

HOW BIG DATA BECAME SO BIG

*A media-archaeological study of transhistorical
ideas about technology & the rising
popularity of Big Data*



Marieke Pots

MA Thesis New Media & Digital Culture
Utrecht University

HOW BIG DATA BECAME SO BIG

*A media-archaeological study of
transhistorical ideas about technology
& the rising popularity of Big Data*

Utrecht University | Faculty of Humanities
MA Thesis New Media & Digital Culture
May 2014

Marieke Pots (3474402)
mariekepots@gmail.com

Abstract

Suddenly, a Big Data revolution seems to be on its way. In both technological and popular discourses, Big Data is often depicted as a new technological development that will undoubtedly bring new and profound insights, answers and knowledge. However, the promises of this rather vague term are highly debatable and have yet to be proven. In this thesis I will therefore inquire how it can be explained that Big Data has given rise to very high expectations and a lot of optimism in a rather short period of time. By building on the work of various media historians, this thesis departs from the proposition that this popular term currently relates to certain age-old and transhistorical ideas, desires and dreams about technology that have been part of our collective and subconscious memory long before the term 'Big Data' was even coined. By investigating both the current optimistic discourses of Big Data, as well as historical discourses that surrounded past media manifestations, it will be argued that Big Data became so big because its discursive construction confronts us with our age-old thoughts, dreams and expectations about technology, and thus constantly embeds Big Data in a discursive mould we already know.

Keywords

Big Data, discursive constructions, media archaeology, technological optimism, transhistorical discourses, discursive moulds

Supervisor: dr. Imar de Vries
Second reader: dr. René Glas



Universiteit Utrecht

Table of contents

Introduction	1
The 'Internet' and the role of popular terms	1
Big Data as emerging buzzword	3
The curious case of Big Data	4
Analysing and explaining today's Big Data discourse	6
1. Big Data, big promises and big expectations	10
1.1. The advent of new and better technologies	10
1.2. Promoting and presenting the possibilities of Big Data	15
1.3. Identifying assumptions and raising questions	23
2. A new vision on the history and development of media	26
2.1. Arguing against a linear approach to media history	26
2.2. Making sense of media archaeology	29
2.3. The rhizome: placing Big Data in a historical framework	33
3. Big Data & deeply ingrained ideas about technology	38
3.1. Technology and the age-old human quest for knowledge	38
3.2. A recurring sense of information overload	42
3.3. The emerging ideal of mechanical objectivity	45
3.4. Today's popularity of Big Data: an explanation and evaluation	48
Conclusion	52
References	57

Introduction

Something odd is happening to our language. Old, trusted words no longer mean what they used to mean; often, they don't mean anything at all. Our language, much like everything these days, has been hacked. Fuzzy, contentious, and complex ideas have been stripped of their subversive connotations and replaced by cleaner, shinier, and emptier alternatives. (Morozov 2013a)

The 'Internet' and the role of popular terms

A lot can happen in just one minute on the Internet. In one random minute in 2012, there were more than two million search queries entered into Google, more than 30 hours of video were uploaded to YouTube, more than 200 million emails were sent and Facebook was viewed more than six million times (Intel 2012). These numbers show how ubiquitous the Internet has become. Moreover, these numbers have undoubtedly gone up at this time of writing. This is not only because present Internet users spend more and more time online (e.g. Lardinois 2012), it is also because more people are still being connected to this worldwide network (e.g. ITU 2012).

But, what then, is the 'Internet' exactly? This question is more complex than it might seem at first. Especially when one starts to look beyond the technical infrastructure of the Internet, it becomes clear that our collective understanding of the 'Internet' is neither unambiguous nor fixed. Rather, our dominant understandings and visions of the Internet, as well as our public debates about it, change over time, and may, at times, even bear very little resemblance to the actual physical infrastructure of the Internet (e.g. Rogers 2012; Morozov 2013b, 14).

The French sociologist Patrice Flichy (2007) has exactly studied the emergence of such dominant understandings and visions of the Internet. In his book *The Internet Imaginaire*, which was originally published in French in 2001, Flichy examines the social imagination and the various collective visions that surrounded the advent of the Internet. He writes in detail how the emergence and shaping of a number of collective visions was often closely related to the terms that were used to talk about the Internet. For instance, the term 'information superhighway' gave rise to an entirely different collective vision of

the Internet than terms such as the 'online community' or 'virtual reality' (ibid.). This implies that our collective understanding of what the 'Internet' is, should or could be, changes over time and correlates with the dominant terms and phrases that are used to talk about it.

Media scholar Marianne van den Boomen (2014) also refers to the term 'information superhighway', as well as to more recent terms such as 'web 2.0' and 'the cloud', and argues similarly that these terms play a big role in the way digital technologies are understood. By emphasising the metaphorical workings of these terms, she claims that we can only articulate, perceive and conceive the digital when it is translated into metaphors (ibid., 12). By building on the work of Lakoff and Johnson (1980), she argues that metaphors do not merely transfer meaning from a source domain onto a target domain, because metaphors also purposely deprecate certain aspects (Van den Boomen 2014, 45). This is why metaphors are never just metaphorical; they generate meaning (ibid., 187). A great example can be found in current understandings of the popular term 'the cloud'. This metaphorical term proves to be so strong that as much as 51 percent of the American people think that bad weather actually affects cloud computing (Yarow 2012). This shows that popular terms, with their often metaphorical workings, affect the way we make sense of technology.

Additionally, similar arguments can be found in scholarly texts about the way new and old media have been discussed in society (e.g. Spigel 1992; Marvin 1988; Douglas 1989). In these discursive, social and cultural studies, it is often argued that our understanding of new media is not solely determined by the actual technological object, but that it is highly dependent on the words we use to talk about it (e.g. Marvin 1988, 8; Lister et al. 2009, 68-69). This approach to media and technology therefore confirms that popular terms play a constructive role in the way a certain understanding and collective vision about technology may emerge.

But why is the role of popular terms and the emergence of certain collective visions worth studying? Van den Boomen (2014) and Flichy (2007) give an indisputable answer. Van den Boomen concludes that popular terms, as discourse metaphors, are "neither good nor bad, but least of all are they neutral" (Van den Boomen 2014, 187). That is because they "can be mobilized, elaborated, reformulated, and criticized in order to achieve particular goals" (ibid., 79). Flichy states in a similar way that the role of these emerging collective visions should not be overlooked because they are capable of mobilising "both designers and users" (Flichy 2007, 6). They are therefore an

“integral part of the development of a technical system” (ibid., 2). This means that there is a non-neutral discursive interplay between the terms we use to talk about technology, the way we subsequently understand a technology, and the way we consequently choose to act upon this. This is thus exactly why popular terms and their emerging collective visions need to be studied.

Big Data as emerging buzzword

After popular terms such as the ‘information superhighway’, ‘virtual reality’, ‘web 2.0’, and many more, it is clear that yet another technological term has conquered a dominant place in our contemporary public debates about technology: Big Data. This term has gotten so much attention and has given rise to such high expectations that there is even talk of a “Big Data revolution” (Mayer-Schönberger & Cukier 2013), a “Big Data era” (Zikopoulos et al. 2012, 4; boyd & Crawford 2012, 663) and an “age of Big Data” (Lohr 2012).

The American research firm Gartner (2013) confirms the enormous popularity of the term by placing it next to gamification, consumer 3D printing and wearable technology at the so-called “peak of inflated expectations” on their 2013 ‘Hype Cycle for Emerging Technologies’. During this phase, new technologies are received with high expectations, a lot of interest and much enthusiasm. Another American research firm, IDG, confirms the popularity of Big Data as well. Their survey found that the average enterprise will spend no less than eight million dollars on Big Data initiatives within the next year and that as much as seventy percent of the surveyed enterprises have either employed or are planning to deploy Big Data initiatives in the near future (Columbus 2014).

The interest for Big Data on the World Wide Web also reflects the vast popularity of the term. At this point of writing, Google’s search results show more than 150 million websites, that have been updated in the past year, with the exact phrase ‘Big Data’ on it (Google 2014a). For comparison, the search results show about 80 million websites for 3D printing, about 7 million websites about gamification, and no more than 2 million websites for the phrase ‘wearable technology’ (ibid.). This means that there are currently more recently updated websites about Big Data than about any of these other popular terms.

Additionally, Google Trends shows that the search volume for Big Data has grown rapidly since the end of 2011 (Google 2014b). During the year 2012, the amount of searches for Big Data tripled, after which this number subsequently doubled in 2013 (Google 2014c). At this point of writing, the

average amount of search queries for Big Data has even multiplied by eight since the end of 2011 (ibid.). This shows that Big Data has become a popular technological term in just a few years time.

With Flichy's (2007) and Van den Boomen's (2014) arguments in mind, it needs to be pointed out that the role of this term should not be overlooked, because it has, just like many other popular terms, become part of the discursive interplay between the way we talk about technology, the way we understand the possibilities of our current technologies and the way we choose to act upon this.

The curious case of Big Data

Despite the high expectations and despite the rapid increase in popularity, the exact meaning of the term remains ambiguous. In popular discourses as well as in technological and scientific fields, we find various descriptions of Big Data. For example, the New York Times refers metaphorically to "a data flood" (Lohr 2012) with "not just more streams of data, but entirely new ones" (ibid.). Here, Big Data is defined as a "shorthand for advancing trends in technology that open the door to a new approach to understanding the world and making decisions" (ibid.). In *Big Data: A Revolution That Will Transform How We Live, Work and Think*, Mayer-Schönberger and Cukier speak of "astronomical quantities" (Mayer-Schönberger & Cukier 2013, 7) of data and claim that Big Data refers to "things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value" (ibid., 6). In technological fields we find more straightforward definitions. IBM claims that the amounts of data are "growing at an explosive rate" (Zikopolous et al. 2012, 129) and refers to the so-called 3V's that distinguish Big Data from just data: volume, variety, and velocity (ibid., 5). The American company O'Reilly Media defines Big Data less in terms of types of data and more in terms of the capacity to analyse this, by claiming that Big Data is "data that exceeds the processing capacity of conventional database systems" (Dumbill 2011). Media scholar Lev Manovich defines Big Data similarly to O'Reilly Media but adds a dimension of time by stating that Big Data is about "data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time" (Manovich 2011). However, media scholars danah boyd and Kate Crawford do not emphasise the size of the data sets and argue that Big Data "is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets" (boyd & Crawford 2012, 663).

Considering these different interpretations and definitions of Big Data, it is striking that this rather vague term has experienced such a rapid increase in popularity. This becomes even more notable if we consider the fact that many promises and expectations of Big Data have yet to be proven and should not be taken for granted. For instance, boyd and Crawford (2012) have critically interrogated today's assumptions and biases by arguing that Big Data tools "have their own inbuilt limitations and restrictions" (ibid., 666), that Big Data is not the same as "whole data" (ibid., 669) and that Big Data enables the practice of seeing patterns where none actually exist (ibid., 668).

The ultimate question I therefore wish to answer in this thesis is: how can it be explained that the rather vague term 'Big Data' has, in a considerable short period of time, gotten so much attention and given rise to such high expectations that there is even talk of a Big Data revolution and era? My proposition is that it is neither *just* a coincidence, nor *just* a well-chosen term, nor *just* a very smart marketing trick that explains the vast popularity of the term; it is rather the way this term, at this current moment in time, relates to certain age-old and transhistorical ideas, desires and expectations about technology that have been part of our collective and subconscious memory long before the term 'Big Data' was even coined.

This proposition may seem daring, but recent scholarly attention proves otherwise. Scholars in diverse fields, from media studies to history, have expressed a sense of *déjà vu* while studying both our present-day responses to new media as well as our previous responses to media that were once new (e.g. Gunning 1991; Marvin 1988; Boddy 1994; Huhtamo 1996). By studying the way various old and new media have been received and talked of, or rather, by studying their discursive constructions, these scholars point to striking similarities, in spite of the very different historical, social and cultural contexts that these media emerged in. By building upon these observations, it is, especially in the field of media archaeology, argued that age-old, recurring and transhistorical ideas about technology provide a prefabricated discursive mould for today's responses, and thus still affect the way we think and talk about new media today (Huhtamo 1996).

In this thesis I therefore aim to study today's popularity of Big Data by investigating the way age-old and transhistorical ideas about technology have paved the way for an optimistic and expectant discourse to emerge about Big Data. With this, I firstly intend to elaborate on boyd & Crawford's (2012) proposal to start questioning the assumptions, values and biases of Big Data.

Secondly, I aim to follow as well as extend Flichy's study. Since Flichy's book was published in 2001, this thesis can be seen as an updated study that explores whether a new collective vision has now emerged for Big Data. However, this thesis is not only an updated study, it is also an extended study. This is because I will not merely describe the emergence of a collective vision, I will also explore a much deeper historical context, because I will argue that it is necessary to study preceding and transhistorical ideas about technology in order to understand the contemporary emergence of a collective vision more fully. My ultimate goal is to provide a broader understanding of today's popularity of Big Data by addressing the entanglement between our current ideas, assumptions and biases of Big Data and past ideas about technology.

Analysing and explaining today's Big Data discourse

To address the main question of this thesis, one important methodological challenge has to be pointed out. That is, this thesis cannot be fully answered by conducting one single research method because, first, this thesis evolves around the expectant discourses of Big Data, and second, this thesis addresses historical discourses about technology in order to get a broader perspective on today's ideas and expectations. This two-stage approach means that it is necessary to conduct both a discourse analysis and a media-archaeological study.

Considering the scope of both research methods, I have chosen not to extensively describe them here in the introduction. Rather, the exact methodological approaches for both research methods will be delineated throughout this thesis. However, at this point I do wish to discuss why I have chosen for these two methodologies, what scholarly (methodological) texts I will refer to, and anticipate two possible comments about combining a discourse analysis with a media-archaeological study.

I have chosen to conduct a discourse analysis because this allows me to examine and describe today's dominant vision of Big Data. I will depart from the work of Michel Foucault, as he is one of the most influential thinkers about how discourses constitute social reality (Deacon et al. 1999, 151). Subsequently, it will be argued that the notion of ideological analysis is rather useful, because such a textual analysis is exactly aimed at revealing the "dominant ideas of an individual, group, class or society" (Brennen 2013, 343). The media-archaeological approach that follows enables me to put today's discourses of Big Data in a broader historical perspective and provides the possibility to address

the role of transhistorical ideas about technology. It is here that I will refer to the work of the influential media archaeologists Erkki Huhtamo (1996; 2011), Siegfried Zielinski (2006) and Jussi Parikka (2011), in order to describe what media archaeology is, and how I aim to conduct a media-archaeological study.

At this point one may wonder why I have presented these two research methods as two entirely different methodologies, because a media-archaeological study is often not that different from a discourse analysis. As a matter of fact, in order to reveal and study discourses that surrounded past technologies, a media archaeologist often conducts a discourse analysis. It is here that I wish to emphasise that my media-archaeological study deviates slightly from most media-archaeological approaches, because it will not only consist of a search for and analysis of primary historical discourses, rather, it will also consist of a search for historical discourses in already existing scholarly work. This means that my approach aims at collecting primary as well as secondary sources and is therefore different from a discourse analysis. This is exactly why I have chosen to distinguish between the two methodologies.

Moreover, this two-fold methodological approach might present itself as somewhat unusual, however, the opposite is true. It is argued that qualitative researchers often tend to use “a variety of different methodologies in their work” (Brennen 2013, 18) in order to “increase the rigor of their analyses and to develop in-depth understandings of social experience” (ibid.). And indeed, developing an in-depth understanding of today’s responses to Big Data is the main reason why I have chosen to combine a discourse analysis with a media-archaeological study. However, before proceeding, I wish to clarify that using these two research methods means that this thesis will not address the practical new technological possibilities of Big Data nor the practical implications of working with Big Data. It will also not enlarge on recent discussions about privacy and associations with ‘big brother’. But, most importantly, combining a discourse analysis with a media-archaeological approach does give me the possibility to place today’s popularity of Big Data in a broader theoretical, comparative, historical, social and cultural framework and address the main question of how Big Data became so big.

In order to do so, I have divided this thesis into three chapters. The first chapter is aimed at studying today’s discourses about Big Data. In this chapter I will describe the methodological guidelines for the discourse analysis, as well as conduct it. The second chapter is dedicated to the methodological question of how to approach and explain today’s discourses from a media-archaeological

point of view. Conducting this media-archaeological study, and proving my proposition, will subsequently be the focus of the third chapter. A more detailed outline for each chapter is as follows.

The first chapter evolves around a discourse analysis aimed at revealing the dominant and optimistic expectations, ideas and ideals about Big Data. To achieve this, I will first place the advent of Big Data in a broader theoretical field concerned with the way previous media have been received. This part serves as a theoretical background of how to analyse the discursive responses to new media and technology. I will show that the rise of each new medium comes with optimistic (and anxious) responses, after which I will describe that today's optimistic responses to Big Data can be analysed with a discourse analysis. After describing the methodological guidelines, I will conduct this discourse analysis and show that Big Data is currently thought of as a new technological development that will bring new and profound knowledge because it is now possible to analyse large amounts of data in an objective and accurate way. At the end of this first chapter, a reflection and overview of this discourse analysis will be given.

In the second chapter, I will address the methodological question of how to explain today's expectations of Big Data. In the first part of this chapter, it will be argued that a non-traditional and non-linear view on the history and development of media can place today's Big Data discourse in a broader historical framework. In the second part, I will subsequently argue that a media-archaeological approach, with its non-linear view on the history of media, is necessary. What follows is an overview of three slightly different media-archaeological approaches described by three influential authors in this field: Erkki Huhtamo (1996; 2011), Siegfried Zielinski (2006) and Jussi Parikka (2011). I will argue that especially Huhtamo's and Zielinski's approaches point to a substantial theoretical explanation for why we react to new media the way we do. That is because age-old ideas about technology, that appear, disappear, reappear, connect and collide in history serve as prefabricated discursive moulds for today's reaction to new media. However, I will also argue that additional methodological guidelines are needed for conducting a media-archaeological study that is exactly aimed at revealing such age-old ideas about technology. I will therefore introduce the concept of the "rhizome" (Deleuze & Guattari 1987) as an alternate image of thought that is capable of providing the necessary additional guidelines.

The third chapter will be dedicated to a historical and media-archaeological quest with which I aim to explain today's popularity of Big Data. In this chapter, I will first show what age-old ideas about technology relate to the discourses that surround Big Data today. It will be demonstrated that today's Big Data discourse is everything but new, because the advent of the printing press, the telescope, the microscope, the calculator, the camera obscura, the camera lucida and the first computer gave rise to very similar discourses with very similar ideas, desires and dreams about technology. In the last part of this chapter I will argue that the popularity of Big Data can be explained with these recurring and transhistorical discourses, because this term is, at this current moment in time, capable of seamlessly reactivating and connecting exactly these past discourses in our subconscious and collective minds. Moreover, with knowledge of these age-old discourses, it is not only possible to explain today's popularity of the term, it is also possible to inquire today's expectations in novel ways.

1.

Big Data, big promises and big expectations

The great American Inventor and his associates were not only making electricity, they were also making meanings. (Flichy 2007, 6)

In the introduction I have argued that today's Big Data discourse needs to be studied because popular technological terms and phrases play a constructive role in the way collective visions and dominant understandings about technology emerge. This chapter is therefore dedicated to a discourse analysis aimed at revealing the collective vision that has emerged around Big Data. In order to do so, I will first place the rise of Big Data in a broader theoretical field concerned with the way previous media have been received. This serves as a theoretical background for describing my methodological approach to today's responses to Big Data. I will subsequently conduct a discourse analysis and examine the dominant ideas about the possibilities of Big Data. At the end of this first chapter, an overview of today's Big Data discourse will be given.

1.1. The advent of new and better technologies

The way that new and old media have been discussed in society is a research topic that has received increasing attention among media scholars in the last few decades. These scholars have in some cases addressed the technical and practical possibilities of media, but have predominantly studied the way new and old media have been received and talked of. A striking finding is that many of these scholars have identified a recurring discursive pattern in the way new media are discussed. This pattern consists of a "recurring sense of optimism and anxiety that each wave of new media calls up" (Lister et al. 2009, 67). On the one hand, there is thus always an expectant and utopian representation of new

technologies, while, on the other hand, there is always a dystopian idea about the negative effects of new technologies.

Focussing on the optimistic side of this pattern, proof of the existence of such responses is not hard to find. For example, with the advent of the radio, many utopian thoughts were expressed about the capability of the radio to make a better future by democratising and improving communication (Douglas 1986). And even though the radio had acquired more characteristics of a mass medium after a while, optimistic futures were still envisioned. For example, Bertolt Brecht suggested in 1932 that the radio was “one-sided when it should be two” (Brecht 1932). By changing the apparatus of the radio from “distribution to communication” (ibid.), “the radio would be the finest possible communication apparatus in public life” (ibid.). Similar optimistic reactions can also be identified around the rise of the Internet. For example, in 1993, Time Magazine presented an optimistic future on the cover by stating that the info highway will bring a “revolution in entertainment, news and communication” (Time Magazine 1993).

Media scholars who have extensively written about the way new media are received, have tried to elucidate this recurring optimistic sense around new technologies. It is argued that our digital media landscape is associated with a “progressivist and utopian spirit” (Robins 1996, 13) in which there is a “renewed confidence in technological solutions to the problems of human culture and existence” (ibid.). In the same light, Lee Worth Bailey speaks of the enchantments of technology, because there is a major “assumption that modern technology has conquered most barriers and is an unstoppable, victorious, utopian historical force” (Bailey 2005, 104). Additionally, media scholar Imar de Vries argues that these recurring optimistic, progressivist and utopian responses are interwoven with our associations of ‘the new’:

It is what ‘the new’ does: it perpetually gives our technological imaginary, our yearning for wholeness and completeness that is projected upon technology, fresh impulses by portraying existing technologies as inadequate, and, in the same sweep, by introducing us to the next big thing as a solution. (De Vries 2012, 165-166)

Moreover, this quotation points to the technological imaginary, a concept that is also concerned with the optimistic responses to new media. Even though there is no indisputable and single definition and approach to the technological imaginary, this concept is often related to the psycho-analytic theories of Jacques

Lacan. Here, the imaginary is one of the three orders of experience, where it is placed alongside the symbolic and the real (Macey 1994, xxi). In this psycho-analytic theory, the imaginary refers “to a realm of images, representations, ideas and intuitions of fulfilment, of wholeness and completeness that human beings, in their fragmented and incomplete selves, desire to become” (Lister et al. 2009, 67). Even though the imaginary is often synonymous with something fictional, illusory or unreal, in a Lacanian or psycho-analytic sense, this is not simply the case because “imaginary identifications can have very real effects” (Macey 1994, xxi). Applied to technology, the technological imaginary then “draws attention to the way that dissatisfactions with social reality and desires for a better society are projected onto technologies as capable of delivering a potential realm of completeness” (Lister et al. 2009, 67). The concept of the technological imaginary thus reveals that the recurring optimistic responses to new media are given in by a primary human desire for fulfilment and completeness.

In sum, these accounts confirm that optimistic and utopian reactions do occur around new media. Moreover, they demonstrate that these reactions are inherent to a prevailing belief, desire and assumption wherein new media are always perceived as technologies that will undoubtedly create and provide for a better future.

Very similar beliefs, desires and assumptions can also be identified around popular terms, phrases and metaphors. Patrice Flichy (2007) shows that the term ‘information superhighway’ led to various utopian ideas in the industrial world. In this collective vision, the information superhighway would “revolutionize the way we relate with one another by enabling to work together, collaborate and access and generate information without regard to geographical boundaries” (ibid., 24). This shows that the same belief in technological progress can be identified in this optimistic response to the information superhighway.

Mirko Tobias Schäfer (2011) confirms this observation by building upon the work of Jan van Dijk (2006) in which popular discourses and utopian notions around terms as ‘teledemocracy’, ‘virtual communities’, ‘the new economy’ and ‘Web 2.0’ are described. Schäfer concludes that these terms “are part of a rhetoric of progress that can be recognized in the representations of new media in popular discourse” (Schäfer 2011, 25). Additionally, Schäfer argues that these terms and metaphors often refer to a “culturally constituted imagination of technological progress” (ibid.) and are often “familiar and thus comprehensible for audiences and easily employable for promoters” (ibid.). Schäfer thus argues that popular terms, phrases and metaphors catch on exactly because they

somehow respond to our already existing and familiar ideas about the future of technology; something I will further explore in the second and third chapter of this thesis.

The most important sub conclusion is that the advent of each new medium and each new popular term gives, next to an anxious response, rise to an optimistic and expectant discourse which is closely related to a prevailing belief in the progressivist nature of technology. The goal of this chapter is to study the optimistic discourse that has undoubtedly emerged with the rise of the popular term 'Big Data'.

In order to do so, I have chosen to conduct a discourse analysis that focusses on the way dominant optimistic, expectant, idealistic and perhaps even utopian ideas about Big Data have emerged. Inspired by the work of Michel Foucault, this discourse analysis departs from the acknowledgement that "discourse is at once controlled, selected, organised and redistributed" (Foucault 1981, 52) and that "discourse is not simply that which translates struggles or systems of domination, but is the thing for which and by which there is struggle, discourse is the power to be seized" (ibid., 52-53). This means that discourses do not merely describe the world around us, but also construct, reinforce or alter power relations through constant struggles over socially constructed meanings. This implies that today's optimistic responses to Big Data do not necessarily represent the 'truth' about this new technological development, rather, the various actors involved in today's discourses actively construct, alter or reinforce its dominant meaning, and with that, construct, alter or reinforce existing power relations.

To be able to study the way a dominant meaning has been socially and linguistically constructed for Big Data, the notion of ideological analysis is very useful. Such a discourse analysis focusses on "the dominant ideas of an individual, group, class or society, the way meanings are socially produced, or even as the false ideas upon which a social, political or economic system is based" (Brennen 2013, 331). The key aspect of an ideological analysis is that it suggests that many dominant ideas often appear to us as "common sense" (ibid., 332), or as "universal" (Fairclough 1989, 33), because they have become "naturalized" (ibid.) through discourse. Similar to a Foucauldian perspective, this analysis acknowledges that power is exercised through discourse, because it is argued that such common sense and universal ideas often "embody assumptions which directly or indirectly legitimize existing power relations" (ibid.).

In agreement with these arguments, I will therefore study today's dominant and optimistic ideas about Big Data by paying particular attention to the way certain ideas are presented as common sense. To do so, I will follow Mike Cormack's (1995) methodological outline for conducting an ideological analysis. With this approach, one is capable of uncovering such common sense ideas with a particular focus on the content of a text, "including the opinions, beliefs, values and other judgments, the vocabulary used, the stereotypes and characterizations of people, and the conflicts, resolutions and other actions within the text" (Brennen 2013, 335). Moreover, the structure (the way an issue is framed in the opening and closing parts of a text), its mode of address, its style, as well as the aspects that one might have expected to be in the text, but that are actually missing, help us understand how a specific social reality is constructed through discourse (ibid.). Additionally, it is important to note that, during this analysis, texts are considered to be "cultural artifacts" (ibid., 316), "things that we use to make meaning from" (ibid.) and especially "traces of socially constructed reality" (ibid.).

But where to find and select these 'textual traces'? It is here that Flichy's (2007) study, that was similarly aimed at revealing dominant thoughts about technology, serves as a valuable example. Flichy selected a number of texts from two main origins. The first origin included documents written by academics and computer scientists, since they were, at that time, "the founding fathers of information highways, the Internet, and virtual reality" (ibid., 12). The second origin included press articles, which consisted primarily of articles written in *Wired Magazine*, the main U.S. magazine for reflection and debate on technology, but also covered articles in popular U.S. news magazines to see how ideas about the Internet "diffused from specialists toward the public at large" (ibid., 13). In short, this shows that Flichy studied the emergence of a collective vision by analysing how the founding fathers promoted the Internet, how technological fields responded to this in *Wired Magazine*, and how this new technology was depicted in popular news magazines.

Even though Flichy does not explicitly explain why he covered these three levels of discourse, I wish to follow this approach because it provides the possibility to gain insight into the way certain seemingly universal and optimistic ideas and assumptions about Big Data may have conveyed from its founding fathers, to technological fields, and to, finally, emerge in popular discourses. A more detailed overview of the way I will study these three levels of discourse for today's Big Data is as follows.

The first level is aimed at revealing how Big Data has been constructed by its founding fathers. It needs to be pointed out that in Flichy's study, the founding fathers were mostly academics and scientists that envisioned the future of the Internet first, but, in the case of Big Data, the founding fathers are rather big enterprises in the IT-industry that have been envisioning and promoting this term first. Therefore, I have chosen to pay particular attention to texts written by IBM, the first big IT-corporation that started specialising in Big Data. By analysing their book *Understanding Big Data* (Zikopoulos et al. 2012), as well as articles and press releases on their website about the possibilities of Big Data, I will describe how this company is not only employing Big Data initiatives, but also making meaning. At the second level, I will, just like Flichy, analyse how Wired Magazine has written about Big Data. By examining the online articles in Wired, I will describe how optimistic and seemingly common sense ideas about Big Data have been discursively constructed. The third level focusses on how Big Data has reached a mass public. It is here that I will study how the biggest popular U.S. newspaper, The New York Times, has written about the popular term. Here, I will, just like I did for Wired, pay particular attention to the way certain optimistic ideas about Big Data may appear as common sense.

1.2. Promoting and presenting the possibilities of Big Data

In the previous part I have described that the advent of each new technology, as well as each new technological term, gives rise to optimistic and anxious responses. By focussing on the optimistic side of this discursive pattern, I have pointed out that this optimism is often part of a utopian and progressivist view on the future of new technologies. Moreover, I have described that today's optimistic responses to Big Data can be studied with an ideological discourse analysis. In this part I will conduct this analysis, and examine how the IT-industry, Wired Magazine and The New York Times have played a part in today's optimistic discursive construction of Big Data.

1.2.1. The IT-industry: promoting and presenting the possibilities of Big Data

Around 2008, IT-company IBM was one of the first companies to adopt Big Data into their business activities. Up to today, it is still one of the leading companies working with Big Data. It is therefore an excellent example to describe how IT-companies market and present their beliefs in the possibilities of Big Data. IBM's

recently published book and their online articles about Big Data provide an illustrative image of the way this term has given rise to very optimistic ideas, as well as how the meaning of this term can be, and has been, constructed through discourse.

In 2012, IBM published a comprehensible and free e-book, *Understanding Big Data* (Zikopoulos et al. 2012), about their business and technological perspectives on Big Data. This book proves to be an excellent example to show how IBM is part of today's discursive construction of Big Data, and how IBM is capable of shaping its meaning through discourse. By examining the structure of the text, it becomes clear that the title and the introduction play a significant role in the way Big Data is framed. Its title seems to refer to Marshall McLuhan's classic book *Understanding Media: The Extensions of Man* (1964) in which a rather deterministic notion about media is expressed. In this book, McLuhan suggests that the characteristics of a medium affect and reshape a society (ibid.). The title of IBM's book therefore seems to imply that Big Data is going to affect and shape our current society in a revolutionary way, and by doing so, IBM directly emphasises the importance of the term. The introduction of the book confirms this. It is here that the authors express their very optimistic belief that the "coming of the Big Data Era" (Zikopoulos et al. 2012, xv) will bring a "chance for everyone in the technology world [...] as this era will bring the biggest opportunity for companies and individuals in technology since the dawn of the Internet" (ibid.).

Through assessing the content of the text, more beliefs, opinions and resolutions can be identified in the arguments that are used to support their claim that a new technological revolution is on its way. For instance, the authors argue that the world has already changed and is still changing because technologies are able to sense, collect and store more data (ibid., 4). Moreover, they point to impressive numbers of the current growth of data and devices, and state that "we're drowning in data" (ibid., 6). With these technological developments, the authors describe that it is now, with Big Data, possible to discover patterns and solve problems that could not be solved before (ibid., 17).

With Foucault's, Fairclough's and Brennen's arguments in mind, it needs to be emphasised that IBM does not merely describe the possibilities of Big Data in this book, rather, they actively construct its meaning. Moreover, it can be argued that they do so through making it appear as common sense that a new technological revolution has arrived and that the great amounts of digital data will bring chances for everyone in the technology world.

This same framing technique, as well as the exact same optimistic belief in the revolutionary impact of Big Data, can also be identified in most of IBM's online articles. The most exemplary article is most certainly 'From Big Bang to Big Data' (IBM 2014a). In this article, IBM describes their project called the 'Square Kilometre Array' (SKA) with which a large number of radio telescopes are being built in South Africa to collect massive amounts of raw data from space. It is especially the style of this article that reveals the ideological dimensions. The article included a large figure in which the incredibly large amounts of data are emphasised. According to this figure, the telescopes generate, on a daily basis, no less than twice the daily amount of data that is generated on the Internet (ibid.). Moreover, by assessing the content of this specific article, it becomes clear that great beliefs are expressed as common sense. For example, it is presented as highly logical that this Big Data project will "unlock some of the mysteries of the Universe, including how galaxies were formed, how they evolve and the ultimate science fiction question, are we alone?" (IBM, 2014a).

This seemingly universal idea that Big Data will lead to more insights and knowledge can continually be recognised in IBM's online articles. For instance, based on past and present data of criminal activities, IBM claims that it assists police officers with predicting crime and improving public safety. In a press release on their website they present it as common sense that "[t]echnology has proven to be a force multiplier that is helping solve crimes more quickly or to prevent them all together, and improve the way citizens are being served and resources are allocated" (IBM 2012). Another distinctive example is the way IBM presents the possibilities of Big Data for the health care industry. Here, IBM aims to predict heart attacks years before they strike. By analysing patterns in electronic health records, such as demographics, medical history, medications, allergies and more, it is assumed that common signals and patterns can be discovered and a future heart attack be recognised (IBM 2013). Moreover, IBM claims in a similar way that Big Data can bring a twenty percent decrease in patient mortality by analysing patient data (IBM 2014b).

At this point it can be argued that it is remarkable that IBM constantly presents these possibilities as common sense, because certain aspects, that one might have expected to be in these articles, are often missing. For instance, it is rarely explained what the term 'Big Data' exactly means. Moreover, questions about the actual existence of such predictive patterns, as well as case studies, results, or examples that proof that such patterns can be found, are often missing. For example, IBM does not give any indication of how many crimes can be, or

have been predicted, nor do they enlarge on the question of how likely it is that the occurrence of future crimes can actually be captured in a pattern, just like they do not provide any in-depth information about how collecting an extremely large amount of raw data from space can lead to answers about the history of the universe and alien life. By taking these missing aspects into account, as well as the way Big Data is continually presented as a highly logical technological revolution, it can be confirmed that IBM is constructing a very optimistic social reality around Big Data through discourse.

This can also be recognised at other major IT-companies. For example, the multinational computer corporation Oracle claims that “Big Data is the electricity of the 21st century—a new kind of power that transforms everything” (Oracle 2014). Moreover, they note that “[w]hile electricity took more than 100 years to transform the world, big data is revolutionizing the way businesses and government operate virtually overnight” (ibid.). Hewlett-Packard (HP), another American multinational technology corporation, describes that “Big Data will build insight, intuition and ideas - and better enterprise decisions” (HP 2014). McKinsey, a consulting firm that also specialises in Big Data, presents similar high expectations of Big Data by noting that it “will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus” (Manyika et al. 2011).

To recapitulate, it has become clear that IBM actively constructs, rather than describes, a very optimistic social reality for Big Data, and with this, they also construct, alter or legitimise existing power relations. They do so by making it appear as common sense that the current amounts of digital data will bring solutions and answers for everyone in the technology world. Even though the possibilities of Big Data (e.g. predicting heart attacks, finding answers to our existence in the universe, predicting crime) can often be considered to be almost heroic, and appear even utopian, they are presented and promoted as if it is common sense that they can now be realised.

1.2.2. Wired: promoting Big Data in technological fields

Wired Magazine is known for its influential articles about how emerging technologies affect culture, economy and politics. It is the main U.S. magazine for reflection and debate on technological developments, and is therefore a perfect source to examine how Big Data has been conceived and discursively constructed

in technological fields. In this part I will show that the articles in this magazine construct, reinforce as well as alter certain optimistic understandings of Big Data.

The first time an article was published about Big Data in *Wired* was already in 2008. In this year, Chris Anderson, American author and entrepreneur, published an article in which he presented many optimistic ideas about the new possibilities of digital data. By paying particular attention to the structure of this text, it becomes clear that this first article was aimed at emphasising the importance of Big Data. In the first paragraph, Anderson sets the tone by stating that we have entered an “era of massively abundant data” (Anderson 2008) in which we can now, finally, explain the world around us. Moreover, by examining the content of this article, many beliefs and opinions can be identified. For example, Anderson also speaks of an “era of big data” (ibid.) as well as a “Petabyte Age” (ibid.), and claims that the massive amounts of data, and the capability to analyse this, change the production of knowledge radically. Traditional scientific methods that are built around pre-formulated and testable hypotheses can now be replaced by superior Big Data methods. Fundamental assumptions can be found in the way Anderson states that it is now possible to “throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot” (ibid.). Moreover, Anderson points to scientific fields such as human behaviour, linguistics, sociology, and psychology, and presents it as common sense that computers can now “track and measure it [human behaviour] with unprecedented fidelity. With enough data, the numbers speak for themselves” (ibid.).

These statements show that Anderson presented it as highly logical that a new technological era had arrived. This becomes even more evident when we consider the way he claims that this current ‘era of Big Data’ is the logical next step in a successive chain of technological events that started sixty years ago with the first digital computer that made information readable. The next step was twenty years ago, when the Internet made this information reachable. Subsequently, ten years ago, the first search engines made it possible to crawl the Internet into a single database. Now, big search engines like Google and other companies are “sifting through the most measured age in history, treating this massive corpus as a laboratory of the human condition” (ibid.). Anderson therefore calls the first digital computers, the Internet and search engines “the children of the Petabyte Age” (ibid.). At this point it is evident that this article did

not merely describe or reflect on Big Data, rather, it was part of the way meaning was created around the possibilities of digital data.

In the years after this article was published, not much was written about Big Data in Wired. However, around the end of 2011, roughly four years after Anderson's article, a number of optimistic articles about Big Data appeared in Wired. An examination of the structure of these articles has revealed that most of these articles framed Big Data as a new technological development that had fully arrived. For instance, in the opening sentence of an article about the new possibilities of Big Data for the health care industry, it was stated that "Big Data is all around us" (Hernandez 2012). Another distinctive example is the title of an article that presented the yearly amounts of Google searches, YouTube video's, Facebook posts, and more: "Information Revolution: Big Data Has Arrived at an Almost Unimaginable Scale" (Pearlstein 2013).

Additionally, an examination of the content of these articles has demonstrated that two opinions, beliefs and values are most often expressed. First, the current large amounts of data are often emphasised in many articles. A typical example can be found in the way Pearlstein supports her argument that Big Data has arrived by stating that there is currently a "nuclear explosion in the collection and storage of digital information" (ibid.). Second, the idea that new insights and knowledge can be gained with Big Data can often be identified. For example, in 2014, an illustrative article was published on how Big Data can bring energy independence in the U.S., because Big Data will give a "more holistic view of where and how to drill and frack in a way that allows us to preempt future problems without creating new ones" (Basu 2014). Moreover, in this article it is presented as common sense that "[c]omputers today can see, hear, and understand — a new world ushered by the fascinating mergers among computational and scientific disciplines — in ways that hasn't been possible to date" (ibid.). Similar beliefs in the way Big Data will bring new insights can also be found in many articles about the new possibilities for the marketing industry. For example, it is argued that new insights for a specific marketing strategy and tactic can now be found by "unlocking the Big Data goldmine" (Steinberg 2013).

Closely related to the possibilities of Big Data within the field of marketing, it has to be pointed out that many critical and pessimistic notions on Big Data have also emerged in Wired over time. For example, it is argued in a common-sensical way that the fictional Big Brother, an all-knowing and omnipresent leader that has total control over a society from Orwell's (1949) dystopian novel *Nineteen Eighty-Four*, has become a reality with the rise of Big

Data (Gallagher 2013). Also, a number of articles have argued against the revolutionary impact of Big Data by pointing to practical and technical obstacles. For example, it is argued that “big data is not magic” (Lin 2013) because machines do not automatically make sense of large amounts of data. This shows that we also encounter discourses in Wired that alter the optimistic discursive construction of Big Data, and that there is indeed, in Foucault’s terminology, a discursive struggle over a socially constructed meaning of Big Data.

To sum up, it can be observed that many optimistic, as well as a number of critical and pessimistic ideas have appeared in Wired. Moreover, this analysis has demonstrated that the optimistic ideas about Big Data are often presented with seemingly universal ideas about how a new technological revolution has fully arrived with the current explosive amounts of digital data, while the critical ideas about Big Data are either presented with seemingly universal ideas about the dystopian effects of this new technological revolution or with arguments against the revolutionary impact of Big Data.

1.2.3. The New York Times: Big Data and popular discourse

In this last part I will examine how and when Big Data has emerged in popular discourses by focussing on the biggest popular U.S. newspaper The New York Times. In this part I will thus analyse how the term has reached a mass public. Moreover, I wish to elaborate on how this popular discourse corresponds to the way IBM and Wired have discursively constructed Big Data as a highly logical technological revolution.

As I have indicated above, the term ‘Big Data’ gained interest in the IT-industry and technological fields around 2008. The first mention of the phrase in The New York Times appeared later. At the end of 2009, the technology director at the newspaper mentioned the term very briefly by stating that “[s]ome of the most interesting research going on right now is in the area of ‘big data’” (Gallagher 2009). The second and third time Big Data was discussed in The New York Times was in 2010, however, these articles were not about the possibilities of Big Data, but were factual news articles about how two companies (Datameer and EMC) entered the data market (Vance 2010a; Vance 2010b).

It was not until late 2011 that bigger and more impactful articles about Big Data appeared in The New York Times. An examination of the content and structure of these articles has demonstrated that very similar ideas about Big Data, as the ones put forward by IBM and Wired, emerged in the popular

newspaper. For example, Steve Lohr, reporter on technology, business and economics, was one of the first authors to write an article about the possibilities of Big Data. In his article, he immediately placed the term in context in the first sentence by speaking of “decision-making insights in the modern data flood” (Lohr 2011). This same idea about gaining insights and knowledge can also be identified in the second optimistic article about Big Data that appeared a month later. In ‘The Big Business of Big Data’, the technology reporter Quentin Hardy writes that benefits will be gained by “cleverly sifting through it [Big Data] to find and exploit new patterns and relationships” (Hardy 2011).

In 2012, many more articles were published about Big Data, often by the same two authors that were mentioned above. Especially the structure of these texts reveals that Big Data is constantly framed as a huge new technological development. For example, Lohr speaks in the title of his article of “The Age of Big Data” (Lohr 2012) and elaborates on this by arguing that “the computer tools for gleaning knowledge and insights from the Internet era’s vast trove of unstructured data are fast gaining ground” (ibid.). Another illustrative example is the way Hardy speaks of a “Big Data revolution” (Hardy 2012) in the title of his article, and subsequently claims that Big Data is about “looking at that information in novel ways to find new patterns for prediction” (ibid.). In these arguments we thus find remarkable similarities with the way Wired and IBM have also presented it as common sense that more data will lead to more insights and knowledge.

In 2013, even more optimistic articles about Big Data were published in The New York Times. Moreover, a distinct category for Big Data was added on Bits, The New York Times blog for business and technology. By examining the beliefs, values and opinions in these articles, it has become clear that the optimistic discursive construction of Big Data is constantly reinforced through discourse. For example, Lohr writes in 2013 that Big Data is broadening beyond the Internet because it is first “a bundle of technologies. Second, it is a potential revolution in measurement. And third, it is a point of view, or philosophy, about how decisions will be — and perhaps should be — made in the future” (Lohr 2013a). However, around 2013, multiple critical notions on Big Data emerge as well. Similar to the topics in Wired, the same two authors also express their concerns about privacy (Lohr 2013b), as well as concerns about technical and practical obstacles (Hardy 2013; Hardy 2014). This confirms that the discursive struggle over a socially constructed meaning can also be found in popular discourses.

Summarising, this analysis has revealed that optimistic visions of Big Data started to appear in popular discourses around the end of 2011. Moreover, it has demonstrated that the ideas constructed around Big Data appear strikingly similar to the ideas already expressed by IBM and in Wired. In all three discourses, it is often presented as common sense that the large amounts of digital data will undoubtedly lead to new and profound knowledge. Whether it is called “the Big Data era” (Zikopoulos et al. 2012, xv) at IBM, or the “Petabyte Age” (Anderson 2008) in Wired, or “the Age of Big Data” (Lohr 2012) in The New York Times, Big Data is, in all three discourses, depicted as a major technological development.

1.3. Identifying assumptions and raising questions

So far I have argued that the advent of new popular terms comes with optimistic and utopian discourses and that these discourses construct our understanding of new technologies. Also, I have examined the discursive responses, expectations and promises of Big Data, and demonstrated that very optimistic ideas have often been presented as common sense. With especially Fairclough’s arguments in mind, it has to be acknowledged that these seemingly universal ideas are part of an ideological and discursive construction, and that they often embody fundamental assumptions (Fairclough 1989, 33). In this part I will therefore summarise the discursive construction of Big Data and show that three underlying and recurring assumptions can be identified. By reflecting on these assumptions briefly, I will argue that they are highly debatable and that it is necessary to inquire how it can be explained that these assumptions have been projected onto the term on such a large scale and in a rather short period of time.

The first recurring assumption is the idea that Big Data is capable of delivering insights with which the greatest problems can be solved and the greatest questions be answered. This idea can be identified in expressions where it is stated that Big Data is a technology that can “unlock some of the mysteries of the Universe” (IBM, 2014a) as well as “find patterns where science cannot” (Anderson 2008). Moreover, this idea can be found in articles wherein it is argued that Big Data is about “looking at that information in novel ways to find new patterns for prediction” (Hardy 2012). The second recurring idea is that the current amounts of data are revolutionary, and often, this is presented in a metaphorical way. Great examples can be found in the way IBM claims that “we’re drowning in data” (Zikopoulos et al. 2012, 6), the way authors in Wired

refer to a “nuclear explosion” (Pearlstein 2013) and a “Big Data goldmine” (Steinberg 2013), as well as the way The New York Times speaks of the “Internet era’s vast trove of unstructured data” (Lohr 2012). The third recurring idea is the assumption that Big Data can offer extremely precise and objective results. In other words, there is a certain presumed idea of technological objectivity about computers and data, and thus about Big Data. This assumption can be found in exemplary expressions in which it is stated that Big Data will bring “better enterprise decisions” (HP 2014), that with Big Data we can track and measure with “unprecedented fidelity” (Anderson 2008), that, “[w]ith enough data, the numbers speak for themselves” (ibid.), that Big Data will give a “more holistic view” (Basu 2014) and that Big Data is about “how decisions will be — and perhaps should be — made in the future” (Lohr 2013a). To sum up, the promises and expectations consist of three central ideas, assumptions and biases about Big Data:

1. Big Data will deliver knowledge, insights and answers about anything;
2. The contemporary growth of digital data is revolutionary;
3. Big Data utilisations are objective and accurate.

A closer look at these three central assumptions reveals that they are highly debatable. This is not only because most expectations have not been realised yet, it is also because scholarly arguments prove the opposite. For instance, boyd & Crawford (2012) argue that important questions have to be asked about the misleading claims of objectivity and the assumption that more data equals better and more representative data. Lev Manovich (2011) is less critical and has optimistic ideas about the future of Big Data, but also argues that important challenges must be met. Most importantly, he states that we need to be “careful of reading communications over social networks and digital footprints as authentic” (ibid., 6). In the same light, the American writer Jaron Lanier claims that Big Data is not an “automatic cornucopia, or a substitute for insights” (Lanier 2013, 211). Data about the weather can to a certain extent be gathered and analysed, but Big Data about people is different: “[i]t doesn’t sit there; it plays against you. It isn’t like a view through a microscope, but more like a view of a chessboard” (ibid., 272). Additionally, the idea that Big Data is accurate and objective, is debatable in the field of software studies. This is because Big Data algorithms have a problematic “technologically-inflected promise of mechanical neutrality” (Gillespie 2013, 15) where the choices that are

made are unjustly “presented both as distant from the intervention of human hands, and as submerged inside of the cold workings of the machine” (ibid.).

These critical observations show that Big Data may deliver new and profound insights, but that this should not be taken for granted. Therefore, we need to ask how it can be explained that these optimistic and perhaps even utopian ideas about Big Data have been projected onto the term on such a large scale and in such a short period of time.

2.

A new vision on the history and development of media

The past is not dead. In fact, it's not even past. (Faulkner, 1950)

In the previous part I have argued that Big Data is often depicted as a revolutionary technology in the IT-industry, technological fields, and popular discourses. It is expected that Big Data will bring answers and solutions to the greatest problems facing society today. Predicting and preventing heart attacks and crime, solving the world's energy problem and finding answers to our existence in the universe, are just a few examples that illustrate the optimistic, and perhaps utopian ideas about Big Data. I have also described that it remains questionable whether Big Data will actually live up to the current expectations, and that it therefore needs to be inquired how it can be explained that these expectations have, in a rather short period of time, become part of a popular and collective vision about technology. In this chapter I will explore how to approach this main question methodologically, introduce media archaeology as a very valuable research field that is capable of providing a theoretical explanation for today's discourses of Big Data, and describe how a media-archaeological study can be conducted. In order to do so, I will first show that new and valuable perspectives can be gained by placing Big Data in a scholarly debate about how to place media developments in a historical context.

2.1. Arguing against a linear approach to media history

Even though history is one of the oldest qualitative academic research fields, it has witnessed two highly different views on how to actually approach the past. On the one hand, there is a linear, progressivist, object-focussed and teleological vision on history. On the other hand, there is a more cultural approach that argues strongly against the teleological vision on history. In this part I will also

argue against a teleological view on the history of media, because this provides the possibility to place today's discourses of Big Data in a broader perspective.

Media scholar Bonnie Brennen (2013) distinguishes between the two types of historians by referring to traditional and cultural historians. As the names imply, the traditional historians may have a longer history, but the cultural historians are currently present in larger numbers. The traditional historians have a "common sense understanding of history" (ibid., 171) with which they often describe a chronological and progressivist reconstruction of the past. Typical of this approach is that history is often presented with "narratives as fact-based objective explanations of events, issues and problems" (ibid.). By contrast, cultural historians argue that history is a "living process, including issues of continuity and discontinuity" (ibid., 172) and that "researchers must go beyond reconstructions of the past to consider people's experiences in culture" (ibid.). Rather than emphasising facts, chronological narratives and events alone, cultural historians thus also take cultural and social developments into account.

In agreement with the approaches of cultural historians, many scholars from diverse research fields have also argued against the linear and teleological viewpoint. The postmodern philosopher Jean-François Lyotard (1984) argued that these teleological histories with grand and goal-driven narratives should be eschewed because they are too reductive and exclusionary. Additionally, the French philosopher Michel Foucault also argued against a linear and chronological notion of history in his 1969 book *L'Archéologie du Savoir*, which was translated into English as *The Archaeology of Knowledge* (1972). Foucault claims that many connections between historical events, that have been drawn and reconstructed afterwards, should not be uncritically taken for granted. He claims that such "pre-existing forms of continuity" (ibid., 25) that are "accepted without question, must remain in suspense" (ibid.). For him, it is especially the "tranquillity with which they are accepted [that] must be disturbed" (ibid.). It is here that Foucault pleads against a reconstruction of history based on mere chronological continuities, and for an approach focussed on continuities as well as discontinuities in historical discourses. This approach will force a researcher to "advance beyond familiar territory, far from the certainties to which one is accustomed, towards an as yet uncharted land and unforeseeable conclusion" (ibid., 39).

Even though many scholars have thus argued against linear and teleological views on history, such viewpoints can still be clearly identified in today's public debates about Big Data. For instance, Anderson (2008) presented

Big Data in *Wired* as ‘the next big thing’ in a successful chain of media developments that started sixty years ago with the invention of digital computers. The progressivist and chronological narrative that is typical for traditional historians can therefore be identified in this rather populist argument about the “children of the Petabyte Age” (ibid.).

Such popular notions in which new media are perceived as the next successive step towards a better future are not unique. On the contrary, it is argued that our western history is dominated by such thinking. Jay David Bolter and Richard Grusin (1999) describe that such linear perspectives are part of a modernist viewpoint that is characterised by a belief in technological progress and radical breaks with the past. This implies that such traditional views on the history of media are also part of the prevailing belief that new media create and provide for a better future, as was identified in contemporary responses to new technologies in the previous chapter.

Additionally, Bolter & Grusin (ibid.) argue in agreement with cultural historians, Lyotard (1984) and Foucault (1972) that such linear and progressive views on the history of media should be rejected. They claim that there are no radical breaks with the past, because “the content of a medium is always another medium” (Bolter & Grusin 1999, 45). With this, they propose that the history of media is a much more complex process in which media always depend upon older media, and in which new media are in constant dialectic with older media (ibid., 50).

Moreover, important observations from cultural media historians also confirm that teleological notions should be eschewed. In the previous chapter I have pointed out that many media historians have expressed a feeling of *déjà vu*, since the advent of each new medium gives rise to both optimistic and anxious responses. But, there is more to this sense of *déjà vu*. Media and art historian Tom Gunning has pointed out that we do not only encounter “the same ambivalence of optimism and anxiety, but the scenarios constructed around these primal ambiguities seem even more clearly legible” (Gunning 1991, 185). This means that specific thoughts and expectations of specific media in very different times, with very different social and cultural contexts, often appear remarkably similar. A specific technological manifestation may thus be entirely new, but the way it is talked of can be part of a recurring discursive phenomenon.

I therefore wish to place Big Data in a broader historical, social and cultural methodological framework that acknowledges exactly these recurring discursive elements and that does not presume that Big Data is simply the next

successive step in computing. I will do this by referring to the emerging research field of media archaeology, because its anti-teleological starting point and its layered view of history can place Big Data in a much broader historical, comparative, social and cultural framework. Moreover, this research field pays particular attention to recurring elements in history, and is, as I will describe next, inherently capable of providing a fundamental theoretical explanation for why we think and talk about new media the way we do. What media archaeology is, how it offers a means of explaining today's reactions, and how a media-archaeological study can be conducted, will be consecutively discussed in the rest of this chapter.

2.2. Making sense of media archaeology

In the past decade, media archaeology has become a ubiquitous term in media studies (Natale 2012, 524). There is however no consensus about either the principles, the terminology or the 'correct' methodological guidelines (Huhtamo & Parikka 2011, 2). Various authors have called their study a media-archaeological study but have developed different approaches to it. How can we then make sense of this emerging research field? To address this question, I will concentrate on three prominent writers in this field: Jussi Parikka, Erkki Huhtamo and Siegfried Zielinski. By doing so, I aim to emphasise the value of this approach for this thesis, as well as address the similarities and the differences between the authors in order to describe how a media-archaeological study can be conducted, and to be able to position myself in this emerging research field.

In *Media Archaeology: Approaches, Applications and Implications*, Erkki Huhtamo and Jussi Parikka (2011) published a selection of articles by various authors aimed at triggering debates about the differences that have emerged in this field. Irrespective of these differences, the interests, goals and findings of media archaeologists are often the same. Based on historical discoveries, "media archaeologists have begun to construct alternate histories of suppressed, neglected, and forgotten media that do not point teleologically to the present media-cultural condition as their 'perfection'" (ibid., 3). Additionally, Huhtamo and Parikka claim that media archaeology is about discovering "textual, visual and auditory archives as well as collections of artefacts, emphasizing both the discursive and the material manifestations of culture" (ibid.). Here we witness an important distinction between a materialistic and a discursive orientation for media archaeologists. That is, on the one hand, a media archaeologist can

construct alternate histories based on characteristics within discourses that surrounded media, while on the other hand, a media archaeologist can construct alternate histories based on material characteristics of media.

The Finnish new media theorist Jussi Parikka is a media archaeologist who inclines towards a materialistic approach. In his book *What is Media Archaeology*, this materialistic orientation can be identified in the way he introduces media archaeology as “a way to investigate the new media cultures through insights from past new media, often with an emphasis on the forgotten, the quirky, the non-obvious apparatuses, practices and inventions” (Parikka 2012, 2). Moreover, this materialistic orientation becomes even more evident when Parikka claims that future media archaeologists should not start their “excavations by going to an archive filled with books and documents, but opens up a PC from the 1980’s, inspects its circuit board, and starts forensics work on the hard drive” (ibid., 88). While Parikka argues validly that this materialistic approach is capable of constructing alternate histories, it needs to be pointed out that this strategy proves to be problematic when one wishes to address the way various apparatuses, practices and inventions may have been part of the aforementioned recurring discursive phenomenon.

Additionally, in Parikka’s approach one can clearly identify an anti-teleological viewpoint on the history of media. He argues that media history is “sedimented and layered” (ibid., 3) and that media archaeology offers insights into “new media and old media in parallel lines” (ibid., 5). The most important goal for Parikka is to challenge our contemporary belief in ‘the new’, and to show that new media are not always as new as one might think (ibid., 13).

Moreover, Parikka addresses a very important methodological question. In a media-archaeological study it is often unclear of ‘where to begin’. If one would start with studying and excavating random past media, there is a danger of drowning in history and never finding anything useful. However, if one would start from our current world of digital media, there is a danger of leaning towards a teleological reconstruction in which today’s media are seen as the ‘end-point’. Parikka therefore proposes to “start in the middle - from the entanglement of past and present” (ibid., 5).

Media historian Erkki Huhtamo approaches media archaeology largely similarly, but with important differences. Just like Parikka, Huhtamo disclaims any linear and chronological approach to history by claiming that such studies are based on an “objectivist fallacy” (Huhtamo 1996). History is not about hard facts, but it is a “multi-layered construct” as well as “basically just another

discourse" (ibid.). For him, media archaeologists should focus on revealing how media were part of "the complex discursive fabrics and patterns reigning in a culture" (ibid.). This means that Huhtamo and Parikka approach the past similarly, but at the same time, we notice a substantial difference since Huhtamo inclines towards a discursive orientation. Huhtamo's approach does thus, unlike Parikka's approach, provide the possibility to illuminate the way various media have been part of a recurring discursive phenomenon.

Additionally, Huhtamo distinguishes his media-archaeological approach strongly by proposing that it consists of a "topos study" (Huhtamo 2011). He borrows the term 'topos', or 'topoi' in a plural sense, from the work of Ernst Robert Curtius in which topoi are conceptualised as commonplaces or clichés (ibid., 29). In a media-archaeological sense, topoi are "building blocks of cultural traditions; they are activated and de-activated in turn; new topoi are created along the way and old ones (at least seemingly) vanish" (Huhtamo 1996). These topoi may thus "(re)appear and disappear and reappear over and over again in media history and somehow seem to transcend specific historical contexts" (ibid.). A media-archaeological study consists in this scenario of "[i]dentifying topoi, analyzing their trajectories and transformations, and explaining the cultural logics that condition their 'wanderings' across time and space" (Huhtamo 2011, 28). Such a topos study is in Huhtamo's view a "tool for cultural critique" that is not only aimed at drawing parallels in history, but also "helps us understand the reverse: the things that are genuinely new and progressive" (ibid., 41).

Moreover, a key aspect in Huhtamo's approach is that topoi "provide 'prefabricated' moulds for experience" (Huhtamo 1996). In this light, Huhtamo claims that the culture industry can 'use' these topoi as "discursive formulas" (Huhtamo 2011, 34) because they "provide advertisers with tried-and-tested formulas that are used to introduce new consumer products by embedding them within moles the customers already know (whether they are aware of it or not)" (ibid., 39). Even though I am inclined to doubt that the culture industry consciously employs such discursive formulas on a large scale, Huhtamo's argument demonstrates that topoi point to certain age-old, recurring and deeply ingrained ideas about media that still, silently, play out today. This, then, is why such recurring ideas about media need to be "pointed out and reconnected with the past by media archaeologists" (ibid., 38).

Finally, yet importantly, the German media theorist Siegfried Zielinski is also one of the prominent writers within the field of media archaeology. Similar

to Huhtamo and Parikka, Zielinski disclaims the linear approach to history by arguing that those approaches are “comforting fables” (Zielinski 2006, 3) in which technology is considered a “universal driving force” (ibid.). Additionally, Zielinski argues similarly that his historical quest is about finding “something new in the old” (ibid., 3) and about rediscovering “forgotten or hitherto invisible layers and events in the historical development of the media” (ibid., 9). It is therefore notable and perhaps even unnecessary that Zielinski disclaims the term ‘archaeology’ and prefers to call his historical quest an ‘anarchaeology’, since he refuses to follow the way archaeology, in the traditional and disciplinary sense, implies an ordering and governing of the ancient or original.

Additionally, Zielinski’s approach shows that he did not hesitate to dive into the richness of our media history, and did not start somewhere in the middle. This becomes evident when he argues that it is especially “the option to gallop off at a tangent, the option to be wildly enthusiastic” (ibid., 27) that gives the possibility to rediscover the deep time of media. Even though this approach presents itself as somewhat arbitrary, Zielinski manages to show fractures in the predictable and turning points in the history of media. By analysing a rich variety of historical records and by taking the reader on a journey, he claims that today’s complex media do not directly proceed from any primitive tool and that “[m]edia are spaces of action for constructed attempts to connect what is separated” (ibid., 7). By showing how ancient machines have made this connection, Zielinski concludes that “technical praxis do not follow in chronological sequence for anarchaeology; on the contrary, they combine at particular moments in time, collide with each other, provoke one another, and, in this way, maintain tension and movement within developing processes” (ibid., 258).

This examination of three media-archaeological approaches shows that all three authors have an anti-teleological starting point, acknowledge the role of imaginary, dead and forgotten media developments, and see history as a more complex layer of various and multiple events. This means that media archaeology is a very valuable approach, because it is capable of placing Big Data in a much broader historical, comparative, social and cultural framework that does not presume that Big Data is the logical next technological step. Moreover, this approach does not only provide a broader perspective, it offers a theoretical means of explaining today’s optimistic discourses of Big Data as well.

This theoretical explanation can especially be found in the work of Huhtamo (1996; 2011) and Zielinski (2006). Both media archaeologists argue that age-old ideas as well as technological manifestations can (re)appear and

disappear in history. For Zielinski, these ideas and manifestations can even combine, collide, connect and provoke one another. Moreover, Huhtamo's topos study implies that there are certain age-old ideas about technology that continue to play out as prefabricated discursive moulds for today's experiences of new media.

This means that it can be argued that Zielinski's and Huhtamo's theories offer a means of explaining how Big Data has gained so much attention and enthusiasm in such a short period of time. With their theories it can be asserted that the term 'Big Data' has, just like many old media and just like many other technological terms, reactivated, combined and connected certain age-old and deeply ingrained ideas about technology. These transhistorical ideas have, in turn, guided and moulded our contemporary responses to the term. This would mean that Big Data has gained so much popularity because its discursive construction constantly embeds the term in a prefabricated discursive mould we already know and thus unconsciously recognise. This theoretical and media-archaeological proposition is exactly what I wish to elaborate on in the rest of this thesis.

2.3. The rhizome: placing Big Data in a historical framework

In the previous part I have argued that a broader understanding of today's popularity of Big Data can be gained by studying the way age-old and transhistorical ideas about technology may have reappeared, connected and combined, and now provide a prefabricated discursive mould for today's responses. But how to do this in a media-archaeological 'correct' way? So far I have discussed the similarities between the three media-archaeological approaches, however, it needs to be acknowledged that differences can also be found. These differences concern the discursive versus the materialistic orientation in history, as well as where the media-archaeological study should 'start'. In this part I will expand further on these two differences, as well as position myself in this field by introducing the "rhizome" (Deleuze & Guattari 1987) as an exceptionally valuable mode of thinking that is capable of providing additional and necessary guidelines for studying the way age-old ideas about technology affect today's responses to Big Data.

The first difference I wish to address is the distinction between a materially and discursively oriented media-archaeological approach. It is here that I will position myself on the discursive side of this continuum, because

recurring and deeply ingrained ideas about technology can especially be found in historical discourses. However, it is worth pointing out that this does not imply that material manifestations are completely neglected in this scenario. On the contrary, recurring, transhistorical and deeply ingrained ideas about technology can especially be found in discourses that surrounded both material and imaginary manifestations of past media. This means that I will study transhistorical and age-old ideas about technology by paying particular attention to the ideas that have been discursively expressed around past realised, imaginary, dead and forgotten media. Indeed, this approach comes close to Huhtamo's approach that is focussed on "the history of ideas" (Huhtamo 1996), a research field concerned with how human thoughts and ideas have developed and changed over time. Does this mean that conducting Huhtamo's topos study is the 'right' choice for studying the role of deeply ingrained ideas about technology in today's popularity of Big Data?

The answer is not *exactly*. Even though Huhtamo would probably not disagree with Zielinski's (2006) finding that technological praxis do not only (re)appear and disappear, but also connect and combine in the course of history; it needs to be pointed out that Huhtamo's topos study is not primarily aimed at investigating such connections and combinations of topoi. That is because topoi are presented as rather separated themes in history that reappear over and over again in Huhtamo's study. This means that this approach provides the possibility to study how different topoi may have appeared, reappeared, disappeared and changed over time, as well as how specific topoi may have been imprinted on past media manifestations, however, it does not enlarge on the way topoi can also combine, connect and collide in the course of history.

Exactly because I wish to elaborate on the idea that today's Big Data discourse may not *just* be a reappearance of age-old and transhistorical ideas, but *also* a combination, connection and collision of these age-old ideas, I will now refer to the philosophical concept of the rhizome as developed by Gilles Deleuze and Félix Guattari (1987), because it emphasises the idea that transhistorical ideas about technology can also combine, connect and collide. Moreover, it is also capable of addressing the second difference between the three media archaeologists concerned with where a media-archaeological quest should start. What follows is a description of the rhizome, after which I will describe how this concept provides the possibility to study the way age-old ideas have (re)appeared, disappeared and connected in time, and how it gives me the

necessary guidelines of 'where to start' my historical quest, while at the same time, staying very close to the media-archaeological approach.

In *A Thousand Plateaus*, Deleuze and Guattari introduce their concept of the rhizome as an alternate "image of thought" (ibid., 16). The concept is based on the characteristics of a botanical rhizome (the rootstock of a plant below the surface) and is aimed at breaking through the dominant thinking in tree and root structures. Unlike the tree and the root, the rhizome resists any notion of causality, chronological lines, origins and end-goals, because it has "no beginning or end; it is always in the middle, between things, interbeing, intermezzo" (ibid., 25), and because it is "perpetually in construction or collapsing" (ibid., 2) with a process that is "perpetually prolonging itself, breaking off and starting up again" (ibid.). Moreover, "[t]here are no points or positions in a rhizome, such as those found in a structure, tree, or root. There are only lines" (ibid., 8).

This means that the rhizome shows striking similarities with thoughts on the history of media in media archaeology. Both the concept of the rhizome and the media-archaeological approach reject any teleological and chronological narrative and argue that there is no predefined beginning or end-goal. However, the rhizome does not only show striking similarities with media archaeology, it also adds two very helpful and necessary notions for studying the way transhistorical and deeply ingrained ideas about technology (re)appear, disappear and connect in the course of history. These notions can be found in three important characteristics of the rhizome.

A first and fundamental characteristic is that rhizomatic lines can, at any point, "be connected to anything other, and must be" (ibid., 7). Moreover, it is important to note that such a connection of lines cannot happen without "the multiplicity changing in nature" (ibid., 8). Additionally, a second important characteristic is that the rhizome "may be broken, shattered at a given spot, but it will start up again on one of its old lines, or on new lines" (ibid., 9). The way rhizomatic lines can thus connect, break off and start up again is similar to how Huhtamo (1996) claims that topoi can be activated and deactivated in time, and similar to how Zielinski (2006) claims that technological praxis can collide and connect in time. This means that a rhizomatic mode of thought provides the possibility to integrate Huhtamo's and Zielinski's approaches and consider ideas about media as just 'lines' that are able to develop over time, break off and start up again, as well as cross each other, combine, connect and multiply.

The third characteristic is that the rhizome is "not amenable to any structural or generative model" (Deleuze & Guattari 1987, 12) and that it is "a

map and not a tracing" (ibid., 2) with "multiple entryways, as opposed to the tracing, which always comes back to the same" (ibid., 12). This means that the rhizome is a perpetually changing map of lines that connect, multiply, break off and start up again in an unpredictable and random fashion. This is exceptionally valuable for studying transhistorical and deeply ingrained ideas about technology because it provides the necessary guidelines for what to research and where to start. That is, in a rhizomatic, as well as a media-archaeological framework, we can view today's discourses of Big Data as part of a large map, in which now a multiplicity of lines, representing transhistorical and age-old ideas about technology, have (re)appeared, started up again and connected. In this light, today's Big Data discourse is thus exactly such a connection of 'lines of thought' in a perpetually changing map.

In my media-archaeological analysis I will therefore aim to fill in a large part of this map. By focussing on discursive elements in history in which ideas about technology were expressed, I will endeavour to map the most fundamental ideas in history that now connect with Big Data. At the same time I wish to map how these ideas have emerged, changed, or have remained unchanged over time. Thus, by examining what lines have now collided and combined, while at the same time paying particular attention to which lines are genuinely new and which lines may have started up again, we are able to reveal what historical discourses provide a discursive mould for today's responses to Big Data.

Considering that the rhizome has many entry points but no beginning or end, I wish to enter this rhizome at the cross-roads of lines that make up Big Data today. This means that I will start my media-archaeological quest indeed "somewhere in the middle" (Parikka 2012, 5), by acknowledging that today's dominant ideas about Big Data are entangled with age-old ideas and discourses that surrounded past media. With this, I will make it my goal, in line with Huhtamo's approach, to reconnect our contemporary ideas with the past (Huhtamo 2011, 38). In order to do so, I have made two important choices.

First, I have chosen to search for and excavate similar past discourses by making a few analytical 'cuts' in history. By doing so, I am forced to dive into the richness of our media history, as well as able to reveal similar ideas about technology in very different historical, social and cultural contexts. It is here that I wish to emphasise that I do not aim at reconstructing teleological and chronological lines that might show how certain ideas may have logically evolved over time, because the opposite is true. Making such clear cuts and exposing recurring ideas is typical for a media-archaeological approach, because

it emphasises “cyclical rather than chronological development, and recurrence rather than unique innovation” (Huhtamo 1996).

I will make these analytical cuts in history based on historical moments in which other media were introduced. This means that I will search for and excavate similar ideas about technology by focussing on the expectations, ideals and beliefs that were expressed around the advent of past media. It is here that I have chosen to concentrate on the discursive responses to a series of great technological inventions: the printing press, the telescope, the microscope, the camera obscura, the camera lucida, the calculator and the computer. That is because I contend that today’s dominant ideas about Big Data can be reconnected with ideas that surrounded the advent of these great inventions, and that I will thus find fundamental parts of the rhizomatic map here. It is important to note that this does not imply that similar ideas cannot be found around the advent of other great technologies such as the television, the radio or the telephone. On the contrary, the recurrence of certain ideas may also be identified here, however, not as evidently as around the other media. Due to the feasibility of studying more media manifestations, I have therefore chosen to leave these media outside the scope of this thesis.

The second choice is about the way I have searched for and selected historical discourses. To prevent myself from drowning in historical records, I have chosen to conduct a media-archaeological study that does not focus on an extensive search for and excavation of primary historical records. Rather, it will mainly build on the work of cultural historians who have studied technological discourses of the past. Even though most media archaeologists rarely include secondary sources, I wish to point out that studying already existing scholarly work does not undermine the validity of this research. That is because I will pay particular attention to secondary sources in which primary historical sources and citations can be found. This means that my findings and conclusions will, just like most media archaeologists, be based on primary historical records.

To sum up, at all times during this media-archaeological study, current ideas about Big Data, as well as similar historical discourses of technology, will be treated as a multiplicity of lines in a rhizomatic map that are capable of connecting, colliding, collapsing and changing over time. By revealing what fundamental ideas about Big Data are part of a cyclical phenomenon in history, and now provide a prefabricated discursive mould for today’s response, I aim to provide a broader understanding of today’s popularity of the term in the following chapter.

3.

Big Data & deeply ingrained ideas about technology

An idea is like a virus. Resilient. Highly contagious. And even the smallest seed of an idea can grow. It can grow to define or destroy you. (Nolan 2010)

This chapter is dedicated to a media-archaeological quest in which I aim to study distinctive past ideas, assumptions, dreams, desires and expectations about technology, and describe how these deeply ingrained ideas still play out in today's responses to Big Data. As I have described in the previous chapter, this media-archaeological quest will start 'somewhere in the middle' by acknowledging that today's ideas about Big Data are entangled with past ideas about media. I will therefore pick up today's Big Data discourse as I left it in the first chapter. Here it was argued that there are three central and underlying assumptions about Big Data today: (1) Big Data will deliver knowledge, insights and answers about anything, (2) the contemporary growth of digital data is revolutionary and (3) Big Data utilisations are objective and accurate. In this chapter I will devote a subchapter to each of these three assumptions and show that they are part of a long discursive history in which similar ideas have been expressed around other technological manifestations. In the last part of this chapter I will connect these historical discourses with today's Big Data and conclude that this term is currently capable of seamlessly connecting these deeply ingrained ideas so that the term becomes highly logical, natural and comprehensible in the blink of an eye.

3.1. Technology and the age-old human quest for knowledge

The first, and perhaps most obvious idea about Big Data today is that it will bring salvation to our most burning questions and our greatest problems. We have witnessed ideas about how Big Data will finally bring answers about the

universe and alien life (IBM 2014a), as well as insights with which we can solve the world's energy problem (Basu 2014), predict heart attacks (IBM 2013) and prevent crime (IBM 2012). The Big Data discourse thus consists of a collective vision in which ideas about a new way of producing new and profound knowledge dominate. By showing that very similar ideas can be found in discourses that surrounded past media, I will argue that there is an age-old desire for knowledge and an age-old idea of gaining knowledge through technology.

However, before addressing a first technological manifestation in history, I wish to refer to an ancient example that illustrates that a desire for answers and knowledge can also be found in a very distant past. This can be found in the Greek myth of the Delphic Oracle. This myth is about Pythia, a priestess that operated as an intermediary between the Greek people and the Olympian god of light and knowledge, Apollo. This Olympian god was capable of knowing more than any human could, since he was, just like the light, capable of seeing through anything. The priestess Pythia could answer in the name of Apollo and get access to all his knowledge. According to the myth, Pythia "responded to the questions of citizens, foreigners, kings, and philosophers on issues of political impact, war, duty, crime, laws—even personal issues" (Broad 2007, 43). Even though there is no mention of technology here, the Delphic Oracle clearly refers to an underlying desire for answers and knowledge about various topics.

This desire for answers and knowledge can be found in other ancient contexts as well. Lee Worth Bailey claims that we have a "primal desire for knowledge of the universe" and argues that the biblical story about the Garden of Eden is an exemplary expression of our age-old questions about the origins of the universe and the meaning of life (Bailey 2005, 142). Discourses in which a desire for new knowledge and new answers were expressed, can thus already be found in a very distant past.

If we leave these age-old times, and fast forward to early modern Europe, we find a similar desire for answers and knowledge in discourses that surrounded the invention of the telescope and the microscope. Focussing on the telescope, it is notable that, even before the actual invention, there were ideas about an imaginary device with which one could "read the smallest letters from an incredible distance, and to count sand, or grains, or blades of grass or any other minute object" (Reeves 2008, 28). Here it is not the possible "inauthenticity of these statements, but rather the longevity of these expectations" (ibid., 29) that points to the desire of seeing and knowing things that are not directly visible to

the eye. Moreover, it is interesting that this desire shows remarkable similarities with the ancient Greek myth of the Delphic Oracle.

The invention of the telescope by the Dutch inventor Hans Lippershey and the improvements made by Galileo Galilei made it ultimately possible to see in a distance. It is argued that Galileo's telescope subsequently "changed the truth" (Burke 2009, 6) because the existing knowledge and understanding of the universe had shifted radically with his invention and work. Historian James Burke writes how "a hundred years later everybody 'knew' that the universe was open and infinite, working like a giant clock" (ibid.). In a sense, Galileo's telescope did thus indeed live up to certain expectations and did finally bring some long-desired answers about the universe.

In 1869, we find similar great expectations around the Great Melbourne Telescope (GMT). The GMT was envisioned and expected to bring answers about the validity of the, still unverified, nebular hypothesis (Gillespie 2009, 2). This hypothesis, formulated by Emanuel Swedenborg in 1734, is concerned with the emergence, formation and evolution of the solar system, and would finally bring definitive answers about the origin of the universe. However, the results were disappointing and the GMT never lived up to these expectations (ibid.). This forgotten technological project thus shows the recurrence of the idea and expectation that technology will bring new and profound knowledge.

Next to telescope, there was also the microscope in early modern Europe that gave rise to exemplary discourses about knowledge and technology. Zielinski writes that in the mid 1800's "human and animal organisms had been thoroughly investigated under the microscope and hooked up to machines and instruments" (Zielinski 2006, 206). Whereas the telescope made it possible to see distant things, the microscope made it possible to see, observe and measure all phenomena of the living organisms, because it was now "possible to make graphic representations of how the blood circulates, how the field of vision is laid out, how the eardrum vibrates, how bodies react in free fall or rotation" (ibid.). This demonstrates that both inventions gave rise to discourses in which we can identify the idea that technology makes it possible to see and know things that were not possible before.

At this point it is possible to place the ancient myths and stories, as well as the discourses that surrounded the invention of the telescope and the microscope in a rhizomatic and media-archaeological map. It has become clear that a desire for answers and knowledge has been present in both times, which means that a rhizomatic line of thought can be identified. However, it has to be

pointed out that this idea has not remained completely unchanged. Whereas this desire for answers and knowledge was especially projected onto religious aspects in ancient times, we have seen that this shifted to projections onto technology with the advent of the telescope and the microscope.

Next to the telescope and the microscope, the first computer also gave rise to exemplary ideas about the way knowledge can be gained through technology. A first and famous example is the imaginary dream machine, the Memex, described by Vannevar Bush in 1945. With this device, a scientist can browse more naturally through his books, records and communications, because it “can be consulted with exceeding speed and flexibility” (Bush 1945). In this way, the Memex is foremost an “enlarged intimate supplement to his memory” (ibid.) with which scientists can come to better conclusions. This dream machine thus reveals the recurrence of the rhizomatic line of thought that more information and knowledge can be gained through technology.

Around the 1960’s, similar ideas can be identified in the work of Ted Nelson. With Project Xanadu, Nelson aimed at ordering information in such a way that one could see all the other articles that relate to the same topic (Wardrip-Fruin & Montfort 2003). This project thus shows the recurrence of the idea that technology can provide access to knowledge and information. In the same decade, in 1964, there was also an imaginary device called the Answer Machine. This machine was presented as a device for children doing their homework and was capable of answering to any question (Novak 2007). For example, if the user would ask what Edison’s first phonograph would look like, the device would answer this particular question with a picture of this machine (ibid.). In 1974, the inventor and science fiction writer Arthur C. Clarke also presented very similar ideas about how computers would provide information and knowledge. He claimed that in the very near future, one can get “all the information he needs for his everyday life” (openculturevideo 2012).

These historical records show that a desire for knowledge can be found in discourses that surrounded the advent of the first computer. This means that the rhizomatic line of thought that was identified around the ancient myths, the telescope and the microscope, can also be identified here. However, by comparing these early ideas about the computer to the ideas expressed around the telescope and the microscope, a subtle difference in this rhizomatic line of thought needs to be pointed out. Whereas the telescope and the microscope were mainly envisioned as technologies that would bring new insights by making things visible that were not visible before; the first computer was more often

envisioned as a technology that would provide easier access to already existing information and knowledge.

However, the idea that technology is capable of bringing new insights and knowledge did not completely disappear around the rise of the computer. It was especially in science fiction that discourses in which the quest for new knowledge, rather than access to information, reappeared. Science fiction is an interesting genre for media archaeology, because it can be “seen as an alternative debate on a new technology” (Lammes 2003, 68). Just like many imaginary, forgotten and dead media as well as technological predictions about the future, science fiction can thus show what discursive ideas about media and technology were expressed in a certain time. In the 1970’s there was one particular science fiction film that reminds us of discourses that surrounded the telescope. In the popular film, *The Hitchhiker’s Guide to the Galaxy* (Adams 1979), we encounter a supercomputer called Deep Thought that is capable of bringing answers to our age-old questions about the universe. After 7,5 million years of calculation, the computer has the answer to life, the universe and everything: 42.

From a rhizomatic and media-archaeological point of view, it can be observed that there is a recurring, transhistorical and age-old desire for answers and knowledge. This desire can be found in religious and ancient stories, as well as in discourses that surrounded the invention of the telescope, the microscope and the computer. Moreover, in discourses that surrounded the microscope, the telescope and the computer, a recurring and deeply ingrained idea about how information and knowledge can be gained through technology can be identified. This means that our current idea about how technology is going to bring new and profound knowledge is a line of thought that has reappeared again, and is thus everything but new.

3.2. A recurring sense of information overload

In the first chapter I have pointed out that the second recurring assumption in today’s Big Data discourse can be found in the idea that the current amounts of digital data are revolutionary. This assumption can be found in several, and often metaphorical, arguments in which there is, for example, talk of a “nuclear explosion” (Pearlstein 2013) of data in which we are currently “drowning” (Zikopoulos et al. 2012, 6). In this part I wish to inquire the newness of these statements and show that similar thoughts about a rapid growth of information have been expressed before.

A remarkable historical moment in which ideas about revolutionary amounts of information were expressed, can be found at the very beginning of early modern Europe. Near the end of the Middle Ages, around 1450, Johannes Gutenberg invented the printing press. This invention made it possible to print ink onto materials such as paper and clothes. Today, the invention of the printing press is often marked as the beginning of early modern Europe, because it revolutionised the way people conceived and described the world by providing the first assembly line and the first mass production of books (McLuhan 1962, 124). And indeed, the invention of the printing press led to a rapid increase in the amounts of books that were printed. It is estimated that there were already 13 million books circulating in a Europe with only 100 million people around the year 1500 (Briggs & Burke 2010, 14).

However, not the great amounts of books that were published, nor the way this invention may have revolutionised the way we describe and conceive the world, are the centre of this media-archaeological quest. Rather, the responses to the invention of the printing press, as well as the responses to the great amounts of printed books, are of interest here. In Asa Briggs's and Peter Burke's book *Social History of the Media: From Gutenberg to the Internet* (ibid.), such reactions to the advent of the printing press can exactly be found. And most interestingly, they write that "[t]he problem of 'information overload' as it is now known, goes back a long way" (ibid., 16).

Briggs & Burke prove this statement by describing and identifying some exemplary reactions to the printing press. For example, around 1550, an unknown Italian writer was complaining that "there were so many books that we do not even have time to read the titles" (ibid., 16). Moreover, the French reformer Jean Calvin called the amounts of printed books a "forest", "ocean" and "flood" in "which the readers had to navigate" and in which "it was hard to escape drowning" (ibid.). These reactions to the rapid increase in the number of printed books, as well as the use of very specific metaphors, thus show remarkable similarities with the way the current growth of digital data is being discussed.

However, it has to be pointed out that it was not only this sense of information overload that dominated the advent of the printing press. In fact, Briggs & Burke describe reactions in which the previously discussed desire for knowledge is apparent as well. For example, the French essayist Michel de Montaigne wrote in the 16th century that "the art of printing will so spread knowledge that the common people, knowing their rights and liberties, will not

be governed by way of oppression" (ibid., 15). Moreover, in 1895, it was stated that print "gave assurance that the work of the Renaissance would last, that what was written would be accessible to all, that such an occultation of knowledge and ideas as had depressed the Middle Ages would never recur, that not an idea would be lost" (ibid., 16). This shows that there were also ideas about the way the printing press would democratise the access to knowledge, as well as ideas about how the printing press would bring an occultation of knowledge. In essence, we thus encounter both a sense of information overload and an optimistic discourse about how this technology will provide knowledge.

Indeed, this sense of information overload seems to go hand in hand with the idea that the newly produced amounts of books or data can also be used to gain knowledge. This correlation is also apparent in a period in which the future of computing and the future of the Internet were envisioned. In the previous part I have pointed towards Bush's Memex, Nelson's Xanadu and Clarke's prediction about the future of computing to argue that these discourses illustrate a desire for information and knowledge. However, in these examples we do not only find references to this desire, we also witness the same recurring sense of information overload.

Bush's text about the Memex reminds us of the aforementioned reaction of the unknown Italian writer during the advent of the printing press. Bush argued that the Memex was a much-needed device at that time because "publication has been extended far beyond our present ability to make real use of the record" (Bush 1945). Moreover, Bush argues that man "has built a civilization so complex that he needs to mechanize his records more fully if he is to push his experiment to its logical conclusion and not merely become bogged down part way there by overtaxing his limited memory" (ibid.). These quotes clearly show that Bush experienced a sense of information overload. In Nelson's project Xanadu we also encounter a similar experience because the name 'Xanadu' refers to a magical world in which nothing is ever forgotten (Flichy 2007, 61). Moreover, Clarke expresses this sense of information overload even more literally as well as familiarly metaphorically in his prediction for the future. He claims that in order "to cope with the information explosion, we may develop a machine for recording information directly onto the brain" (Lucidl 2010).

This demonstrates that the Memex, project Xanadu and Clarke's predictions were not merely given in by the idea that technology is capable of bringing knowledge; they were also given in by the idea that there was, at that time, too much information. It can therefore be argued that these two ideas, or

rather, these two lines of thought in a rhizomatic and media-archaeological map, have not only (re)appeared around the invention of the printing press and the advent of the first computer; they have also connected and combined in both times.

However, by comparing this sense of information overload during the rise of the computer to this same sense during the advent of the printing press, one important difference in this recurring rhizomatic line of thought has to be pointed out. Whereas the printing press was thought of as a technology that produced great amounts of books, the computer was not envisioned as a device that would produce information. Rather, the computer was mainly envisioned as a machine that would bring order in this information explosion by providing easier access to information.

If we now consider today's discourses that surround Big Data, it is most interesting that both of these ideas reappear. In the first chapter we have seen that, on the one hand, digital devices are currently seen as technologies that are capable of producing massive amounts of data, and on the other hand, they are also seen as technologies that are capable of analysing and ordering this information.

To recapitulate, this rhizomatic and media-archaeological quest has demonstrated that our current sense of information overload is everything but new. Very similar ideas and metaphors can be identified in discourses that surrounded the rise of the printing press and in discourses that surrounded the emergence of the first computer. These two historical discourses in which technology was depicted as a producer of new information during the advent of the printing press, as well as an orderer in the abundance of information during the emergence of the first computer, both reappear in today's ideas about Big Data.

3.3. The emerging ideal of mechanical objectivity

This third and last media-archaeological quest is based on current assumptions of objectivity and accuracy that surround today's Big Data utilisations. In the first chapter I have described that these assumptions can be found in articles in which it is stated that we can just "throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns" (Anderson 2008). The same assumptions can also be found in discourses in which it is stated that "[c]omputers today can see, hear, and

understand a new world" (Basu 2014) or where it is stated that there is a "drift toward data-driven discovery and decision making" (Lohr 2012). In this media-archaeological quest I will therefore search for similar ideas of a presumed technological objectivity and accuracy, and show that these ideas and ideals have emerged as well as faded in the course of history.

In *The Objective Image*, Peter Galison (2010), professor in the field of history of science and technology, has exactly traced the dominant ideas about objectivity in relation to technology. Galison argues that a big transformation and inversion of dominant ideas about objectivity and subjectivity took place around the 18th century. Where the subjective intervention by an author used to be "the most important feature to be prized in scientific representation, it becomes a veritable epistemic vice" (ibid., 7). It was no longer the subjective touch that was preferred; it was only through superhuman self-restraint that the author could aspire to let nature "speak for itself" (ibid.).

From a media-archaeological point of view it is interesting that this discursive discontinuity was strongly correlated with the advent of the camera obscura and the camera lucida. That is because photographs eliminated the "possibility for subjective alteration" (ibid., 18) and because "the photograph is the actual section" (ibid.) with no "artist's interpretation in the reproduction of structures" (ibid.). This means that the formerly preferred personal, subjective, artistic and interpretive aspects were, with the advent of the camera obscura and the camera lucida, being replaced by a striving for "direct transfer of nature onto the picture" (ibid.). According to Galison, it was here, around the photographic, that the ideal of mechanical objectivity emerged; "an objectivity defined by its moralised and automatic status beyond the reach of the artist's hand" (ibid.).

Media archaeologist Siegfried Zielinski also discovered and described this distinction between a preference for the interference of the subjective, as opposed to a preference for the technological objective. Zielinski claims that there was a small group that still preferred subjective alterations in the 17th century. These so-called 'dioptricians', like Giovanni Battista Della Porta, were interested in the "illusionizing potential of projection" (Zielinski 2006, 86) and believed that "nature could only reveal and develop its hidden powers through the intervention of the researchers" (ibid.). On the other hand, there was a much larger group, the so-called 'catoptricians', with Kepler, Galilei, Descartes and Newton, who believed in the "physics of the visible" (ibid.) and of "seeing through something in the sense of insight or understanding" (ibid.). This shows

that a small group may have still advocated for subjective interference, but that the majority strived for insights based on technological representations of nature.

This demonstrates that substantial viewpoints about how to capture the world changed during the rise of the camera obscura and the camera lucida. Jonathan Crary, professor of modern art and theory, confirms this observation. He argues, in line with Zielinski (2006) and Galison (2010), that the camera obscura was not simply a neutral piece of equipment, but that it was often possessed by philosophers and natural scientists in order to trigger philosophical reflections on the nature of perception and the nature of knowledge (Crary 1992, 27). It can therefore be said that these first cameras were not only a precursor of cinema and film, as we often view them today; they were also part of fundamental debates about objective and subjective ways of capturing and describing the world.

If we would place these historical discourses in a rhizomatic and media-archaeological map, it can be observed that the idea of a technological objectivity is a line of thought that appeared around the invention of the camera obscura and camera lucida. But, it was not only the invention of the camera obscura and the camera lucida that gave rise to discourses in which ideas of a technological objectivity appear. The introduction of the calculator also gave rise to exemplary dream machines in which the ideal of a technological objectivity can be recognised. In the 17th century, the German mathematician and philosopher Gottfried Wilhelm Leibniz worked on two calculators; the Calculus Rationator and the Universal Logic Machine. The Calculus Rationator was a machine with which Leibniz tried to demonstrate that reasoning could be mechanised (Abbas 2006, 48). The Universal Logic Machine builds upon this idea and was envisioned as a machine that is capable of making calculations to the most complex social and philosophical questions, and arrive at fair and objective solutions and answers (Mayer 1999, 4). For Leibniz, this machine would therefore provide the foundation for a more just society (ibid.). This shows that both concepts reveal a desire to create a machine that would bring objective answers with which various issues could be solved.

This means that these historical discourses that surrounded the invention of the calculator, the camera obscura and the camera lucida, show that the idea and ideal of a technological objectivity is a line of thought that has been expressed around these early mechanical manifestations. But what happened to this recurring line of thought when the first computer started to emerge?

Galison argues that another transformation had taken place in the 20th century. He describes that we are far from the “idealizing sage” (Galison 2010, 29) and the “self-abnegating” (ibid.) scientist in pursuit of a technological objectivity. In addition, he argues that we have not gone back to preferring and trusting the subjective touch of the scientist (ibid., 30). Rather, “a new way of being a scientist” (ibid.) has emerged where there is no longer the ideal and striving for “re-production; instead it is one of production, of making things, whether they are new forms of nanoscale matter or productive new structures” (ibid., 29).

In sum, this rhizomatic and media-archaeological quest has identified the emergence of a line of thought that represents the ideal of a technological objectivity in early modern Europe. However, this ideal seems to have faded away in the 20th century because science was no longer about reproducing the world. But, today, with Big Data, we clearly see the reappearance of this ideal in the assumptions and connotations around data analysis, computers and algorithms. This means that this rhizomatic line of thought, representing the idea and ideal that technology is capable of reproducing the world in an objective way, has started up again and reappeared in today’s discourses.

3.4. Today’s popularity of Big Data: an explanation and evaluation

In the previous three parts I have conducted a media-archaeological study in which I have revealed what historical discourses remind us of today’s expectations and assumptions about Big Data. Old myths like the Delphic Oracle, discourses that surrounded the invention of the telescope and the microscope, and early thoughts about the future of computing proved that there is an age-old and recurring idea and desire for answers, as well as a deeply ingrained idea about how technology is capable of unlocking information and knowledge. Also, early responses to the printing press and early thoughts about the purpose of computers showed that a sense of information overload is everything but new. Moreover, discourses that surrounded the invention of the camera lucida, the camera obscura and the calculator demonstrated that the idea and ideal of technological objectivity has emerged in the course of history. This means that today’s dominant ideas about Big Data are everything but new. In fact, today’s optimistic discourse consists of many recurring, transhistorical and age-old ideas, desires and expectations about technology.

But, what then, do these past discourses exactly tell about today’s optimistic and expectant reactions to Big Data? In this last part I will show that

these historical discourses give a broader understanding of today's Big Data discourse in two ways. First, these old ideas, dreams and expectations are capable of explaining today's popularity of Big Data. Second, these historical discourses provide the possibility to inquire today's expectations in new ways.

But first things first. How do these historical discourses explain today's popularity of Big Data? In the second chapter I have indicated that the field of media archaeology, and especially the work of Erkki Huhtamo (1996; 2011) and Siegfried Zielinski (2006) offered a substantial proposition about the way ideas about technology travel in time and how these old ideas may still play out today. Zielinski argued that technological praxis are capable of appearing, disappearing, connecting and colliding in the course of history, and Huhtamo argued that these recurring themes in history serve as prefabricated discursive moulds for today's responses to new technologies. At this point, we can take these statements and the media-archaeological quest together, and understand today's responses to Big Data with a two-stage explanation.

First, it is exactly because today's discursive construction of Big Data is characterised by the idea that technology now provides the possibility to gain new and profound knowledge by analysing revolutionary amounts of data in an objective and accurate way, that this term is currently capable of reactivating our age-old ideas, desires, expectations and dreams about technology. The various actors involved in today's discourses of Big Data, from promoters in the IT-industry to journalists writing about the term, all, whether they are aware of it or not, made use of the interpretative flexibility of the term and actively constructed and reinforced its meaning by constantly embedding the term in a discursive mould we already know. Today's discourses about Big Data thus tacitly confront us with our age-old thoughts and dreams about how technology will bring knowledge, how technology can produce and order revolutionary amounts of information, and how technology is capable of capturing the world (and the universe) in an objective and accurate way. With this, these discursive constructions subsequently lure us into believing that it is only logical that Big Data will, in this age of technological progress, undoubtedly realise these expectations. This means that these deeply ingrained ideas, with their invisible and unconscious functioning, as well as their seemingly commonplace character, have played a powerful role in the way very optimistic, and perhaps even utopian ideas, have been seamlessly projected onto the term in such a short period of time and on such a large scale.

Second, by building on the previous argument, it can be argued that the popularity of Big Data can be explained because it is not just one deeply ingrained idea about technology that has reappeared. Rather, during my media-archaeological study I have demonstrated that three recurring rhizomatic lines of thought can be identified in today's discourses on Big Data. This means that the various age-old and deeply ingrained ideas that were identified around the advent of the telescope, the microscope, the printing press, the camera obscura, the camera lucida, the calculator and the computer, have now, in this very specific moment in time with today's discursive construction of Big Data not only reappeared or started up again; they have also combined, connected and collided. By drawing upon the work of Deleuze and Guattari who argued that such a collision "necessarily changes in nature as it expands its connections" (Deleuze & Guattari 1987, 8), it can be argued that this connection of transhistorical ideas about technology is more than the sum of its parts, and has thus strengthened the concept of Big Data.

Thus, by conducting a media-archaeological study from a rhizomatic point of view, I have reconnected today's ideas with the past, as well as revealed and drawn connections between historical and contemporary discursive events that prove that today's enthusiasm and popularity for Big Data is not explainable because it might be *just* a well-chosen term, a coincidence or a smart marketing trick, but that it is rather a term that has been discursively constructed in such a way that it constantly and unconsciously (re)activates and connects transhistorical, age-old and deeply ingrained ideas about technology.

Next to giving an explanation for today's popularity of Big Data, this media-archaeological quest makes it also possible to evaluate and inquire today's Big Data discourse in new ways. The aforementioned conclusion that recurring and deeply ingrained ideas about technology serve as prefabricated moulds for today's reaction, gives rise to the question whether it is likely that Big Data is now, finally, after all these great attempts in the past, going to realise these age-old ideas, expectations and desires.

Current media-archaeological studies argue that it rarely happens that such recurring, transhistorical and age-old ideas, expectations, and desires are truly realised. For instance, the media archaeologist Eric Kluitenberg argues that "imagination and realization rarely coincide" (Kluitenberg 2011, 67). Additionally, media scholar Imar de Vries argues that "despite reality proving many utopian claims wrong, new communication technologies have been and still are presented and perceived as providing opportunities for finally realising

true democracy, cultural and social unification, the unlocking of all the world's knowledge, and so forth" (De Vries 2012, 165). Thus, by recognising that today's Big Data discourse consists of many age-old and deeply ingrained ideas, expectations and desires about technology, it needs to be pointed out that it is unlikely that Big Data will live up to the great expectations, however likely, probable and realistic these new possibilities may be presented today.

However, this does not imply that Big Data is just a useless buzzword without any valuable utilisations. On the contrary, the great amounts of digital data, as well as the great possibilities to collect, store, combine and analyse data today, will undoubtedly offer valuable and groundbreaking utilisations. As a matter of fact, we may even discover *some* new insights about the universe, as well as that we may predict *some* heart attacks and *some* criminal activities, just like we may make *some* better decisions based on data. However, Big Data will not deliver the long-desired and holistic knowledge that humanity has dreamed of for centuries. It is here that we find the "painful paradox of modern technology" (Bailey 2005, 199), because indeed, Big Data will just like the printing press, the telescope, the microscope, the camera obscura, the camera lucida, the calculator, and the computer succeed wildly, but yet, it will also fail miserably.

Conclusion

But if thought corrupts language, language can also corrupt thought. (Orwell, 1949)

This thesis was a journey that began with the current high expectations of Big Data, subsequently continued to be about the entanglement between the present and the past, about methodological choices and challenges, about age-old technological dreams and desires, and was completed with a media-archaeological explanation for today's popularity of Big Data as well as a critical evaluation of today's expectations.

My ultimate goal was to provide a broader perspective on how it can be explained that Big Data has, in a rather short period of time, experienced such a rapid increase in popularity that there is currently even talk of a Big Data revolution, a Big Data era or the age of Big Data. To achieve this, I have conducted a discourse analysis in the first chapter that evolved around the expectant discourse that has emerged about Big Data. I have argued that today's Big Data discourse is characterised by the idea that Big Data will provide profound new answers, insights and knowledge about almost anything, because it is now possible to analyse the revolutionary amounts of digital data in a very objective and accurate way. In the second chapter, I have inquired how a broader perspective on this expectant discourse can be gained. I have argued that a media-archaeological approach, with its anti-teleological starting point, as well as its acknowledgement that there is a recurring pattern in the way we respond to new media, offers a profound theoretical explanation for the popularity of Big Data. In this field it is argued that we respond to new media the way we do because recurring, age-old and deeply ingrained ideas about technology provide a prefabricated discursive mould for today's reaction. However, I have also pointed out that there were some differences in this field, after which I have argued that the rhizome, as an alternate image of thought, introduced by Deleuze & Guattari (1987), offered the necessary additional guidelines, while at the same time, staying very close to the media-archaeological approach. With the rhizome, I have conceptualised today's expectant Big Data discourse as part of a large rhizomatic map filled with lines that represent deeply ingrained ideas about

technology that are capable of connecting, colliding, breaking off and starting up again over time.

In my media-archaeological quest I have subsequently examined what deeply-ingrained ideas have now, with Big Data, (re)appeared, started up again, connected and collided. It was argued that today's Big Data discourse is everything but new since the telescope, the microscope, the printing press, the camera obscura, the camera lucida, the calculator and the first computer gave rise to very similar discourses with incredibly similar ideas, assumptions, expectations, dreams and desires about technology.

Most importantly, I have argued that these historical discourses explain today's popularity of Big Data. That is because the discourses that surround this term, at this current moment in time, are capable of seamlessly (re)activating and connecting these age-old and deeply ingrained ideas about technology, so that we unconsciously recognise and project them onto the term in the blink of an eye. This means that Big Data became so big because its discursive construction anticipates and responds to these dateless ideas, expectations, assumptions, dreams and desires about technology that have been part of our collective and subconscious memory long before the term 'Big Data' was even coined.

Moreover, I have concluded that, once we are aware of these deeply ingrained ideas about technology, it is possible and necessary to inquire today's expectations of Big Data in new ways. Exactly because it rarely happens that such age-old ideas are truly and finally realised, I have argued that it is highly unlikely that Big Data is now, finally, after all these great technological attempts in the past, going to bring the long-desired insights, answers and knowledge about the universe, alien life, crime, heart attacks and much more.

This explanation and evaluation of today's popularity of Big Data has made me, more than ever, realise the powerful role of language in creating a dominant understanding of the next big technological thing in two ways. The first, and the one I have already described for a large part, is the way that words do not only capture the world around us, but also actively construct it. The term 'Big Data' and the words that have been used to talk and write about it, did not merely describe a 'true' and 'real' representation of a new technological development, rather, these words have also constructed a dominant meaning of Big Data, and with that, a collective understanding of a new technological development.

Moreover, if one would compare Big Data to previous technologies such as the computer, the television or the radio, it can be argued that Big Data even

reinforces this discursive power. Unlike the computer, the television or the radio, Big Data does not seem to refer to a tangible or ontological object, and can therefore be particularly shaped and constructed through discourse. As a matter of fact, one might even wonder to what extent Big Data actually exists outside of discourse. With this, I do not imply to inquire the existence of the current large amounts of digital data, nor the existence of software tools that are used to analyse data. I imply to inquire to what extent the Big Data revolution, the Big Data era, and the age of Big Data, and thus Big Data's socially and linguistically constructed reality can be said to actually exist. This is in agreement with the way the media archaeologist Eric Kluitenberg argues that imaginary media are, at times, "not recognised as imaginary by their inventors, producers, or users, who perceive the imaginary constructs as 'realized' and 'actual' even if no evidence exists to support the functional characteristic ascribed to them" (Kluitenberg 2011, 54). With this, it can be argued that the term 'Big Data' may even show characteristics of an imaginary medium, because its seemingly realistic promises predominantly consist of a collection of, yet unrealised, transhistorical and deeply ingrained ideas, desires and expectations about technology.

This brings me to the second way in which language has proven to play a powerful role in the way dominant understandings about technology may emerge. The fact that age-old and transhistorical ideas, and thus past discursive constructions of media, can still function as prefabricated moulds for the way we think and talk about new media today, demonstrates that our understanding of new technological developments is not only constructed by the language we use today, but that it is also shaped by the language that has been used to talk about technology forever.

It is at the same time remarkable that these age-old, transhistorical and deeply ingrained ways of thinking and talking about media are often not consciously recognised, but somehow make very optimistic, and perhaps even utopian promises appear as common sense. It is here that our modern culture, with its progressivist spirit, as well as its assumptions of radical breaks with the past, makes us perhaps blind for the paradoxical fact that the oldest of the old is presented as the newest of the new with Big Data. Reconnecting these seemingly new ideas with the past, and pointing out what "modern culture has repressed and disavowed" (Robins, 1996, 161), is therefore an increasingly important task, and has proven to be a very valuable and fertile strategy for explaining as well as evaluating today's popularity of Big Data.

However, by critically reflecting on this chosen research strategy, two important points cannot be left undiscussed. First, reconnecting current ideas with the past is inevitably a practice of studying historical events with knowledge of the present, and therefore always comes with the danger of unintentionally implying a linear and teleological reconstruction of the past. It is here that I have attempted to reduce this risk by employing a media-archaeological and rhizomatic framework that has conceptualised today's ideas about Big Data as part of a large map, with no beginning or end, that is filled with lines of thought that do not logically evolve, but connect, start up and break off in an unpredictable and seemingly random fashion. Second, I wish to point out that the central theoretical and media-archaeological proposition about the way recurring and transhistorical discourses provide prefabricated discursive moulds for today's responses to new technologies, may perhaps contain a too deterministic and totalistic hypothesis about the role and existence of collective and unconscious ideas. A future empirical research that investigates the powerful role of transhistorical ideas, expectations, dreams and desires about technology would therefore contribute to a more thorough understanding of the way we logically and naturally project these transhistorical ideas onto new technologies today.

Additionally, I wish to conclude this thesis by adding one final note and widen the scope of this thesis, because it is not only capable of explaining and inquiring the popularity of Big Data. Rather, this thesis is also capable of gaining a broader understanding of many more popular terms, in both the past and the future. It can be argued that all popular catch phrases actually catch on because the newest of the new is constantly packaged in the oldest of the old, and with that, relates to, and reactivates our deeply ingrained ideas about technology. By continually recognising such transhistorical and deeply ingrained ideas about technology, rather than uncritically and unconsciously projecting them onto these terms, we are able to assess a newly presented technology at its true value. These deeply ingrained ideas can often be identified every time we encounter a sense of *déjà vu* and a sense of 'this-is-too-good-to-be-true'. This means that, every time discourses are encountered wherein new popular terms are depicted as new technologies that will either realise mass participation, true democracy, perfect communication, social equality, or any other ideal, however realistically this may be presented, we need to be aware of the fact that certain deeply ingrained ideas provide a discursive mould. It is therefore necessary to keep reflecting on these complex discursive constructions as well as inquire the often unrealistic

promises, because, in the end, these terms play a big and non-neutral role in the way we make sense of new technologies and the way we subsequently choose to act upon this.

References

- Abbas, Niran Bahjat. 2006. *Thinking Machines: Discourses of Artificial Intelligence*. Münster: LIT Verlag Münster.
- Adams, Douglas. 1979. *The Hitchhiker's Guide to the Galaxy*. Pan Books.
- Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired*. http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.
- Bailey, Lee Worth. 2005. *The Enchantments of Technology*. Illinois: University of Illinois Press.
- Basu, Atanu. 2014. "What the Frack: U.S. Energy Prowess With Shale, Big Data Analytics." *Wired*. <http://www.wired.com/2014/01/big-data-analytics-can-deliver-u-s-energy-independence/>.
- Boddy, William. 1994. "Archaeologies of Electronic Vision and the Gendered Spectator." *Screen* 35 (2): 105–122.
- Bolter, Jay David, and Richard Grusin. 1999. *Remediation: Understanding New Media*. Cambridge (MA): MIT Press.
- Boomen, van den, Marianne. 2014. *Transcoding the Digital: How Metaphors Matter in New Media*. Amsterdam: Institute of Network Cultures.
- boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15 (5).
- Brecht, Bertolt. 1932. "The Radio as an Apparatus of Communication" ['Der Rundfunk Als Kommunikationsapparat']." *Blaetter Des Hessischen Landestheaters* (16).
- Brennen, Bonnie S. 2013. *Qualitative Research Methods for Media Studies*. New York City: Routledge.
- Briggs, Asa, and Peter Burke. 2010. *Social History of the Media: From Gutenberg to the Internet*. Cambridge UK: Polity.
- Broad, William J. 2007. *The Oracle: Ancient Delphi and the Science Behind Its Lost Secrets*. New York City: Penguin Press.
- Burke, James. 2009. *The Day The Universe Changed*. New York City: Hachette Book Group.

- Bush, Vannevar. 1945. "As We May Think." *The Atlantic Monthly*: 101–108. <http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>.
- Columbus, Louis. 2014. "2014: The Year Big Data Adoption Goes Mainstream In The Enterprise." *Forbes*. <http://www.forbes.com/sites/louiscolumbus/2014/01/12/2014-the-year-big-data-adoption-goes-mainstream-in-the-enterprise/>.
- Cormack, Mike. 1995. *Ideology*. Ann Arbor: University of Michigan Press.
- Crary, Jonathan. 1992. *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*. Cambridge (MA): MIT Press.
- Deacon, David, Michael Pickering, Peter Golding, and Graham Murdock. 1999. *Researching Communications: A Practical Guide to Methods in Media and Cultural Analysis*. 2e ed. Bloomsbury: Hodder Arnold.
- Deleuze, Gilles, and Félix Guattari. 1987. "Introduction: Rhizome." In *A Thousand Plateaus: Capitalism and Schizophrenia*, edited by Brian Massumi, 1–25. Minneapolis: University of Minnesota Press.
- Dijk, Jan van. 2006. *The Network Society. Social Aspects of New Media*. London, Thousand Oaks, CA: Sage.
- Douglas, Susan J. 1986. "Amateur Operators and American Broadcasting: Shaping the Future of Radio." In *Imagining Tomorrow: History, Technology, and the American Future*, edited by J. Corn, 35–57. Cambridge: MIT Press.
- . 1989. *Inventing American Broadcasting, 1899-1922*. Baltimore: Johns Hopkins University Press.
- Dumbill, Edd. 2012. "What Is Big Data? An Introduction to the Big Data Landscape." *O'Reilly Media*. <http://strata.oreilly.com/2012/01/what-is-big-data.html>.
- Durepos, Gabrielle, and Albert J. Mills. 2012. "Actor-Network Theory, ANTi-History and Critical Organizational Historiography." *Organization* 19 (6): 703–721.
- Fairclough, Norman. 1989. *Language and Power*. New York: Longman Inc.
- Faulkner, William. 1950. *Requiem for a Nun*. New York: Random House.
- Flichy, Patrice. 2007. *The Internet Imaginaire*. Cambridge (MA): MIT Press.
- Foucault, Michel. 1972. *The Archaeology of Knowledge*. New York City: Pantheon Books.

- . 1981. "The Order of Discourse." In *Untying the Text: A Post-Structuralist Reader*, edited by Robert Young, 48–78. Boston: Routledge.
- Galison, Peter. 2010. "The Objective Image."
- Gallagher, David F. 2009. "Talk to The Times: Deputy Technology Editor David Gallagher." *The New York Times*. http://www.nytimes.com/2009/10/28/business/media/28askthetimes.html?pagewanted=all&_r=0.
- Gallagher, Fred. 2013. "Is Big Data the Next Big Brother?" *Wired*. <http://www.wired.com/insights/2013/03/is-big-data-the-next-big-brother/>.
- Gartner. 2013. "Gartner's 2013 Hype Cycle for Emerging Technologies Maps Out Evolving Relationship Between Humans and Machines." *Gartner.com*. <http://www.gartner.com/newsroom/id/2575515>.
- Gillespie, Richard. 2009. "The Great Melbourne Telescope: A Contextual History." <http://extras.springer.com/2010/978-90-481-2232-5/Gillespie-The-Great-Melbourne-Telescope-A-Contextual-History.pdf>.
- Gillespie, Tarleton. 2013. "The Relevance of Algorithms." <http://governingalgorithms.org/wp-content/uploads/2013/05/1-paper-gillespie.pdf>.
- Google. 2014a. "'Big Data' - Google Search." https://www.google.com/?gfe_rd=cr&ei=ntwAU_f7LsuvOrq0gIgJ&gws_rd=cr#q='Big+Data'&tbs=qdr:y.
- . 2014b. "Google Trends - Web Search Interest: Big Data - Worldwide 2004 - Present." *Google Trends*. <http://www.google.com/trends/explore#q=Big+Data>.
- . 2014c. "Google Trends - Interesse in Google Zoeken: Big Data - 2008, 2011, 2012, 2013, Feb 2014 - Wereldwijd." *Google Trends*. <http://www.google.nl/trends/explore#q=Big+Data&date=1/2008+12m,1/2011+12m,1/2012+12m,1/2013+12m,2/2014+1m&cmpt=date>.
- Gunning, Tom. 1991. "Heard over the Phone: The Lonely Villa and the de Lorde Tradition of the Terrors of Technology." *Screen* 32 (2): 184–196.
- Hardy, Quentin. 2011. "The Big Business of 'Big Data.'" *The New York Times*. <http://bits.blogs.nytimes.com/2011/10/24/big-data/>.
- . 2012. "Big Data's Grass-Roots Revolution." *The New York Times*. http://bits.blogs.nytimes.com/2012/03/02/big-datas-grassroots-revolution/?_r=0.

- . 2013. "Why Big Data Is Not Truth." *The New York Times*. <http://bits.blogs.nytimes.com/2013/06/01/why-big-data-is-not-truth/>.
- . 2014. "Big Data Shrinks to Grow." *The New York Times*. <http://bits.blogs.nytimes.com/2014/01/01/big-data-shrinks-to-grow/>.
- Hernandez, Daniela. 2012. "Big Data Is Transforming Healthcare." *Wired*. <http://www.wired.com/2012/10/big-data-is-transforming-healthcare/>.
- HP. 2014. "Harness the Power of Big Data." *HP.com*. <http://www8.hp.com/us/en/business-solutions/big-data.html>.
- Huhtamo, Erkki. 1996. "From Kaleidoscomaniac to Cybernerd: Towards an Archeology of the Media." In *Electronic Culture. Technology and Visual Representation*, edited by T Druckrey, 297–303. New York: Aperture.
- . 2011. "Dismantling the Fairy Engine: Media Archaeology as Topos Study." In *Media Archaeology. Approaches, Applications and Implications*, edited by E Huhtamo and J Parikka, 2–47. Berkeley: University of California Press.
- Huhtamo, Erkki, and Jussi Parikka. 2011. *Media Archaeology: Approaches, Applications and Implications*. Berkeley: University of California Press.
- IBM. 2012. "Charleston Police Department Aims to Reduce Crime Using IBM Predictive Analytics." *IBM.com*. <http://www-03.ibm.com/press/ie/en/pressrelease/37985.wss>.
- . 2013. "\$2 Million Awarded to Sutter Health, IBM and Geisinger Health System to Study Heart Failure Prediction." *IBM.com*. <http://www-03.ibm.com/press/us/en/pressrelease/42156.wss>.
- . 2014a. "Big Data in Action." *IBM.com*. <http://www-01.ibm.com/software/data/bigdata/industry.html>.
- . 2014b. "From Big Bang to Big Data." *IBM Research*. <http://www.research.ibm.com/news.shtml>.
- Intel. 2012. "What Happens In An Internet Minute?" *Intel.com*. <http://www.intel.com/content/www/us/en/communications/internet-minute-infographic.html>.
- ITU (International Telecommunications Union). 2013. "Key ICT Indicators for Developed and Developing Countries and the World (totals and Penetration Rates)." http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2013/ITU_Key_2005-2013_ICT_data.xls.

- Kluitenberg, E. 2011. "On the Archaeology of Imaginary Media." In *Media Archaeology. Approaches, Applications and Implications*, edited by E Huhtamo and J Parikka, 48–69. Berkeley: University of California Press.
- Lakoff, George, and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lammes, S. 2003. "The Gendering Televisual Technologies in Science Fiction (1925-1939)." In *Mapping the Margins: Identity, Politics, and the Media*, edited by K. Ross and D. Derman, 67–79. Cresskill, NJ: Hampton Press.
- Lanier, Jaron. 2013. *Who Owns the Future?* New York City: Simon & Schuster.
- Lardinois, Frederic. 2012. "Forrester: 84% Of U.S. Adults Now Use The Web Daily, 50% Own Smartphones, Tablet Ownership Doubled To 19% In 2012." *Techcrunch.com*. <http://techcrunch.com/2012/12/19/forrester-84-of-u-s-adults-now-use-the-web-daily-50-own-smartphones-tablet-ownership-doubled-to-19-in-2012/>.
- Lin, Thomas. 2013. "Big Data Is Too Big for Scientists to Handle Alone." *Wired*. <http://www.wired.com/wiredscience/2013/10/big-data-science/>.
- Lister, M, J Dovey, I Grant, and K Kelly. 2009. *New Media: A Critical Introduction*. London, New York: Routledge.
- Lohr, Steve. 2011. "Big Data: Sorting Reality From Hype." *The New York Times*. <http://bits.blogs.nytimes.com/2011/09/30/big-data-sorting-reality-from-the-hype/>.
- . 2012. "The Age of Big Data." *The New York Times*. <http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?pagewanted=all>.
- . 2013a. "Sizing Up Big Data, Broadening Beyond the Internet." *The New York Times*. <http://bits.blogs.nytimes.com/2013/06/19/sizing-up-big-data-broadening-beyond-the-internet/>.
- . 2013b. "Big Data and a Renewed Debate Over Privacy." *The New York Times*. <http://www.nytimes.com/2013/03/24/technology/big-data-and-a-renewed-debate-over-privacy.html?pagewanted=all>.
- Lucid, L. 2010. *BBC Horizon (1964) with Arthur C. Clarke (Part 1 of 2)*. http://www.youtube.com/watch?feature=player_embedded&v=KT_8-pjuctM.
- Lyotard, Jean-François. 1984. *The Postmodern Condition: A Report on Knowledge*. Minneapolis: University of Minnesota Press.

- Macey, David. 1994. "Introduction." In *The Four Fundamental Concepts of Psycho-Analysis*, edited by Jacques Lacan. London: Karnac Books.
- Manovich, Lev. 2011. "Trending: The Promises and Challenges of Big Social Data." http://www.manovich.net/DOCS/Manovich_trending_paper.pdf.
- Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. 2011. "Big Data: The next Frontier for Innovation, Competition, and Productivity." *McKinsey & Company*. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.
- Marvin, Carolyn. 1988. *When Old Technologies Were New: Thinking about Electric Communication in the Late Nineteenth Century*. Oxford: Oxford University Press.
- Mayer, Paul. 1999. *Computer Media and Communication*. Oxford: Oxford University Press.
- Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Eamon Dolan.
- McLuhan, Marshall. 1962. *The Gutenberg Galaxy: The Making of Typographic Man*. Toronto: University of Toronto Press.
- . 1964. *Understanding Media: The Extensions of Man*. New York: McGraw-Hill.
- Morozov, Evgeny. 2013a. "The Meme Hustler: Tim O'Reilly's Crazy Talk." *The Baffler*. http://thebaffler.com/past/the_meme_hustler.
- . 2013b. *To Save Everything, Click Here: Technology, Solutionism and the Urge to Fix Problems That Don't Exist*. London: Allen Lane.
- Natale, Simone. 2012. "Understanding Media Archaeology." *Canadian Journal of Communication* 37: 523–527.
- Nolan, Christopher. 2010. *Inception*.
- Novak, M. 2007. "The Answer Machine (1964)." *Paleofuture*. <http://www.paleofuture.com/blog/2007/5/23/the-answer-machine-1964.html>.
- Openculturevideo. 2012. *Arthur C. Clarke Predicts the Internet & PC*. <http://www.youtube.com/watch?v=OIRZebE8O84>.
- Oracle. 2014. "The Power to Transform Your Business." *Oracle.com*. <http://www.oracle.com/us/technologies/big-data/index.html>.
- Orwell, George. 1949. *Nineteen Eighty-Four*. London: Secker and Warburg.

- Parikka, Jussi. 2012. *What Is Media Archaeology*. Cambridge: Polity Press.
- Pearlstein, Joanna. 2013. "Information Revolution: Big Data Has Arrived at an Almost Unimaginable Scale." *Wired*. <http://www.wired.com/magazine/2013/04/bigdata/>.
- Reeves, Eileen Adair. 2008. *Galileo's Glassworks: The Telescope and the Mirror*. Cambridge MA: Harvard University Press.
- Robins, Kevin. 1996. "The Touch of The Unknown." In *Into the Image: Culture and Politics in the Field of Vision*, 11–36. New York City: Routledge.
- Rogers, Richard. 2012. "Mapping and the Politics of Web Space." *Theory, Culture & Society* 29 (4-5): 193–219.
- Schäfer, Mirko Tobias. 2011. *Bastard Culture! How User Participation Transforms Cultural Production*. Amsterdam: Amsterdam University Press.
- Spigel, Lynn. 1992. *Make Room for TV: Television and the Family Ideal in Postwar America*. Chicago: University of Chicago Press.
- Steinberg, David. 2013. "Unlocking the Big Data Goldmine for SMBs." *Wired*. <http://www.wired.com/insights/2013/04/unlocking-the-big-data-goldmine-for-smbs/>.
- Time Magazine. 1993. "The Info Highway: Bringing a Revolution in Entertainment, News, and Communication." *Time Magazine*, April. http://www.atlanticfinancial.com/about-atlantic-financial/time_magazine_information_highway.htm.
- Vance, Ashlee. 2010a. "EMC Takes a Big-Data Swipe at Oracle, H.P. and I.B.M." *The New York Times*. <http://bits.blogs.nytimes.com/2010/10/13/emc-takes-a-big-data-swipe-at-oracle-h-p-and-i-b-m/>.
- . 2010b. "Start-Up Goes After Big Data With Hadoop Helper." *The New York Times*. <http://bits.blogs.nytimes.com/2010/04/22/start-up-goes-after-big-data-with-hadoop-helper/>.
- Vries, de, Imar. 2012. *Tantalisingly Close: An Archaeology of Communication Desires in Discourses of Mobile Wireless Media*. Amsterdam: Amsterdam University Press.
- Wardrip-Fruin, Noah, and Nick Montfort. 2003. "From Computer Lib / Dream Machines." In *The New Media Reader*, edited by Noah Wardrip-Fruin and Nick Montfort, 301–338. Cambridge (MA): MIT Press.

- Yarow, Jay. 2012. "51% Of People Think Stormy Weather Affects 'Cloud Computing.'" *Business Insider*. <http://www.businessinsider.com/people-think-stormy-weather-affects-cloud-computing-2012-8>.
- Zielinski, Siegfried. 2006. *Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technical Means*. Cambridge (MA): MIT Press.
- Zikopoulos, Paul C., Chris Eaton, Dirk DeRoos, Thomas Deutsch, and George Lapis. 2012. *Understanding Big Data*. New York: McGraw-Hill.