

Bedankt voor het downloaden van dit artikel. De artikelen uit de (online)tijdschriften van Uitgeverij Boom zijn auteursrechtelijk beschermd. U kunt er natuurlijk uit citeren (voorzien van een bronvermelding) maar voor reproductie in welke vorm dan ook moet toestemming aan de uitgever worden gevraagd.

Boom

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch door fotokopieën, opnamen of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van kopieën uit deze uitgave is toegestaan op grond van artikelen 16h t/m 16m Auteurswet 1912 jo. Besluit van 27 november 2002, Stb 575, dient men de daarvoor wettelijk verschuldigde vergoeding te voldoen aan de Stichting Reprorecht te Hoofddorp (postbus 3060, 2130 KB, www.reprorecht.nl) of contact op te nemen met de uitgever voor het treffen van een rechtstreekse regeling in de zin van art. 16l, vijfde lid, Auteurswet 1912.

Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16, Auteurswet 1912) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten, postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/pro).

No part of this book may be reproduced in any way whatsoever without the written permission of the publisher.

info@boomamsterdam.nl
www.boomuitgeversamsterdam.nl

ARTICLES

Big data, grand challenges: On digitization and humanities research¹

*José van Dijck**

Introduction

When I started working on my PhD, in 1987 at the University of California, San Diego (US), I wanted to examine how a public debate on a controversial issue results in consensus. My PhD thesis resulted in a book on the course of the public debate around in vitro fertilization in the US news media between 1978 and 1985 (Van Dijck, 1995). Due to the enormous supply of newspapers and journals it was infeasible to retrieve all information from this debate, so selection was necessary. Fortunately, I encountered a (private) archive of an institute that had documented very systematically (though not exhaustively) clippings about this subject. Audio-visual sources were almost impossible to collect and, even if I would have had them, I would have lacked the time to plow through all of them. Thus my corpus was limited, and within this limitation, I had to show my mastery. The interpretative approach I chose turned out to be an excellent exercise in analyzing a public debate. The most important lesson from that aptitude test, now twenty-five years ago: available data determine the nature of the research question as well as the set of instruments with which you can query the sources.

But times have changed. Over the past ten years, we, as humanities researchers, have been getting increasing numbers of, more diverse, data and ever larger databases at our disposal: digitization has added an altogether new dimension to the pre-existing materiality of sources, as a result of which we can do research on a much larger scale, encompassing many more different types of sources. This not only means that we can adjust our research questions, but also that we have to develop new instruments to answer those questions. The reverse also holds: new instruments enable questions that we previously could not ask due to physical limitations. As a matter of fact, that is not a new phenomenon in science. Without the Hubble telescope, astronomers could not have come up with certain questions about the stars; without the particle accelerator there would be no Higgs particle; and without DNA sequencers, the quest for the human genome would probably have proven fruitless.

* Prof. dr. José van Dijck is professor of Media Studies at the University of Amsterdam and currently President of the Royal Netherlands Academy of Arts and Sciences. Email: Jose.van.dijck@knav.nl.

New questions, new instruments

Humanities scholars have always been researching human culture. They pose fundamental questions, such as: why have some regions of the world been rich for such a long time and others poor? Why does persistent imaging of certain minorities continue to exist in public debates? How does language change under the influence of migration? For decades, those questions have been asked and answered by historians, media scientists, linguists, and many other researchers. Humanities scholars are very good at interpreting content, in particular of separate data, each in their own field of research. Historians work with data from archives and with structured data, for instance emanating from municipal archives or institutions such as Statistics Netherlands. Linguists draw from large textual and oral datasets. Media experts use textual and audiovisual material from newspapers, journals, radio, television, and, increasingly, Internet sources and social media.

Humanities scholars, one could say, each study in their own way building blocks of culture and patterns of culture change. The building blocks they have traditionally been working on (text, images, sound, and historical data) were – and are – numerous and fragmented. Because of that, many humanities scholars focus on a single piece of the puzzle in order to interpret and analyze that as well as possible. For instance, the work of a single painter, the novels of a single author, the figures from municipal archives of a single historical period, or the language use of a single social group. In my own PhD thesis I did precisely that: focusing my research on a single source (written media texts) from a limited time period. After all, the available data and the limited time I had available prompted my choice for a qualitative approach to the public debate, because a large-scale study of source materials was simply not an option.

The interesting aspect of digital search engines is that they stimulate complex questions. My own limited question about IVF and the US news media between 1978 and 1985 was based on a much broader curiosity, namely: how do public debates about controversial issues lead to consensus or normalization? Such a complex question requires a coherent insight into socio-historical developments, imaging, shifting norms, values, and laws over a longer time period. Not as separate phenomena, but as a complex whole. To tackle such a question, I could hardly limit myself to my own field of expertise; I would have to expand the scope and diversity of sources, because of which I would no longer be able to do the work on my own. Humanities scholars have traditionally been used to working with sources that can be evaluated and interpreted. We still do not have much of a tradition of cooperating with multi-disciplinary teams in which a larger diversity of sources and methods are put on the table.

Digital Humanities and the digital turn

Over the next few years, researchers and heritage institutions (archives, libraries, sound and vision, knowledge centers) will be faced with a common challenge. The size of digitized files has increased exponentially. In addition, new 'born digital'

sources have come into being, such as blogs, webpages, and social media – all expressions of culture that we cannot ignore if we want to study culture or culture change seriously. The amount of ‘data’ or digital content has increased to such an extent that we have started to speak of Big Data – no matter how problematic this term is.² To mine this wealth of material, new instruments need to be developed: instruments to query data for meaningful content. With this, not only the objects of research change, but also the methods of humanities scholars. In recent years, we have been speaking of ‘Digital Humanities’ (DH) when we are talking about the digitization of sources and the adjustment of our research methods to these developments. The term DH includes many disciplines, is broad as well as specialized, and has evoked both euphoria and resistance. I would like to dwell on a couple of comments.

Now that more and larger information files can be searched in an automated fashion, it becomes possible to ask questions relating to longer time periods and more types of sources. More data does not by itself mean more knowledge or better insights. In fact, it mainly means: more interpretation and the possibility to combine different methods. We want to be able to ask new meaningful questions and to substantiate possible answers with a range of sources. To return to my previous example: if I were to conduct the same research into the public opinion around IVF in the Netherlands at this moment in time, I would have a much richer range of data at my disposal: digitized newspapers and journals at the Royal Library of the Netherlands; audiovisual files at the Netherlands Institute for Sound and Vision; but for example also the minutes of the Dutch parliament, where decisions were taken concerning reproductive techniques – decisions that had a political nature and therefore caused much debate.³

But to mine all those databases and to see the interpretations of these sources in conjunction, I need new instruments. In fact, I also need the help of colleagues: not just colleagues from within the humanities who know all there is to know about textual, imaging, and other data, but also from computer scientists, so as to be able to search and query the sources; and from social scientists, for the use of analytical methods, such as discourse and network analysis. In order for the pieces of the puzzle to fit – pieces from the field of language, pictures, moving image, sound, and historical data – experts must learn from one another how they can mine these databases for their research.

CLARIAH

Over the past three years, a number of important instruments have been developed in the different branches of the humanities; last year a number of researchers from the humanities conceived the plan to develop a common infrastructure. CLARIAH (common lab research infrastructure for the arts and humanities) is a common project initiated by a core team of scientists – supported by a consortium of forty knowledge and heritage institutions, public organizations, and companies – that the Netherlands Organisation for Scientific Research (NWO) recently rewarded with a subsidy of twelve million euros. With this money for a common infrastruc-

ture, humanities scholars can not only develop digital instruments to mine large databases; by making these instruments ‘communicate’, humanities scholars also learn to cooperate to answer those complex questions. Three areas play a leadership role in CLARIAH: linguistics, media studies, and socio-economic history. Linguists concentrate primarily on the mining of digital text files. Media studies experts mainly develop tools for the interpretation of audiovisual sources. And socio-economic historians focus on structured databases from archives. But the tools to be developed must be useful for *all* researchers who work with different types of digital data. Linguists use audiovisual sources for examining changing patterns in spoken language. And when I research public debates, I have to deal with textual, audiovisual, as well as structured data.

As an infrastructure, CLARIAH aims to contribute something essential to the larger scientific questions, both within and beyond the humanities. The project is also designed to deliver building blocks complementary to the work of natural and social scientists in the field of data mining. Where computer scientists excel in the design of search algorithms, and social scientists want to know everything about the behavior of users, the power of humanities scholars lies in interpreting human messages in digital content. Big Data in the humanities are primarily rich data: they are full of noise, just like culture is full of noise. Figures on poverty are not facts, but ask for interpretation. Opinions in a public debate are numerous but also diffuse – they have differing densities and impact. And images or texts can be ironic or ambiguous. Whoever studies culture knows that content needs interpretation, and that messages only gain meaning in their conjunction. To understand such complexity of content – therein lies the contribution of humanities scholars to the research into large amounts of digital data. As such, CLARIAH implies an even more intense cooperation between humanities, social science, and natural science where the understanding of cultural complexity is concerned.

Challenges and critical comments

A project such as CLARIAH provides humanities scholars with great challenges, and at the same time it raises important questions about the nature, utility, and necessity of our research. Digital humanities, whatever this means exactly, is *not* a revolution and it does not in and by itself offer solutions for a better world or even better science. Each time period develops the instruments that are necessary to understand the world at that moment. That is why I want to elaborate on four important challenges, which at the same time raise critical questions:

- 1 the ‘digital turn’ and the ‘push’ of automated, quantitative research;
- 2 the necessary combination of qualitative and quantitative methods;
- 3 the dilemma of multidisciplinary cooperation;
- 4 the ideological question of why the humanities should be concerned with computers and digitization at all, rather than exclusively with archives, books, and the content thereof.

1. Quantification and the digital turn

Challenge number one concerns the quantification and automation of humanities research. Digitization has caused large quantities of information to become available and searchable; all this source material requires new and supplementary research methods. Because so many more data are available, we can search patterns over longer time periods, from more kinds of sources. We can distill certain stylistic patterns or characteristics of authorship from large text data sets. Analyzing structured data on painters, buyers, and traders in 17th-century Amsterdam, we can reconstruct networks in order to find out how this ‘creative industry’ functioned and influenced cultural production at the time. Another example concerns the changing public image of minorities over the past five decades, for which we can analyze large numbers of audiovisual data from the National Archive of Sound and Vision and national newspaper archives.

As it happens, a number of colleagues has already developed instruments to pursue this latter type of research. In various pilots, Jasmijn van Gorp and Pieter Vijn have demonstrated how the archives of the Netherlands Institute for Sound and Vision can be searched for specific debating themes. With the help of TROVe, they analyze the spread of contemporary news through various media (TV, radio, online newspapers, blogs, and Twitter), while AV Researcher XL enables digital content analysis by searching through TV subtitles and newspapers.⁴ Both instruments were applied by Jasmijn van Gorp to examine the course of East European migration, more specifically the public images of Poles and Romanians in the debate about labor migration. With these tools, the analysis of that debate becomes comprehensive, including both audiovisual and textual media sources, so the analysis of a debate is less dependent on an arbitrary sample, the way I was in the 1990s. Moreover, and this is really new, using a search engine such as TROVe it becomes instantly clear who the most important players are in such a debate because we can now directly connect content and context.

In order to make large amounts of (audiovisual and textual) data searchable, a number of instruments already exist, but this is only the beginning: much still needs to be developed and customized. Those new methods of research are often quantitative or computational. Critics of ‘Digital Humanities’ often remark that the digitization and quantification of sources and methods may in fact restrict humanities research: after all, the model-based fashion in which digital sources must be searched determines the kind of questions that can be asked. Historians Piersma and Hiddens (2013) argue that querying large quantities of digital sources is (too) strongly oriented toward making hypotheses testable or automatable. That critique may be partially justified, but not fully: computational tools do exist that are specifically focused on qualitative analysis.⁵ Some humanities scholars fear that quantitative or automated methods of the ‘Digital Humanities’ crowd out other (qualitative) approaches in the humanities, but, and this brings me to my second point, that creates a fallacious contradiction.

2. The necessary combination of quantitative and qualitative methods

In the field of ‘digital humanities’, computational methods are very often combined with qualitative methods, especially in the various stages of research. Quantitatively obtained data can be interpreted with qualitative methods, such as text analysis, ‘close reading’, or image analysis. Especially in the explorative phase of a research project it can be useful, for instance, to take a representative sample of the material and to juxtapose this to all available data and visualize the results. Questions and instruments are never ready-to-use; they are always developed in relation to each other. And, as was always the case, qualitative interpretation is indispensable in the use of digital methods. In the first place, this happens by applying sharp source criticism to both the tools and the underlying data, and to the conjectures that underlie both. Databases do not speak for themselves; they are not thermometers of society. As every archivist knows, knowledge about the origin of every collection is essential for weighing and understanding its content, in particular when those data are ‘born digital’.⁶ Counting words without knowing the difference in density between, let’s say, an opinion article from 1972 and an advertorial from 2008, disqualifies a researcher. Being able to recognize the ambiguity of a concept such as ‘*verzuiling*’ (‘pillarization’, or societal compartmentalization) in different decades of the last century is as important as recognizing fifty shades of grey for painters or fifty meanings of snow for Eskimos.

In the era of Big Data, interpretation may be more important than ever before. The instruments with which you gather and prepare your dataset are all but value-free: you have to know what precedes source selection and disclosure. New sources and instruments create new possibilities and restrictions; strengths and weaknesses of old *and* new instruments must be better aligned. Maybe a comparison with medical science helps illustrate this argument. The invention of the MRI scanner made the inside of the human body accessible in a three-dimensional fashion to the eye of the physician. That did not mean in any way that the X-ray, the CT scanner, or ultrasound became superfluous; each device enabled a different diagnostic. Moreover, interpretation of those scans was anything but automatic: on the contrary, years of interpretation, comparison, and adjustment of the instruments were spent on the fine-tuning of instrument, methods, and interpretation (Van Dijck, 2005). Or, as my colleague Julia Noordegraaf observed: we need both the telescope and the microscope to address such fundamental questions about human culture.

The connection of quantitative and qualitative methods, of computational and interpretive instruments, poses a novel challenge to the humanities. The digitization of a rich diversity of sources in no way means that we homogenize or equalize all methods (Svensson, 2012). We continue to use text interpretation and network analysis alongside TROVe or CLIO Infra. Together those diagnostics of the humanities deliver a spectrum of instruments, which we need to investigate the increasingly complex multimedial articulations. But the use of each of those instruments also raises critical questions: why do we use which instrument in which cases? And what does this contribute to the grander challenge to the humanities?

3. Multidisciplinary cooperation

This brings me to the third point of my argument: the cooperation between various disciplines within and outside of the humanities. I have already said a number of things about the cooperation between humanities scholars in the context of CLARIAH. But let me here focus specifically on multidisciplinary cooperation outside the humanities, and then in particular that with computer scientists. There is a kind of indeterminate fear among humanities scholars that the humanities will (also) be taken over by natural scientists once we take the ‘digital humanities’ turn. As some argue, computational thought – thinking in code, programming languages, and algorithmic reasoning – would be incompatible with critical-analytical thought, and the latter threatens to be dominated by the first. Although I thought we had moved beyond the ‘two cultures divide’ since 1959, one can see C.P. Snow’s phantoms re-appear on stage at least once every decade. In the context of the ‘digital humanities’ debate, critics such as Stanley Fish (2012) exorcize the computer science demons by sending them back ‘into their cages’, and admonish humanities scholars, urging them to resist the digital turn. But it is nonsensical to draw sharp boundaries between the two cultures – the computational and the critical hermeneutic. I cannot put it better than Federica Frabetti, so I cite:

“[C]omputation and the humanities cannot be thought as two separate entities whose relations can be defined once and for all. ... In fact, the ability to question inherited conceptual frameworks regarding technology might be one of the digital humanities’ strengths, which is pivotal to the production of new knowledge.” (Frabetti, 2011, p. 2)

What Frabetti states here is fundamental to conceptualizing the cooperation between humanities scholars and computer scientists: it is not about a fusion of each other’s methods or questions, but about an articulation of common curiosity. That curiosity is driven by interest in each other’s expertise and each other’s way of querying the world. Over the years I have experienced interesting discussions between computer scientists and humanities scholars, for instance by looking together at data and the patterns that we distilled from them. Sometimes they led to very different insights and at such a moment you force each other to make presuppositions explicit: why do I see what I see and do you see something else? Are those data really what they seem? Why are other or multiple interpretations possible? And what can we deduce from that? But it is precisely through those discussions that we arrived at ideas for new or adjusted questions. To be honest, I never found a deaf ear with computer scientists when we proposed a qualitative approach of a research hypothesis in addition to a quantitative or algorithmic one. And the reverse: by working together with computer scientists, I learned why interpretative questions form the basis of computational thought and in turn lead to novel interpretations. The value of search instruments customized specifically for your research can deliver golden insights.

I cannot state it any clearer than Frabetti, but there is no such thing as an independent humanities framework from which we can query technology or computer

science. This statement, more or less, was eloquently argued by Peter Paul Verbeek in his book *Op de vleugels van Icarus. Hoe techniek en moraal met elkaar meebewegen* (2014) (*On the wings of Icarus. How technology and morality move together*). Algorithmic configurations are, as Foucault (1980) aptly states, “technologies to construct reality” (technologies of truth). Technologies underlie human communication and its storage, whether we talk about writing or about the cataloging or digitization of sources. If you really want to understand what sources say, you must know something about the ‘apparatus’ with which they came into being. The chasm between humanities scholars and computer scientists will not be bridged instantly, I fear. Humanities scholars (including those that have enthusiastically embraced the digital turn) still often argue that they have to bring in a developer or computer scientist ‘for the technical part’ of their research. Cooperation does not mean that the computer scientists become a kind of assistants for the humanities; that is just as nonsensical as claiming that the humanities threaten to be incorporated by information scientists. Researchers in computer science want to cooperate with humanities researchers to allow computers to be able to approach ‘human’ interpretations as closely as possible. And this image brings me to the final critical objection I want to discuss: the question of why the humanities should in the first place be concerned with computers and digitization rather than dedicate themselves fully to archives, books, and their content.

4. Digital heritage: old *and* new sources

Sometimes the discussions about the digital turn in the humanities morph into a polemic between exegetes of the Old versus the New Testament. Either you, as a humanities researcher, join the fashion of the new media, or you stay faithful to materialities such as paper sources or books, and restrict yourself consequently to established methods such as content analysis and source criticism. In fact, the term ‘Digital Humanities’ itself is a bad omen: in other disciplines I have never encountered this prefix. Have you ever heard of ‘digital chemistry’ or ‘digital social science’? For one reason or another, the digitization of sources has been accompanied by raising a barrier between ‘old’ and ‘new’ sources and ditto researchers. What is the basis for this schism and what does it say about the future of the humanities? The distinction between old (material) and new (digital) sources is not only theoretically nonsensical, it is even practically damaging for the academic profession. Let me begin with the theoretical part. The inseparable connection between *techné* and *epistémè* – between technology and knowledge – already existed long before Plato objected to the rise of writing as a replacement of memory. The replacement theory has since never disappeared; from writing to the typewriter and from the printing press to the computer: new inscription devices and storage technologies were invariably seen as a replacement of, or a threat to, the old. And while typewriters and printing presses are becoming virtually extinct in daily communication between people these days, it is nonsensical to argue that collections of writings, pictures, film tapes, and other non-digitized sources have become superfluous. The conclusion that older media or collections derived from ‘old media’ can be discarded

because the ‘contents’ now exist in digitized form somewhere in the world is like saying that we no longer have to conserve paintings because of the invention of photography. After all, not just the materiality of the source matters, but also the indissoluble tie between materiality, production, and distribution technology, and the selection of sources at a particular historical moment.

The digitization of sources is not just a technological issue; it deals with production and curating of content. Still, the possibility of converting all kinds of sources into digital files triggers two contrary impulses: the one extreme is to want to store, from now on, every single digitized or digital born utterance; and the other extreme is to discard all ‘original’ sources once they have been digitized. Both extremes are implications of the replacement theory: the belief that we can record everything with computers as a result of which everything non-digital becomes superfluous. Whoever believes in the possibility of complete inscription and storage of every single utterance should learn from history that this idea(l) has recurred over the past five centuries.

Automatized search and storage machines, such as Google Search and Google Scholar, are by definition selective; the software and hardware that supports storage and retrieval is based on selection and ranking mechanisms (Rieder & Sire, 2014; Van Dijck, 2010). Every academic who uses digitized sources ought to know how the apparatus directs selection and informs interpretation, even if this is sometimes very difficult to find out. Archivists and curators know that the writing of history only becomes possible by selecting and sorting. As keepers of our collective history they weigh the importance of both quality and quantity of sources. This right of selection, the right to store and to forget, may no longer be restricted to professional archivists, but it is equally misleading to assume that these professional activities have suddenly become superfluous now that computer systems can store everything and make it searchable. Ideological questions of selection and retrieval inform search algorithms and storage machines. Knowing how these algorithms inform selections and choices is highly relevant now that libraries and archives are facing the choice to digitize sources and/or discard ‘old’ collections because of lack of space or funding.

Archivists and heritage conservators, together with humanities scholars and computer scientists, need to continue asking fundamental questions about the curating of sources of heritage, whether the issue is digitization and selection of sources by Google’s search engines, the public accessibility and availability of information, or the instruments with which we search and query data. The materiality of culture will keep changing continuously and because of this the profession moves along.

Big data, grand challenges

Those who think that the ‘digital humanities’ are only about searching and making searchable large digital databases, overlook something essential. Digital humanities ask for a radical engagement with this new materiality as well as the preparedness to experiment with it. It is precisely those experiments with larger research questions about cultural complexity and cultural change, applied to larger databases,

that hopefully lead to insight into, and critical reflection on, the sources we have been using in daily life. To return to my previous example: if I were to restart my PhD research into the public debate about IVF and reproductive technologies, my curiosity would no longer be satisfied by querying the few sources I had at my disposal in 1991, no matter how valuable they were for that purpose at that point in time. The *condition technologique* of our present time provides me, as a researcher, with access to much more diverse source material, allowing me to expand, focus, and broaden my research question.

That does not mean that I ‘surrender’ to a new methodological paradigm and, with that, leave all the old behind me. On the contrary, more than ever I feel challenged to confront those new sources and methods with critical interpretation and qualitative analysis. Not just that: by experimenting with digital methods, by getting to know, and work with, digital sources and by delving into the ‘secrets’ of algorithmic and computational thought, I can better understand which dilemmas are being raised by the digitized society. By experimenting with digital methods, humanities scholars learn more about the role of Big Data in our (future) society or about the necessity of public accessibility of data. I would like to conclude with the thesis that the humanities cannot afford *not* to engage with ‘digitality’. Or let me put it even more firmly: society direly needs the expertise of humanities scholars – their critical insights, analytical acuity, and knowledge of ambiguity and diversity – to make sense of a digital culture that permeates and directs our daily life. As academic guardians of the arts, culture, language, heritage, and the traditions of humanities thinking, we will have to engage in multifarious ways with the interrelatedness of digital technology in all kinds of cultural practices.

Notes

- 1 This article is an original translation of José van Dijck’s *Ketelaarlezing*: ‘Big Data, Grand Challenges. Over digitalisering en het geesteswetenschappelijk onderzoek’, organized by the *Nationaal Archief* and the *Koninklijke Vereniging van Archivarissen in Nederland* at December 10, 2014. www.kvan.nl/files/Ketelaarlezing/Ketelaar12_2014-DEF.pdf.
- 2 For some humanities researchers, the term ‘Big Data’ is problematic; particularly where historical research is concerned, such data are not always ‘big’, except for instance when it comes to searching large numbers of newspaper pages.
- 3 Kaptein, Marx, and Kamps (2009) show, for example, how argumentation structures in minutes of the Dutch parliament can be digitally reconstructed, in order to not just uncover the content, but also the positions of a debate (who said what against whom?).
- 4 Both TROVe and AVResearchXL are still in a pilot phase. These tools are not just helpful for scientists, but also for journalists and other researchers of public debate or imaging. See, for example, the recent workshop for journalists that used AV Research XL for, among others, an analysis of the Islam debate www.clariah.nl/blogs/221-onderzoeks-tool-av-researcher-xl.
- 5 See, for instance, the research of Paul Dijstelberge’s visualization of anatomic drawings from various books, in his NWO KIEM project *Metabotnik*, which is primarily explorative and thus heuristically interesting. He uses visualization tools that show thousand images on a page, so that these drawings through the ages can be compared and explored. But

examples also exist of applying these kinds of tools for style analysis, such as Dijstelberge's research into the development of decorative initials in European books.

- 6 Qualitative research methods and critical analysis are even more important in the case of big files derived from social media or blogs. Apart from knowledge about the origin and context of these data, it is necessary to have an eye for the technical characteristics of this content: you have to know something about the underlying mechanisms (algorithms, user interface) of for instance Twitter or Facebook to understand how opinions are massaged and channeled through these platforms. Twitter is no thermometer of public debates in society, as some claim: the Twitter flow stands in continual dialogue with mass media – digital, paper, and audiovisual. This 'technicity' of the content demands as much interpretation as the utterances themselves (Niederer & Van Dijck, 2010).

References

- Fish, S. (2012). The Digital Humanities and the transcending of mortality. *The New York Times*, September 1.
- Foucault, M. (1980). Truth and power (Original 'Intervista a Michel Foucault'). In G. Gordon (Ed.), *Michel Foucault, power/knowledge: selected interviews and other writings 1972-1977*. New York: Pantheon Books.
- Frabetti, F. (2011). Rethinking the Digital Humanities in the context of originary technicity. *Culture Machine*, 12, 1-22.
- Kaptein, R., Marx, M. & Kamps, J. (2009). Who said what to whom? Capturing the structure of debates. In J. Allan, J.A. Aslam, M. Sanderson, C.-X. Zhai, & J. Zobel (Eds.), *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 831-832). New York: ACM Press.
- Niederer, S. & Van Dijck, J. (2010). Wisdom of the crowd or technicity of content? Wikipedia as a socio-technical system. *New Media & Society*, 12(8), 1368-1387.
- Noordegraaf, J. (2014). *De digitale erfenis – enter en return*. Inaugurele rede, University of Amsterdam, 7 February 2014. Amsterdam: Amsterdam University Press.
http://www.oratiereeks.nl/upload/pdf/PDF-6027weboratie_Noordegraaf.pdf.
Consulted 2015-04-10.
- Piersma, H. & Ribbens, K. (2013). Digital historical research. Context, concepts and the need for reflection. *BMGN Low Countries Historical Review*, 128(4), 78-102.
- Rieder, B. & Sire, G. (2014). Conflicts of interest and incentives to bias: A microeconomic critique of Google's tangled position on the Web. *New Media & Society*, 16(2), 195-211.
- Svensson, P. (2012). Envisioning the Digital Humanities. *Digital Humanities Quarterly*, 1(1).
- Van Dijck, J. (1995). *Manufacturing babies and public consent. Debating the new reproductive technologies*. New York: New York University Press.
- Van Dijck, J. (2005). *The transparent body. A cultural analysis of medical imaging*. Seattle: University of Washington Press.
- Van Dijck, J. (2010). Search engines and the production of academic knowledge. *International Journal of Cultural Studies*, 13(6), 574-592.
- Verbeek, P.P. (2014). *Op de vleugels van Icarus. Hoe techniek en moraal met elkaar meebewegen*. Rotterdam: Lemniscaat.