Where are we headed in business analytics? A framework based on a paradigmatic analysis of the history of analytics

Nik Rushdi Hassan
University of Minnesota Duluth, nhassan@d.umn.edu

Where Are We Headed in Business Analytics? A Framework Based on a Paradigmatic Analysis of the History of Analytics

Completed Research Paper

Nik Rushdi Hassan
University of Minnesota Duluth
1318 Kirby Drive, LSBE 385A
Duluth, MN USA
nhassan@d.umn.edu

Abstract

The explosion of interest in business analytics (BA) comes with multiple problems. With as many as eleven distinct disciplines teaching analytics, it is not clear which areas of study constitute the BA field. If the information systems (IS) field is to exert a significant influence in analytics, what the IS researcher and practitioner need to focus on has to be made clear. Using a paradigmatic historiographical analysis of the field of analytics this study provides evidence for the bifurcation of analytics into data science and BA as founding disciplines of computer science, mathematics and statistics, machine learning and IS contribute to the analytics movement. The results from this analysis also identify a set of conceptual foundations for BA that takes advantage of both the intellectual strengths of the IS field without sacrificing the necessary depth of data science.

Keywords: information systems theory, historical paradigms, data science, big data, business analytics

Introduction

The information systems (IS) field is at the forefront of the explosion of interest in data analytics in business. The Association to Advance Collegiate Schools of Business (AACSB International) estimates that worldwide there are over 400 business analytics (BA) programs in over 220 business schools (Davenport, 2019). This number has significantly increased from the first program at North Carolina State University in 2007 to 131 programs by 2012 (Wixom et al., 2014). The numbers for data science programs, the school of science counterpart of BA, is no less impressive. A GitHub resource (Swanstrom, 2019) lists 604 programs as of April 17, 2019, which includes existing BA programs. The University of Michigan announced its Michigan Institute for Data Science (MIDAS, https://midas.umich.edu/about/) in 2015 with a donation of $100 million and 25 faculty (Donoho, 2017). Today, MIDAS works with over 230 faculty from across the University of Michigan system. The University of Virginia received $120 million to establish a school of data science (Hester, 2019) while the University of California San Diego received $75 million for its data science institute. The *Wall Street Journal* (Belkin, 2018) reported on the fastest growing class at the University of California Berkeley, the Foundations of Data Science class, with 1,295 students and an enrollment of 1,200 majors (Harcourt, 2018). It is not uncommon to find both BA and data science programs in the same university. Virginia Tech's Center for Business Intelligence and Analytics operates in the Pamplin College of Business while its Computational Modeling and Data Analytics degree program is housed in its Academy of Integrated Science that combines four departments, statistics, mathematics, computer science and physics.

As Wixom et al., (2014) noted, this explosion of interest for BA education did not occur without problems. The analytics movement appears to be split between BA in the business school and data science in the college of sciences (Aasheim, William, Rutner and Gardiner, 2015; Phelps and Szabat, 2017). Especially for the BA program, the dearth of any guidelines or model curricula raises fundamental questions around the number of courses that constitute the BA major or minor, the relationship of BA with other majors and especially with the fields of statistics and computer science, and how best to evolve the BA program. With as many as eleven unique disciplines teaching analytics (Wixom et al., 2014) ranging from Marketing and Finance to IS, computer science and mathematics, it is no surprise that the inter-disciplinary approach applied by the programs in the college of sciences has been very successful for data science. The same may not apply for BA. Studies of the curriculum (Gorman and Klimberg, 2014; Phelps and Szabat, 2017) for BA shows a broad range of emphases ranging for example, from no coverage at all in statistics to over 50% of the courses in the program, and from no coverage for IS courses to 100%. Most BA programs are designed to incorporate a database or data warehousing course, visualization, big data, data modeling and data mining courses with a higher-level statistics course, but generally do not require programming courses. Data science programs offered by either computer science or mathematics and statistics academic units, or as an interdisciplinary program, generally require programming courses, visualization, big data, data modeling, and data mining courses, and may not require database or warehousing courses. Generally, data science programs tend to require more statistics courses and a higher level of math competency than BA programs (Aasheim et al., 2015; Phelps and Szabat, 2017).

It may be clear to most within the IS field that our researchers are not expected to invent the next Hadoop or MapReduce, or even to write the next classification or clustering algorithm. If those technologies are not where our efforts should be expended, what exactly is the role of the IS researcher, and by extension, the practice of BA that is most relevant to IS? Is the IS researcher left with the trite and uninspiring task of researching the adoption or acceptance of big data analytics? Or can the IS researcher, as Dhar (2013) proposes, provide interesting answers to questions that we do not yet know? Or even better, as Pentland (2014) claims, we can solve macro-level problems using the micro-level big data that are being analyzed and "build a society that is better at avoiding market crashes, ethnic and religious violence, political stalemates, widespread corruption, and dangerous concentrations of power" (p. 17). If these claims are true, BA presents to the IS field, the potential towards establishing its relevance. The question is: Given the wide breadth of the analytics field and the seemingly separate domain of data science, which areas should the IS field be focusing on to achieve these goals, assuming they are at least being considered? If such focus areas can be identified, what exactly is the role of the BA version of the data scientist? Or is the data scientist no different from the BA expert? At the organizational level, where will the BA expert reside and what can be expected from that expert? Which journals should IS researchers publish their BA research? These are some of the questions that will be addressed with the help of an historical analysis of the analytics movement.

It is interesting to note that when writing about BA, Davenport and colleagues (2007; 2010; 2018) generally use the generic "analytics" term and do not use the "business" qualifier before "analytics," except in a 2005 Babson College research report (Davenport, Cohen and Jacobson, 2005) that predated Davenport and Harris (2007), where the term "business analytics" is specifically mentioned. Using just "analytics" avoids the complications associated with choosing which particular reference to analytics the writing is about. But that leaves the analytics concept being addressed in ambiguity, which is not ideal, especially for rigorous research work. After performing a paradigmatic historiographical analysis of analytics, this essay will propose a set of conceptual foundations for BA that can be traced back prior to the introduction of decision support systems and business intelligence, and with this solid foundation, help engender a unique and fecund research tradition and build an equally rewarding profession.

The Alternative Historiographical Paradigms of Analytics

As Bryant et al., (2013) explain, the process of researching and writing the history of IS will help address most, if not all of the concerns of the field with regard to its direction, identity, subject matter and relevance. For what identity would nations have if not for the history about them? The subject matter of the arts is defined in part by the field of art history. Anything of significance is recognized and becomes relevant when we say that it "will go down in history!" In other words, historical analyses provide the necessary foundations for many fields. The Association for Information Systems (AIS) itself rewards its scholars with the LEO award, which is based on J. Lyons & Co. that used to own teashops and hotels, and supplied bread, cakes and pastries in the United Kingdom in the 1950s. What made J. Lyons & Co. significant to the IS field was their sponsorship of the EDSAC project, one of the earliest computers invented, their implementation of an EDSAC-derivative system, the LEO (Lyons Electronic Office) that automated the company's office tasks and developed what would be called a decision-support system (DSS) for management decision making (Land, 2010, 2015), and their prototypical Chief Information Officer (CIO), John Simmons, who architected the system (Simmons, 1962). Such a network of related events makes the LEO not only an historic event to the IS field, but a defining framework for the whole field (Hassan, 2018).

By analyzing the historiographical paradigms (Hassan, 2018) associated with analytics, we can identify all the intellectual, technical, social, and political traditions and influencers of BA and suggest not only what would be most suitable for the IS field, but also project possible future directions that are most productive for both our researchers and practitioners. Many IS authors claim that analytics evolved from decision support systems (DSS) or business intelligence (BI), two well-founded areas within the IS field (Watson, 2011, 2014; Chen, Chiang and Storey, 2012; Holsapple, Lee-Post and Pakath, 2014; Sharma, Mithas and Kankanhalli, 2014; Abbasi, Sarker and Chiang, 2016; Mikalef, Pappas, Krogstie and Giannakos, 2018). However, an historiographical paradigmatic analysis of analytics may uncover something quite different. For example, the term "business intelligence," although well-known within IS circles, was actually coined by prolific IBM researcher and inventor Hans Peter Luhn, who was working on indexing, classifying and encoding documents in order to expedite the search for information. He developed this concept to address the growing volume of organizational data and information that needed to be better managed (Luhn, 1958). In other words, the original business intelligence had to do more with information retrieval (or today's text analytics) than it did with what is understood by IS researchers today.

Table 1: Historiographical Paradigms (adapted from Hassan, 2018)

Historical paradigm	***Narrative***	***Relevance to BA***
Enlightenment paradigm	Exceptionalism in present history, rationalistic, practical spirit freeing mankind from superstition	The study of the potential of big data analytics to contribute to the advancement of society by relying on the rationalistic principles of the mathematical sciences
Romantic paradigm	Literary, heroic leadership, spiritual values, teleological, optimism in people's potential, appreciation for particularism and contextualized study (historicism)	Study of the human spirit behind analytics, the visionaries and champions of analytics who laid the foundations and led the analytics movement
Rankean paradigm	Scientific method, order, objectivity, organic whole, rigorous reconstruction, content and scientometric historiography	Objective, scientific study of the analytics movement finds a bifurcation of analytics into data science and BA
Social science paradigm	Philosophy and methodology of social science, focus on society, politics, economy and culture, social structures (structuralism), theoretical abstractions, interpretivism, generalizing categories	Sociological study of how different disciplines appropriate analytics for their benefit and how they exert influence within the analytics movement.
Critical paradigm	Human consciousness, class struggle, social action, dialectical process, domination, emancipation	A critical examination of how the analytics movement dominates the business narrative resulting in a disproportionate emphasis on data at the expense of other aspects of society

Postmodern paradigm	Post-structuralism, idiographic, genealogy, textual analysis and hermeneutics, deconstruction, cultural anthropology and ethnography	A more relativistic, deeper interpretive approach to studying analytics that takes into consideration the unique context, non-absolutist, symbolic view of its development

Using the historiographical paradigms proposed by Hassan (2018), Table 1 summarizes the narratives associated with each paradigm. The example of Luhn leading the field of information retrieval, coining "business intelligence" and its historicistic implications would qualify as studying the history of BA based on the Romantic paradigm. The social science paradigm uncovers the eventual transformation of business intelligence from its IBM roots to something that the IS field is more familiar with in terms of data warehousing, online analytical processing (OLAP) and decision support systems (DSS). For the sake of brevity, not all of the paradigms listed in Table 1 will be elaborated in this article. However, the paradigms that had the most influence on analytics will be discussed.

Romantic Historiographical Paradigm of Analytics

Other than Luhn, many other visionaries and champions led the way of what was to become today's analytics movement, including Frederick Taylor, William Fair, Earl Isaac, John Tukey, Edgar Codd, Howard Dresner, Arthur Samuel, Peter Naur, Herbert Simon, Michael Scott-Morton, Gregory Piatetsky-Shapiro, Usama Fayyad, Jeff Wu, William Cleveland, and Robert Heicht-Nielsen. Other visionaries such as Hal Varian and Thomas Davenport build on the accomplishments of these earlier champions. The Romantic paradigm views these champions as leaders who believed in their causes and the potential of humanity to improve through learning and education. Unlike the empiricist sciences, the Romantic paradigm does not seek empirical generalizations; instead, the Romantic paradigm places importance on the cautious, rigorous and contextualized interpretation of history – what historians called historicism – to uncover the unique paths that those champions took given their context. Out of the 18 visionaries and champions listed, only four, Herbert Simon, Michael Scott-Morton, Howard Dresner and Thomas Davenport, had any association with DSS or business intelligence as understood in the IS field. The rest worked within the disciplines of computer science, mathematics and statistics, and artificial intelligence. As mentioned earlier, the term "business intelligence" was coined by Luhn (1958, p. 314), a computer scientist working in IBM who defined it as:

> The communication facility serving the conduct of a business (in the broad sense) … The notion of intelligence is also defined here, in a more general sense, as "the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal."

Unlike DSS, Luhn's business intelligence processed unstructured text and would become the precursor to today's text analytics. At about the same time, William Fair, a mathematician with degrees from Cal Tech, Stanford and Berkeley, and Earl Isaac, an electrical engineer, set up a management consulting company in 1956 to refine the complex credit checking process with the help of mathematics discourse and computing power. They launched their credit scoring system in 1958, which was probably based on multivariate analyses of some kind involving payment history, loan utilization, credit history and credit mix to predict future likely behavior with regard to payments. Improvements in computer technology made it possible for their credit scoring system to be automated and by excluding the use of age, sex or race in the algorithm, it gave a sense of objectivity not found in other scoring systems like the one used by the Retail Credit Company at the time (Long, Jacques and Kepos, 2000a). Fair and Isaac would later rebrand their company, the Fair Isaac Corporation as FICO, one of the largest companies in credit scoring.

The same kind of leadership can also be found in another tradition, the operations research (OR) and management science (MS) tradition, that provided the methods and techniques applied by DSS. Many authors trace the development of analytics back to Frederick Taylor's scientific management (Gorman and Klimberg, 2014; Mortenson, Doherty and Robinson, 2015; Power, Heavin, McDermott and Daly, 2018) although they acknowledge that analytics is not the same as scientific management. Leaders in OR such as Ackoff (1989) have always hinted towards how data could contribute to knowledge and wisdom in what is currently called the hierarchy of data-information-knowledge-wisdom (DIKW) (Rowley, 2007). Much earlier, it was World War II and the invention of the radar by the British Royal Air Force by leaders such as chemist Henry Tizard (Clark, 1965) and physicist Albert Rowe (1948) that formed the context from which "operational research" as we know today emerged (McCloskey, 1987). It is not an understatement that OR helped the allies win the war and this OR tradition left two major legacies that continue to be the bedrock of analytics today: (1) data-driven analysis and decision-making, and (2) analytics is not merely mathematics and statistics, it is an empirical science. Both were taken up by visionaries such as John Tukey.

As a statistician in the early 1960s, Tukey (1962) published "The future of data analysis," to reflect on his discipline of statistics that he felt needed a new direction despite its accomplishments. He emphasized that "**data analysis is intrinsically an empirical science**" (p. 63, original emphasis) that needed to be taught in laboratories like biochemistry, and which will be significantly impacted by the development in computing technology. Like a science, data analysis also applies models to answer questions, focuses on discovery in an open-minded fashion rather than on blind confirmation, intuitively seeks explanation for specific findings, and at every stage, reviews the results before continuing in an interactive, trial-and-error, feedback-guided manner. Tukey and Wilk (1966, p. 695) defined what would today become known as analytics:

> … to seek through a body of data for interesting relationships and information and to exhibit the results in such a way as to make them recognizable to the data analyzer and recordable for posterity. Its creative task is to be productively descriptive, with as much attention as possible to previous knowledge, and thus to contribute to the mysterious process called insight.

Although a statistician himself, Tukey wanted his new data analysis to be kept separate from statistics: "Data analysis can gain much from formal statistics, but only if the connection is kept adequately loose" (Tukey and Wilk, 1966, p. 696). It was during Tukey's time that the semblance of data science, data analysis as a science, took shape, as he described analytics as doing experiments rather than as mathematical and statistical processes in experiments. Tukey (1977) published *Exploratory Data Analysis*, to summarize all of his thinking surrounding his proposed "new data analysis" field.

At about the same time, independently, Peter Naur (1974), published a book titled *Concise Survey of Computer Methods*, where in the first chapter "Data and their Applications," he wrote a section titled "A Basic Principle of Data Science." In this section he emphasized the need to focus on the transformation of the data using data processing tools available, and the difficulties that ensue in representing, converting, forming "new yet unknown data" and finding relevance to that field of interest. By the 1970s, the amount of data was already stretching the capabilities of the existing file-based network and hierarchical databases that stored the data at the time and Edgar Codd (1970), working in IBM, found a solution in the form of the relational model that allowed data to be queried, updated, and deleted in a consistent fashion. The relational database model quickly became the industry standard as new databases like IBM's DB2 and Oracle's V2 implemented it. In some ways, relational database technology actually increased the rate by which data and information grew making data even more unwieldy. Responding to these limitations, Codd et al., (1993) proposed creating a middleware functionality he called online analytical processing (OLAP) that would preprocess data from the relational database into multiple dimensions to be accessed more readily by user-interfaces. By doing so, data from both online transaction processing (OLTP) and historical decision-support could be synthesized to provide even better information. OLAP boosted the capabilities of existing DSS that already built similar models from historical data for decision making.

It was these capabilities that industry consultants like Bill Inmon (1992) and Ralph Kimball (1996) leveraged to sell the concept of the "data warehouse" and reintroduce Luhn's notion of "business intelligence" to organizations, albeit, not in the same form. Howard Dresner, who was working at DEC and moved to Gartner around 1989, is commonly credited for making the data warehouse version of business intelligence a must-have for any firm that wanted to leverage data for strategic decision-making. The definition for business intelligence that is spread on the Internet attributed to Dresner is "concepts and methods to improve business decision making by using fact-based support systems," which closely aligns with the DSS discourse. According to a practitioner magazine news story from Nylund (1999), Dresner (with Ralph Kimball) helped a Xerox Palo Alto Research Center spinoff called Metaphor Computer Systems build an early data warehousing system for Proctor & Gamble, which linked their sales information and scanner data from marketing giant A.C. Nielsen to their products and customers. This project became the prototypical data warehouse that implemented his vision of business intelligence.

Independently, within the computer science discourse, artificial intelligence and machine learning technologies were making progress and notions of supervised and unsupervised learning, (Sebestyen, 1960) were taking shape. Inspired by these machine learning capabilities and the possibility of extracting knowledge from data, Gregory Piatetsky-Shapiro (2000) organized a workshop in 1989 on how to mine knowledge from databases. He called it Knowledge Discovery in Databases (Piatetsky-Shapiro, 1991; Piatetsky-Shapiro and Frawley, 1991). That annual workshop grew into the first conference on Knowledge Discovery and Data Mining (KDD-95 https://aaai.org/Conferences/KDD/kdd95.php) in 1995 and would become the most sought-after event for the data mining community. The term "data mining" became well-known as industry found opportunities to "mine" the data (Stonebraker et al., 1993, p. 4) for "interesting" information. It was at this point that the process of analytics, which includes the KDD process, was articulated by Fayyad et al., (1996). This discourse became part of the data science movement that made its entry again as Jeff Wu (1997), the Carver Chair and Professor of Statistics at the University of Michigan, advocated statistics be renamed "data science" and statisticians be called "data scientists." John Chambers (1999) and William Cleveland (2001), both inspired by Tukey's "new data analysis," launched data science as an academic field with their work in software and their writings. All of this became mainstream when Davenport and colleagues published "Competing on Analytics" (Davenport, 2006; Davenport and Harris, 2007) and when Hal Varian (2009), Google's chief economist, reminded everyone that:

> I keep saying the sexy job in the next ten years will be statisticians. People think I'm joking, but who would've guessed that computer engineers would've been the sexy job of the 1990s? The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids. Because now we really do have essentially free and ubiquitous data. So, the complimentary scarce factor is the ability to understand that data and extract value from it.

By that time, universities were already taking the cue and Michael Rappa (2007) from MIT established the first graduate program in analytics at North Carolina State University in 2007.

While the academic community was just getting excited about the prospects for data science, industry was quietly milking the benefits of OLAP, data mining and artificial intelligence. By the time consumer credit cards were issued in the 1950s, the Retail Credit Company, which later became Equifax, had over 250 branch offices across the country (Long et al., 2000a). FICO, launched in 1956, achieved revenues of over a $1 billion by 2018, which is only 7% of the world's credit bureau market (Long, Jacques and Kepos, 2000b) that applies the same OLAP, data mining and artificial intelligence techniques. For example, Robert Hecht-Nielsen developed a neural network application for his company, HNC Software, Inc., in 1986 to detect credit card fraud. The data mining system called Falcon Fraud Manager was so successful, it helped bring down fraudulent transactions from 0.18% to 0.05% (Horan, 2014). HNC Software would expand its solution into other markets including preventing workers' compensation fraud, reviewing medical charges for property and casualty claims, processing new customers, merchandising management, bank loan processing, and data warehousing solutions such as sales forecasting and inventory management. Because of its potential, it was acquired by FICO for nearly a $1 billion in 2002. In the meantime, in the hardware and file systems area, the problems of large data sets or big data that could not be loaded into memory were being addressed by two NASA scientists, Michael Cox and David Ellsworth (1997), and highlighted in industry as the three Vs (Laney, 2001). It did not take long for companies like Yahoo!, and Google to offer solutions to these problems in the form of big data technologies like Hadoop (Cutting, 2006) and MapReduce (Dean and Ghemawat, 2004).

This Romantic historiographical study suggests that two major directions emerged within the analytics movement – data science and business analytics (BA). The BA direction led by visionaries such as Hans Peter Luhn, William Fair, Earl Isaac, Edgar Codd, Howard Dresner, Ralph Kimball, Herbert Simon, Michael Scott-Morton, Robert Heicht-Nielsen and Thomas Davenport promote more of the application of data science in the business context, while at the same time linking back to modify the data science that needs to meet business requirements. One difference for BA as Kohavi et al., (2002) suggest is that the key consumer of those insights is the business user unlike the consumers for data science, which are scientists.

The gap between the analytical knowledge of the business user and the data scientist creates a unique challenge for BA that prompts the need for a separate field of study from that of data science or data mining. The data science direction is led by visionaries such as John Tukey, Arthur Samuel, Peter Naur, Gregory Piatetsky-Shapiro, Usama Fayyad, Jeff Wu, William Cleveland and Hal Varian, although it can be argued that John Tukey and Peter Naur clearly had the concerns of the business user in mind in their vision of data science.

How distinguishable these two directions are is unclear. Saar-Tsechansky (2015) thinks there is a "data science community" in the IS field, hence the existence of "business data science" which she proposes to be positioned as a form of design science. Writing in an IS journal, Chen et al., (2012) assume that BA is just data science in business. But they do not address which areas of data science are most relevant to the IS field? Which discourse takes advantage of the strength of the intellectual discourse of the IS field? A company can hire a statistician to crunch the numbers, but can the company expect the statistician to weigh into the business deliberations in the business units those numbers originate from? In the academic context, which journals IS authors publish their analytics studies will determine the future of BA. How does one distinguish a BA article from a data science article? For journal editors, how should they recruit reviewers when evaluating different analytics papers? And what primary set of criteria should be used to evaluate these papers? Some answers are provided in the next sections that analyze the Rankean and social science paradigms of analytics historiography.

The Rankean Historiographical Paradigm of Analytics

The Rankean or scientific paradigm of historical analysis critically examines historical documents to find the causal nexus or organic whole of the subject matter being analyzed. While accepting the particularistic analysis of historicism, the Rankean paradigm also attempts, through strict objectivity and the rigorous study of multiple historical sources, to understand (verstehen) what essentially happened. Power et al., (2018) undertook a study of this kind to define BA with the help of tools such as Google Trends and Google Ngram, text analyses of 15 definitions of BA, text analyses of descriptions of Master's degrees in BA programs, and a traditional search of article databases for the key term "business analytics." The study found no single widely known definition of BA, and inferred that BA is a term used primarily for political purposes of the related stakeholder: software vendors use it to sell their software, consultants and industry use it to create context for their services, academic administration define their BA program so as to increase their reputation as being leading edge, and journals use it to attract readers. Academic textbooks use a definition that is closest to the background of their authors. For example, Power et al., (2018) cite Sharda et al.'s (2014, p. 393) definition for BA as "the application of models directly to business data. Business analytics involve using DSS tools, especially models, in assisting decision makers." The deference given to DSS reflects the authors' background.

A Google Trends analysis (2019) for the search of the three terms, "data science," "business analytics" and "big data (Figure 1) shows that the terms big data and data science began to separate from BA around early 2012. By 2018, both data science and big data receive roughly the same number of searches. This trend corroborates the earlier analysis of the Romantic paradigm suggesting that the analytics movement has bifurcated into two directions: data science and BA. Aasheim et al., (2015) studied the differences between data science and data analytics curriculum at the undergraduate level. "Data analytics" undergraduate programs were selected from business schools, so they can be assumed to be BA programs, while the data science programs were in computer science, mathematics and statistics or from interdisciplinary academic units. The study finds that both BA and data science cover data mining and data preparation, and most cover modeling and analytics techniques as well as visualization. However, data science programs include additional coverage of mathematics and statistics, as well as additional coverage of programming, and lack courses on decision-making skills, data governance, data capture, storage, security, communication skills and the use of case studies. Conversely, BA programs focus less on mathematics, statistics and programming. Additionally, BA programs place greater emphasis on the evaluation of tools and techniques, while data science programs tend to emphasize the algorithms underlying the techniques and their implementation. What is notably absent in both BA and data science are courses on the ethical or legal issues of big data.

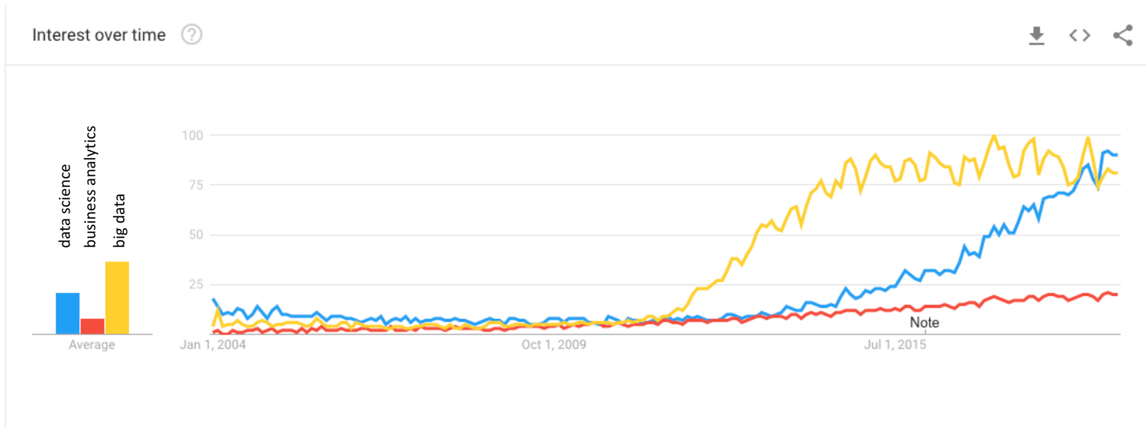

Figure 1:Google Trends Analysis of Search Terms

The Social Science Historiographical Paradigm of Analytics

The social science paradigm of historical analysis studies the theoretical questions related to the meanings of human actions in its historical setting. It searches for abstract analytical concepts that explain the historical forces unique to the character of the society and culture, uncovering the dynamics of social influences, technological changes, and other human science forces at work. It explains, for example, why despite the similarities between OR models and techniques and those in analytics, OR remains limited in its influence. As early as 1950s, concerns were raised about how OR lacked communication with business and industry, a consistent description of what the field was about, and lacked a repertory of professional standards (Rinehart, 1954). A policy change in the AACSB accreditation standards in 1991 left out OR as a requirement for business schools, and as a result OR never recovered. Many universities dismantled their OR departments from their colleges of business (Grossman, 2003), and as a branch of study, OR found itself absorbed into other fields such as management, information systems, transportation and engineering (Corbett and van Wassenhoff, 1993; Fildes and Ranyard, 1997). Currently, OR maintains its relevance within IS as can be seen by the high status enjoyed by the *Information Systems Research*, published by INFORMS. With the advent of analytics and big data, OR is hopeful of making a comeback (Hazen, Skipper, Boone and Hill, 2018), however what direction that comeback needs to take is unclear and remains a work-in-progress (Liberatore and Luo, 2010; Mortenson et al., 2015).

The computer science discourse appears to be sociologically well-positioned for analytics, followed closely by the mathematics and statistics discourse. Using Abbott's (2001, p. 140) concept of the "axis of cohesion," which essentially represents the discipline's central principles, we find that the computer science discourse supports many axes of cohesion that are well-aligned for analytics and data science. First, the axis of data and its transformation forms the bases for computer science especially since the time of Peter Naur's "data science" and focus on data and its programming. Within the computer science discourse, several axes struggled for dominance. For example, Luhn's (1958) "business intelligence" was overtaken by Codd's (1970) relational database technology and was recategorized as part of information retrieval. Later, business intelligence would be associated with data warehousing and DSS. Second, the axis of artificial intelligence and machine learning created the KDD community that rallied the computer science community to support data science. The 2018 KDD conference received more than 3,377 attendees, 1,480 submissions, with 486 submissions in the applied data science track alone. Piatetsky-Shapiro also started the KDNuggets newsletter in 1993 which grew into a website (www.kdnuggets.com) that became the main resource for the data mining and data science community in terms software, jobs, academic positions, calls for papers, courses, datasets, education, meetings, publications, webcasts and data mining competitions (SIGKDD, 1997). Third, computer science and artificial intelligence have an immediate "practical clientele" (Abbott, 2001, p. 140) for its data science movement in the form of numerous successful industry applications as shown by many companies such as HNC that offered numerous industry solutions. Academic and industry stakeholders also constructed several standards and processes designed to discipline data science activities including Fayyad et al.'s (1996) KDD process, Cross Industry Standard Process for Data Mining or CRISP-DM (Chapman et al., 2000), and SAS's (2017) Sample, Explore, Modify, Model and Assess (SEMMA).

The mathematics and statistics discourse also exerted considerable influence in the development of data science, but its influence was closely intertwined with the progress of computing technology. Tukey realized that what he envisioned for his new data analysis was too laborious and time-consuming for the technology at his time and saw a future where the formal theories of statistics and its techniques would be supported by accelerating developments in computers and display devices (Tukey and Wilk, 1966). This vision was carried by others inspired by him including Rick Becker and John Chambers, who created the S language in 1976, the precursor to today's R programming language (Becker, 1994). The cultural differences between statistics and Tukey's vision of the new data analysis is clearly described as the emergence of what John Chambers (1999, p. 83) quoted of Tukey as: "the peaceful collision of statistics and computing" and its predictive culture (Breiman, 2001; Donoho, 2017).

The main axes of cohesion the IS discourse offers for analytics are related to DSS and BI and this foci of BA can be seen in its textbooks (Sharda et al., 2014; Sharda, Delen and Turban, 2018). Arguably, such foci place limitations on BA experts, if they are to interact with data scientists or to apply data science in the business context (Chen et al., 2012). Based on the analyses of both the Romantic paradigm and the Rankean paradigm for BA, the DSS and BI axis is but one of at least four other axes of cohesion for the analytics discipline: Luhn's unstructured text analysis, the computer science axis, the artificial intelligence and machine learning axis, and the mathematics and statistics axis. Although elements of these axes are found in DSS and BI, they are deemphasized, and as a result, BA itself may lack substance and the necessary intellectual foundations. Leaving BA in this state of reliance on the technologies of DSS and BI that were developed in the 1980s and 1990s may not bode well for its future. A clear set of foundations need to be defined for BA so that its future direction need not rely completely on data science or its related disciplines of computer science, mathematics and statistics, and it can start building its own traditions and theories.

Holsapple et al., (2014) suggested such a unified foundation for BA. However, their proposed foundation that includes six definitional perspectives of BA as a movement, collection of practices and technologies, a transformational process, a set of capabilities, activity type set or decisional paradigm does not distinguish BA from data science. The same proposed foundation can be applied to data science without any modification. This makes it difficult to build a unique BA tradition based on this framework. Also, critical elements such as the creation of theory and conceptual development necessary to provide interesting answers to questions that we do not yet know (Dhar, 2013), or build a society capable of "avoiding market crashes, ethnic and religious violence, political stalemates, widespread corruption, and dangerous concentrations of power" (Pentland, 2014, p. 19) are left out from or not foregrounded in the framework.

Proposed Framework for Business Analytics: Developing the Unicorn

In 2015, the CEO and co-founder of the Silicon Valley firm Answerlab, Amy Buckner Chowdhry, described the analytics graduates her company wanted to hire as "unicorns," those mythical creatures that do not exist (Noon and Gilbert, 2015). Specifically, Chowdhry was looking for graduates who are able to manage projects and interact with clients, and in addition to technical skills, they needed to have "soft" teamwork, leadership, writing and communication skills. On top of those skills, they needed to have business expertise, subject matter expertise in data analysis, intellectual curiosity and creativity. Given the limited number of years any student has to accumulate these skills, it is little wonder they are unicorns to most businesses. The difficulty of locating the right kind of analytics talent remains a challenge for companies, and developing that talent, an equally difficult challenge for higher education.

The kind of BA talent that would be useful for a company like Answerlab was suggested by McKinsey and Company in their *Age of Analytics* (Henke et al., 2016) report. The BA expert, which they call the "business translator" (p. 38) is distinct from the data scientist. The high-end data scientist, whose job is to research and advance the most cutting-edge algorithms are limited in number and cost typically between $5 million to $10 million for businesses to acquire. McKinsey and Company estimates a shortfall of 250,000 data scientists. On the other hand, they estimate a demand for between two million to four million business translators. We will call this business translator the data analyst. They have both technical and domain- or function-specific business knowledge, enabling them to turn analytical insights into business value. One of their important skills include visualization which is considered vital to the challenge of discovering value. Kohavi et al., (2002) suggest other skills including the ability to deliver insights rapidly due to the shorter cycle time for BA, providing a sufficient depth of the analysis, managing expectations when aligning with business goals, being able to meet the different needs related to data collection and data transformation, delivering insights in different formats and the need to integrate multiple sources of data (i.e. complex ETL processes).

Based on the paradigmatic historiographical analysis above, we propose the following axes of cohesion that combine the strengths of the data scientist with the business acumen required for BA to support the development of a new BA discipline, profession and the training of the future data analyst. By doing so we simultaneously propose specific focus areas in analytics that best fit the intellectual strengths of the IS field.

Axis 1: Naur's Computing as a Human Science

Naur's (1992) philosophy of computing as a human activity establishes certain foundations for BA. His emphasis on data and its role in transforming reality is consistent with the foundations of other major IS topics such as systems analysis and design and conceptual modeling. This axis distinguishes BA from data science, which also targets human concerns, but does not place as much emphasis on the human sciences. An example that is not foregrounded in data science is the impact and implications of big data analytics for society. The increasing use of big data is expected to create many issues that need to be addressed, such as legal and ethical dilemmas stemming from the intrusion of privacy and other data-related problems. The case of Equifax (formerly the Retail Credit Company) losing the data of half of the population of the United States (Berghel, 2017) highlights the need to find more innovative solutions that can only come from multidisciplinary efforts within BA. The people-related issues that arise from the gap between the expertise of the data scientists and that of the business manager require new forms of governance and implementation strategies that can be developed within BA.

Rather than limiting BA to the traditional positivist science, BA as a human science has the potential to transcend positivism and capture genuine human experience. As Dilthey (1883, p. 66) argues, the human sciences are not alternatives to the natural sciences, they include and complement them since "the human sciences do encompass natural facts and are based on knowledge of nature." An emphatic understanding of human behavior (*verstehen*) is necessary to capture the "knowledge of the forces that rule society, of the causes that have produced its upheavals, and of society's resources for promoting healthy progress [that] has become of vital concern to our civilization" (p. 56). This is where the theory building potential in BA comes into play. Naur views computing as a process of theory building through computing. Analogously, the job for the data analyst in BA will be to envision the theory of how businesses work (or in the case of Equifax, how it failed), especially in their different vertical markets, and to create specialized BA applications in those vertical markets. Using Ryle's (1949, p. 286) notion of having a theory to mean to be "prepared to state it or otherwise apply it … to give a good answer … to deliver … an intelligible statement of the conclusions of the theory, the problems which they solve and … the reasons for accepting" it, the data analyst becomes more than just an instrumental cog in the machinery of the organization, but with the help of the underlying theories that are uncovered, becomes a responsible developer and manager of the business. This axis addresses certain misconceptions about big data ushering the "end of theory" (Anderson, 2008). There is much that can be developed theoretically from a deeper understanding of how data will change the way we think about business, health, politics, education and all aspects of life in the years to come (Mayer-Schönberger and Cukier, 2013). Scientific research need not rely exclusively on causal mechanisms because big data provides an alternative approach to science (Shmueli and Koppius, 2011; Dhar, 2013; Agarwal and Dhar, 2014). Analyzing *all* of the data in its messiness have been found to be quite effective in delivering useful insights. Processing this data does require models, which are essentially components of theory, so theory cannot be divorced from analytics.

Axis 2: Tukey's New Data Analysis

BA can learn much from Tukey's efforts to transform statistics into a more relevant science. In the preface to *Exploratory Data Analysis*, Tukey (1977, p. v) emphasizes a principle that does not receive much attention in data science or in any of the existing analytics-proposed life cycle – "**It is important to understand what you CAN DO before you learn to measure how WELL you seem to have**

DONE it (Original emphases). For example, the KDD process (Fayyad et al., 1996) begins with data selection and proceeds on to data cleaning, preprocessing, and transformation before going on to applying the chosen data mining technique. A tutorial on big data analytics only briefly mentions descriptive analytics (Watson, 2014). Within the category of deriving knowledge in the big data information value chain (Abbasi et al., 2016), data exploration is mentioned as part of the Cross-Industry Standard Process for Data Mining (CRISP-DM) process model (Chapman et al., 2000). However, the connection between learning first what can be done and how the direction of the analysis would proceed, as Tukey (1977) emphasized, is largely ignored. For BA, this axis implies that exploratory data analysis (EDA) is not just one of many stages within the BA life cycle, it is the phase that determines the direction of the BA life cycle itself. The foci of exploratory data analysis are to (1) simplify, i.e., to help make any body of data more easily understood and digestible by analysts and the targeted audience of the analysis, and (2) uncover insights that lie below the surface of the data. Both these goals of simplifying and uncovering insights are accomplished in part by visualization, the craft of visual evidence, visual reasoning, and visual understanding (Tufte, 1983). In BA, visualization takes center stage, and by avoiding statistical jargon that makes no sense to the business user, it delivers the same valuable information in visual format. For example, interactions, statistical significance and heteroscedasticity can all be represented in different ways, but more persuasively in visual format, while it "retains the information in the data" (Deming (1985) as cited by Cleveland (1993)). This axis of BA addresses Kohavi et al.'s (2002) concern of the gap that exists between the data scientist and the business user.

Axis 3: Luhn's Unstructured Data Analytics

As the above analyses show, the earliest analytics efforts were about unstructured data (Luhn, 1958). However, the DSS and BI discourse historically process the same structured data that IS departments produce daily. The same can be said about data science. Even the "traditional" division of analytics into descriptive, predictive and prescriptive analytics (Lustig, Dietrich, Johnson and Dziekan, 2010) specified focusing on structured data while using unstructured data to support the analysis of structured data. The irony in all this is that the vast majority of big data is unstructured (Grimes, 2008). This reality of the overwhelming dominance of unstructured data is consistent with Dilthey's (1883, p. 66) argument that the human sciences encompass the natural sciences, which traditionally rely on structured data. This axis promises a large potential for BA to establish itself as an influential force in the growth of analytics knowledge. The significance of unstructured data is already bearing fruit, as can be seen in the case of question and answering technologies like IBM's Watson (High, 2012), sentiment analysis from social media sources and other data sources (Pang and Lee, 2006; Liu, 2012), speech and audio analytics, video and surveillance analytics, and social media and social network analytics, to mention a few.

Axis 4: The Management of Analytics and Its Deployment

An important aspect of analytics that is left out in the data science movement is the management of the data science effort and its deployment. Both these areas have not been studied to any significant extent. In an article titled "The science of managing data science," a vice-president of engineering at a research startup for data mining and machine learning describes the difficulty she had in explaining to the company's executives what their data scientists were working on daily, or what projects should take priority, and how to deploy their research to quickly benefit customers (Matsudaira, 2015). The CRISP-DM process defines the deployment stage as "applying 'live' models within an organization's decision making processes … [which] can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise" (Chapman et al., 2000, p. 11). It consists of planning the strategy for the deployment as well as its monitoring and maintenance, producing the final report and reviewing what went wrong, what was done well and what needs to be improved. Both the management of the data science effort and its deployment is where data science interfaces with the social component of the socio-technical enterprise and where expertise concerning both technology and human needs intersect. This is an area where BA, with the help of the IS discourse, offers a long and respected tradition.

The development of the deployment stage of CRISP-DM is also in its infancy. The progenitors of the descriptive-predictive-prescriptive model of analytics (Lustig et al., 2010) suggest that deployment is prescriptive analytics and define it in terms of optimization (or how can the best outcome be achieved?), that is, choosing the best response or action given the limited resources of the organization. Their paradigm for prescriptive analytics, however, is limited to OR-type decision-making or problem-solving with the help of computers running mathematical algorithms that would propose the best course of action. This approach which views the organization as a machine (Morgan, 1986) may not be the best option for successful deployment of the results from the analysis. ISD management knowledge (Hassan and Mathiassen, 2017) that describes the coordination, organization, and practice of ISD among participating groups provides an avenue to refine the CRISP-DM process.

Future Research Directions

If we can assume that our researchers are not expected to invent the next Hadoop or MapReduce, or even to write the next classification or clustering algorithm, what direction should future research in BA take? The framework proposed above provides such a direction that goes beyond the traditional DSS- or BI-type solutions that have become the staple for organizations in the past. The following subsections describe several future research directions that can help accomplish what Pentland (2014) envisioned to be the role of the IS field in analytics, which is to address intractable societal problems by solving macro-level problems using micro-level big data.

The Human Science of Analytics

The proposed framework for analytics suggests a growing gap of skills in analytics that is very different from what the typical "data scientist" possesses. This gap is representative of an area of study surrounding societal and human concerns stemming from the impact and implications of big data analytics that are under-researched. Consequently, a lot of work is being done in analytics to build and enhance the algorithms that segment, cluster and predict customer behavior, purchases, online preferences and even political leanings, but very little research is done to study the impact of the deployment of those algorithms and their implications on the privacy and security of the citizens that produce and own the data. As the case of Cambridge Analytica (Cadwalladr and Graham-Harrison, 2018) shows, data and profiles of nearly 100 million Facebook users were misused to benefit political agendas while at the same time enrich private coffers. Research that focuses on the human sciences which encompass both societal concerns and the algorithmic and technological concerns addresses this neglected area and takes advantage of the strengths of the IS field.

Business Analytics Foundations and Theories

Closely related to the proposed research area in the human sciences of analytics is the development of the theoretical foundations of BA itself. Research and practice in analytics and data science have always relied on the existing theoretical foundations first laid down centuries ago in the disciplines of mathematics, statistics, and more recently in the disciplines of operations research, computer science and artificial intelligence. BA research in the IS field follows a similar path to the point that IS researchers become ambivalent to theory in analytics (Dhar, 2013). This ambivalence leads to a situation where BA as a field will be lacking foundations that are not only necessary for distinguishing BA as its own intellectual field of study, but is also required for building theory required for a deeper understanding of how data will change the way we think about business, health, politics, education and all aspects of life in the years to come.

The development of insights from exploratory data analysis (EDA)

Tukey's (1977) recommendations to focus on the insights from the exploration of data have only recently received the attention of the analytics community, especially after the popularity of Tufte's works (2001, 2006) and tools like Tableau. Research on the role of EDA to guide the direction of the analysis is sorely lacking and emerges as a gap between the data scientist and the business user (Kohavi et al., 2002; Henke et al., 2016). The IS field that straddles human concerns and technological capabilities is perfectly positioned to address this gap and discover how insights can be better extracted from big data analytics. Accomplishing this goal fulfills Tukey's vision of being "productively descriptive, with as much attention as possible to previous knowledge, and thus to contribute to the mysterious process called insight" (Tukey and Wilk, 1966, p. 695).

Unstructured Data as Primary Thrust of BA Research

Currently, most of the work in unstructured data like text mining has been exploratory (Kobayashi et al., 2018). Following from the potential for BA to build theory, an initial challenge for BA would be to test existing socio-technical theories as well as build new ones. Such studies are already taking place especially using blogs and other social media. Yarkoni (2010) analyzed over 100,000 words per blog from 576 bloggers and found a relationship between personality and language use. Using just word counts and previous psychometric research, Coppersmith et al., (2015) were able to link Twitter postings with the users' mental health condition. With more research in this area of unstructured data, especially in combining different data sources of varying formats, handling "messy, unstructured data" (Davenport and Patil, 2012, p. 70), BA can help resolve society's most intractable problems.

Research in Deploying Analytics

The tradition of IS development (ISD) overlaps with project management and includes people management, method management, performance management, project organizing, quality assurance, risk management, stakeholder management and supplier management (Hassan and Mathiassen, 2017), all of which are applicable to the deployment of the results from data science and analytics. This ISD tradition provides a framework for BA to manage analytics projects and deploy them successfully within the organization. By placing greater emphasis on meeting the expectations and requirements of the customers of the analytics projects, research in this area will enhance existing CRISP-DM and SEMMA deployment processes.

Conclusion

By undertaking a paradigmatic historiographical analysis of the analytics movement, this essay answers the question "Where are we headed in business analytics?" and uncovers several of its intellectual, technical, social and political influencers. Using the detailed analysis of the Romantic, Rankean and social science paradigms, the analytics movement is shown to have bifurcated into two directions – the data science direction and the BA direction, both still in flux and each building its traditions. The data science tradition appears to be consolidating its foundations, with its well-aligned axes of cohesion, its academic and industry accomplishments, and the reception it is receiving in terms of financial support and numbers of enrolled majors. The BA tradition appears to be lagging behind, with its foundations relying on data science, and arguably held back by the legacy of its purported origins in DSS, data warehousing and BI of the 1980s and 1990s. Although some modest efforts have been made, the societal needs for analytics have outpaced developments within the IS field, and a new set of axes of cohesion is required to enable BA to be an influential force. Using the insights from the champions and visionaries who inspired the data science movement, this essay proposes four axes of cohesion that take advantage of the depths and strengths of data science, but at the same time distinguishes BA such that it will be possible for BA to grow its own unique foundations, theories and applications that will be able resolve many of society's most intractable problems.

References

Aasheim, C. L., S. William, P. Rutner and A. Gardiner. (2015). "Data analytics vs. data science: A study of similarities and differences in undergraduate programs based on course descriptions." *Journal of Information System Education*, *26*(2), 103–115.

Abbasi, A., S. Sarker and R. Chiang. (2016). "Big data research in information systems: Toward an inclusive research agenda." *Journal of the Association for Information Systems*, *17*(2), i–xxxii.

Abbott, A. D. (2001). *Chaos of Disciplines*. Chicago, IL: University of Chicago Press.

Ackoff, R. L. (1989). "From data to wisdom." *Journal of Applied Systems Analysis1*, *16*(1), 3–9.

Agarwal, R. and V. Dhar. (2014). "Big data, data science, and analytics: The opportunity and challenge for IS research." *Information Systems Research*, *25*(3), 443–448.

Anderson, C. (2008). "The end of theory." *Wired*, *16*(7), 71.

Becker, R. A. (1994). "A brief history of S." In: P. Dirschedl & R. Ostermann (Eds.), *Computational Statistics* (pp. 81–110). Berlin: Springer-Verlag.

Belkin, D. (2018, November 2). "U.S. News: At Berkeley, big data on campus -- enrollment in science of

digital information soars as Silicon Valley can't hire fast enough." *Wall Street Journal*, p. A.3.
Berghel, H. (2017). "Equifax and the latest round of identity theft roulette." *IEEE Computer*, *50*(12), 72–76.
Breiman, L. (2001). "Statistical modeling: The two cultures." *Statistical Science*, *16*(3), 199–231.
Bryant, A., A. Black, F. Land and J. Porra. (2013). "Information Systems history: What is history? What is IS history? What IS history? ... and why even bother with history?" *Journal of Information Technology*, *28*(1), 1–17.
Cadwalladr, C. and E. Graham-Harrison. (2018). "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach." *The Guardian*. London, UK.
Chambers, J. (1999). "Statistical computing and graphics computing with data : Concepts and challenges." *The American Statistician*, *53*(1), 73–84.
Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth. (2000). "CRISP-DM 1.0." SPSS. Retrieved from ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf (visited on April 5, 2019)
Chen, H., R. H. L. Chiang and V. C. Storey. (2012). "Business intelligence and analytics." *MIS Quarterly*, *36*(4), 1165–1188.
Clark, R. W. (1965). *Tizard*. Cambridge, MA: MIT Press.
Cleveland, W. S. (1993). *Visualizing Data*. Summit, NJ: Hobart Press.
Cleveland, W. S. (2001). "Data science: An action plan for expanding the technical areas of the field of statistics." *International Statistical Review*, *69*(1), 21–26.
Codd, E. F. (1970). "A relational model of data for large shared data banks." *Communications of the ACM*, *13*(6), 377–387.
Codd, E. F., S. B. Codd and C. T. Salley. (1993). "Providing OLAP (Online Analytical Processing) to User-Analysts." Retrieved from www.uniriotec.br/~tanaka/SAIN/providing_olap_to_user_analysts.pdf (visited on March 9, 2019)
Coppersmith, G., M. Dredze, C. Harman and K. Hollingshead. (2015). "From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses." In: *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. Denver, CO, June 5.
Corbett, C. and L. van Wassenhoff. (1993). "The natural drift: what happened to operations research?" *Operations Research*, *41*(4), 625–640.
Cox, M. and D. Ellsworth. (1997). "Managing big data for scientific visualization." In: *ACM SIGGRAPH '97 Conference* (pp. 27–34). Los Angeles, CA, Aug 3-8.
Cutting, D. (2006). "Scalable computing with Hadoop." Apache Foundation. Retrieved from https://docs.huihoo.com/apache/hadoop/yahoo-sds.pdf (visited on April 25, 2019)
Davenport, T. H. (2006). "Competing on analytics." *Harvard Business Review*, *84*(1), 98–107.
Davenport, T. H. (2018). "From analytics to artificial intelligence." *Journal of Business Analytics*.
Davenport, T. H. (2019). "Analyzing the Future." *BizEd Magazine*, *18*(1), 21–24.
Davenport, T. H., D. Cohen and A. Jacobson. (2005). *Competing on Analytics*. Working Knowledge Research Center, Babson College: Babson Park, MA.
Davenport, T. H. and J. G. Harris. (2007). *Competing on Analytics: The New Science of Winning*. Boston: Harvard Business School Press.
Davenport, T. H., J. G. Harris and R. Morison. (2010). *Analytics at Work: Smarter Decisions, Better Results*. Boston, MA: Harvard Business School Publishing.
Davenport, T. H. and D. J. Patil. (2012). "Data scientist: the sexiest job of the 21st century." *Harvard Business Review*, *90*(10), 70–76.
Dean, J. and S. Ghemawat. (2004). "MapReduce: Simplified data processing on large clusters." In: *OSDI '04: 6th Symposium on Operating Systems Design and Implementation* (Vol. 51, pp. 137–149). USENIX Association.
Deming, W. E. (1985). "Letter to William S. Cleveland."
Dhar, V. (2013). "Data science and prediction." *Communications of the ACM*, *56*(12), 64–73.
Dilthey, W. (1883). *Introduction to the Human Sciences*. Princeton, N.J.: Princeton University Publishers.
Donoho, D. (2017). "50 Years of data science." *Journal of Computational and Graphical Statistics*, *26*(4), 745–766.
Fayyad, U., G. Piatetsky-Shapiro and P. Smyth. (1996). "From data mining to knowledge discovery in databases." *AI Magazine*, *17*(3), 37–54.
Fildes, R. and J. C. Ranyard. (1997). "Success and survival of the operational research groups: a review."

Journal of the Operational Research Society, *48*(4), 336–360.

Google Trends. (2019). "Google trends analysis: Data science, business analytics, big data." Google, Inc.: Mountain View, CA. Retrieved from https://trends.google.com/trends/explore?date=all&q=data science,business analytics,big data (visited on April 27, 2019)

Gorman, M. F. and R. K. Klimberg. (2014). "Benchmarking academic programs in business analytics." *Interfaces*, *44*(3), 329–341.

Grimes, S. (2008). "Unstructured data and the eighty percent rule." Retrieved from http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/ (visited on April 10, 2019)

Grossman, T. A. (2003). "Up to standard." *OR/MS Today*, *30*(4), 18–21.

Harcourt, R. (2018). "Celebrating Berkeley's first data science majors." Berkeley Division of Data Sciences. Retrieved from https://data.berkeley.edu/news/celebrating-berkeleys-first-data-science-majors (visited on April 17, 2019)

Hassan, N. R. (2018). "Taking IS history seriously." In: R. D. Galliers & M.-K. Stein (Eds.), *The Routledge Companion to Management Information Systems* (pp. 5–29). London: Routledge.

Hassan, N. R. and L. Mathiassen. (2017). "Distilling a body of knowledge for information systems development." *Information Systems Journal*, *28*(1), 175–226.

Hazen, B. T., J. B. Skipper, C. A. Boone and R. R. Hill. (2018). "Back in business: operations research in support of big data analytics for operations and supply chain management." *Annals of Operations Research*, *270*(1–2), 201–211.

Henke, N., J. Bughin, M. Chui, J. Manyika, T. Saleh, B. Wiseman and G. Sethupathy. (2016). "The age of analytics: Competing in a data-driven world." McKinsey Global Institute. Retrieved from https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world (visited on April 17, 2019)

Hester, W. P. (2019). "UVA plans new school of data science; $120 million gift is largest in university history." UVAToday, University of Virginia. Retrieved from https://news.virginia.edu/content/uva-plans-new-school-data-science-120-million-gift-largest-university-history (visited on April 17, 2019)

High, R. (2012). *The era of cognitive systems: An inside look at IBM Watson and how it works*. *IBM Redguides for Business Leaders*. Armonk, NY: IBM Corporation.

Holsapple, C., A. Lee-Post and R. Pakath. (2014). "A unified foundation for business analytics." *Decision Support Systems*, *64*, 130–141.

Horan, T. J. (2014). "Evolution of fraud analytics - An inside story." Retrieved from https://www.kdnuggets.com/2014/03/evolution-fraud-analytics-inside-story.html (visited on March 11, 2019)

Inmon, W. H. (1992). *Building the Data Warehouse* (1st ed.). New York: Wiley & Sons.

Kimball, R. (1996). *The Data Warehouse Toolkit : Practical Techniques for Building Dimensional Data Warehouses*. New York: John Wiley & Sons, 1996.

Kobayashi, V. B., S. T. Mol, H. A. Berkers, G. Kismihók and D. N. Den Hartog. (2018). "Text mining in organizational research." *Organizational Research Methods*, *21*(3), 733–765.

Kohavi, R., N. J. Rothleder and E. Simoudis. (2002). "Emerging trends in business analytics." *Communications of the ACM*, *45*(8), 45–48.

Land, F. (2010). "The use of history in IS research: an opportunity missed?" *Journal of Information Technology*, *25*(4), 385–394.

Land, F. (2015). "Early history of the information systems discipline in the UK: An account based on living through the period." *Communications of the Association for Information Systems*, *36*(1 (Art 26)), 563–575.

Laney, D. (2001). "3-D Data Management: Controlling Data Volume, Velocity and Variety." META Group: Stamford, CT. Retrieved from https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (visited on March 11, 2019)

Liberatore, M. J. and W. Luo. (2010). "The analytics movement: Implications for operations research." *Interfaces*, *40*(4), 313–324.

Liu, B. (2012). "Sentiment analysis and opinion mining." *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Long, S., D. Jacques and P. Kepos (Eds.). (2000a). "Equifax." In: *International Directory of Company Histories* (pp. 117–121). Farmington Hills, MI: St. James Press.

Long, S., D. Jacques and P. Kepos (Eds.). (2000b). "Fair Isaac Corporation." In: *International Directory of Company Histories* (pp. 168–171). Farmington Hills, MI: St. James Press.

Luhn, H. P. (1958). "A business intelligence system." *IBM Journal of Research and Development*, *2*(4), 314–319.

Lustig, I., B. Dietrich, C. Johnson and C. Dziekan. (2010). "The analytics journey." Retrieved from http://analytics-magazine.org/the-analytics-journey/ (visited on April 10, 2019)

Matsudaira, K. (2015). "The science of managing data science." *Communications of the ACM*, *58*(6), 44–47.

Mayer-Schönberger, V. and K. Cukier. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.

McCloskey, J. F. (1987). "The beginnings of operations research: 1934-1941." *Operations Research*, *35*(1), 143–152.

Michael Rappa. (2007). "Institute for Advanced Analytics." North Carolina State University: Raleigh, NC. Retrieved from https://analytics.ncsu.edu/?page_id=7 (visited on April 25, 2019)

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

Morgan, G. (1986). *Images of Organization*. Beverly Hills : Sage Publications.

Mortenson, M. J., N. F. Doherty and S. Robinson. (2015). "Operational research from Taylorism to Terabytes: A research agenda for the analytics age." *European Journal of Operational Research*, *241*(3), 583–595.

Naur, P. (1974). *Concise Survey of Computing Methods*. Lund, Sweden: Studentlitteratur AB.

Naur, P. (1992). *Computing: A Human Activity*. New York: ACM Press/Addison-Wesley.

Noon, C. and K. Gilbert. (2015). "Academic analytics program." *OR/MS Today*, *42*(1), 34–35.

Nylund, A. (1999). "Tracing the BI family tree." *Knowledge Management*, *7*(July), 2.

Pang, B. and L. Lee. (2006). "Opinion mining and sentiment analysis." *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135.

Pentland, A. (2014). *Social Physics: How Good Ideas Spread -- the Lessons from a New Science*. New York: Penguin Press.

Phelps, A. L. and K. A. Szabat. (2017). "The current landscape of teaching analytics to business students at institutions of higher education: Who is teaching what?" *American Statistician*, *71*(2), 155–161.

Piatetsky-Shapiro, G. (1991). "Knowledge discovery in real databases: A report on the IJCAI-89 Workshop." *AI Magazine*, *11*(5), 68–70.

Piatetsky-Shapiro, G. and W. J. Frawley. (1991). *Knowledge Discovery in Databases*. Menlo Park, CA: AAAI Press.

Piatetsky-Shapiro, G., Q. U. Sydney, M. S. Bu, L. Kershberg, R. Quinlan and P. Langley. (2000). "Knowledge Discovery in Databases : 10 years after Applications and concluded with a summary panel discussion by." *SIGKDD Explorations Newsletter*, *1*(2), 59–61.

Power, D. J., C. Heavin, J. McDermott and M. Daly. (2018). "Defining business analytics: an empirical approach." *Journal of Business Analytics*, *1*(1), 40–53.

Rinehart, R. F. (1954). "Theats to the growth of operations research in business and industry." *Journal of the Operational Research Society of America*, *2*(3), 229–233.

Rowe, A. P. (1948). *One Story of Radar*. Cambridge, UK: Cambridge University Press.

Rowley, J. (2007). "The wisdom hierarchy: Representations of the DIKW hierarchy." *Journal of Information Science*, *33*(2), 163–180.

Ryle, G. (1949). *The Concept of Mind*. London, UK: Hutcheson.

Saar-Tsechansky. (2015). "The business of business data science in IS journals." *MIS Quarterly*, *39*(4), 3–6.

SAS. (2017). "SAS Enterprise Miner." SAS Institute: Cary, NC. Retrieved from http://documentation.sas.com/ (visited on April 19, 2019)

Sebestyen, G. S. (1960). *Classification decisions in pattern recognition*. Massachussetts Institute of Technology (MIT).

Sharda, R., D. Delen and E. Turban. (2014). *Business Intelligence: A Managerial Perspective on Analytics* (3rd ed.). New York: Prentice Hall.

Sharda, R., D. Delen and E. Turban. (2018). *Business Intelligence, Analytics and Data Science*. New York: Pearson.

Sharma, R., S. Mithas and A. Kankanhalli. (2014). "Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations." *European Journal of Information Systems*, *23*(4), 433–441.

Shmueli, G. and O. Koppius. (2011). "Predictive analytics in information systems research." *MIS Quarterly*, *35*(3), 553–572.

SIGKDD. (1997). "KDD Cup 1997: Direct marketing for lift curve optimization." Retrieved from https://www.kdd.org/kdd-cup/view/kdd-cup-1997/ (visited on March 11, 2019)

Simmons, J. R. M. (1962). *LEO and the Managers*. London, UK: Macdonald.

Stonebraker, M., R. Agrawal, U. Dayal, E. J. Neuhold and A. Reuter. (1993). "DBMS research at a crossroads: The Vienna update." In: *Proceedings of the 19th International Conference on Very Large Data Bases, Aug 24-27* (pp. 688–692). Dublin.

Swanstrom, R. (2019). "Awesome data science colleges." GitHub. Retrieved from https://github.com/ryanswanstrom/awesome-datascience-colleges (visited on April 17, 2019)

Tufte, E. R. (1983). *Visual Display of Quantitative Information* (1st ed.). Cheshire, CT: Graphics Press, LLC.

Tufte, E. R. (2001). *Visual Display of Quantitative Information* (2nd ed.). Cheshire, CT: Graphics Press, LLC.

Tufte, E. R. (2006). *Beautiful Evidence*. Cheshire, CT: Graphics Press, LLC.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley Pub. Co.

Tukey, J. W. and M. B. Wilk. (1966). "Data analysis and statistics: An expository overview." In: *AFIPS '66 (Fall), Nov 7-10* (pp. 695–709). San Francisco: ACM Press.

Varian, H. (2009). "Hal Varian on how the Web challenges managers." McKinsey & Company. Retrieved from https://www.mckinsey.com/industries/high-tech/our-insights/hal-varian-on-how-the-web-challenges-managers (visited on April 17, 2019)

Watson, H. J. (2011). "Business analytics insight: Hype or here to stay?" *Business Intelligence Journal*, *16*(1), 4–8.

Watson, H. J. (2014). "Tutorial: Big data analytics: Concepts, technologies, and applications." *Communications of the Association for Information Systems*, *34*(1), 1247–1268.

Wixom, B., T. Ariyachandra, D. Douglas, M. Goul, B. Gupta, L. Iyer, … O. Turetken. (2014). "The current state of Business Intelligence in academia: The Arrival of Big Data." *Communications of the Association for Information Systems*, *29*(Art 1), 1–13.

Wu, J. C. F. (1997). "Statistics = Data Science?" Atlanta, GA. Retrieved from http://www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf (visited on January 3, 2019)

Yarkoni, T. (2010). "Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers." *Journal of Research in Personality*, *44*(3), 363–373.