Human data analysts employed statistical techniques on small data sets to provide summaries and generate reports for management executives. When the quantity of data grew exponentially in the 1980s, however, business practition-

---

[49] Usama M Fayyad, "Data Mining and Knowledge Discovery: Making Sense out of Data," *IEEE Expert: Intelligent Systems and Their Applications* 11, no. 5 (1996): 23.

[50] William J Frawley, Gregory Piatetsky-Shapiro, and Christopher J Matheus, "Knowledge Discovery in Databases: An Overview," in *Knowledge Discovery in Databases*, ed. William J Frawley and Gregory Piatetsky-Shapiro (AAAI Press Menlo Park, CA, 1991), 2.

[51] Niels Kerssens, "The Database 'Revolution': The Technological and Cultural Origins of the Big Data-Based Mindset in American Management. 1970s-1980s," *TMG - Journal for Media History* 21, no. 2 (2018).

ers and researchers increasingly realized that human-driven methods of analyses were no longer sufficient to effectively and efficiently infer business intelligence from large amounts of data.[52] In result, a select amount of researchers and business practitioners developed considerable interest in automating the data analytical process through the use of computers and data mining techniques.[53]

KDD researchers defined data mining as an inductive and automated method of data analyses, applying 'specific [discovery] algorithms for extracting patterns from [typically large amounts of] data.'[54] Here inductive referred to the use of algorithmic processes for discovering relationships and patterns from large data sets that had not already been described in pre-established hypotheses—a definition that is still used for the idea behind the mining of big data.[55] The use of this concept of data mining in KDD was a legacy from the field of statistics—no conscious or strategic choice of researchers to make sense of an abstract and complex process such as computational analyses in terms of the more familiar and concrete concept of mining for *natural* resources. In other words, natural resource extraction did not (yet) provide a target domain for making sense of computational data practices.

Within statistics, the term data mining was used since at least the 1960s 'to describe the process of trawling through data in the hope of identifying patterns' without any a-priori hypothesis to verify the findings.[56] In this context of statistics, the term carried many negative connotations, as many statisticians argued that patterns detected by these methods could 'simply be a product of random fluctuations' that didn't 'represent any underlying structure.'[57] One of the reasons for why computational methods for mining databases developed in fields outside statistics,[58] such as the sub-branch of Artificial Intelligence research known as machine learning. Research in machine learning focused on 'the automation of inductive learning processes' by modeling the inductive learning methods of 'hu-

---

[52] William J Frawley, Gregory Piatetsky-Shapiro, and Christopher J Matheus, "Knowledge Discovery in Databases: An Overview," *AI magazine* 13, no. 3 (1992); Ronald J Brachman et al., "Mining Business Databases," *Communications of the ACM* 39, no. 11 (1996).

[53] Jan M Zytkow and John Baker, "Interactive Mining of Regularities in Databases," in *Knowledge Discovery in Databases*, ed. William J Frawley and Gregory Piatetsky-Shapiro (AAAI Press Menlo Park, CA, 1991), 31.

[54] Fayyad, Piatetsky-Shapiro, and Smyth, "From Data Mining to Knowledge Discovery in Databases," 39.

[55] Dyche (2012) thus argues that 'mining Big Data reveals relationships and patterns that we didn't even know to look for.' (in Kitchin, *The Data Revolution*.)

[56] David J Hand, "Data Mining: Statistics and More?," *The American Statistician* 52, no. 2 (1998): 112.

[57] Hand, "Data Mining: Statistics and More?"

[58] Leo Breiman, "Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author)," *Statistical science* 16, no. 3 (2001).

mans and other intelligent creatures' in the machine so they can be performed by a computer.[59] It developed as a branch of Artificial Intelligence research starting somewhere in the late 1950s. In the late 1980s, KDD researchers started employing theories and techniques advanced in machine learning (e.g. neural networks, decision trees, genetic algorithms) for developing algorithmic methods dedicated to the task of inferring knowledge from large data sets. KDD scientists adopted the mining metaphor from the field of statistics, using it to refer to computational pattern detection methods, while presenting the use of these methods as the solution to a (corporate) 'need to find the knowledge adrift in the flood of data'.[60]

Discussing and explaining the value and contents of data mining, KDD researchers did associate data mining processes with practices of *extracting* resources. For example, in the foreword of a founding publication of the field, KDD researcher John Ross Quinlan wrote, '[s]uch collections [one containing many thousands of records] are potential lodes of valuable knowledge, but in order to extract the ore, we must have efficient mining tools.'[61] Additionally, leading KDD researchers Gregory Piatetsky-Shapiro and William J. Frawley defined data mining as 'procedures to *extract* [emphasis added] knowledge from data.'[62] By use of the extraction metaphor, it seemed as if these authors treated data mining as a tool for withdrawing some sort of latently existent value from data. Putting forward a notion of value and meaning as latently existing within databases also occurred by evoking the concealed property of natural resources. For instance, such as the case when defining data mining as 'the search for relationships and global patterns that *exist* [emphasis added] in large databases, but are "hidden" among the vast amounts of data'.[63]

However, although mining metaphors were sometimes used in KDD literature, any explicit reference with the extraction of natural resources (e.g. oil, gold) was missing. Moreover, their meaning-making power did not exceed the surface of academic discourse. That is to say, *mining as the extraction of knowledge from data* did not meet the conditions of a 'discourse metaphor', and had little or no impact on how KDD researchers conceptualized the process of turning data into knowledge. It was the concept of *knowledge discovery* rather than *data mining* which steered their conceptualization. Researchers defined knowledge discovery as an iterative, dynamic and constructive process of 'discovering useful knowl-

[59] Marcel Holsheimer, *Data Mining: The Search for Knowledge in Databases* (Amsterdam, the Netherlands: Centrum voor Wiskunde en Informatica, 1994), 8.

[60] Frawley, Piatetsky-Shapiro, and Matheus, "Knowledge Discovery in Databases: An Overview," 2.

[61] John Ross Quinlan, "Foreword," In *Knowledge Discovery in Databases*, x.

[62] William J Frawley, Gregory Piatetsky-Shapiro, and Christopher J Matheus, "Knowledge Discovery in Databases: An Overview," In *Knowledge Discovery in Databases*, 15.

[63] Holsheimer, *Data Mining: The Search for Knowledge in Databases,* 1.

edge from data.'[64] They further conceptualized the process as one in which data *accumulated* value and meaning through a step-by-step procedure that involved data mining through algorithmic analysis, but also 'many decisions made by the user' to ensure the usefulness and trustworthiness of the knowledge produced.[65]

Importantly, KDD researchers deliberately chose knowledge discovery, not data mining, as the label for their research to emphasize that knowledge formed the end-product of a *knowledge-making* process—not the starting point of a process directed at its extraction from a database. Additionally, they made special effort to emphasize that data mining and knowledge discovery were not to be seen as synonyms. With knowledge discovery, as mentioned, they referred to 'the overall process of discovering useful knowledge from data.'[66] Data mining referred to an essential, yet particular, step in this process. Other steps that were considered essential parts of the process of turning data into knowledge were, for example: developing understanding of the application domain and the relevant prior knowledge; identifying the goal of the process from the customer's viewpoint; data preparation, selection and cleaning; data reduction and projection in function of the goal of the process; matching the goal of the process with a particular data mining method (e.g. summarization; classification; regression; clustering); choosing data mining algorithm(s) and selection method(s) to be used for searching for data patterns; interpreting mined patterns (and 'returning to any of steps 1 through 7 for further iteration'); and, finally, acting on the discovered knowledge by 'incorporating it into another system for further action'.[67]

KDD researchers developed data practices, methods and techniques on the basis of a concept of knowledge discovery that acknowledged how bias and subjectivity were integral to any data- and algorithm driven type of knowledge production, to ensure such variables had as little effect as possible on the knowledge produced. That KDD researchers were strongly aware of the pitfalls of their automated approach to the analyses of large data sets is also reflected by their attention for issues associated with: the imperfect nature of the data (e.g. garbled and missing data);[68] the inductive nature of its methods of analysis (e.g. finding patterns that appear to be statistically significant but, in fact, are not)[69]; and algorithmic

[64] Fayyad, Piatetsky-Shapiro, and Smyth, Fayyad, Piatetsky-Shapiro, and Smyth, "From Data Mining to Knowledge Discovery in Databases," 39.

[65] "The Kdd Process for Extracting Useful Knowledge from Volumes of Data," *Communications of the ACM* 39, no. 11 (1996).

[66] Fayyad, Piatetsky-Shapiro, and Smyth, "From Data Mining to Knowledge Discovery in Databases," 39.

[67] Fayyad, Piatetsky-Shapiro, and Smyth, "From Data Mining to Knowledge Discovery in Databases," 42.

[68] Quinlan, "Foreword," In *Knowledge Discovery in Databases*, x.

[69] Fayyad, Piatetsky-Shapiro, and Smyth, "From Data Mining to Knowledge Discovery in

bias.[70]In sum, I argue, KDD researchers agentialized and de-neutralized data- and algorithm driven knowledge production—representing knowledge as output of a process through which data accumulated value and meaning through the potential subjective mediation of algorithmic and human agencies.

The mining metaphor as discursive framing device in corporate media discourse

> There is a gold-rush on. Only this time, the hunt is not for gold of the shiny metallic type, but the digital variety made up of bytes of information.[71]

> There's gold in your data, but you can't see it. It may be as simple (and wealth-producing) as the realization that baby-food buyers are probably also diaper purchasers. It may be as profound as a new law of nature. But no human who's looked at your data has seen this hidden gold. How can you find it?[72]

As discussed, research in KDD had strong ties with the corporate world and developed in response to an issue that particularly in the business world was perceived as both a problem and opportunity—the (in)ability to efficiently and effectively draw knowledge from increasingly large data sets. Although data mining methods and techniques were still at an extremely early stage of development in the mid-1990s, various data mining technologies developed for business appear on the market during this period. Examples include *Darwin* developed by Thinking Machines, *MineSet* developed by Silicon Graphics and the *Intelligent Data Miner* developed by IBM. The trade press started reporting on the nascent data mining industry at an early stage, closely following technological advances in what seemed to be a most promising field. To the dismay of KDD researchers, data mining developed into a true media hype in the mid-1990s. Hype was supported by a business-oriented rationale of reporting, set out to explain how data mining formed a huge opportunity to maximize profit by leveraging idle data resources. The *data mining is the extraction of natural resources* metaphor (extraction of gold in particular) embodied the business rationale. From the mid-1990s onwards,

---

Databases," 40.

[70]Quinlan, "Foreword," In *Knowledge Discovery in Databases*, x.

[71]Michael Luck, "Nuggets of Information: Exploiting Data Resources," *Computing* (1992).

[72]Edmund X DeJes us, "Data Mining: Turn Computers Loose on Your Data, and You Don't Know What They'll Come up with-That's the Whole Point," *Byte* 20, no. 10 (1995).

the metaphor developed into a key device for conceptualizing, envisioning and promoting the economic value of data mining for business.

In order to properly substantiate the role of this *mining for gold* metaphor in meaningfully framing data-driven value production in the corporate world, it is important to first briefly outline the context of reporting in which the metaphor was used. Here it is important to understand that the trade-press did not just report developments in KDD research and data mining applications in business—it also played an important role in shaping and feeding the need of business organizations for employing the new technologies in the first place. Firstly, corporate media discourse emphasized that volumes of data amassed and stored by business organizations were 'exploding,'[73] yet at the same time it portrayed such 'vast amount[s] of data […] being collected' as 'lying idle', having little purpose or value for organizations.[74]

Importantly, news articles promoted data mining as the means to relatively easily turn such worthless data into economic value. The presumed relationship between data mining and financial gain is strongly reflected in titles of news articles such as: 'Mining for Dollars' (1996);[75] 'Datamining unearths dollars from data' (1997);[76] 'Strike it rich' (1997);[77] and 'Poppin fresh dough' (1997).[78] Even though there were few actual applications of data mining, examples of successful applications were frequently cited to make concrete how data mining could increase profit. Here emphasis was placed on the use of data on customers and buying patterns—as exemplified in the article 'Unearthing underground data' (1996):

> Grocery chains have analyzed customers' baskets of purchases and learned that cosmetics buyers typically also purchase greeting cams. They've subsequently increased sales in both product categories by redesigning store layouts to ensure that the two product lines were positioned in the same aisle.[79]

Additionally, the trade press emphasized the need for a rapid response by promoting data mining not only as the essential means for businesses to increase

---

[73] Sara Reese Hedberg, "The Data Gold Rush: Smart Data Miners Are Cashing in on Valuable Information Buried in Private and Public Data Sources," *Byte* 20, no. 10 (October 10 1995): 83.

[74] Luck, "Nuggets of Information: Exploiting Data Resources."

[75] David Aubrey, "Mining for Dollars," *Computer Shopper* 16, no. 8 (1996).

[76] Eva Freeman, "Datamining Unearths Dollars from Data," *Datamation* 43, no. 7 (1997).

[77] Vance McCarthy, "Strike It Rich! Thanks to New Datamining Tools, Companies Are Unearthing Valuable Information About Their Customers. The Result? Closer Customer Relationships and a Healthier Bottom Line," *Datamation* 43, no. 2 (1997): 44.

[78] Michael J. Tucker, "Poppin' Fresh Dough," *Datamation* 43, no. 5 (1997): 50-60.

[79] Cheryl D Krivda, "Unearthing Underground Data," *LAN magazine* 11, no. 5 (1996).

financial gain, but also as a way to improve their competitive position. As journalist Sara Reese Hedberg wrote in the computer magazine Byte in 1995:

> The amount of information stored in databases is exploding. From zillions of point-of-sale transactions and credit card purchases to pixel-by-pixel images of galaxies, databases are now measured in gigabytes and terabytes. In today's fiercely competitive business environment, companies need to rapidly turn those terabytes of raw data into significant insights to guide their marketing, investment, and management strategies.[80]

Such news items on data mining in the 1990s, I argue, shaped the meaning of data mining in service of a business rationale of efficiency and economics, which, as I will argue, facilitated the unquestioned adoption and implementation of data mining technologies by American business.

Mining for gold

> While the practice of corporate data warehousing and data mining is receiving a lot of hype, there's considerable confusion about what it is and who should be using it. The vague, even proprietary terminology for describing this sophisticated approach to storing and retrieving database information hides the serious technology that comprises this practice.[81]

The use of the mining for gold metaphor by the American trade-press played a crucial role in the production and promotion of a business rationale of data mining that emphasized its managerial and financial benefits. As mentioned in the excerpt above, the press paid little attention to explaining the workings of data mining technology, which, I argue, enforced the power of natural resource metaphors to meaningfully frame data practices discussed in corporate media discourse. The mining for gold metaphor, I argue, produced a neutralized and de-agentialized understanding of data mining—that is, it shaped and reinforced a corporate perception of automated methods of data analyses as intermediary and objective vehicles for extracting valuable insights from data. The mining for gold metaphor conceptualized data mining as such by mapping two interrelated

---

[80] Hedberg, "The Data Gold Rush."
[81] Aubrey, "Mining for Dollars."

concepts drawn from the natural resource domain of gold mining onto data practices, that is: (1) the idea that valuable insights already have a latent existence in data; and (2) the idea that data mining methods manifest (rather than construct) this value through extraction.

The latent existence of value in mountains of data

> There's gold in them there [sic] databases! That's the rallying cry of business analysts and database administrators who have discovered the techniques and benefits of data mining.[82]

> It's in there. The discovery, the fact, the one piece of the puzzle that will blow away the competition, propel your company to the top, and stick a "VP" after your name. It's right there, in your database. But you can't see it. Yet.[83]

In the historical examples from news items on data mining cited above, information is treated as something that always already *exists within* the databases. Analogous to nuggets of gold, valuable insights ("the one piece of the puzzle") are presented as pre-existing the method of their excavation, as somehow already present "right there, in your database". Making use of the same analogy with the mode of existence of natural resources, news articles describe how such insights are buried deep within layers of data, hidden from human perception. Exemplified by the excerpts below, data is treated as some kind of ore—analogous to the ores in which gold occurs in nature—providing a containing and protecting body that conceals valuable pieces of information:

> Deep within the pulsating mass of bits and bytes strung throughout the enterprise lie answers to the most perplexing problems of any business. Which customers will turn to competitors? Which offers will prompt customers to buy more? What are the signs of fraudulent activity?[84]

> You know it's there. Buried in gigabytes of marketing data or point-of-sale transactions lies the key information about an important customer trend or a successful product launch. Now all you have to do is extract it in a way that informs the decision at hand.[85]

---

[82] Barbara DePompa, "There's Gold in Databases: New Tools Will Help Companies Extract Valuable Information," *Information week*, no. 561 (1996).

[83] Hedberg, "The Data Gold Rush."

[84] Julie Bort, "Data Mining's Midas Touch; Data Mining Tools Are Helping Companies Reap Rewards from Their Data Warehouses," *Info World* (1996).

[85] Alan Simon, "Without the Proper End-User Tools for Data Access and Analysis, Your Data Warehouse May Be Shuttered. Better Clients, Better Decisions," *Byte* 22, no. 1 (1997).

> Data mining will likely have many network users digging in their organization's databases for buried treasure.[86]

> Sophisticated new tools … help to remove the information "ore" buried in corporate files or archival public records.[87]

Valuable insights formulated as 'answers to the most perplexing problems of any business' or 'key information about an important customer trend' are presented as existing buried within a data ore—'pulsating mass of bits and bytes'; 'gigabytes of marketing data or point-of-sale transactions' —that just needs to be removed for these insights to appear in the self-contained form of a business asset that can be used to drive managerial decision-making.

The problem of the analogy with gold mining, as pointed out by Puschman and Burgess, lies in 'inferring certain properties of the target that the source domain possesses but that do not map onto the target'.[88] For example, the *given* character of a natural resource is not in accordance with the *constructed* nature of data. Due to the analogy with a physical source domain so familiar to our notion of how value is created from given resources such as gold, the idea of data as a resource given by nature appears as quite natural to us. As Gitelman and Jackson point out, '[a]t first glance data are apparently before the fact: they are the starting point for what we know […]'.[89] Yet, they point out, the result of treating data as starting point is 'an unnoticed assumption that data are transparent, that information is self-evident, the fundamental stuff of truth itself'.[90] Gitelman and Jackson attempt to make clear in their book that data has no ground of existence independent of the human imagination of data, which, they argue, is always grounded in an 'interpretative base'.[91] And these interpretations, they emphasize, already play an important role in generating and producing data. Rather than being extracted, all data is created by humans to be recorded by machines. Approaching data as analogous to the natural resource of gold thus naturalizes—presents as naturally given—something that cannot exist without human interventions in its production—the reason why Gitelman and Jackson consider the phrase 'raw data' an 'oxymoron'. By disabling human agencies on the production side of data, the metaphor is also misleading because it conceals the representational nature of stored data—the fact that it is always an abstraction (sample) of what it repre-

---

[86] Cheryl D. Krivda, "Datamining Dynamite: Blasting Loose Those Buried Nuggets of Information Requires Clean Data, Warehousing Strategies, Powerful Parallel Processors, and Heaps of Hard Disk Space," *Byte* 20, no. 10 (October 10 1995): 97.

[87] Scott Mace, "Data Turns into Gold for Information Detectives," *Info World* (1994).

[88] Puschmann and Burgess, "Big Data, Big Questions."

[89] Gitelman and Jackson, "Introduction," 2.

[90] Gitelman and Jackson, "Introduction," 2.

[91] Gitelman and Jackson, "Introduction," 3.

sents, and not reality itself (population). This also means the metaphor ignores issues related to possible human (selection and observational) bias involved in the creation of data, which can affect the reliability and integrity of datasets.

The problem of treating the value of data as naturally given—something that always already exists—within the databases, however, goes beyond suppressing the role of human and technological agencies in the production of data. There is another property natural resources such as gold possess, which does not map onto the relation between data and information. Gold actually *exists* as a thing, a nugget ready-formed in the earth. It thus has an actual ontological status in the sense of it pre-existing within the earth a-priori to its method of extraction, already harboring an inherent economic value. Unlike gold, information doesn't exist in a nugget-like form prior to the process of mining the data. The mining for gold metaphor, however, plays an important role in what van den Boomen refers to as *ontologizing*—treating information, or business intelligence, as if it composes a stable matter of fact that has an independent existence. As van den Boomen points out, '[i]n the act of ontologizing dynamic processes get substituted with their results'.[92] Here it means that through ontologizing business insights are treated as things with an ingrained economic value that latently exist within a data ore, waiting for the moment that this value is discovered. This is misleading, as rather than being inherent to data, value is accumulated by, and actively assigned to, data through algorithmic analysis and human interventions in the preparation, selection, and interpretation of data tools and their outputs—an agentic notion of *knowledge discovery* that, as discussed, was actively promoted within the KDD research community.

Manifesting value through extraction

> Data mining lets the power of computers do the work of sifting through your vast data stores. Tireless and relentless searching can find the tiny nugget of gold in a mountain of data slag.[93]

> Data mining is the act of drilling through huge volumes of information in order to discover relationships, or to answer specific questions, that are too broad in nature for traditional query tools.[94]

---

[92] van den Boomen, *Transcoding the Digital,* 37.

[93] Edmund X. DeJesus, "Data Mining: Turn Computers Loose on Your Data, and You Don't Know What They'll Come up with-That's the Whole Point." *Byte* 20, no. 10 (October 10 1995): 83.

[94] Julie Bort, "Data Mining's Midas Touch."

> Varghese believes the big payoff from data warehousing and mining will not come with new tools, but with training users to dig deep into the vast store of business data.[95]

> Datamining is the industry's latest solution to the problem of un-earthing data that has been carefully squirreled away.[96]

The natural resource metaphor is embedded within the extraction metaphor—the latter referring to the process of withdrawing the resource from the earthly body that contains it. In the excerpts cited above, mining is sometimes explicitly referred to as extraction, or implicitly through the use of associated metaphors that treat data mining as a process of removing the 'information nuggets' from its container (e.g. sifting, digging, drilling, unearthing). Through analogy with these extraction verbs, I argue, data mining techniques are treated as *vehicles*—that is, intermediary processes that act as a means for unlocking and transferring value and meaning from its concealed and worthless existence in data, to a more worthy and actionable existence above ground in the world of corporate management:

> Data mining (DM) … is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data nuggets.[97]

> In reality, data mining is the process of sifting through vast amounts of information in order to extract meaning and discover new knowledge.[98]

> The whole point of data mining is to reveal hidden information for prompt decision-making and action.[99]

> Mining for data: A powerful new query ability lets users drill down through millions of records to grab golden nuggets of information that, when combined, yield creative answers to questions no one ever thought to ask.[100]

> A raft of data mining tools offers a wide range of features for digging up business opportunities.[101]

---

[95] DePompa, "There's Gold in Databases."

[96] Deborah Asbrand, "Is Datamining Ready for the Masses?," *Datamation* 43, no. 11 (1997).

[97] Hedberg, "The Data Gold Rush."

[98] Hedberg, "The Data Gold Rush."

[99] Cheryl D Krivda, "Datamining Dynamite."

[100] Louis Connor, "Mining for Data: A Powerful New Query Ability Lets Users Drill Down through Millions of Records to Grab Golden Nuggets of Information That, When Combined, Yield Creative Answers to Questions No One Ever Thought to Ask," *Communications Week*, no. 596 (1996).

[101] Herb Edelstein, "Mining for Gold: A Raft of Data Mining Tools Offers a Wide Range of Features

> Oftentimes, datamining can unearth fundamental facts about a company's business.[102]

Illustrated by the examples above, data mining is presented as a value *extraction* rather than a value *creation* process—one in which data mining techniques 'extract meaning'; 'reveal hidden information'; 'grab golden nuggets of information; 'dig[…] up business opportunities; and 'unearth fundamental facts'. To demonstrate how such extracted information could directly be of value for corporate organizations, news articles linked them to a variety of purposes and results:

> Data mining is the process of sifting through mountains of data for patterns usually buying patterns that could be useful for marketing or other purposes.[103]

> What WalMart and several other MPP system buyers are most interested in doing is "data mining," digging through mounds of data in search of buying patterns and other nuggets of information on which to build marketing strategies and create new products.[104]

> Have you learned anything new from your data lately? Datamining will help you find subtle, unexpected patterns hidden in your database, which could lead to increased sales and healthier profits.[105]

> Strike it rich! Thanks to new datamining tools, companies are unearthing valuable information about their customers. The result? Closer customer relationships and a healthier bottom line.[106]

Illustrated by these excerpts, corporate discourse emphasized the results and benefits of data mining—'marketing'; 'creat[ing] new products'; 'increased sales and healthier profits'; and 'closer customer relationships'. While emphasizing its corporate benefits, most of the articles paid little attention to the advanced technology involved in data mining, which certainly strengthened the power of metaphor to make data mining sound as easy as using a word processor or a spreadsheet.

Again, the problem with the analogy between data mining (target) and the type of

---

for Digging up Business Opportunities. Here's How to Find the Best Product for You.," *Information Week*, no. 627 (1997).

[102] Vance McCarthy, "Strike It Rich!"

[103] Kevin Fogarty, "Data Mining Can Help to Extract Jewels of Data," *Network World* 11, no. 23 (1994).

[104] Michael Alexander, "Mine for Gold with Parallel Systems," *Datamation* 40, no. 22 (1994).

[105] Eva Freeman, "Datamining Unearths Dollars from Data," *Datamation* 43, no. 7 (1997).

[106] Vance McCarthy, "Strike It Rich!"

extraction processes involved in the mining for natural resources (source) is that the properties of the source domain do not match with the properties of the target domain on which they are mapped. The metaphor of extraction directs focus to an understanding of data mining as a non-agential transfer process—merely *transporting* meaning and value from its data container to a more actionable existence within the organization. This obscures the fact that data doesn't *possess* value and meaning but has to *acquire* value and meaning through human practices and technological processes that help to contextualize it.[107] What is portrayed as an extraction procedure is actually a meaning-making process in which computational mining techniques and operations—such as classification, vectorization, optimization, probabilization and pattern recognition—are employed to make sense of the data and to *create* a 'nugget' of information. In other words, through data mining value and meaning is actively assigned to data, which means such methods have an agential capacity in a process of making meaning.

These methods, as discussed, are not neutral or value-free, but developed on the basis of particular theories of how to infer value from finite data sets—theories that all have their own assumptions and concepts of the relation between data and information, meaning the idea of bias underlies all of them. In that respect, the data mining as extraction metaphor is misleading because it distances business intelligence from all technological and human agencies involved in data's meaningful contextualization. Placing focus on business insights as things subject to excavation, the metaphor denies the agential capacity and role both algorithmic technologies and humans play in what is actually a mediated process of *constructing* (not extracting) knowledge from data, thereby automatically cancelling out a moment of critical reflection on the nature of the knowledge produced.

Conclusions

Using a historical perspective, this article traced a shift in the meaning-making power of the data mining metaphor in the 1990s, comparing KDD and corporate media discourses. Here I followed recent works arguing for the importance of studying discourses *surrounding* data practices and algorithmic processes,[108] including metaphors,[109] because, as claimed, discursively constructed concepts of such real-world phenomena are also key factors of 'their potential power'. I

---

[107] Acker, "Toward a Hermeneutics of Data."

[108] Beer, "The Social Power of Algorithms."; "Envisioning the Power of Data Analytics."; Couldry and Yu, "Deconstructing Datafication's Brave New World."

[109] Puschmann and Burgess, "Big Data, Big Questions."

studied how a concept of data mining was variously constructed within these different discourses, directing particular attention to a significant change in the role of the mining metaphor as conceptualization device shaping understandings of value production through data practices. As argued, the mining metaphor developed from a device with marginal impact on the meaningful construction of data practices within KDD discourse, into a full-blown discourse metaphor when natural resource extraction suddenly appeared as a source domain in corporate discourse. The metaphor framed such media discourse, and played a very significant part in discursively shaping a dominant concept of these practices. This metaphor-driven conceptualization, I argue, led to critical misrepresentations of data and data practices, which continue to exist and affect current interpretations.

Within the field of KDD, the metaphor of discovery rather than mining functioned as a framing device for data practices. Metaphors associated with the domain of natural resource extraction had a presence in KDD discourse, yet had little or no impact on how KDD researchers conceptualized their efforts. KDD researchers conceptualized data mining as the application of (algorithmic) methods for finding useful patterns in data. They inherited this more technical understanding of the concept from the field of statistics. In contrast to later corporate media discourses, the extent to which the concept derived meaning from the domain of natural resource extraction was negligible. That is to say, concepts from the domain of natural resource extraction didn't provide any explanatory bridge for the conceptualization of data practices, nor was data discussed analogous to a natural resource (e.g. oil, gold). Instead, KDD researchers employed the discovery metaphor to communicate their preferred understanding of these practices. As a linguistic device, however, this metaphor had little impact on how data practices were conceptualized in KDD discourse. Conceptualization occurred by accurately defining the relevant terms, and by splitting the data to knowledge trajectory into a series of component processes and practices. Knowledge discovery was specified as a multi-step process for generating knowledge from data, involving a combination of human and algorithmic agencies, all considered essential to data practices. KDD discourse, in that sense, *agentialized* rather than *de-agentialized* data and algorithm-driven knowledge production. It portrayed data practices in terms of an agentic notion of knowledge discovery—as a process of *knowledge-construction* in which value is accumulated by, and actively assigned to, data through a combination of algorithmic agencies (e.g. mining patterns through summarization; classification; regression; clustering) and human agencies (e.g. data selection, cleaning; reduction and projection of data in function of objective; interpretation of mined patterns).Importantly, by detailing rather than obscuring these agencies, KDD discourse foregrounded the potentially non-neutral, partial, biased, and subjective nature of its efforts.

In the corporate media discourses that developed from the mid-1990s onwards, however, the mining metaphor rapidly developed into a discursive framing device that shaped collective understandings of data practices. Moreover, it did so by meaningfully connecting the notion of mining to the domain of natural resource extraction rather than the field of statistics. In corporate media discourse, I argue, the metaphor *data mining is the extraction of natural resources* helped produce and promote a business rationale of data-driven knowledge production highlighting financial and organizational benefits of leveraging abundant data resources. While the metaphor helped to promote data mining's benefits for business, it did so through the discursive construction of a de-agentialized concept of knowledge production through data mining. Through the analogy with the process of extracting natural resources such as gold, data-driven knowledge production was neutralized—cancelling out all agential forces (e.g. algorithmic decisions and human interpretations) contributing to a process of data *accumulating* meaning and value. In effect, metaphorical representation constructed an image of data mining as natural, more specifically, objective and neutral process.

The dominant presence of metaphors associated with natural resource extraction in contemporary big data discourses only gives evidence to the fact that metaphors continue to play a dominant role in shaping and coloring the corporate perception of the data to knowledge trajectory. The role of discourse in promoting such a value-neutral business rationale is problematic, maybe even dangerous, as many critical media scholars have pointed out. By associating big data practices with a misleading image of objectivity, such discourses facilitate the unquestioned acceptance of big data in the business world.[110]Moreover, they protect the corporate world from ethical questioning.[111]The same applies also to other domains, such as science and government, in which data mining techniques are increasingly used to produce knowledge or inform decision-making. Deconstructing the discursive devices (such as metaphors) through which data analytical processes are conceptualized, helps to show how dominant conceptions do not neatly map onto practices and processes at hand, involving critical misrepresentations. Data mining, as many media scholars have argued, is not a neutral vehicle of value transportation, and, as stressed in the article, so aren't the conceptual metaphors we 'mine' by. To ensure that all people currently working with big data analytics (in business science, or government) maintain critical agency, it is crucial that we keep on deconstructing the celebratory discourses currently in power of (mis)shaping our collective understanding of data practices.

To prevent the unquestioned acceptance of these practices in the worlds of busi-

---

[110]Beer, "The Social Power of Algorithms," 7.
[111]Couldry and Yu, "Deconstructing Datafication's Brave New World."

ness, science and government, it is important that we provide representations that employ more accurate framings that do more neatly map onto data practices and processes at hand rather than ones constructed through the use of easy and attractive natural resource metaphors. Using a historical perspective, this article exposed contingencies in the connection between explanations of data practices and the domain of natural resources, and showed that in the absence of this connection, KDD discourse provided alternative and more accurate representations of data practices. I assert that we can learn from KDD discourse, and how it conceptualized data mining as a component step in a broader process of knowledge discovery. As a starting point, we can (re)position the notion of *knowledge discovery* rather than *data mining* at the center of discourses surrounding data practices. In doing so, the discursive construction of the data to knowledge trajectory will depart from the end-product—knowledge—rather than a presumed (natural) starting point—data. This will encourage us to conceptualize *discovery* as a multi-agential process of *making* knowledge from data, instead of conceptualizing the trajectory as a de-agentialized practice of exploiting a data resource through extraction, that is mining.