University of Groningen

Digital Humanities and Media History

Wijfjes, Huub

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Huub Wijffes

Digital Humanities and Media History

A Challenge for Historical Newspaper Research[1]

Abstract

Digital humanities is an important challenge for more traditional humanities disciplines to take on, but advanced digital methods for analysis are not often used to answer concrete research questions in these disciplines. This article makes use of extensive digital collections of historical newspapers to discuss the promising, yet challenging relationship between digital humanities and historical research. The search for long-term patterns in digital historical research appropriately positions itself within previous approaches to historical research, but the digitization of sources presents many practical and theoretical questions and obstacles. For this reason, any digital source used in historical research should be critically reviewed beforehand. Digital newspaper research raises new issues and presents new possibilities to better answer traditional questions.

Keywords: Media History, Political History, Mediatisation of Politics, Digital Humanities, Historic Newspapers

Using digital newspaper collections in historical research is quite new, but some of the problems and possibilities connected to this kind of research can actually be quite old. This article aims to explore this theme in the broader context of the rise of digital humanities, especially digital history. The big question here is if we are facing a revolution in humanities or a clash of innovations and traditions that can be fruitfully reconciled. This also raises questions about the need for digital literacy in historical science. Zooming in on the more specific digital potentials for newspaper history, some theoretical and practical problems will be discussed. A closer look is dedicated to a specific example of digital newspaper research in historical context. This 'Pidemehs'-project tried to uncover the interaction of politics and newspapers in a long period of Dutch history between 1918 and 1967. The findings stress the need to see digital history as a complimentary approach, rather than one that can replace the traditional historical approaches. Digital newspaper research raises new types of questions and offers new ways to answer traditional questions.

Clashes in Digital Humanities and Digital History

Although the first handbook on digital humanities was published in 2004, it builds on traditions in using computers in historical research going back to the rise of computer aided research in the late 1940s.[2] Digital humanities nowadays is still an experimental but fast growing field of academic research and education, connecting traditional humanities methodologies (for

example historical hermeneutics) to tools that researchers can use to curate or access online collections and to analyse big data sets. Research of this kind has triggered mixed responses, especially in historical sciences.

In a special issue of *BMGN - Low Countries Historical Review* in 2013 several historians debated the possibilities, problems and pitfalls of 'digital history' without coming to some sort of agreement about its value. That seems logical because relatively little historical research using digital sources has been performed, tested and properly evaluated. Although some historians practise computer-aided research since the nineteen sixties, digital history is still at the beginning of its development. Fundamental questions about the availability and controllability of sources and about the new methods required for digital research still need answers. Furthermore, a functional and openly accessible infrastructure for digital humanities research and research presentation is not operational in most countries. Still, despite all technical and methodological problems and obstacles, digital humanities bear great opportunities for new research that in nature is 'global, trans-historical and trans-media' and has led to impressive claims about its potential impact. Roughly speaking, these claims divide the world of humanities in enthusiastic fans and hesitant critics. In relation to the historical profession it has been said that 'the digital' has divided the profession between 'stalwart believers and underwhelmed agnostics.'[3]

The agnostics tend to say that until now the digital revolution didn't create a real paradigmatic revolution, but is a 'practical revolution' at heart, making relatively simple keyword searches in singular online sources far easier.[4] 'Stalwart believers', like Rens Bod in his 2012 inaugural lecture at the University of Amsterdam, claim that they are going to revolutionise humanities to an all-encompassing version 3.0. He stated that after the establishment of hermeneutical and critical traditions of humanities 1.0 in the nineteenth and twentieth century, we are now involved in finding historical patterns in digital big data in humanities 2.0. That is roughly similar to what media historian Bob Nicholson calls 'the digital turn in cultural history 2.0.' Advocates of this idea say that modern media historians should be looking for patterns and developments rather than performing traditional, interpretative research of separate and specific mediahistorical cases.[5] For the future, Bod sees the big challenge in finding a combination of 1.0 and 2.0 in humanities 3.0: a stage where critical hermeneutical traditions are combined with digital approaches that are able to map encompassing patterns and developments.[6]

This idea of phases in the development of humanities or historical sciences that are determined by the nature and availability of sources (analogue or digital) and the goal of historical research (interpreting unique events in narrative forms or reconstructing and analysing 'patterns') reignites an old fundamental split in historical science. On the one hand there are the historians producing narratives on the basis of detailed study of a small sample of exemplifying sources. On the other hand historians are aiming to analyse long-term developments based upon a varied set of (almost) complete or representative sources, providing conclusions that cover a big time span.

The latter find new arguments in 'the digital society' with its seemingly endless possibilities in shaping and connecting information and knowledge, any place and any time. In the discussions accompanying this rise of 'digital society' a sharp division can be seen between people who envision a totally new society where the political, economic, technological and social

relations will be shaped on a totally different basis, and people who stress the power of traditional culture to adjust to these challenges. It's a split between technological and cultural determinists.[7] This clash between technological determinism (sometimes also called 'solutionism' or 'belief in the technological sublime') and cultural criticism is somewhat artificial, because a lot of researchers are open to dialogue. But the 'hyperbolic discourse surrounding digital media' isn't very fruitful in inviting culturally orientated academics that want to be convinced of the practical value of digital research methods.[8]

More specifically, the clash can be seen in historiography. In their provocative *Historical Manifesto,* Armitage and Guldi show, for example, the typical technological determinist combination of worrisome language about out-of-date analogue traditions, and the unlimited promises of 'big data' that can be 'mined' to reconstruct 'patterns' and create something of a scholarly paradise. They claim nothing less than 'the power of big data to illuminate the shadow of history.'[9] Most cultural historians see this kind of ambitious claims for redefining historical research around 'the digital paradigm' or 'the digital turn' as a threatening takeover by quantitative scientist with an unlimited belief in technological rationality. In their eyes, the 'mechanisation' of the heuristic process threatens to repress a critical attitude and devaluate cultural, contextualised analysis.[10]

Actually, the call of Armitage and Guldi to 'save' historical science by shifting the research focus from unique details towards generalised patterns is not totally new. In some respects it can be seen as a digital revival of the *Annales*-movement. This French born, but decisively international movement inspired generations of historians since the nineteen thirties. The central idea was to approach history as a *longue durée*, a long-term development that can be found in social and economic life, but also in culture and mentality. *Annales*-historians were seeking for overarching metanarratives, using a combination of quantitative historical trend data and qualitative micro histories that illustrated the trends on a different level. In the vision of Armitage and Guldi, a revival of this idea is a way to keep pace with the growing influence of economists and social scientists in the current and future public debates. It also offers the possibility of keeping historical sciences in tune with the ways new and future generations of scholars formulate research questions, perform searches and interactively connect the presentation of results to the online world.

The debate about 'the digital turn' in historical science shows the old ideological question if history should hermeneutically focus on understanding and contextualising unique events or on analysing structure and patterns based on quantifiable units and data. In the nineteen seventies, this recurring debate could be seen in historical discussions about the need to integrate sociological and economic theory and methodology in historical research. It was considered a shift in research that could prove at last that history was 'a real science' with falsifiable hypotheses and verifiable methods and models.[11] The questions in this theoretical debate relate directly to the more practical problem if historians should use 'documents' or 'data', or, in other words, should interpret and tell stories or provide quantitative evidence for hypotheses.[12] According to Rieder and Röhle digital methods actually raise the question: do statistics and algorithms reach a higher level of objectivity than human interpretation? A second question is about the domination of visual output in digital humanities research. A lot of this research seems to flourish thanks to the spectacular 'infographics' and 'shock and awe'

animations. Are these kind of results of more importance than other output? Visualisation is of course tempting, because it gives us a (sometimes animated) image of patterns in history, and for some people visual material (often called 'evidence') is more powerful than evidence in words, which is often called 'argumentative'.[13]

Josh Begley, Every NYT front page since 1852. *Example of a 'shock and awe' animation based on digitised newspaper material.*

Interpretative storytellers such as cultural historians tend to think that we cannot understand complex historical or cultural processes without a notion about what constitutes and drives culture. In their opinion, sole use of quantitative data, the quest for 'patterns', and turning history into a social science therefore are too limited, or even misleading. In the classic words of cultural historian Robert Darnton: 'the social scientists live in a world beyond the reach of ordinary mortals, a world perfectly organised in perfect patterns of behaviour, peopled by ideal types, and governed by correlation coefficients that exclude everything but the most standard of deviations.' Such a world can never be joined with, what Darnton calls, 'the messiness of history.'[14] This critique is familiar to the critique on 'algorithmic culture' that is formulated in digital society. Critics say that this reliance on code, computer languages and

algorithmic reasoning is problematic for, or even incompatible with, the critical interpretative approach that still is at the basis of most humanities research.[15]

In this heated debate, there is a danger for unconstructive mutual condemnation. Rather than stressing the unbridgeable technological and cultural determinism, it is much more fruitful to conceive the divergent approaches as a set of methodological and practical issues that need to be addressed and solved in concrete research and should be subject to constant methodological evaluation. The critical scepticism about digital history creates an artificial antagonism between quantitative and qualitative methods or – to say it more harshly – between 'scientific, digital' and 'interpretative, analogue' historical research.[16]

However, in the research practices usually both perspectives and methods are used side by side in a complementary way.[17] Fears of cultural historians that their ownership of the historical field will be stolen or washed away by a digital flood, doesn't demonstrate a lot of self-confidence. If the historical debate about the *Annales*-methodology for example shows anything, it is that the structuralist and quantitative approaches didn't replace, but in the long run strengthened cultural, political, biographical and other qualitative or interpretative historical approaches.

In historical research, the nineteen nineties even gave rise to a 'cultural turn' as a response to the rise of quantitative methods coming from social and economic history. This could for example be seen in media history. From focusing on big processes in institutional media production and societal and political developments, attention shifted to the media content and its meaning in the specific historical context of media reception by publics, each with a different cultural background.[18]

This all indicates that 'the digital turn' does not necessarily mean squandering the strengths of cultural approaches. Progress can be made if we understand what digital cultural data are, what digital tools exactly do and how the results can be fitted and contextualised in broader ensembles of historical sources. As Berry asserts in an edited volume with reflections on digital humanities: 'Computationally supported thinking doesn't have to be dehumanising (...) but can give us greater powers of thinking and larger reach for our imaginations...'.[19] Of course one must acknowledge that there is a difference between the traditional close reading of a limited amount of texts and the 'distant reading' of large amounts of data. Historians however should not become what they aren't: computer scientists. They should use new methods to expand their horizon and possibilities to answer questions of historical value.

On the other hand, digital historians should be more aware that there is a big and understandable difference between statistical or algorithmic significance that computers and software engineers subscribe to, and the cultural or historical significance that historians are attached to as a way of contextualising history. Generally speaking 'the way in which computers work is not automatically compatible with the way historians work.'[20] Not automatically indeed, but compatibility can be achieved by acknowledging the strengths of both sides. Historical research cannot exclusively be the algorithmic processing of big data sets, no matter how sophisticated the methods are or will be.[21] It also needs research based on the critical interpretation of hybrid information from multiple and varied sources.

Literacy and source criticism in Digital History

Of course, digital history creates research dilemmas, especially about the balance between digital methods and historical interpretation. Digital historical research often concentrates on technological possibilities and the shrewdness of digital tools as such.[22] This implicitly creates a new dominant paradigm about history to be understood not as a set of unique social and cultural phenomena largely determined by distinction, deviance and coincidence but as a cohesive culture that can be understood just by using shrewd algorithms and present the results in spectacular 'shock and awe visualisations'.[23] Data analysts also acknowledge that 'there is a risk that we look more carefully at the technical components of the datasets than the historical context of the information that they represent.'[24]

But digital history is more than that. Since the increasing importance of digital communication and digitised historical sources from the nineteen nineties onwards, interest in what this means for historical sciences is obviously growing.[25] Looking at the practical results of digital history one should say that expectations about 'a revolution' should not be too high. Most historians still see the digital world just as a convenient place for fast and efficient browsing in the rich information sources available and not as a vital environment for historical analysis. Digital history is sometimes seen as an effort to give history meaning in a new environment and create interactive historical debates on the Internet. Characteristically, one of the first books dedicated to digital history, dating from 2006, focused on 'the Gathering, Preserving and Presenting the Past on the Web'.[26]

Still scarce are historians who seriously explore the possibilities of analysing digital historical data and integrate results in a broader historical debate. The reason for this may be the pressing need to understand the nature of big data and the many techniques and tools for data storage and analysis, like text mining, topic and concept modelling, network analysis and visualisations. In order to look at historical big data through a 'macroscope' it is required for a historian to get a grip on these data, techniques, methods and tools.[27]

Big question here is to what extent historians need to understand software and digital techniques. Are they digitally literate enough for this task? Of course, every specific research effort requires deep understanding of the methods used for delivering answers, but fully understanding digital methods is challenging for humanities scholars because it requires specialised knowledge of statistical modelling, programming languages, and the way algorithms are used for 'data mining'. This knowledge generally is restricted to insiders; for most historians the necessary computational knowledge and software is a step too far and the technical side of data collection remains a black box process that is hard to assess.[28] Because of their insufficient insight in the algorithmic logic driving these black box processes, historians run the risk of making themselves dependent on a computational logic they do not fully understand, having to rely on professionals in different and often distant fields, such as computational linguistics, information and computer science, who, in turn, lack the domain specific expertise that historians bring to the table.[29]

Another question that historians are faced with, is whether we can understand history just by looking at and analysing digital sources. For an understanding of our dominantly digital contemporary culture one cannot deny the indispensable relevance of digitally born sources. But what about history that is created in analogue forms, like handwriting, manuscripts, print and analogue audiovisual material? You can of course say that the problem will be solved when these forms will be digitalised, but that moment is still far away. As we shall see in the review of digital newspaper research, the lack of digital historical sources can be a real problem, that should be tackled on the basis of classic source critique: the need to evaluate the reach and restrictions that relevant sources (or the lack of them) offer for answering specific historical research questions.

In this respect it is of utmost importance to acknowledge that most archival sources are not digitised yet and shall not be digitised and made publicly accessible in the coming decades because of the enormous costs and copyright problems. Solely relying on digital analysis is therefore too limited in scope and even dangerous because it feeds the idea that only information that is instantly available online is relevant. That creates 'digital laziness' which is a direct threat to the historical need to critically evaluate all relevant surviving sources and not only the digitally available. In this kind of evaluation constant acknowledgement is necessary that every source only gives a very specific picture of historical reality.[30] The importance and relevance of this is provided in research showing the sensibility of media historical researchers for the availability of data and tools. Research questions and strategies can change fundamentally in this 'data-driven research'.[31] If data are not digitally available, you just turn to data that are and fit the questions to this environment.

This also directs us to the problem of a distinct and properly facilitated digital infrastructure for performing digital historical research. Enormous sets of digital historical data have already been gathered in data archives, sometimes together with digital tools to analyse the data. On this foundation, research projects have been set up, generally bringing together historians with computer scientists. This research effort doesn't seem to root in an urgent need for different views on history, but in the awareness that digital data and software are increasingly guiding our contemporary world and can therefore also be decisive for historical knowledge and understanding. Or as Lev Manovich wrote about 'softwarised culture': 'software plays a central role in shaping both the material elements and many of the immaterial structures which together make up *culture*.'[32] If it is true that the digital is determining our contemporary culture, it is also determining how we should perform historical research.

Close cooperation of specialists in both fields is the obvious solution, but generally speaking the digital techniques dominate a lot of the current cooperations. Maybe that is logical because of the many technical problems that must be solved, but historians have important problems to solve as well. Although real interdisciplinary research efforts are still at the very start of development, the combined use of digital and more traditionally stored historical sources has become a more or less normal part of the professional historical field. The big challenges therefore not only lie in the analysis of digital sources, but in developing a professional attitude as a historian in the digital world.[33]

A digital turn in newspaper history

How did media historical research, especially newspaper research develop in this emerging digital infrastructure? For an answer we must return to 'the cultural turn' in media history since the nineteen eighties. As stated before, the focus in research shifted from the history of institutional and political background of media institutions to the cultural meaning of media content for publics.[34] In this respect, the availability of content sources like newspapers, films and broadcasting programmes were increasingly vital. Methods to analyse this content were too.

Traditionally, a lot of experience was already built in historical media content analysis. In historical newspaper analysis for example tailor-made approaches were developed in the context of every specific research. Media historian Frank van Vree for example analysed the content of four major Dutch newspapers in relation to their attitude towards Nazi Germany between 1933 and 1939. The sections on the historical context of the press in this period are just as long as the actual content research that can be characterised as a historical discourse analysis strongly focusing on opinion articles and background stories in the four newspapers. Because of the labour intensive work of this sort of analysis not the entire content of the newspapers could be included. Nor could vital sections of the Dutch press in this period be included, like the national neutral or regional press. So questions can be raised about the representativeness of this research for the interpretation of 'public opinion'.[35] In a later study into the cultural transformation of the leading national newspaper *De volkskrant* in the nineteen sixties and seventies, Van Vree's focus was also restricted to certain carefully selected sections of the newspaper. In comparable studies of similar developments in newspapers, the same restrictions were characteristic for the research.[36]

More recently, methods in historical newspaper research have been developed to look more systematically at the long-term development of journalistic practices or genres. In the Netherlands, media historian Marcel Broersma kicked off this research by making a longitudinal analysis of the content of one newspaper for 250 years. Style and genre analysis were integrated in thoroughly contextualised research of the institutional and political development of this newspaper.[37] Following the same lines, but with more emphasis on a single genre within several (international) newspapers was the research of Frank Harbers, who analysed the development of the reportage in newspapers in Great Britain, the Netherlands and France between 1880 and 2005. Rutger de Graaf also employed a quantitative content analysis to reconstruct the intertextual connections between the content of pamphlets and newspapers in nineteenth century Dutch society.[38]

The principal aim of these studies was not to analyse digital data, but shed light on long term trends in newspaper content in relation to societal and political development. The data itself was mainly gathered by manually conducting a large-scale quantitative content analysis, using specific coding schemes and testing for intercoder agreement to ensure the reliability of the research. The advantage of these methods is that the coding is tailored to answering very specific historical questions. The disadvantage was, of course, the still limited amount of research material that could be examined and the risk of subjectivity of the coding decisions.

Generally speaking only samples were taken every ten or twenty years, for instance two constructed weeks to represent a particular sample year. As long as there is no sound method of automating the search for a specific and complex historical entities like 'reportage' or 'comment article', manually conducted research relying on smaller samples of the research material will remain necessary.

The cultural, interpretative tradition in newspaper history shows the value of textual research, but also the critical importance of contextualisation of this type of research. Strictly focusing on the text itself can be very useful, in linguistic studies for example, but in media history the context is indispensable for a meaningful interpretation of the past. In the digital environment this is crucial too. An example of the necessity of contextualising digital research questions is shown in an exploratory study of the theoretical concept of 'pillarisation' in Dutch history. A research project called 'Verrijkt Koninkrijk' aimed to analyse the digital texts of historian Loe de Jong in relation to 'pillarisation', a long term process of societal and political segmentation characteristic of Dutch culture roughly between 1900 and the 1960s. It showed that De Jong in his fourteen-volume book about the Netherlands during the Second World War did not write about concepts like 'zuilen' (pillars) and 'verzuiling' (pillarisation), but referred to related concepts like 'volksdelen' (sections of the national community). Researchers also found that these words were not used with the same and uniform connotations. So alternative queries had to be developed, taking into account that pillar is a broad concept with different meanings on different levels. To get a grip on that, contextualised research is necessary. A researcher should also look at the sentiment in which the more detailed concepts were used. All this requires sufficient historical expertise to frame the problem in historically correct proportions and digital expertise to produce sophisticated search methods and tools.[39]

For newspaper research digital approaches seem to offer more possibilities than 'old, analogue' methods, like selectively browsing through newspapers, reading some selected and relevant content and interpreting that in relation to other sources for historical knowledge. Browsing through and closely reading historical newspapers in this manner, gives opportunities to see historical context of newspaper content more clearly. So any suggestion that digital history research can best be performed in a closed digital environment with the big data as the only source, would be a misunderstanding of the value of 'analogue' research forms like browsing and in depth analysis of singular sources.[40]

Undoubtedly, new text and data mining methods bear a promise as they can overcome some manual browsing limitations. In principle all texts are available for fast computer-aided analysis, no longer dependent on indexing or coding and with possibilities for unlimited combinations of keyword searches.[41] Expectations sometimes are so high that historians like Joris van Eijnatten argue that 'manual browsing and sampling in various forms (...) are no longer necessary.'[42] Yet, the same author also casts doubt on these expectations by concluding that 'text mining techniques will displace but not replace traditional hermeneutic methods.'[43]

That may be comforting for the traditionalists, but above all it accentuates that digital history is here to stay. Almost all historians working with historical media sources agree that the greatest potential in working with digital sources lies in reconstructing long term connections between contents that till now could not be connected. New software techniques

for historical data mining facilitate historians who are looking for patterns in large amounts of texts like newspapers. An example offers a content analysis of millions of articles published in British periodicals since 1800 aiming to detect specific events, like wars, epidemics, coronations, or conclaves.[44] With the use of refined artificial intelligence techniques, the researchers were able to move beyond counting words by detecting references to named entities. These techniques showed both a systematic underrepresentation and a steady increase of women in the news during the 20th century and the change of geographic focus for various concepts. They could also detect the dates when electricity overtook steam and trains overtook horses as a means of transportation, both around the year 1900, along with observing other cultural transitions.

An example offers the research project 'Transatlantis' of Utrecht University, that maps debates about the supposed Americanisation of European culture in the twentieth century. The theoretical concept used in this research is 'reference culture', defined as 'spatially and temporally identifiable cultures that offer a model to other cultures and have exerted a profound influence in history.' This concept is researched in a set of digital historical sources like newspapers, creating a network of references to the United States in the Netherlands between 1890 and 1990.[45]

Tracing 'patterns' like this is indeed a goal of digital humanities research in general. But most historical researchers stress that these patterns only get real meaning if they are combined with contextualised research, for example qualitative interpretation of specific texts, words or visuals. With digital newspaper research we can trace the development and intensity of influential events and persons, but for the interpretation of how these constructions were made in different periods we need to take a closer look at the content in its media and cultural context.

To make the problem more concrete on an international level: with digital newspaper sources we may be able to trace the complete newspaper coverage of the Dreyfus-affair in French society in the twentieth century (supposing all newspapers are digitised, which isn't the case). Yet, in order to say something about *how* this event was constantly redefined in different contexts, we need to look at single newspapers in connection to a broad cultural and political context of its time. For this we need digital research too, because it can allow us to zoom in on content that in a traditional way could only be found by time consuming browsing of newspapers or viewing many hours of broadcasting material.

Putting theory to practice: opportunities, challenges and problems

Historical newspaper research offers a relevant insight in the practical and methodological problems of digital history. The growing digital collections of newspapers everywhere in the world promise a lot, but experiences in analysing newspaper content in historical research also confronts us with practical problems that cannot be solved easily and immediately.

First of all it must be stressed that an entirely centralised storage of all digital newspapers on a national level doesn't exist, even in countries with a powerful national library infrastructure, like most Western European countries. In these countries the collections are held by national institutions, such as the British Newspaper Archive (subscription), Library of Congress (free), ProQuest Historical Newspapers and Newspaper Archive Library Edition (subscription), the Delpher collection of the National Library of the Netherlands (free), Zefys of

the Staatsbibliothek in Berlin (free), Gallica of Bibliothèque Nationale de France (free) and the Trove collection of National Library of Australia (free).

Instruction video for Delpher online database (in Dutch).

Next to these big digital newspaper archives all kinds of specialised – regional, local, thematic - collections pop up in the online world. Each of these collections can make use of specific interfaces, standards and/or tariffs for accessibility and use. Most of them are publically funded; some are private initiatives that can reach high quality of services. The American based 'Media History Digital Library' for example digitises and hosts full and free access to complete collections of classic media periodicals, mainly magazines on broadcasting, film, and communication technique and policy. This online library is supported by owners who loan their magazines for scanning. Voluntary donors contribute the funds to cover the cost of scanning.[46]

Because there is no standardised rule for adding metadata in these digitisation processes, connections between the metadata sets of all these separate collections are hard to establish. That complicates really new digital search methods like text mining and network analysis. In addition to that, some important collections like the commercial Lexis-Nexis Academic Newspaper database are based on text only and therefore totally ignore the visual dimension of news, a fundamental problem for certain research questions.[47]

That problem is comparable to other problems surrounding the statistical analysis of the digital data behind the newspaper itself. This metadata, containing all the words, tags, dates, titles and other relevant bits of information, are also used to make segmentations in the newspapers, for example on basis of articles, visual elements, advertorials etcetera. Metadata and segmentation can be the basis for statistical analysis. But for that purpose the data should be uniform, quantifiable and preferably also complete. The uniformity and calculability cannot be guaranteed in public search engines such as Delpher, Zefys, Gallica and Trove. These search engines are designed for relatively simple search queries and making connections between the content of newspapers, magazines, journals and – in some cases – even in books. They seem ready made for researching long term and complex interrelated 'patterns'.[48]

But for making statistical calculations they are not very well suited. For statistical analysis the metadata behind the search engines can be useful, but metadata in most cases are not

publically accessible. For research reasons they sometimes can be consulted on request. But more convenient would be an infrastructure that is especially designed for research. Preferably all heritage institutions that have media historical collections would cooperate in this infrastructure. A good, but still experimental example is 'Europeana Newspapers', a project of eighteen European libraries creating full-text versions of about ten million newspaper pages.[49] It also detects and tags millions of single articles with metadata and named entities (information identifying people, locations etcetera).

This kind of projects offers advantages in developing useful tools and expertise on the collections itself, but in the long run they can also provide opportunities to connect databases of different origin together. In order to shed some light on the historical development of the public spaces for example, one can imagine that we need to connect the content of journalistic magazines, newspapers, and radio and television with other reality sources, like proceedings of parliament, general magazines, scientific and special interest journals, films, books and new media content.

Next to this general infrastructural problem (that really must be solved to improve the value of digital media historical research) practical problems call for solutions. First of all, and most prominent, is the problem of incompleteness. The digitisation of sources and the preservation of original (analogue) sources come with considerable costs. Making complete digital versions of analogue sources therefore takes a lot of time. Since the beginning of the twenty-first century big projects have started to digitise collections of newspapers. The National Library of the Netherlands for example has invested in a project with the aim to digitise every newspaper in their huge collection that overarches the period from 1618 to 2000. In 2015 more than nine million pages originating in 1700 newspaper titles and containing approximately eighty million articles were digitised (Figure 1). These figures are impressive, but still only fifteen percent of the total collection of newspapers is covered. With eighty-five percent still to go, digitising all newspapers is indeed a long-term project.[50]
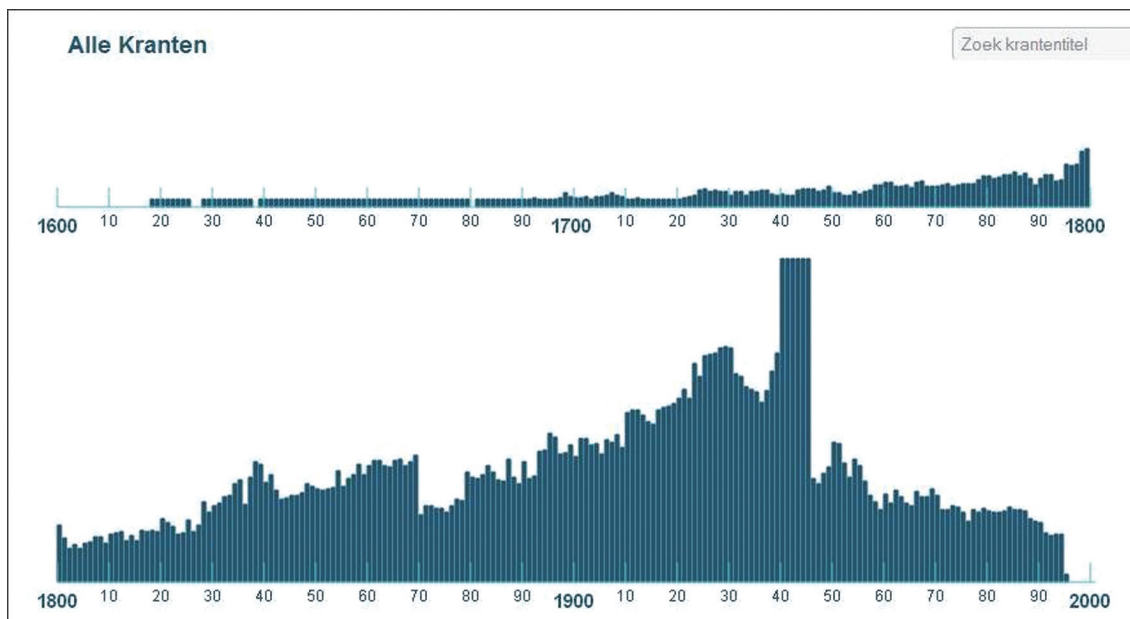

Figure 1. *Amount of digitised newspapers per year, available in Delpher collection of the National Library of the Netherlands, 1600-2000. Reference date: January 2017. Source: the National Library of the Netherlands, The Hague: http://www.delpher.nl/nl/kranten#krantenoverzicht. The figures in the graph are continuously updated.*

Obviously, with the digital newspaper collection now available, big gaps can be seen. While circulation figures of the Dutch press show a considerable growth between 1945 and 2000, in contrast the digital collection shows a considerable decrease. The reason is that newspaper titles younger than seventy years can only be digitised and made publicly accessible with permission of copyright holders. The consequences are demonstrated in figure 1. For the period after 1945 most newspapers are not publicly available for digital research. It can be said that we are facing an enormous black hole in the digital collection of historical newspapers. From a historical point of view, avoiding this problem by focusing on the available newspapers can be an irresponsible and unjustifiable solution – emphasising the need for researchers working with these collections to always demonstrate their accountability and the awareness that they are basically working with a 'convenience sample'.

The depth of this problem of incompleteness was shown concretely in the historical research project 'Pillarization and Depillarization Tested in Digitised Media Historical Sources' (Pidemehs).[51] The universities of Groningen and Amsterdam performed this project between 2014 and 2016, in close cooperation with the Netherlands eScience Centre, the National Library of the Netherlands and NIAS. It aimed at reconstructing long-term patterns in the historical relationship of Dutch political and newspaper cultures on the basis of available digital newspaper collections and digital political sources, like party political programs and proceedings of parliament. Presentation of the results is forthcoming in another publication, so here only some findings about the research practice are presented.[52]

Pidemehs first of all showed the necessity of thorough preparation (including critical source evaluation) and controlling digital search queries on the basis of contextualised historical research. Before starting such a historical research in digital newspapers some consideration had to be made about the nature of the digital data sources. In what way and to what depth are these data constructed, assembled or stored and how representative are they for the total of newspaper sources produced in certain periods? An important question related to this, is what metadata are connected to the data and how this data relates to the automated segmentation of newspaper content in articles, visuals, advertorials, etcetera.

The project showed the huge limitations created by the relative scarcity of digital sources, gaps in collections and technical failures connected to the digitisation process. These problems limited the research to the period in which a representative and relevant set of digital newspapers could be guaranteed: 1918–1967. The original setup that stretched out from the period until 2000, was impossible to realise due to copyright problems.

The availability or lack of digital newspaper titles showed to be vital for tackling certain research questions within the Pidemehs-project. For an analysis of the long-term relationship between newspaper content and political identity for example, digital copies of the newspapers were needed that are known for their political or religious identity and those who called themselves 'neutral' or 'not partisan'. It appeared that both could be lacking. In the newspaper collection of the National Library of the Netherlands for example no complete digital set of the most important protestant newspaper between 1870 and 1940 – *De standaard* – is kept, probably because of a lack of money to digitise the complete set. Furthermore, at the time of

this research project a complete set of liberal newspapers like *NRC* and *Algemeen handelsblad* was lacking; only certain parts of the interwar years are digitised and made accessible.[53] Similarly, at that time, a digital copy of the most important catholic newspaper *De volkskrant* from 1919 until now was not available because of copyright problems.[54] All in all, the available data limited the research to an analysis of socialist, catholic and neutral groups and newspapers.

The incompleteness of available data is the biggest practical problem, but not the only. Lack of uniformity in data is another. Effective historical data mining builds upon uniform data. For example, if you're looking for the intensity of newspaper attention for a political party named RKSP, how can you be sure you'll retrieve all relevant data? One problem is that newspapers don't make it a habit to standardise names and concepts, so a search query needs to include all name varieties. Building on expertise knowledge about political history and existing documentation of political parties, a list can be made with all varieties the party RKSP (and its predecessor) used in a period between 1918 and 1940. That list looks like this: 'ABRKKV; BRKKV; Algemeene Bond van Rooms-Katholieke Kiesvereenigingen; Bond van Katholieke Kiesvereenigingen; Katholieke Kiezersbond; R.K.S.P.; RKSP; Roomsch-Katholieke Staats-Partij; Rooms-Katholieke Staatspartij; Katholieke Staatspartij; kath. Staatspartij; R.K. Staatspartij, onze Staatspartij, onze partij'. The same procedure was followed in connection to other party names.

Searching for names of persons (leading politicians in this case) can create the challenging problem of how to isolate exactly one relevant person and exclude persons bearing the same name. Working with searches that combine the name with the proximity of relevant names, titles or concepts (party leader, prime minister, politician etc.) can help, but this requires some carefully performed trial and error operations. It all stresses the importance of specialised context knowledge needed when performing this kind of digital historical newspaper research.

While reconstructing the historical relationship of prominent political persons (ministers, party leaders etcetera) to newspaper content in the Pidemehs-project, it is shown that restriction to the quantity of mentioning these persons in newspapers raises questions. In Dutch context you will find that politicians dominating a distinct period like the interwar years (Colijn, De Geer) or the nineteen fifties (Drees, Romme) are mentioned more than average, not only in press that is loyal to their policies. That gives a clear indication that pillarisation is not only a question of loyalty restricted within one's own ideological group; it is also about the need for a competitor or enemy. This calls for more qualitative research into the way politicians are depicted in certain newspaper content. This can also be researched digitally, using sentiment mining techniques.

The above demonstrates that in order to efficiently excavate in big data you need tools that only highly skilled data-engineers can use or develop. Close cooperation with language specialist and/or historians is vital here.[55] The heritage institutions can have a role in developing such tools to analyse their digital collections in cooperation with universities and research institutes. Some experience has for example been built up with open source mining technology in research of historical newspapers. In the historical 'sentiment mining' programs WAHSP and BILAND word clouds are created based on relative frequencies in the retrieved selection of documents in the corpus. A word cloud can highlight negative or positive

connotation, but this still needs further historical contextualisation because connotation constantly changes in time.[56] A tool like Texcavator – developed by university of Utrecht and Netherlands eScience Centre in order to trace patterns in public discourse – is also coping with this problem.[57] Developing complex and tailor-made digital search methods that can tackle specific problems forms one of the big challenges of digital media history. This is especially valid to the problem how to retrieve and analyse visual or iconic elements within newspapers, like photographs, cartoons, maps and graphics. The search for the proliferation of iconic photographs in public debates for example has just begun.[58]

‘Pidemehs’ and other digital humanities projects show how copyright problems can create severe limitations of use, especially for late twentieth century newspapers. Retrieval and consultation in a shielded research environment (using a proxy-server for example) may offer a solution, but then the publication of results in an open access environment can become problematic. If scholars can only read about results without the possibility to check and verify them in the original research data, the scientific historical routine is threatened.

This does not mean that completeness and full accessibility are reached for the newspapers dating from the period before roundabout 1940. In the digitisation processes of newspapers priority selections have been made, generally on basis of advice given by researchers. Unavoidably, that creates gaps in the digital collection. Specialised research has shown that even for the seventeenth century, where copyright problems are not an issue and the total amount of newspapers is relatively small, fifty-two percent of all surviving hard copy newspapers between 1618 and 1650 are ‘lost in digitisation’. From the 750 surviving copies of the oldest Dutch newspaper – the *Courante uyt Italien, Duytschlandt &c* published by Jan van Hilten – until now only 199 copies have been digitised and made publicly accessible in Delpher.[59]

It needs historical expert knowledge to understand the depth of this problem and possibly create solutions. But maintaining expertise about the context of the original sources and the handling of digital bearers not only costs a lot of money, but also requires understanding of the relationship of the original analogue newspaper and the digital form. ‘When we digitise a newspaper, it is fundamentally changed (...) sources are remediated and not just reproduced,’ historian Bob Nicholson rightly remarked.[60] Tagging of articles with metadata categories like ‘advertorials’, ‘family advertisements’, ‘news lead’ or ‘news reports’ for example, facilitates research considerably, but these tags can be anachronistic because the connotation of these kind of concepts change over time.

This historical source awareness is growing steadily. So maybe the problem of cost is more pressing. Who will pay for the digitisation of all newspapers? In general one can only say that creating facilities for scientific research in Western Europe is in principle publicly funded. But the public interest clearly clashes with private interests on the issue of copyrights. And the copyright problem really is decisive for the lack of completeness in media historical sources of the twentieth century like newspapers, magazines, films and broadcasting material.

Next to the incompleteness in quantity, problem are also created due to OCR-mistakes. It is still unclear how stable and precise the technology of digital bearers is, but experience in digital projects clearly shows unreliability in the relation of the original analogue and the new digital bearer. The accuracy and quality of Optical Character Recognition (OCR) in scanned documents

can seriously influence the segmentation and the amount of mistakes in the digital search possibilities, especially in documents that require specialised knowledge to read or interpret.[61] OCR-mistakes are for example a special problem in almost all texts produced before 1850, because of the inconsistency in typographic form and layout in the older periods.[62]

One can see the consequences in the digitised collection of historical newspapers in the National Library of the Netherlands. It is shown that the accuracy level of the OCR increases considerably in time: the older the original bearer the more mistakes it contains. It is estimated that this can run up to more than eighty percent for some seventeenth and eighteenth century newspapers that have peculiar layout features or use unique fonts. For seventeenth century newspapers with a regular layout with gothic lettering and vertical text layout the failure rate is estimated between fifteen and twenty percent.[63]

It is not absolute to say that the failure rate in newspapers with modern, standardised lettering and layout is negligible or even non-existent. A search for the use of a relatively new Dutch word like 'verzuiling' (pillarisation) in historic newspapers demonstrates this. Historical context research has shown that 'verzuiling' was developed as a concept to interpret Dutch political culture in the nineteen fifties of the twentieth century. But this neologism shows up two times in eighteenth century Dutch newspapers available through the search engine Delpher of the National Library of the Netherlands. In the nineteenth century thirty-three results show up as 'verzuiling' while in the original newspapers are mentioned: verzameling, vervulling, verzetting, verzoeking, verzoening, verzorging, vergoding and verzanding. In the twentieth century period before the first proper use of 'verzuiling' in 1952, more than thirty-five OCR-mistakes pop up.

Carolyn Strange and other American press historians also point at OCR-errors and other technical obstacles in their historical research like the lack of expert metadata at document level in historical American newspapers. Their conclusion on basis of a clearly outlined selection of nineteenth century newspaper research, is that correction of OCR-failures (in their data set: around twenty percent) is 'desirable but not essential' in this kind of topical research, supposing there is enough time to check what exactly the failures do in specific search queries.[64] That is of course different with failure-rates running up to more than eighty percent in older newspapers with peculiar typographical features. And it is different if statistical analysis is one of the research tools, because statistical programs or algorithms generally do not automatically discount OCR-mistakes.

There are several methods for OCR-failure correction – which cannot be discussed in detail within the scope of this article – but none have yet developed into a definite solution. Ideal is reducing failures, preferably by double manual correction or even crowd sourcing. Crowd sourcing is promising, but despite the success of crowd sourced knowledge databases like Wikipedia and the positive experiences with some crowd sourcing projects at cultural heritage institutions, there is still some doubt about the value and reliability for scientific purposes.[65] Technicians predict that self-learning software can solve the problem in the long run, but this requires human input to 'instruct' the software of what is correct and what is not. And although there are scholars claiming that crowds of annotators can produce better, more reliable results in adding or correcting metadata than annotators with expert knowledge, curators of heritage institutions remain cautious.[66]

These institutions still have a vital intermediate function and some experiment with increasing the reliability of metadata and segmentation. British Newspaper Archive and National Library of Australia allow users to correct OCR-errors and add tags they think are relevant for the article in question.[67] Together with the Meertens Institute, the National Library of the Netherlands works with a large group of volunteers to re-type the articles in the digital collection of seventeenth century newspapers on basis of the OCR.

Conclusion

The digitisation of historical newspapers undoubtedly has stimulated research, but eagerness to use the sources sometimes takes away from the awareness of new problems accompanying these approaches; especially since the storage and retrieval of and the access to the data are still highly problematic.[68] Storage and free access are of course classical problems. From the perspective of historical research free availability of complete and uniform sources has always been vital. The historical infrastructure that was built in the nineteenth and twentieth centuries is the result of this endeavour: publicly accessible archives, concise and extensively annotated source publications, heritage institutions guarding complete and contextualised collections, and long term research projects.

These cultural endeavours get a new dimension in the digital world. Finding proper solutions for a fruitful infrastructural combination of analogue and digital sources is in full development. For researchers reflection on the value and use of digital sources is necessary. Analysing historical newspapers is getting a different dimension when we see this as analysing big data. Manually browsing through newspapers (on paper or using microfilms) automatically used to give some historical context to the content of articles, the position in relation to other content, the cultural forms and media genres to be found in these sources. When analysing digital newspaper data however, a researcher should be aware that he is doing decontextualised research. One should also get used to the idea that scarcity of sources is replaced by relative abundance.[69]

But this abundance is relative, because it is clear that not all analogue sources are digitally available. It has been shown in this article that in a digital environment completeness and uniformity cannot be guaranteed. Although millions of euros have been invested in digitisation projects, still only a fraction of historical newspapers are accessible for research purposes. OCR and other technical problems also afflict the quest for optimal source accessibility and applicability. Lack of money, but also the scattering of collections and especially the copyright problems still are decisive for the success of research efforts.[70] So, a researcher who wants to work with complete newspaper data needs to be able to organise, improvise and negotiate. There is also need for funding of digitisation of the necessary sources, which can be too substantial for a single research project. Last but not least, a researcher needs to realise that good preparation is more than half of the work; it is almost all of the work.

Historical research in digital newspapers needs well-equipped heritage institutions that create and maintain an effective infrastructure. It is not only a question of storing and organising digital data, making them accessible and developing digital tools for analysis. It is also about guarding the original and maintaining expert knowledge of all newspaper sources,

digital and analogue alike. And it is about making a serious effort in solving the copyright problem by putting the interest of public consultation high on the agenda. So, media heritage institutions should continue with the digitisation of sources with the ultimate goal to reach completeness. Doing this they should be constantly aware that historians and digital scientist both need complete and uniform data, but they also raise different questions and use different methods.

For researchers it raises the question of what value they attach to certain components of digital history research: software and data handling techniques, contextualisations, methodological operationalisation, analysis and interpretation. All these components should be in balance and be critically evaluated in the light of the specific historical research question. Just as the assumptions of historians formulating research questions are not neutral, the assumptions of digital toolmakers and analysts aren't too. 'Theory is already at work on the most basic level when it comes to defining units of analysis, algorithms, and visualisation procedures.'[71]

In overview we must conclude that the existing digital humanities research cannot live up to the claims of some digital humanities and information science scholars that we are experiencing a revolution. We are facing important methodological and practical problems that need to be solved in order to make compelling breakthroughs in historical research. Breakthroughs not strictly in theoretical sense but in performing concrete historical newspaper research for example. In close cooperation with digital scholars, media historians should be able to connect long-term developments in digital sources to exemplary historical events. Performing source critique and formulating questions on the basis of historical agendas are crucial. Formulating new research agendas on basis of digital sources can only be useful if acknowledging that analogue sources and contextualised knowledge are vital. The traditional historical guidelines to look carefully and critically at the unique materiality and historical context of sources and not to rely on just one source or method are still relevant, probably more relevant than ever.

Notes

1.　This text is part of the research project "Pillarization and Depillarization Tested in Digitized Media Historical Sources" (Pidemehs), performed by University of Amsterdam and University of Groningen. The project is made possible thanks to the generous support of the National Library of the Netherlands, the Netherlands Institute of Advanced Studies NIAS and Netherlands eScience Centre.

2.　Frédéric Clavert and Serge Noiret, "Digital Humanities and History. A New Field for Historians in the Digital Age," in *Contemporary History in the Digital Age*, ed. Ibidem (Brussels: Peter Lang, 2009), 15–26; ed. David M. Berry, *Understanding Digital Humanities* (Houndmills: Palgrave Macmillan, 2012); Susan Schreibman, *A Companion to Digital Humanities* (Malden: Blackwell, 2004).

3.　Joris van Eijnatten, Toine Pieters, Jaap Verheul, "Big Data for Global History. The Transformative Promise of Digital Humanities," *BMGN - Low Countries Historical Review* 128, no. 4 (2013): 55–77, there 57. The general discussion on 'the Digital inflecting humanities fields and disciplines' in: ed. Patrik Svensson and David Theo Goldberg *Between Humanities and the Digital* (Cambridge, MA: The MIT Press, 2015), for the historical field especially pp. 17–33.

4.　Bob Nicholson, "The Digital Turn. Exploring the Methodological Possibilities of Digital Newspaper Archives," *Media History* 19, no. 1 (2013): 59–73.

5.　Nicholson, "The Digital Turn", 63.

6.　Bod, Rens, *Het einde van de geesteswetenschappen 1.0*. Inaugural lecture, University of Amsterdam, 14 December 2012; "Forum: the End of Humanities 1.0," *BMGN - Low Countries Historical Review* 128, no. 4 (2013): 145–180.

7.   Jim Macnamara, *The 21st Century Media (R)evolution: Emergent Communication Practices* (New York: Lang, 2014); Jo Bardoel and Huub Wijfjes, "Journalistieke cultuur in Nederland: een professie tussen traditie en toekomst," in *Journalistieke Cultuur in Nederland*, ed. Ibidem (Amsterdam: Amsterdam University Press, 2015), 11–29.

8.   Extensive analysis of this 'hyperbolic debate' gives: Paul Gooding, *'Search all about it!' Historic Newspapers in the Digital Age* (Abingdon: Routledge, 2017), 22–48. Compare: Alan Liu, "Where is Cultural Criticism in the Digital Humanities?," in *Debates in the Digital Humanities,* ed. Matthew K. Gold (Minneapolis: University of Minnesota Press, 2012), 490–509; Evgeny Morozov, *To Save Everything, Click Here: the Folly of Technological Solutionism* (New York: Public Affair Books, 2013); Andreas Fickers, "Towards a New Digital Historicism? Doing History in the Age of Abundance," *View. Journal of European Television History and Culture* 1, no. 1 (2012), http://www.viewjournal.eu/index.php/view/article/view/jethc004/4. Andreas Fickers, "Veins Filled with the Diluted Sap of Rationality. A Critical Reply to Rens Bod," *BMGN - Low Countries Historical Review* 128, no. 4 (2013): 155–163.

9.   David Armitage and Jo Guldi *The History Manifesto* (Cambridge (UK): Cambridge University Press, 2014), 117.

10.   Fickers, "Veins Filled"; Liu "Where is Cultural Criticism".

11.   Kees Bertels, *Geschiedenis tussen structuur en evenement* (Amsterdam: Wetenschappelijke Uitgeverij, 1973); R.W. Fogel, "'Scientific History' and 'Traditional History,'" in *Which Road to the Past? Two visions of History*, ed. R.W. Fogel and G.R. Elton, G.R (New Haven, NJ: Yale University Press), 7–70.

12.   Hinke Piersma and Kees Ribbens, "Digital Historical Research. Context, Concepts and the Need for Reflection". *BMGN - Low Countries Historical Review* 128, no. 4 (2013): 78–102, there 82–85.

13.   Bernhard Rieder and Theo Röhle, "Digital Methods: Five Challenges," in *Understanding Digital Humanities*, ed. David M. Berry (Houndmills: Palgrave Macmillan, 2012), 67–84.

14.   Robert Darnton, *The Kiss of Lamourette. Reflections in Cultural History* (New York: W.W. Norton, 1990), 60. Compare: Fickers, "Veins Filled".

15.   Stanley Fish, "The Digital Humanities and the Transcending of Mortality," *The New York Times*, January 9, 2012, http://opinionator.blogs.nytimes.com/2012/01/09/the-digital-humanities-and-the-transcending-of-mortality/; José van Dijck, "Big data, grand challenges. Over digitalisering en het geesteswetenschappelijk onderzoek," Ketelaar-lezing 12 (2014), http://www.clariah.nl/files/publicaties/Ketelaarlezing_2014.pdf; Ted Striphas, "Algorithmic Culture," *European Journal of Cultural Studies* 18, no. 4–5 (2015): 395–412.

16.   Fogel, "'Scientific History'"; Shawn Graham, Ian Milligan and Scott Weingart, *Exploring Big Historical Data. The Historian's Macroscope* (London: Imperial College Press, 2016), 1–35. The concept of 'analogue humanities' in: Jonathan Stern, "The Example: Some Historical Considerations" in *Between Humanities and the Digital*, ed. Patrik Svensson and David Theo Goldberg (Cambridge, MA: The MIT Press, 2015), 17–33.

17.   Van Dijck, "Big Data".

18.   Huub Wijfjes, "Perspectief in persgeschiedenis," *BMGN - Low Countries Historical Review*. 114, no. 2 (1999): 223–235, doi: http://doi.org/10.18352/bmgn-lchr.4949; Donald G. Godfrey ed., *Methods of Historical Analysis in Electronic Media* (Mahwah, NJ: Lawrence Erlbaum, 2006); Michele Hilmes, *Only Connect. A Cultural History of Broadcasting in the United States* (Belmont: Thompson Wadsworth, 2010); John Hartley, *Digital Future for Cultural and Media Studies* (Chichester: Wiley-Blackwell, 2012), 27–58.

19.   Berry, *Understanding Digital Humanities*, 12.

20.   Ibid.

21.   Piersma and Ribbens, "Digital Historical Research", 57.

22.   Toni Weller, "Introduction," in *History in the Digital Age*, ed. Ibidem (London: Routledge, 2013).

23.   Armitage and Guldi, *History Manifesto*, 122.

24.   Prescott, "The Deceptions of Data", as cited in: Gerben Zaagsma, "On Digital History," *BMGN - Low Countries Historical Review* 128, no. 4 (2013): 3–29, there 24. Also: Andrew Prescott, "An Electric Current of the Imagination: What the Digital Humanities Are and What They Might Become," *Journal of Digital Humanities* 1, no. 2 (2012), http://journalofdigitalhumanities.org/1-2/an-electric-current-of-the-imagination-by-andrew-prescott/.

25.   About the history of history and computing see: Zaagsma "On Digital History".

26.   Daniel J. Cohen and Roy Rosenzweig, *Digital History. A Guide to Gathering, Preserving and Presenting the Past on the Web* (Philadelphia: University of Pennsylvania Press, 2006).

27.   Extensive exploration of these aspects in: Graham, Milligan and Weingart, *Exploring Big Historical Data*. For humanities in general an instructive manual is: Richard Rogers, *Digital Methods* (Cambridge, MA: The MIT Press, 2013).

28.   Rieder and Röhle, "Digital Methods".

29. Frank Pasquale, *The Black Box Society. The Secret Algorithms that Control Money and Information* (Cambridge (MA): Harvard University Press, 2015).

30. Michiel van Groesen, "Digital Gatekeeper of the Past: Delpher and the Emergence of the Press in the Dutch Golden Age," *Tijdschrift voor Tijdschriftstudies* 38 (2015): 9–19, there 17, doi: http://doi.org/10.18352/ts.340.

31. M. Bron, J. van Gorp and M. de Rijke, "Media Studies Research in the Data-Driven Age: How Research Questions Evolve," *Journal of the Association for Information Science and Technology* 67, no. 7 (2016): 1535–1554, doi: http://10.1002/asi.23458.

32. Lev Manovich, *Software Takes Command* (2008), 15, http://softwarestudies.com/softbook/manovich_softbook_11_20_2008.pdf.

33. Zaagsma, "On Digital History", 17–18.

34. Wijfjes, "Perspectief in persgeschiedenis"; Godfrey, *Methods of Historical Analysis*; Hilmes, *Only Connect*; Hartley, *Digital Future.*

35. Frank van Vree, *De Nederlandse pers en Duitsland. Een studie over de vorming van de publieke opinie* (Groningen: Historische Uitgeverij, 1989).

36. Frank van Vree, *De metamorfose van een dagblad. Een journalistieke geschiedenis van de Volkskrant* (Amsterdam: Meulenhoff, 1996); Gerard Mulder and Paul Koedijk, *Lées die krant! Geschiedenis van het naoorlogse Parool 1945–1970* (Amsterdam: Meulenhoff, 1996); Mariëtte Wolf, *Het geheim van de Telegraaf* (Amsterdam: Boom, 2009).

37. Marcel Broersma, *Beschaafde vooruitgang. De wereld van de Leeuwarder courant 1752–2002* (Leeuwarden: Friese Pers, 2002).

38. Rutger de Graaf, *Journalistiek in beweging. Veranderende berichtgeving in kranten en pamfletten (Groningen en 's-Hertogenbosch 1813–1899)* (Amsterdam: Bert Bakker, 2010); Frank Harbers, *Between Personal Experience and Detached Information. The Development of Reporting and the Reportage in Great Britain, the Netherlands and France, 1880–2005* (dissertation, University of Groningen, 2014).

39. Piersma and Ribbens, "Digital Historical Research", 91–95.

40. Marcel Broersma, "Nooit meer bladeren? Digitale krantenarchieven als bron," *Tijdschrift voor Mediageschiedenis* 14, no. 2 (2011): 29–55.

41. Van Eijnatten et.al. (2013), 73; Bingham, Adrian (2012). "Reading Newspapers: Cultural Histories of the Popular Press in Modern Britain". *History Compass* 10/2, 140–150.
Hart, Roderick P. and Lim, Elvin T. (2015). "Tracking the Language of Space and Time, 1948–2008". *Journal of Contemporary History* 46/3, 591–609.

42. Van Eijnatten, Pieters and Verheul, "Big Data", 73.

43. Ibid., 75.

44. Thomas Lansdall-Welfare, Saatviga Sudhahar, James Thompson, Justin Lewis, FindMyPast Newspaper Team and Nello Cristianini, "Content Analysis of 150 Years of British Periodicals," *PNAS* 114, no. 4 (2017): 457–465; published online January 9, 2017, doi: http://10.1073/pnas.1606380114.

45. Van Eijnatten, Pieters and Verheul, "Big Data", 69; http://translantis.wp.hum.uu.nl.

46. http://mediahistoryproject.org.

47. David Deacon, "Yesterday's Papers and Today's Technology. Digital Newspaper Archives and Push Button Content Analysis," *European Journal of Communication* 22, no. 1 (2007): 5–25; Broersma, "Nooit meer bladeren"; N. Maurantonio, "Archiving the Visual. The Promises and Pitfalls of Digital Newspapers," *Media History* 20, no. 1 (2014): 88–102.

48. Maarten van den Bos and H. Giffard, "The Grapevine: Measuring the Influence of Dutch Newspapers on Delpher," *Tijdschrift voor Tijdschriftstudies* 38 (2015): 29–41, doi: http://doi.org/10.18352/ts.342.

49. http://www.europeana-newspapers.eu.

50. An overview of available titles in this digital KB-collection offers: https://www.kb.nl/sites/default/files/docs/Beschikbare_kranten_alfabetisch.pdf.

51. https://www.esciencecenter.nl/project/pidimehs.

52. The technical setup of Pidemehs is shown in: P. Bos, H. Wijfjes, M. Piscaer and Voerman, "Quantifying Pillarization: Extracting Political History from Large Databases of Digitised Media Collections," in *Proceedings of the 3rd HistoInformatics Workshop, Krakow, Poland*, ed. M. Düring, A. Jatowt, J. Preiser-Kapeller and A. van den Bosch (Aachen: CEUR Workshop Proceedings, 2016), 52–57, http://ceur-ws.org/Vol-1632/. Forthcoming is: H. Wijfjes, G. Voerman and P. Bos, *Meten van verzuilde media. Een digitale benadering van politiek in dagbladen 1918–1967.*

53. Currently, both newspapers have been added to the digital collections in Delpher.

54.   It is expected that, from august 2017 *De volkskrant* (and other titles in the portfolio of the media company *De Persgroep*) will be available in Delpher.

55.   Carolyn Strange, Josh Wodak and Ian Wood, "Mining for the Meanings of a Murder. The Impact of OCR Quality on the Use of Digitised Historical Newspapers," *Digital Humanities Quarterly* 8, no. 1 (2014), http://www.digitalhumanities.org/dhq/vol/8/1/000168/000168.html.

56.   Van Eijnatten, Pieters and Verheul, "Big Data", 61; http://biland.science.uva.nl/wahsp/.

57.   Joris van Eijnatten, Toine Pieters and Jaap Verheul, "Using Texcavator to Map Public Discourse,", *Tijdschrift voor Tijdschriftstudies* 35 (2014): 59–65; https://www.esciencecenter.nl/project/texcavator.

58.   Martijn Kleppe, "Wat is het onderwerp op een foto? Kansen en problemen bij het opzetten van een eigen fotodatabase," *Tijdschrift voor Mediageschiedenis* 14, no. 2 (2012): 73–107.

59.   Van Groesen, "Digital Gatekeeper", 19.

60.   Nicholson, "The Digital Turn", 61, 64.

61.   Charles Jeurgens, "The Scent of the Digital Archive," *BMGN - Low Countries Historical Review* 128, no. 4 (2013): 30–54, there 34.

62.   Thomas Smits, "Problems and Possibilities of Digital Newspaper and Periodical Archives," *Tijdschrift voor Tijdschriftstudies* 36 (2014): 139–146, there 141.

63.   Van Groesen, "Digital Gatekeeper", 17.

64.   Strange, Wodak and Wood, "Mining for the Meanings".

65.   Johan Oomen and Lora Aroyo, "Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges," *Proceedings of the Fifth International Conference on Communication and Technologies* (New York: ACM, 2011), 138–149; Daren C. Brabham, *Crowdsourcing* (Cambridge, MA: The MIT Press, 2013); Gregory D. Saxton, Onook Oh and Rajiv Kishore, "Rules of Crowdsourcing: Models, Issues, and Systems of Control," *Information Systems Management* 30, no.1 (2013): 2–20, http://dx.doi.org/10.1080/10580530.2013.739883.

66.   Lora Aroyo and Chris Welty, "Truth is a Lie: Crowd Truth and the Seven Myths of Human Annotation," *Artificial Intelligence Magazine* 36, no. 1 (2015): 15–24; Mia Ridge ed. *Crowdsourcing our Cultural Heritage* (Abingdon/New York: Routledge, 2016).

67.   Nicholson, "The Digital Turn", 64.

68.   Deacon, "Yesterday's Papers"; Broersma, "Nooit meer bladeren".

69.   Broersma (2011), 35–37.

70.   Vgl. Karel Berkhout, "Het Digitale Drama," *NRC handelsblad*, 10 september 2011.

71.   Rieder and Röhle, "Digital Methods", 70.

Biography

Huub Wijfjes (1956) is associate professor in Journalism Studies and Media History at University of Groningen and professor in History of Radio and Television at University of Amsterdam (department of Media Studies). He is the author of numerous books and articles on media history, political history and journalism. He wrote comprehensive books on the history of Dutch Public Service Broadcasting: *VARA, biografie van een omroep* ('VARA, biography of a public broadcasting association'; Amsterdam 2009, including a website) and the history of Dutch Journalism: *Journalistiek in Nederland 1850–2000. Beroep, organisatie en cultuur* ('Journalism in the Netherlands 1850–2000. Profession, organisation and culture'; Amsterdam 2004). In 2009 he edited (with G. Voerman) the volume *Mediatization of Politics in History* (Peeters Leuven). In 2015 and 2016 he was research fellow at the Netherlands Institute of Advanced Studies (NIAS) for a research into the dynamic historical relationship of politics and newspapers in modern Dutch history on basis of digital sources.