Senior Leadership in the Age of Data Analytics: Exploring Organizational Expectations

Although data analytics leadership by senior managers is considered crucial for ensuring the successful implementation of data-driven initiatives, there has been some ambiguity regarding what such leadership actually entails. In this paper, we explore: What are the expected responsibilities and positions of data analytics leaders? To answer this question, we collect job ads for senior managers with titles pertaining to “data analytics”, “digital”, and “information technology”. To analyze the contents of these job ads, we apply a topic modeling algorithm and iterate between the resulting patterns and the job descriptions. This allows us to identify four data-related responsibilities: “data infrastructure”, “data control”, “applied analytics”, and “data privacy” that characterize leadership in the age of data analytics. Through subsequent analyses, we observe that while job ads for “data analytics” positions often provide explicit information regarding data-related responsibilities, positions with titles pertaining to “information technology” and “digital” are much less likely to be responsible for providing data analytics leadership. We also find that most “data analytics leaders” will be responsible for ensuring the creation of insights, products and services from data through analytics, whereas the privacy implications of data analytics are marginally discussed in the job ads for data analytics leaders. We discuss how these findings help us advance our understanding of leadership in the age of data analytics.

An earlier version of this paper has been co-authored by dr. M. H. Rezazade Mehrizi and dr. J. Van Angeren.
4.1 Introduction

For some years now, data and analytics have been among the top priorities on senior managers’ agendas (Gartner, 2013; 2018-1; Kappelman et al., 2013). Increasingly, managers see strategic opportunities in data (Hartmann et al., 2016; Woerner & Wixom, 2015; Wixom & Ross, 2017; Dallemule & Davenport, 2017) and seek to “exploit their growing data and computational power to get smart, and get innovative, in ways they never could before” (LaValle et al., 2011, p. 3). To this end, managers have progressively been investing resources into acquiring and developing data analytics capabilities (Gartner, 2013; Gartner, 2018-1). The International Data Corporation (2018) predicts that these investments will only continue to grow, and yield $260 billion in global revenues by 2022.

However, senior managers arguably have to do more than invest resources in data analytics. Indeed, scholars have argued that senior managers play a key role in implementing data analytics capabilities and ensuring that the associated benefits are realized. Multiple studies frame “top management support” and “leadership” as mediating variables or key success factors for the effective implementation of data-driven initiatives (Chen et al. 2015; Davenport et al., 2010; Gao et al., 2015; McAfee & Brynjolfsson, 2012; Seddon et al., 2017). Beyond providing resources and support, studies highlight that managers of organizations and business units need to “take leadership of initiatives or projects” (Seddon et al., 2017, p. 243; Davenport et al., 2010) and should be actively involved in all stages of data-related initiatives (Gao et al., 2015).

Although there seems to be a consensus that data analytics leadership by senior managers is important, there has been some ambiguity regarding what such leadership actually entails. For one, it remains unclear—from the wide range of technological, social, and practical aspects that senior managers may be responsible for (Hales, 1986; Fells, 2000)—which responsibilities are related to data analytics, and how these may be specific to leadership in the age of data analytics. Additionally, scholars are debating which positions are likely to be responsible for providing data analytics leadership—are they, for example, “Chief Information Officers”, “Chief Digital Officers”, or perhaps “Chief Data Officers” (Galbraith, 2014; Lee et al., 2014; Tumbas et al., 2017)? By exploring which responsibilities are currently associated with data analytics leadership, and which positions are typically responsible for providing data analytics leadership in practice, we may be able to advance our understanding of leadership in the age of analytics and how it differs from other types of leadership.

In this study, we ask: What are the expected responsibilities and positions of data analytics leaders? To answer this question, we collected a total of 9145 unique job ads for chief officers, presidents, and directors of which the titles pertain to “data analytics”, “digital”, and “information technology”. We
collected these job ads from Indeed.com—a popular job website in the U.S. On 6997 of those jobs ads that explicitly mention a data analytics-related term in the description, we applied topic modeling to explore three specific questions: 1) What are the data-related responsibilities that characterize data analytics leadership? 2) Which positions will be responsible for providing data analytics leadership? and 3) Which (combinations) of responsibilities are stressed as the most important for “data analytics leaders”? 

Our analysis reveals four data-related responsibilities of senior managers: “data infrastructure”, “data control”, “applied analytics”, and “data privacy”. We contribute to the literature on data analytics leadership by reflecting on the extent to which these may be contemporary responsibilities that are characteristic of leadership in the age of data analytics. Additionally, the data suggest that job ads for “data analytics” positions often provide explicit information regarding data-related responsibilities, which implies that organizations do not just signal these terms in the job titles (Nelson et al., 2014). Conversely, positions with titles pertaining to “information technology” and “digital” are much less likely to be responsible for providing data analytics leadership. We discuss what this implies for our understanding of data analytics leadership and future research on the topic. Finally, we find that most “data analytics leaders” in our sample will be responsible for applied analytics, while the privacy implications of data analytics are marginally discussed in their job ads. We consider what this means in light of recent debates regarding the implications of data analytics for consumers and society (Clarke, 2016; Newell & Marabelli, 2015; Markus, 2015; Zuboff, 2015).

In the following sections, we first further explicate the ambiguities in the literature on data analytics leadership. Then, we describe how we collected job ads for senior managers that might become “data analytics leaders”, and explain how we analyzed the job descriptions. After, we summarize and discuss our main findings and explicate the boundary conditions of our study.

4.2 Theoretical Background

Scholars and practitioners have long considered the importance of senior leadership in implementing technology-related initiatives, thereby often focusing on the role of senior “Information Systems (IS) leaders” (e.g., Karahanna & Watson, 2006; Applegate & Elam, 1992; Todd et al., 1995) such as Chief Information Officers (e.g., Chun & Mooney, 2009; Peppard et al., 2011; Ross & Feeney, 1999; Weill & Woerner, 2013) and Chief Technology Officers (e.g., Kappelman et al., 2013). Like any manager, these positions may be concerned with leading a business unit; forming and maintaining contracts; dealing with information; allocating resources; dealing with disturbances; negotiating; innovating; planning, and directing employees (Hales, 1986; Fells, 2000). But next to such more
classical leadership responsibilities and tasks, “IS leaders” are also specifically responsible for aspects such as IT strategy development and execution; Business-IT alignment; IT costs and investments; IT governance and management; IT architecture and infrastructure; IT support and operations; IT risks, information security, and privacy; IT system development and maintenance; IT-driven products and services, and IT-driven innovation (Applegate & Elam, 1992; Chun & Mooney, 2009; Kappelman et al., 2013; Peppard et al., 2011; Ross & Feeney, 1999; Weill & Woerner, 2013).

In the specific context of data analytics, the importance of senior leadership has not gone unnoticed either. The general assumption is that “[t]he greater the extent of analytic leadership in the organization, the greater the organizational benefits” (Seddon et al., 2017, p. 254; Davenport et al., 2010). Although theoretically, data analytics leadership may be provided by champions residing at any organizational level (Seddon et al., 2017), scholars and practitioners often consider the role of senior managers as crucial (e.g., Chen et al., 2015; Fitzgerald, 2014; Lee et al., 2014; McAfee & Brynjolfsson, 2012; Gao et al., 2015; Gartner, 2017-1). Compared to lower-level managers, senior managers may be in a better position to gain a holistic view of enterprise-wide data analytics efforts, and have higher authority to make decisions regarding data-driven initiatives (Fitzgerald, 2014; Lee et al., 2014).

So far, the literature on data analytics leadership has largely focused on explicating the importance of leadership (e.g., Gao et al., 2015; McAfee & Brynjolfsson, 2012; Seddon et al., 2017), distinguishing between different types of leaders (e.g., Lee et al., 2014; Tumbas et al., 2017), and making suggestions for what makes a “strong” data analytics leader (e.g., Davenport et al., 2010). For example, Lee et al. (2014) suggest that data analytics leadership roles differ depending on the envisioned value impact (service vs. strategy), the type of data that leaders focus on (existing vs. new) and their collaboration direction (inwards vs. outwards). The literature also suggests that data analytics leaders need to be able to effectively communicate with people; drive analytics strategies; promote data use and data-driven results; manage a portfolio of diverse projects; build a network; be persistent and committed while also being aware of the limitations of their domain; know how to hire the right people, and teach and lead by example (Davenport et al., 2010).

Although the insights provided by the literature on data analytics leadership are insightful, there is also some ambiguity. First, the literature is ambiguous about what concrete responsibilities may characterize data analytics leadership. Scholars have suggested that when organizations actively try to realize value from data, areas of leadership responsibilities may emerge that traditional

27 In this study, we use “data analytics leadership”—as opposed to only analytic leadership—to refer to the type of leadership that is crucial for ensuring the successful implementation of all data-driven initiatives, including those without an analytical component.
leaders have not been responsible for (Galbraith, 2014; Lee et al., 2014; Sharma et al., 2014). For example, managers might find that they need someone to be responsible for developing “data governance” capabilities (Sharma et al., 2014; Tallon et al., 2013-14; Davenport et al., 2010). Yet, we have limited knowledge of the range of data-related responsibilities that senior managers may be responsible for in practice, nor about the extent to which these responsibilities are actually contemporary responsibilities that uniquely characterize leadership in the age of data analytics. Capturing the full range of data-related responsibilities of senior managers may be the first step towards understanding what data analytics leadership entails, and how it may be different from more traditional forms of IS leadership.

We also see some debates in the literature regarding which positions are responsible for providing data analytics leadership. Because data analytics has its roots in technology, organizations may initially consider traditional IS-related positions, such as “Chief Information Officers”, to be responsible for data analytics (Fleckenstein & Fellows, 2018; Galbraith, 2014; Gartner, 2017-2). For example, Galbraith (2014) describes a case where the Chief Information Officer was wearing “two hats” and was responsible for both traditional technology-related tasks and for implementing data analytics. He argues that this can be useful when data analytics is seen as a “competence-enhancing innovation”, rather than something that drives radical transformation (Galbraith, 2014, p. 4). Recently, scholars and practitioners have also been pushing for new positions such as “Chief Data Officers” and “Chief Analytics Officers” (e.g., Aiken & Gorman, 2013; Fitzgerald, 2014; Gartner, 2017-1; Lee et al., 2014; Zhang et al., 2017). They worry that more traditional IS-related leadership positions might “run out of time or lack the ability to handle all aspects of data” (Zhang et al., 2017, p. 1). Other scholars have suggested that so-called “Chief Digital Officers” may be held responsible for providing data analytics leadership, as data are inherently “digital” artifacts (Tumbas et al., 2017; Galbraith, 2014; Zhan & Mu, 2016). Literature also acknowledges that in practice, data analytics leaders may not formally carry “chief officer” titles at all—e.g., when organizations do not find this necessary (Lee et al., 2014). To theorize about data analytics leadership, it would be helpful to consider which positions are actually expected to be responsible for providing such leadership in practice.

4.3 Methods

Because of the novelty of the field of data analytics and the ambiguity surrounding our specific topic of interest, we adopt an exploratory approach to answer our research question. Specifically, we collect job ads for senior managers with titles pertaining to “data analytics”, “digital”, and “information technology”, and analyze their contents by means of topic modeling (Berry et al., 2007). Job ads
have been used as data sources for academic studies (e.g., Todd et al., 1995; Kennan et al., 2009) and “are particularly relevant to areas where job roles are increasingly non-traditional and not connected to a well-defined profession” (Harper, 2012, p. 30). This seems very applicable in our context, given that data analytics is still a relatively new phenomenon. Organizations that publish job ads often aim to find candidates that fit well with their business and can fulfil the role that is expected from them (Todd et al., 1995). Therefore, job ads provide a rich source for understanding what data-related responsibilities are associated with particular positions.

4.3.1 Collecting Job Advertisements

As part of an ongoing research effort, we scraped job ads from the popular job website Indeed.com. Indeed has been claimed to hold “the largest market share among employment sites” (Schlee & Karns, 2017, p. 71). The website is perceived as “one of the most comprehensive job-posting websites in the United States”, partly because Indeed also scrapes job ads from other webpages, digital newspapers, and company websites (Alfaro & Watson-Manheim, 2015, p. 39). To maintain consistency and avoid patterns emerging due to cultural bias, we decided to only scrape job ads from the U.S.-version of the job website.

To find job ads for senior managers, we used Indeed’s advanced search option and scraped job ads that included the terms 1) “chief” and “officer”; 2) “avp”, “vp”, “svp”, “evp”, or “president”, or 3) “director” in the title. By including these different titles, we account for the fact that senior managers who provide data analytics leadership may not always carry the title of “chief officer” (Lee et al., 2014). Although the actual levels of seniority of “directors”, “presidents” and “chief officers” can differ depending on the type and size of the organization, we assumed that these titles are granted to individuals who are, at least, responsible for an organizational (sub)unit.

To find job ads for senior managers who may be held responsible for data analytics, we explicitly searched for job ads with any of the terms “data”, “analytics”, “digital”, “information”, and “technology” in the title28. We included the terms “information” and “technology” because data has traditionally been considered the responsibility of IT (Todd et al., 1995), and scholars have acknowledged that managers with these terms in the title may also be held responsible for data analytics in some cases (Fleckenstein & Fellows, 2018; Galbraith, 2014). We included “digital”, because this term has recently also been associated with data-related responsibilities and because data are inherently “digital” resources (Tumbas et al., 2017; Galbraith, 2014; Zhan & Mu, 2016).

---

28 Indeed’s search engine also automatically searched for different forms of these terms, e.g., technologies and analytic.
We included “data” and “analytics” to capture roles such as the “Chief Data Officer” and “Chief Analytics Officer” (e.g., Aiken & Gorman, 2013; Fitzgerald, 2014; Gartner, 2017-1; Lee et al., 2014; Zhang et al., 2017). We collected job ads on a weekly basis between May, 29, and August, 21, 2018.

4.3.2 Selecting and Preparing Job Advertisements for Analysis

For each job ad, we scraped the title (which reflects the position) and the job description (which describes the responsibilities). Initially, our searches resulted in a total of 13191 ads with unique identifiers for the given period. However, we found that even though all job ads had unique identifiers, there were still many duplicate texts (roughly 30%) in the dataset. We suspect that organizations re-upload the same job ad under a different identifier at later points in time to ensure that the vacancy receives more attention. It might also be that Indeed.com collects the same job ad from different sources, and publishes the individual copies under unique identifiers. We used the cosine distance between two weighted word-vector representations of ads to decide whether they are duplicates and kept one version (See Appendix, Table A.4.1 for details on filtering procedures). Additionally, we excluded job ads that resulted from our search queries, but were not actually for chiefs, presidents, or directors. This includes ads for assistants (to the chief, president or director) and other employees that are part of the “office of” the chief, president, or director. For consistency, we also excluded job ads for chief officers whose titles do not relate to one of the areas that we searched for, such as a “Chief Finance Officer” and “Chief Operations Officer”, and job ads that we came across that were clearly not relevant (e.g., a Massage Therapy Program Director at a School of Business and Technology). Finally, we excluded job ads of less than 500 characters, because ads that small are typically not elaborate enough to gain insights from their description, and to minimize the effect of size on our analysis. After this initial filtering, 9145 job ads remained, published by 4719 organizations.

Left with a set of job ads reflecting the roles and responsibilities of senior managers with positions pertaining to “data analytics”, “digital”, and “infor-

---

29 We scraped on a weekly basis and repeated our searches with different filters to minimize the chance of missing job ads due to a failed scrape attempt.
30 By implication, job ads for which the issuing company largely copied texts from job ads by other companies are also considered duplicates. Similarly, job ads are marked as duplicates when organizations copy the same job ad for extremely similar positions (e.g., the same position in a new location). New versions of the same job ad, however, may be considered unique.
31 These resulted from the search query when the title was, for example “chief finance officer for a technology company”.
32 We count the number of “organizations” by capturing the number of unique company and department names as presented in the metadata of the job ads. By implication, when an organization explicitly specifies the department or uses a slightly different brand name, it is also counted as a distinct “organization”. An organization could in practice also be a recruiter for several other organizations.
mation technology”, we cleaned them such that we could automatically analyze their contents. Cleaning the descriptions consisted of a number of steps (for details about the algorithms and parameters used, see Table A.4.2 in the Appendix). To begin, for each job description, we removed disclaimer sentences\(^{33}\) and URLs, standardized abbreviations, and dealt with punctuation and numbers. Next, for each job ad, we filtered out the company name, so we would not capture company names as dimensions in our analyses. For similar reasons, we removed corporate titles, U.S. states, and the names of cities from the descriptions, because we captured this information in the job ads’ metadata. After, we lemmatized words (i.e., brought them back to a base form), removed stopwords (e.g., do, have, he, she), and removed “words” that consisted of a single character. We also removed words that are common to job vacancies (e.g., job, position, requirement, organization, skill, ability), words that denote time periods (day, month, year), and words that are common to company settings, including corporate suffixes (e.g., inc, co, company). After cleaning each job description, we used an algorithm to find bigrams, that is, pairs of words that often occur together, such as “Information Technology” and “Business Intelligence”. Because bigrams are references to concepts in and of themselves, we replaced the two words by their concatenated version in each job ad.

After cleaning the job ads, we examined which job ads included “data” or “analytics”\(^{34}\) in the title, or otherwise mentioned “data”, “analytics”, “business intelligence”, “artificial intelligence”, or “machine learning” in the description. We mainly did this to filter out irrelevant job ads, thereby assuming that job ads that do not mention any of these terms are unlikely to describe roles of data analytics leaders. We found that 6997 job ads (published by 3693 organizations) fit this criterium and proceeded to analyze this collection of job ads.

### 4.3.3 Topic Modeling of Job Advertisements

Topic modeling is a type of text analysis that can be applied to identify a number of “topics”, or latent features, that describe the texts in a corpus (collection of texts) (Schmiedel et al., 2018). The intuition behind topic modeling is that topics cannot directly be observed in documents; rather, topics are represented by a weighted combination of words (Chang et al., 2009). Popular topic modeling techniques include, for example, Latent Semantic Analysis (LSA) (e.g., Evangelopoulos et al., 2012), Latent Dirichlet Allocation (LDA) (Blei et al., 2003; Blei, 2012), and Nonnegative Matrix Factorization (NMF) (Berry et al., 2007).

In this paper, we chose to apply Nonnegative Matrix Factorization

---

\(^{33}\) For example: “We are an equal opportunity employer. We evaluate qualified applicants without regard to race, color, religion, sex, sexual orientation, gender identity, genetic information, national origin, disability, veteran status, or any other protected characteristic”.

\(^{34}\) Including “analytic”, given that indeed also automatically searched for different forms of the terms.
(NMF), which essentially is a technique to reduce the dimensionality of data 
(Berry et al., 2007). It takes a certain matrix $A$ as input—in our case a docu-
ment-term matrix—and decomposes this matrix into two nonnegative matrices 
$W$ and $H$ (Berry et al., 2007). This nonnegativity is one of the benefits of using 
NMF over other techniques such as LSA, that are known to produce negative 
values (Choo et al., 2013). The characteristic of nonnegativity allows scholars 
to “derive meaningful interpretations” from the resulting matrices (Shahnaz et 
al., 2006, p. 374), which is especially helpful when one wishes to interpret the 
topics and analyze documents and their corresponding topics.

Another advantage of NMF is that it allowed us to apply a weighting 
scheme to the document-term matrix. We used “Term Frequency Inverse Doc-
ument Frequency” (TFIDF) weighting before feeding the jobs ads into the topic 
model (Salton & Buckley, 1988). TFIDF is based on the premise that “the best 
terms for document content identification are those able to distinguish certain 
individual documents from the remainder of the collection” (Salton & Buckley, 
1988, p. 324). The assumption is that when a word occurs very frequently in a 
single job ad relative to the other job ads, it may be a useful word to represent 
that ad. A TFIDF weighting allows us to capture those job ads that are truly 
about data analytics, as opposed to those job ads that more casually refer to 
data and analytics (Salton and Buckley, 1988). Being able to apply a weighting 
scheme is one of the benefits of using NMF over, for example, a probabilistic 
model such as LDA (O’Callaghan et al., 2015)—when applying probabilistic 
models such as LDA, frequent terms are likely to show up in many of the top-
ics because of their high frequency and co-occurrence with many other terms. 
When we transform a document-term matrix using TFIDF, the output is a doc-
ument-term matrix where each term is weighted: more “common” words in the 
corpus of job ads are given less weight, whereas more distinguishing words are 
given more weight (Salton & Buckley, 1988).

The topic modeling algorithm required us to estimate the number of 
topics that might be found in the documents beforehand. Choosing the num-
ber of topics is a key challenge when applying any topic modeling algorithm, 
and often comes down to finding a balance between getting meaningful topics, 
getting useful topics, and getting many topics (Chang et al., 2009). In our case, 
this involved running NMF multiple times with varying numbers of topics. We 
found that models with the number of topics ranging between 50 and 60 gen-
erally provided us with interesting and coherent topics\textsuperscript{35}. Eventually, we chose 
a model that presented us with a total of 55 topics. An overview of all 55 topics

\textsuperscript{35} Each time we ran a model for a certain number of topics, we also ran it for a number of 
different random seeds and saw that model outputs could differ. As Berry et al. (2007) argue: “in 
practice, even local minima can provide desirable properties such as data compression and feature 
extraction” (p. 156). Once we settled on the number of topics, we tried numerous different seeds 
and chose a model that provided us with conceptually coherent and distinct data-related topics 
that helped us to answer our research question.
and the top 10 words for each topic can be found in the Appendix (Table A.4.3).

After choosing a model, and following (Choo et al., 2013), we normalized the resulting matrices, thereby arriving at both topic-word allocations and advertisement-topic allocations (comparable to the output of an LDA topic model). We first column-normalized the topic-word matrix by dividing all values in a row by the sum of values in that row, such that the word distributions for each topic add up to one36. After, we column-normalized the advertisement-topic matrix by dividing all values in a row by the sum of values in that row. Note, that this step “does not affect the interpretation of each document in terms of its relative relationship to topics” (Choo et al., 2013, p. 1994). However, it does affect how we interpret top ads per topic, because the column cells now represent the relatedness of a job ad to that topic relative to all other topics in that job ad. Rather than representing a raw value, the topic weight now signifies, on a scale of 0 to 1, how representative the topic is for that document.

To interpret the topics resulting from NMF, we first examined the top 10 words for each topic and assigned an initial label capturing the concept described by those 10 words. Then, we examined at least the top 10 job ads that scored highest on each topic—in doing so, we highlighted the top words corresponding to the topic to see how these words are used in the job descriptions, and how well our label actually represents the topic. We then categorized topics into three main categories, based on whether the topics say something about the job, the application domain, or the industry. The remaining topics may be categorized as “vacancy-related” or “other”. Within these main categories, we inductively grouped similar topics into general themes (e.g., “data analytics”, “information technology”, “marketing”). Given our research question, we created more specific subthemes for those topics that are specifically related to “data analytics”. For example, based on closer examination, we combined the topics “data governance” and “data management” under the subtheme “data control”, because they appeared to be closely related. Similarly, we combined topics “business intelligence” “analytics”, “data science”, and “analysis” under the subtheme “applied analytics” (see also Appendix, Table A.4.4).

To distinguish between different positions of senior managers, we extracted both the function (i.e., “chief officer”, “president” or “director”) and the main subject area (i.e., “data analytics”, “digital”, or “information technology”) from the title of the job ads. In some cases, job titles constituted a combination of such terms. In these cases, we favored “data analytics” over “digital”; and “digital” over “information technology”. In other words, if a title included both the terms “digital” and “data”, we would categorize it as a “data analytics” position. Similarly, we favored “chief officer” over “president”, and “president” over “director”.

36 While keeping the ratios the same, such that the product of matrices is still an approximation of the input matrix.
4.4 Findings

Before we had filtered out job ads that do not mention any data analytics-related term (see section 4.3.2), we had collected 9145 unique job ads for “directors”, “presidents”, and “chief officers” in the areas of “data analytics”, “digital”, and “information technology”. Looking at the positions as reflected in the titles of these job ads already gives an overview of the demand for positions with these specific titles on Indeed.com.

![Figure 4.1 Distribution of job ads over positions.](image)

Figure 4.1 shows that, of all the unique job ads that we initially found on the basis of our search queries, 37% have “data” or “analytics” in the title, compared to 22% that have “digital” in the title, and 41% that primarily have “information” or “technology” in their title. These numbers suggest that while “digital” and “data analytics” positions have both been highlighted in the literature as emerging and important positions, “data analytics” positions were in much higher demand than “digital” positions at the time of searching. Although this relatively high number of “data analytics” positions does not necessarily indicate that these positions are in high demand by all the different organizations. On further inspection, we find that 55% of the companies and departments in our dataset published at least one job ad for a senior “information technology” position, compared to 36% that published a job ad for a “data analytics” position, and 28% that published a job ad for a senior “digital” position (see also Appendix, Table A.4.5, for the top 10 publishing organizations for these positions).

Figure 4.1 shows that we found only very few job ads for “chief officer data analytics” positions; most of the senior “data analytics” positions are “directors” and “presidents”. In total, we found only 29 unique job ads for “chief officer data analytics” positions, which suggests that the actual titles of “Chief Data Officer” and “Chief Analytics Officer” are not often granted to newly hired
senior managers. Figure 4.1 shows that we found an even lower number of job ads for “chief officer digital” positions, only 14 job ads in total. By comparison, we found 627 job ads for “chief officer information technology” positions—which includes “Chief Information Officers”, “Chief Information Security Officers”, and “Chief Technology Officers”. Thus, even if organizations generally have more presidents and directors than they have chief officer positions, the number of “chief officers data analytics” and “chief officers digital” is still strikingly low compared to the number of “chief officers information technology”.

4.4.1 Identifying Data-Related Responsibilities

In order to filter out job ads that are very unlikely to describe “data analytics leaders”, we considered only those job ads that included “data” or “analytics” in the title, or otherwise mentioned “data”, “analytics”, “business intelligence”, “artificial intelligence”, or “machine learning” in the description. On these, we applied topic modeling to see which themes have been discussed in them. Our topic modeling approach on 6997 job ads resulted in a total of 55 latent topics (see Appendix, Table A.4.3). Based on our examination of the top words and documents for each topic, we identified nine topics as being specifically related to data analytics (See Appendix, Table A.4.4 for an allocation of topics to themes). In Figure 4.2, we visualize the top 10 words for these data-related topics. The size of the word is proportionate to the weight of the word to the topic.

![Figure 4.2 Top words per data-related topic.](image)

Based on a closer examination of the job descriptions of job ads that score high on data-related topics (see Appendix, Table A.4.6 for the top 10 positions), we further grouped data-related topics into four themes based on the conceptual similarities among them: “data infrastructure”, “data control”, “applied analytics”, and “data privacy”. We consider these themes to be representative for the range of data-related responsibilities of senior managers.
Data Infrastructure

The first data-related responsibility is “data infrastructure”, which comprises both “data sources” and “big data infrastructure”. Job ads that score high on “data sources” describe senior managers who will be responsible for the setup, development and maintenance of electronic data sources, such as databases and data warehouse environments. This includes, for example, a Director of Enterprise Data Architecture that will be responsible for “the delivery of reliable, timely, and consumable data for operational use and downstream analytics” and generally “all data assets under business functions” (Software Company). Other positions are, for example, directors and presidents of Data Engineering, Data Sourcing, and Data Warehouses. Job ads that score high on “big data infrastructure” lay more emphasis on the experiences that applicants should have in working with cloud solutions (public, private, and hybrids) and sophisticated big data technologies such as “Hadoop, Spark, Hive, Impala, HBase, Cloudera Manager, Sqoop, Flume, Pig, etc” (AVP-Big Data & Analytics Product Manager; Financial Services Company). Among the job ads that score high on “big data infrastructure” are, for example, directors of Big Data Infrastructure, Big Data Platforms, Cloud Data Strategies, and Big Data Engineering.

Job ads that generally highlight “data infrastructure” as a data-related responsibility describe many technology-related tasks associated with building a (big) data infrastructure, such as data acquisition, data sourcing, data access, data modelling, data engineering, data cleaning, data quality, data migration, data integration, data transformation, data validation, data storage, data operations, and data architecture. The focus in these job ads is generally on how organizations may, from a technical perspective, effectively store and disseminate data such that they are available to and can effectively be used by many different stakeholders. Given the universal focus on the technological aspects of (big) data analytics, applicants are generally expected to have technical backgrounds and “hands-on” experience with (big) data technologies. However, they may also be expected to frequently communicate with non-technical stakeholders, such that they can effectively implement “high-level architectural decisions and designs based on business priorities and requirements” (Director, Big Data Infrastructure; Large Retail Company).

Data Control

The second responsibility that we associated with data analytics based on our analysis is “data control”. This responsibility subsumes two topics. The first topic, “data governance”, is defined in one of the job ads as “the overall management of the availability, usability, integrity and security of data used in an enterprise” (Director Enterprise Data Governance Data Quality; Finan-
cial Services Company). Job ads that highlight data governance suggest that it is about designing data standards and policies; facilitating compliance with data standards and policies; designing roadmaps and governance strategies; ensuring alignment of data roadmaps with the business strategies; allocating data-related responsibilities; deciding on decision rights, accountabilities, and ownership; establishing data stewardship, and other governing tasks to ensure that data can be exploited to create business value.

According to the job descriptions, “data management” concerns the management activities that are needed to prepare data for (future) use and analysis. Job ads describe managers who will monitor and improve data quality and integrity; develop and execute policies, and generally, “develops systems for organizing data” (Associate Director, Clinical Data Management; Biopharma). Applicants are expected to already have data management experience, including knowledge of database development, management, administration, testing, and release. Given that these managers need to adhere to many standards, knowledge about policies and regulations may be a prerequisite. Additionally, these senior managers may benefit from having time management skills, being organized, and being able to pay attention to detail, such that they can “effectively direct the activities [...] projects, and staff in a manner that ensures all functions, projects, and tasks are handled with the highest quality, on time, and within budget” (Director, Data Management; Clinical Research Organization).

Based on our qualitative examination of the job ads, we argue that data governance is more about designing frameworks, policies and standards, whereas data management is more about actually implementing frameworks and ensuring that policies and standards are followed in practice. This is in line with how scholars distinguish between data governance and management (Khatri & Brown, 2010). Still, the topics are conceptually very closely related and sometimes used interchangeably in the job ads. Both topics suggest that leaders in the age of data analytics should care about controlling data access, quality, and use within and beyond the enterprise and business units.

**Applied Analytics**

Another responsibility that we associated with data analytics is “applied analytics”. We refer to this responsibility as “applied analytics” to reflect that it is really about facilitating and initiating the analysis of data through the application of (sophisticated) algorithms and tools. We associated four topics with this responsibility: “business intelligence”, “analytics”, “data science”, and “analysis”.

Job ads that score high on “business intelligence” and “analytics” generally describe senior managers who will be responsible for the development and creation of analytics capabilities, tools, and data-driven (strategic) insights.
Job ads mention a whole range of different tools, such as Tableau, Qlikview, PowerBI, Google analytics, and Adobe analytics. Next to having knowledge of existing tools, senior managers may also be expected to “contribute to the development of new analytics tools and techniques” (VP of Analytics; Recruitment Company). Job ads scoring high on “business intelligence” tend to focus on the development of reporting solutions and facilitating the analysis of data for decision making. For example, senior managers may be expected to “advocate for the effective use of data and drive a culture of data-informed decision making across the company” (Director, Business Intelligence and Analytics; Auto Parts Provider). Job ads that score high on “analytics” may also consider analytics solutions more generally, and highlight the need for building analytics capabilities, driving an analytics culture, guiding analytics-driven change trajectories, and acting as the “voice of analytics for the business” (VP Digital Analytics, Sports Company).

Judging from the job ads, “data science” is associated with more sophisticated analytics and algorithms, big data, and machine learning techniques. For example, organizations may look for senior managers who have experience with “Logistic regression, naive bayes, SVM, decision trees, neural networks, and random forests” (Vice President-Data Science; Financial Services Company). Consequently, job ads that highlight this topic ask for advanced degrees (even PhDs) in exact domains such as computer science and statistics. The focus of these positions is not exclusively on analytics for decision-making, but also on developing data-driven solutions in the form of products and services, as part of a longer-term data science strategy. Next to having solid technical and analytical skills, applicants may be expected to translate insights into actions and act “as a liaison between business, data science, and engineering to monitor and implement projects through completion” (Director of Data Science; Marketing Company).

The topic that we labelled as “analysis” represents hands-on analytical responsibilities. The associated words—which include words such as model, modeling, and statistical—suggest that this topic is about actually doing analytics. If we examine the content of job ads that highlight “analysis”, we see that companies are indeed looking for people who can actually perform analyses and create models. Types of analyses include predictive modeling, natural language processing, and speech recognition. Companies that strongly voice this theme in their positions have quite high expectations in terms of analytical skills, sometimes even requiring a PhD in statistics, mathematics or computer science. Applicants should at least be able to work in programming languages such as Python, R, and Matlab and be familiar with tools such as Qlikview, Tableau.

37 We also identified several topic models in which “machine learning” is highlighted as a distinct topic. However, we found that “machine learning” is also often associated with “data science” and decided to choose a model in which these concepts are combined into one topic.
leau, and Spotfire. Communication skills are also needed to share findings and insights for the decision-making process.

We can conclude that there is a noticeable heterogeneity between the topics that we combined under “applied analytics”. For example, “data science” is generally associated with more sophisticated analytical techniques, whereas job ads scoring high on “business intelligence” may refer to more traditional reporting. Still, the four topics all revolve around the analysis of data through the application of (sophisticated) algorithms and tools. This suggests that data analytics leaders can be held responsible for ensuring that insights can be gained from data, and that analytics efforts result in new solutions and value-creating actions.

**Data Privacy**

We labelled the fourth and final responsibility “data privacy”. Although “data privacy” may be related to “data control” in that both responsibilities are about compliance, policies, and regulations to a certain extent, we argue that it is also distinct because of its ethical connotation. Whereas “data control” is more about controlling data access and use, “data privacy” is more specifically about protecting the data and the people.

Senior managers responsible for data privacy generally need to ensure that data-driven initiatives comply with local and global privacy regulations and that personal data—either owned by the company or by third-parties—are “protected”. They are expected to develop and lead a data privacy program; develop and monitor adherence to privacy policies, standards and procedures; educate stakeholders about privacy best practices; control damage when data are leaked and privacy is breached, and follow and prepare for changes in privacy regulation. They may need to act as a “data privacy advocate and liaise as a trusted advisor with business leaders to increase awareness and stakeholder engagement throughout all departments, whilst delivering support to business led projects and initiatives” (Director, IT Data Privacy; BioTechnology Company). Although we scraped job ads from the U.S.-based version of Indeed.com, applicants are expected to also have thorough knowledge of foreign privacy laws, such as the European Union General Data Protection Law (GDPR) and the Canadian Ontario Act. Applicants need to be adept at conflict resolution and maintaining relationships with peers, customers, and partners. Interestingly, some companies request a formal education in technology-related fields, whereas others emphasize the need for law school degrees.

Organizations may highlight in their job ads that “data privacy” is about more than regulation. For instance, an IT service provider acknowledges that “technology advances much more quickly than legislation” and is looking for
someone who will “be on the cutting edge of new thinking, new innovations, and acting as a leader in the future of privacy and data protection” (Chief Privacy and Data Ethics Officer). Another company, a data services provider, is looking for a “Privacy and Data Protection Evangelist” who will enact a “forward-looking vision regarding privacy and data protection”. These examples illustrate that organizations may expect senior managers to actively consider the potential social implications of data analytics use, and may even see “data privacy” as an area that they can differentiate on.

4.4.2 Identifying Data Analytics Leaders

Now that we have gained insights into the range of data-related responsibilities that senior managers may be responsible for, we can use these insights to see which positions will typically be held responsible for providing “data analytics leadership”. To do this, we first defined a measure of “data-relatedness” that indicates—on a continuous scale from 0 to 1—how much data-related responsibilities represent the job ad. We operationalize data-relatedness by taking the sum of the weights of all data-related responsibilities for each job ad. Figure 4.3 below shows a frequency histogram of data-relatedness for the job ads. We also visualized a number of measures to describe the distribution, including the mean, the first quartile, the median, and the third quartile.

Figure 4.3 Distribution of data-relatedness.

Figure 4.3 shows that, even though all job ads mention at least one single data analytics-related term, many of the job ads score relatively low on data-relatedness. The 25th percentile is 0.03, which means that 25% of the job ads score below this value on data-relatedness. In other words, for those job ads, 97% of their topic weight can be explained by other responsibilities or topics that are not specifically related to data analytics. We also see in the figure that the median for data-relatedness is close to 0.08, which means that for half of the job ads, still around 92% of the weight can be explained by other responsibilities or topics. There are several reasons for why job ads have low data-relatedness scores—even when they do mention at least one single data
analytics-related term. For example, job ads may have mentioned data-related terms in a different context (e.g., a CIO of a school that needs to have “dexterity for the purpose of using a telephone and data entry”). Similarly, job ads may consider data-related responsibilities, but only very marginally (e.g., as just another IT component) or next to so many other responsibilities or elements that the data-related responsibilities are pushed to the background. Another reason why job ads may score low on data-relatedness is when the responsibilities and requirements described are so general that they may be applicable to many (non-data-analytics-related) jobs—even when the title suggests that the role is about data analytics.

Now that we have a way of expressing how much data-related responsibilities represent the job ads, we can see how this differs for senior manager positions with different titles. In Figure 4.4, we show the data-relatedness of job ads for different positions.

![Data-Relatedness per Position](image)

**Figure 4.4 Data-relatedness per position.**

The figure shows that those job ads that do not have “data” or “analytics” in the title generally score low on data-relatedness. This seems especially true for “digital” positions: we find that 90% of all the job ads for chief officers, presidents, and directors digital score below the median (0.08) on data-relatedness, and hardly any of them score above the third quartile (0.22). This is sur-
prising, because the literature has pointed to senior “digital” positions as those who may be held responsible for providing data analytics leadership (Tumbas et al., 2017; Galbraith, 2014; Zhan & Mu, 2016). One would expect that data-related responsibilities are highlighted in the job ads for these senior managers. Of the job ads for senior managers that primarily have “information” or “technology” in their titles, about 78% score below the median on data-relatedness. Job ads that have forms of “data” or “analytics” in their title do highlight data-related responsibilities in the descriptions: 85% of them score above the median on data-relatedness.

We may assume that job ads for “data analytics leaders” are those that highlight data-related responsibilities in the description, and therefore score above a certain threshold on data-relatedness. We choose the exact median as our threshold, which represents the start of the turning point—the “elbow” of the curve—in Figure 4.3. In total, 3499 job ads (published by 1836 organizations) score above this threshold. We classify these job ads as job ads for “data analytics leaders” that provide us with explicit information about data-related responsibilities. In Table 4.1, we summarize for each position: how many job ads we had initially collected, how many of them mentioned a data-related term, and how many job ads could be considered ads for “data analytics leaders” based on their data-relatedness scores.

<table>
<thead>
<tr>
<th>Position</th>
<th>Initial</th>
<th>Mention data-related terms (% of initial)</th>
<th>“Data analytics leaders” (% of initial)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chief officer data analytics</td>
<td>29</td>
<td>29 (100%)</td>
<td>26 (90%)</td>
</tr>
<tr>
<td>Chief officer digital</td>
<td>14</td>
<td>11 (79%)</td>
<td>1 (7%)</td>
</tr>
<tr>
<td>Chief officer information technology</td>
<td>627</td>
<td>426 (68%)</td>
<td>79 (13%)</td>
</tr>
<tr>
<td>Director data analytics</td>
<td>2571</td>
<td>2571 (100%)</td>
<td>2166 (85%)</td>
</tr>
<tr>
<td>Director digital</td>
<td>1663</td>
<td>1102 (66%)</td>
<td>114 (7%)</td>
</tr>
<tr>
<td>Director information technology</td>
<td>2537</td>
<td>1481 (58%)</td>
<td>341 (13%)</td>
</tr>
<tr>
<td>President data analytics</td>
<td>768</td>
<td>768 (100%)</td>
<td>635 (83%)</td>
</tr>
<tr>
<td>President digital</td>
<td>354</td>
<td>242 (68%)</td>
<td>26 (7%)</td>
</tr>
<tr>
<td>President information technology</td>
<td>582</td>
<td>367 (63%)</td>
<td>91 (16%)</td>
</tr>
<tr>
<td>Total</td>
<td>9145</td>
<td>6997 (77%)</td>
<td>3499 (38%)</td>
</tr>
<tr>
<td># publishing organizations</td>
<td>4719</td>
<td>3693 (78%)</td>
<td>1836 (39%)</td>
</tr>
<tr>
<td>Analysis</td>
<td>Exploring the demand for positions with different titles</td>
<td>Topic Modeling to identify data-related responsibilities</td>
<td>Analysis of important responsibilities</td>
</tr>
</tbody>
</table>

Table 4.1 Overview of job ads selected for each analysis step.

4.4.3 Most Important Responsibilities of Data Analytics Leaders

Now that we have gained insights into the range of data-related responsibilities that characterize data analytics leadership, and have identified job ads for “data analytics leaders”, we can explore which (combinations of) data-related responsibilities organizations stress as the most important for these data analytics leaders. In Figure 4.5, we show which combinations of data-related

38 By implication, we excluded job ads that may describe “data analytics leaders”, but that do not highlight the data-related responsibilities.
responsibilities are more frequently discussed in the job ads for data analytics leaders, and which data-related responsibility these job ads highlight the most. Hereby, we assumed that a data-related responsibility is “discussed” in a job ad when the combined weight of the underlying topics is more than \( \frac{1}{\text{total number of topics}} \), where the total number of topics equals 55. The responsibility that is highlighted the most is defined as the one with the highest weight.

A frequently occurring combination of responsibilities, as shown in Figure 4.5, is the combination of “applied analytics” with “data infrastructure” (39%). This might not be surprising considering that those who are responsible for analytics also often need to work with data and have knowledge of (big) data infrastructures. Job ads in this cluster describe, for example, directors of Data Science and Analytics who need to be experts of big data technology platforms, as well as directors of Data Engineering and Technology who are expected to have experience with deploying analytics solutions. Figure 4.5 also shows that it is not unlikely for senior managers to be somehow responsible for (big) data technologies and platforms, analytics solutions, and data governance and management procedures—this combination represents the second-largest cluster of job ads (17%).

Figure 4.5 shows that by far, most data analytics leaders will be responsible for “applied analytics”: 82% of the job ads for data analytics leaders discuss “applied analytics”, and almost 63% of the job ads most dominantly highlight this responsibility. The third-largest cluster (15%) even consists of job ads that primarily discuss this responsibility and none of the other data-related responsibilities. This includes managers who will be responsible for analytics next to being responsible for non-data-related tasks, but also, for example, VPs and directors of Data Science and Analytics whose primary job will be doing things such as acting “as a thought leader in data science and machine learning for the entire company” (Healthcare Software Company) and working on a “long-term vision of data science and the company” (Pharmaceutical Company).
Figure 4.5 shows that “data privacy” is by far the least frequently discussed data-related responsibility. Figure 4.5 indicates that only about 14% of the job ads for “data analytics leaders” discuss “data privacy” at all, and in only about 6% of these ads is “data privacy” the most dominantly highlighted data-related responsibility. Moreover, we already observed that one of the biggest clusters consists of job ads that talk about everything but data privacy. Thus, even though we identified “data privacy” as a data-related responsibility, this responsibility is only marginally represented and discussed in the job ads for “data analytics leaders”. If we look at the job ads that do dominantly highlight data privacy, we see that they are often more traditional directors and presidents of “information security” and “information management”. Other job ads that dominantly highlight data privacy describe very specialized positions, such as a “Director Privacy” and “Chief Privacy and Data Ethics Officer”.

Given that there is still some variation in the titles of job ads for “data analytics leaders”, we may gain complementary insights into our research question by exploring, for data analytics leaders with different positions: 1) how many job ads discuss a responsibility, and 2) how many job ads dominantly highlight a responsibility?

<table>
<thead>
<tr>
<th>Position</th>
<th>For each position, how many job ads discuss a responsibility and for how many ads is a responsibility the most dominantly highlighted?</th>
</tr>
</thead>
<tbody>
<tr>
<td>chief officer data analytics (n = 26)</td>
<td>57.7% 19.2% 69.2% 30.8% 26.9% 7.7% 96.2% 43.2%</td>
</tr>
<tr>
<td>chief officer digital (n = 1)</td>
<td>0.0% 0.0% 100.0% 0.0% 0.0% 0.0% 100.0% 100.0%</td>
</tr>
<tr>
<td>chief officer information technology (n = 79)</td>
<td>48.1% 17.7% 22.8% 10.1% 55.7% 39.2% 60.8% 32.9%</td>
</tr>
<tr>
<td>director data analytics (n = 2186)</td>
<td>87.8% 69.9% 34.7% 12.1% 8.6% 1.6% 76.8% 16.4%</td>
</tr>
<tr>
<td>director digital (n = 114)</td>
<td>78.9% 56.1% 21.1% 2.6% 11.4% 7.0% 70.2% 34.2%</td>
</tr>
<tr>
<td>director information technology (n = 341)</td>
<td>57.5% 34.3% 31.1% 7.9% 39.6% 24.9% 64.5% 32.8%</td>
</tr>
<tr>
<td>president data analytics (n = 635)</td>
<td>86.6% 67.4% 36.4% 13.7% 12.0% 1.9% 70.6% 17.0%</td>
</tr>
<tr>
<td>president digital (n = 26)</td>
<td>73.1% 61.5% 15.4% 3.8% 15.4% 11.5% 65.4% 23.1%</td>
</tr>
<tr>
<td>president information technology (n = 91)</td>
<td>54.9% 26.4% 27.5% 9.9% 34.1% 17.6% 73.6% 46.2%</td>
</tr>
</tbody>
</table>

Figure 4.6 Responsibilities in job ads for “data analytics leaders” with different positions.

Previously, we showed that many of the job ads for senior “digital” positions scored lower than the threshold on data-relatedness. Figure 4.6 suggests that for the group of senior managers “digital” that may be expected to be “data analytics leaders”, the focus is on data infrastructure and applied analytics. These include, for example, directors of Digital Marketing who are expected to apply analytics to optimize their digital marketing strategy; presidents and
directors of Digital Products who may be responsible for commercializing data analytics efforts, and Digital Platform directors and presidents who will build platforms while leveraging data science and machine learning techniques.

We also previously showed that many of the job ads for senior “information technology” positions do not highlight data-related responsibilities sufficiently enough so that they can be considered ads for “data analytics leaders”. Figure 4.6 shows that the focus of those job ads that do highlight data-related responsibilities varies. Data analytics leaders with titles pertaining to “information technology” may be slightly more concerned about “data privacy” than data analytics leaders with other titles. As argued before, job ads that highlight data privacy often describe senior managers of “information management” and “information security”, who are expected to have strong knowledge of privacy laws and are generally responsible for managing and securing information. Senior “information technology” positions responsible for “data infrastructure” include, for example, directors of (Enterprise) IS Architecture that are mostly concerned with data storage and the potential of big data technologies, and generally senior managers of Information Technology that are involved in the development and implementation of data infrastructures. Senior “information technology” positions responsible for “applied analytics” include, for example, directors of Marketing Technology and Supply Chain Technology that will rely on analytics solutions to optimize organizational processes.

Finally, we see that “directors” and “presidents” of “data analytics” are also often held responsible for “applied analytics”. In more than 69% of the job ads for directors and presidents “data analytics”, this is the most highlighted responsibility. Other positions are, for example, directors and presidents of Data Governance who are mainly responsible for data control, and directors and presidents of Data Engineering who will be responsible for building a (big) data infrastructure. As argued before, we also see some positions—albeit relatively very few—that focus specifically on data privacy, such as a “Chief Data Privacy and Ethics Officer”, and directors of Data Privacy, Protection, and Compliance. Interestingly, Figure 4.6 suggests that “chief officer data analytics” positions are more frequently than others responsible for “data control”. We should note that we found only few of these positions. But the fact that those few are frequently responsible for data control might signify that they reside at the corporate level, where they are expected to ensure that data access and use are facilitated across the entire organization (Lee et al., 2014).

4.5 Conclusion and Discussion

The aim of our study was to examine what data analytics leadership entails. To this end, we collected job ads for “chief officers”, “presidents”, and “directors” in the areas of “data analytics”, “digital”, and “information technology”, and ex-
explored what organizations expect from these senior managers when it comes to data analytics. Our analysis of the job ads resulted in three main findings that have implications for how we understand and study leadership in the age of data analytics.

### 4.5.1 Characterizing Leadership in the Context of Data Analytics

By applying a topic modeling algorithm to the contents of the job ads and iterating between the patterns and the qualitative descriptions, we were able to identify four data-related responsibilities: “data infrastructure”, “data control”, “applied analytics”, and “data privacy”. Now, critical readers may argue that (some of) these data-related responsibilities that we found have traditionally been associated with IS leadership, and are therefore not new. In the following paragraphs, we want to discuss the possibility that the data-related responsibilities presented in this study are actually contemporary responsibilities that may uniquely characterize leadership in the age of data analytics.

Regarding “data infrastructure”, one might say that this does not uniquely characterize leadership in the age of data analytics, because data infrastructure is an intrinsic element of “IT infrastructure”, which has long been one of the main responsibilities of senior IS leaders (Chun & Mooney, 2009). Indeed, data infrastructure and IT infrastructure are unequivocally related.

However, we did identify several job ads that highlight data infrastructure as a distinct topic and as a much more integral and important element of the IT infrastructure than traditionally may have been the case. Organizations are increasingly considering how sophisticated data-related technologies will help them access, combine, and share all types of data. We saw that the focus in job ads that score high on “data infrastructure” is not necessarily on how data can best be fed into systems and applications, but rather on how and where organizations may, from a technical perspective, effectively store data such that they are available to and can effectively be used by many different stakeholders.

With regard to “data control”, one might argue that senior IS leadership has traditionally covered the governance and management of IT, and that this includes data governance and data management. Recent studies have acknowledged that IT governance and management and data governance and management are closely related, but also highlight that there are quite substantial differences between them (Tallon et al., 2013-14; Khatri & Brown, 2010). Unlike physical artifacts, data may continuously be used and reused across many different contexts (Lycett, 2013), which poses unique opportunities and challenges to ensure that the right data can be accessed and used at the right

---

39 We also identified numerous topic models in which the topic “big data infrastructure” is merged with the topics “architecture” and “engineering”, which are more general IT infrastructure-related topics. This suggests that they are very closely related.
time (Tallon et al., 2013-14; Khatri & Brown, 2010). Studies have warned that we may “need to exercise caution in extrapolating what we know about physical IT artifacts to governing nonphysical or information artifacts” (Tallon et al., 2013-14, p. 143), which may also apply to responsibilities in the context of data analytics leadership.

The responsibility of “applied analytics” may less obviously be linked to one of the responsibilities that have traditionally been associated with senior IS leadership. Until the start of this century, senior managers have not explicitly been held responsible for the creation of insights, products and services from data. In 2004, Polansky et al. (2004) argued that a “21st Century CIO will be expected to provide actionable business intelligence” (p. 31). Based on our analysis of the job ads, we argue that organizational leaders nowadays are held responsible for more than “business intelligence”. We saw that data analytics leaders are expected to ensure the generation of data-driven insights, products and services through the application of sophisticated big data technologies, machine learning techniques, and in some cases even artificial intelligence. Organizations are actively looking for senior data analytics leaders who can help them gain strategic value from data through applied analytics.

Privacy-related responsibilities have previously been associated with “information security” (Peppard et al., 2011)—a domain that has been considered part of IS leadership since the 2000’s (Chun & Mooney, 2009). In the job ads, we observed that indeed, data privacy is closely related to information security and more generally information management. However, we also identified a number of job ads that treat “data privacy” as a distinct data-related dimension. We saw that organizations are looking for “Chief Data Privacy and Ethics Officers”, and “Privacy Evangelists” to help them differentiate on the basis of privacy propositions. In line with this observation, the literature on data-driven value realization has recently emphasized that organizations should not underestimate privacy concerns in the age of data analytics (Clarke, 2016; Newell, 2015; Zuboff, 2015). In fact, literature has suggested that “data privacy” issues may be considered part of a much wider dimension that covers a range of social implications of data analytics. For example, scholars have argued that organizations should not only consider data privacy issues, but should also consider how data analytics may affect our freedom of choice and our abilities to learn (Clarke, 2016; Newell & Marabelli, 2015; Zuboff, 2015). In line with this literature, we argue that “data privacy” may become part of a new leadership dimension that captures a whole range of social and ethical implications of data analytics.

In sum, like any leader, data analytics leaders may be concerned with allocating resources, dealing with disturbances, negotiating, innovating, planning, and directing employees (Hales, 1986; Fells, 2000). Based on our findings
and reflection, we suggest that data analytics leadership might be uniquely associated with the development of a (big) data infrastructure that facilitates data access and synthesis; the governance of highly complex and dematerialized data artifacts; the creation of insights, products and services from data, and finally, the range of privacy and other social implications of data analytics. We explicated a number of nuanced differences between the responsibilities that we found and the responsibilities that IS leaders have traditionally been held responsible for. Future research is needed to assess to what extent these responsibilities also raise new and unique challenges for senior managers, and how they may effectively deal with these potentially unique challenges in such a way that they can fulfill their important roles. This would require future studies to observe how data analytics leaders enact each of the data-related responsibilities that we identified in practice.

4.5.2 New Data Analytics Leadership Positions

As argued in the theoretical background, there are some controversies regarding which positions may be responsible for providing data analytics leadership. In this study, we examined to what extent organizations expect senior managers with different titles to become responsible for data-related responsibilities. Hereby, we focused our attention on senior managers that were likely candidates according to the literature, that is, senior managers (directors, presidents and chief officers) whose titles pertain to “data analytics”, “digital”, and “information technology” (e.g., Galbraith, 2014; Lee et al., 2014; Tumbas et al., 2017). Our search efforts resulted in a relatively large number of ads for senior managers in the specific area of “data analytics”—many of which also provide very explicit information about data-related responsibilities. This suggests that the job ads genuinely represent formal “data analytics” roles, and that those organizations that publish these job ads are not just signaling to the outside world that they are thinking about this phenomenon by including “data analytics” in the job titles (Nelson et al., 2014). Conversely, we observed that those positions that do not have “data” or “analytics” in the title are much less likely to be held responsible for data analytics.

We may assume that senior manager positions with titles pertaining to “data analytics” are relatively new. Similar to when new IS positions at the level of senior management emerged as organizations began to see strategic opportunities in information technologies (Chun & Mooney, 2009), new senior data analytics positions may now be emerging because organizations take data analytics seriously at the strategic level. Importantly, the fact that organizations are creating new positions to be responsible for data analytics leadership implies that the introduction of data analytics to organizations may do more than simply causing existing organizational structures to “evolve” (Shar-
ma et al., 2014); rather, existing organizational structures may be fundamentally reshaped through the introduction of such new positions. Future studies that focus on understanding how organizations may implement data analytics leadership should thus also consider how organizational structures, processes, and relations change as organizations create new data analytics leadership positions; how different data-related roles may best be structured in organizations, and how organizations may prevent leadership roles from overlapping—as this could have negative financial consequences (Zhan & Mu, 2016). Future research may also observe whether some of these positions might disappear or merge over time.

We should point out that we found only few job ads for “chief officer data analytics” positions, and that this number was also strikingly low compared to the number of job ads for “chief officers information technology”. This might be partly because organizations already have these types of officers in place, or because they tend to search candidates for these positions internally, rather than publishing job ads online (Zhang et al., 2017)—or on Indeed.com. Given that we did find significantly more job ads for “chief officer information technology” positions, it also seems likely that organizations are not (yet) convinced of the need for a dedicated data analytics role that carries the actual “chief officer” title. Future research is needed to validate these potential explanations, and to see if data analytics leadership may indeed primarily be in the hands of those who are called “director” or “president”. It would also be interesting to further explore where these senior managers tend to reside in organizations (e.g., which business units and reporting to whom), how directors and presidents of data analytics can exercise (different forms of) power and authority to provide data analytics leadership, and when organizations may truly need to create “chief officer” positions for data analytics (Lee et al., 2014).

4.5.3 Data Analytics Leadership Focus

After identifying data-related responsibilities and job ads for “data analytics leaders”, we explored which data-related responsibilities these leaders are expected to focus on. We found that “applied analytics” is the most highlighted responsibility, which affirms a trend wherein organizations are increasingly seeking to create value from data through analytics (Gartner, 2017-1). This might also be an indication that organizations are maturing when it comes to data analytics (Davenport et al., 2010); not only are they looking for senior managers who can help them manage or govern data, but they are also looking for leaders who can lead the creation of insights, products, and services from data through applied analytics. We also observed that “data privacy” was by far the least popular responsibility and was only marginally discussed in job ads for “data analytics leaders”.

100
The fact that “data privacy” is marginally discussed in job ads for “data analytics leaders” seems somewhat alarming, given that scholars have increasingly been pointing out the numerous risks for consumers and society when organizations exploit data (Clarke, 2016; Newell & Marabelli, 2015; Markus, 2015). Note that we are by no means claiming that organizations are generally not concerned about privacy implications. As argued before, it might be that organizations consider privacy—and perhaps other ethical issues—as a dimension of information security. It might also be that organizations have dedicated “privacy officers” in place elsewhere in the organization, and that we did not capture these positions through our search queries. Still, we may question to what extent data analytics leaders will be expected to “fully consider the context of their data through the lens of privacy and unique privacy interests” (Everson, 2017, p. 32) and evaluate how “datafication of our everyday lives and the associated algorithmic decision-making […] will affect society” (Newell & Marabelli, 2015, p. 13) when “data privacy” is not explicitly mentioned as a part of their job.

In light of recent debates, our findings call for future research that focuses on how organizations ensure that responsibilities related to the social implications of data analytics are an integral part of “data analytics leadership”, especially when data privacy and related concerns are considered part of another domain. For example, future studies may examine how organizations can ensure that those who are working with the data are also aware of the social implications their solutions might have; and how organizations can ensure that those who are responsible for the potential ethical implications of data analytics are sufficiently aware of all the different data-analytics practices.

4.5.4 Implications for Practice

For practitioners, the findings that we presented in this study provide a starting point for thinking about how data analytics leadership may be structured in their organization. We presented four data-related responsibilities that organizations should consider: “data infrastructure”, “data control”, “applied analytics”, and “data privacy”. Practitioners may learn from our descriptions of the underlying topics and examples from the job descriptions. As organizations allocate these responsibilities, they may consider the creation of new senior manager positions that can become fully dedicated to realizing value from data—this seems to be an emerging trend in our data. We would like to highlight, however, that managers should carefully consider how this impacts existing structures and existing roles, and that not much is known yet about how these people may become effective data analytics leaders. Finally, we urge organizations to think about how responsibilities related to data privacy and other social implications of data analytics are currently allocated within their organizations, and how they incorporate this as an integral part of data analytics leadership.
4.5.5 Boundary Conditions and Additional Directions for Future Research

The study is subject to a number of boundary conditions that provide opportunities for future research. First, the findings in this study are based on an analysis of a cross-sectional sample of job ads that we constructed between May, 29 and August, 21, 2018 as part of an ongoing research effort. As such, our findings provide a snapshot of the responsibilities of data analytics leaders and associated positions. Future research may exploit the longitudinal nature of these data to see how these responsibilities and roles develop over time. Doing so can be especially helpful for gaining insights into whether some of the responsibilities and positions that we identified might disappear or merge over time, or whether additional responsibilities and positions may emerge.

In this study, we collected job ads that were issued by companies that are based in the United States. This was a deliberate choice to minimize the influence of cultural bias. Future research may collect job ads from other parts of the world, to pinpoint similarities and differences between data-related roles and responsibilities of U.S.-based senior managers and those based in other countries. For example, future studies may consider if there are differences regarding the extent to which “data analytics leaders” in other countries are explicitly responsible for data privacy concerns, given that there are differences between regulation in different parts of the world.

Future research is also needed to explore patterns across industries and types of organizations—after all, not all organizations have the same needs when it comes to creating value with data (Lee et al., 2014). Studies may, for instance, capture the demand for “data analytics leaders” for each individual organization, in order to identify patterns related to organizational size and specialization. This would require researchers to identify job ads of many different types of organizations and gather high-quality metadata about these organizations, which they may collect from other job websites and complementary sources.

Future studies may also consider how other senior managers whose titles do not explicitly include any of the terms “data”, “analytics”, “digital”, “information”, and “technology” are involved in providing data analytics leadership. For example, studies may add a range of related keywords (e.g., “business intelligence” or “artificial intelligence”) and use abbreviations (e.g., IT, BI, AI) to capture a larger sample of the total population. Studies may also explore how other positions (e.g., Chief Marketing Officers and Chief Operating Officers) are involved in providing data analytics leadership. As the field of data and analytics touches many aspects of a business, these senior managers might also need to provide some form of “data analytics leadership” and be involved in the implementation of data-driven initiatives (Sharma et al., 2014). We may question
how their leadership roles will change in the age of data analytics.

A key assumption of our study is that the prevalence of data-related topics in job ads is representative of the extent to which positions are ought to be responsible for and involved with realizing value from data. This assumption is based on the notion that job ads reflect what organizations are actually looking for in terms of responsibilities (Todd et al., 1995). In practice this might not always unequivocally be the case. For example, it might be that organizations over- or underplay certain expectations to increase their chances of finding the right candidate. Alternatively, job ads may also simply be poorly written. Given the volume of job ads we collected, it seems reasonable to assume that the patterns that we uncovered are indicative for the current trend with regard to data analytics leadership. Still, future research may consider alternative data sources, such as internal function descriptions, or may perform case studies to gain more detailed insights into why and how some of the patterns that we identified occur.

In conclusion, the current study provides insights into what “data analytics leadership” entails. We presented four data-related responsibilities that characterize leadership in the age of data analytics; examined which positions are typically expected to be responsible for providing data analytics leadership, and explored which data-related responsibilities are highlighted as the most important. We discussed the findings in light of recent debates and provided several avenues for future research. As such, this study provides the groundwork for future theories on data analytics leadership—one of the most important factors for organizations aiming to realize value from (big) data (e.g., Chen et al., 2015; Fitzgerald, 2014; Lee et al., 2014; McAfee & Brynjolfsson, 2012; Gao et al., 2015; Gartner, 2017-1).

ACKNOWLEDGEMENTS

We would like to acknowledge Stijn Heldens for developing the tool that allowed us to scrape the data, and for his insights regarding topic modeling algorithms. We gratefully acknowledge a European postal service organization for providing the financial support for this research. We would like to thank Mostafa Dehganì, Corné de Ruijt, and Ruben van de Geer for their suggestions. Finally, we would like to thank the reviewers of AOM 2019 for their feedback.