The 1971 Census:

An Analysis of One of the First Dutch Privacy Debates

MA Thesis

New Media & Digital Culture

Ruud Dielen

5484359

Thesis supervisor: Dr. Mirko Tobias Schäfer

Second reader: Dr. Jasper van Vught

Number of words: 9644

Chicago Manual of Style

Table of Contents:

The 1971 Census: An Analysis of One of the First Dutch Privacy Debates

INTRODUCTION

In our modern society, one often hears or reads discussion on privacy, security and access to our personal data. Data breaches occur regularly, and big media conglomerates are always under suspicion of misusing the data of their customers. New laws, like the European GDPR for data protection, are implemented to guarantee the safety of our personal information. Because of the omnipresence of social media and technological devices, we fear for the protection of our data. We fear that – in this digital day and age – corporations, governments or other powerful groups might use our data for their own interest. These fears are however not a unique burden on a millennial's shoulder: they have been around for a long, long time.

This thesis investigates a pre-Internet data debate, and the fears accompanying it. As each ten years, in 1971 the Dutch government gathered personal data from their citizens, in cooperation with the CBS, the Central Bureau for Statistics. A *volkstelling* (a national census) was held. The census of 1971 was the first time administration of the population would be put into a computer. As the following citation from a newspaper article shows, the central administration of personal data caused controversy:

> The knowledge that a computer will be used, has released opposing forces not seen before within national censuses, including the one of 1960.Ten years ago, it was only Jacques Gans who had to pay 25 guilders for his refusal to participate in the state administration. He was suspected of making a mountain out of a molehill, but this time such accusations are misplaced. The mountain, in the form of the computer, is amidst us.[1]

This citation is only the tip of the iceberg. People felt threatened by the central administration and wanted change. This thesis investigates the census, and analyzes the sentiments expressed in the debate. It ultimately argues that there are many similarities between the data debate of 1971 and our contemporary debates. The main research question for this thesis is: *How does the (data) privacy debate within Dutch newspapers on the Volkstelling of 1970 relate to our contemporary privacy discussion?* The sub questions posed are: *What are the privacy concerns expressed in Dutch newspaper, covering the Volkstelling of 1971? What role does technology play in the privacy debate of 1971?* By looking at the way newspaper have written about the debate, this thesis formulates an answer to these questions. The first chapter of this thesis discusses a theoretical history of privacy discourses. The second chapter is an elaboration on this thesis' methodology and corpus. The third chapter analyzes the debate on the census of 1971 and answers the sub questions. Finally, the fourth chapter concludes by answering the main question of this thesis.

[1] *Trouw*, "Verzet tegen volkstelling uit angst voor computers," November 14, 1970, translated by the researcher.

Chapter 1: Privacy & Technologies

This chapter serves as a theoretical elaboration on the argument this thesis ultimately makes: the discourse on privacy, data and technology is not a uniquely contemporary phenomenon, but is rooted in a long tradition of different instances in which the desire for privacy and the presence of technology collide. This chapter starts out with a short description of what the *volkstelling* was, followed by the early instances of debates on privacy, data and technology. The chapter closes with some contemporary conceptualizations of privacy. From the different discourses discussed in this chapter, it becomes apparent that privacy discourses are intrinsically connected through emerging media forms and their flows of information.

De Volkstelling of 1971

In the Netherlands, *volkstellingen* were firstly held on national scale in 1795. In 1897 the *Volkstellingenwet* (National Census Law) established that each ten years those censuses would return. The national census was an inquiry by the Dutch government to capture data of their citizens, who were obliged to participate. *Tellers* (counters) were hired to go from door-to-door to pick-up the census surveys people filled in or to help them answer the questions in those surveys. According to legal scholar Jan Holvast, each census asked more intrusive questions than its predecessor. Slowly, resistance grew, especially for the census of 1971.[2] A great distrust of the government, and a fear of computers and automatization, led to protests and resistance regarding the national census.[3] The 1971 census was the last census held in the Netherlands, and its data debate eventually led to the first Dutch privacy laws.[4] People felt their privacy being violated, but this was not the first, and would not be the last time, in which privacy, data and technology would conflict.

Privacy & Technologies

This theoretical elaboration on the history of privacy starts in October 1882. Lawyer Samuel D. Warren became engaged to Mabel Bayard, daughter of a high-profile USA politician. From then on, Warren encountered newspapers reporting on his – and more often his family-in-law's – life. One article described Warren as the 'bridegroom who need not to be mentioned' at his own wedding, another article covered the deaths and funerals of his wife's mother and sister and contained highly personal and medical information.[5] Legal scholar William Prosser argued that an annoyance for this gossip led Warren to write, in cooperation with his law partner Louis D. Brandeis, *The Right to Privacy*, one of the earliest articles arguing against the sensational nature of newspapers and asking for the protection of privacy.[6]

---

[2] Jan Holvast, *De Volkstelling van 1971: Verslag van de Eerste Brede Maatschappelijke Discussie over Aantasting van Privacy* (Zutphen: Uitgeverij Paris, 2013), 32 - 47.
[3] "De burger in kaart: de Volkstelling in 1971," *Andere Tijden*, October 2011, https://anderetijden.nl/aflevering/156/De-burger-in-kaart, [1:33 - 2:55].
[4] Holvast, *De Volkstelling van 1971,* 9.
[5] Amy Gajda, "What If Samuel D. Warren Hadn't Married a Senator's Daughter?: Uncovering The Press Coverage That Led To 'The Right to Privacy,'" *Michigan State Law Review* 35 (2008): 44–54.
[6] William L. Prosser, "Privacy," *California Law Review* 4, no. 3 (1960): 383–384.

Whether it is only speculated that personal frustration was Warren's motivation for writing the article, the news coverage of Warren's family is exemplary for what he and Brandeis noted in their influential article:

> instantaneous photographs and newspaper enterprises have invaded the sacred precincts of private and domestic life; and numerous mechanical devices threaten to make good the prediction that "what is whispered in the closet shall be proclaimed from the house-tops." [7]

Through the argumentation that the press is overstepping its boundaries by reporting and gossiping on personal matters, and by showing how the existing laws failed to protect against the newspaper's power, Warren and Brandeis argued for a law protecting a *right to privacy*, a right to not be exposed.[8] In the same year, law journalist E.L. Godkin argued that newspapers had turned detrimental gossip into a printed, harmful commodity that could be spread over miles and miles, also arguing for the regulation of the invasion of private life by the press.[9] It is telling that the first conceptualizations of privacy have been initiated as reaction to a former 'new' medium (the newspaper) posing threats to our private lives.

Within the field of law, the protection of privacy is oftentimes discussed in relation to the emergence of new media forms. 30 years after Warren and Brandeis' article, legal scholar Sam Elson argued that the intrusion of private life by photographs and the press, and the accompanying emotional hurt it brought, was not taken seriously.[10] Twelve years later, legal scholar Louis Nizer discussed the still absent recognition of privacy protection, and also linked this to the operations of other new media forms, amongst others: motion-pictures, radio technologies, telephony, telegraphy, and television. Nizer warned for a future full of devices that "may soon be possible to know everything about everybody everywhere".[11]

Based on earlier discussion of privacy and new media forms, it seems almost natural that in the late 60s concerns grew for the protection of privacy in relation to the new medium of the computer. Again within the field of law, Arthur R. Miller both acknowledged the computer's potential in processing digital information as well as the danger for its omnipresence. Miller noted that computers can "become the heart of a surveillance system that will turn society into a transparent world in which our homes, our finances, and our associations will be bared to a wide range of observers".[12] This two fold in views was also recognized by social psychologist Robert Lee. By conducting surveys he identified two attitudes: firstly people tend to see the computer as beneficial to human purposes and secondly as an entity which can perform human thinking, which often made people feel inferior to the device.[13] Psychology professor Timothy B. Jay even developed a concept for the fear and negative attitudes towards computers: *computerphobia*. His definition of computerphobia generally takes three forms: resistance to talk and

[7] Samuel D. Warren and Louis D. Brandeis, "The Right To Privacy," *Harvard Law Review* 4, no. 5 (December 1890): 195.
[8] Idem, 213.
[9] E. L. Godkin, "The Rights of the Citizen: To His Own Reputation," *Scribner's Magazine* 8, no. 1 (1890): 66.
[10] Sam Elson, "Recent Developments in the Right of Privacy," *St. Louis Law Review* 14, no. 3 (1929): 306–314.
[11] Louis Nizer, "Right of Privacy - A Half Century's Developments," *Michigan Law Review* 39 (1941): 526–60.
[12] Arthur R. Miller, "Personal Privacy in the Computer Age: The Challenge of a New Technology in an Information-Oriented Society," *Michigan Law Review* 67, no. 6 (April 1969): 1089–1092.
[13] Robert S. Lee, "Social Attitudes and the Computer Revolution," *The Public Opinion Quarterly* 34, no. 1 (1970): 53–59.

think about computer technology, fear and anxiety about computers, and hostile and aggressive thoughts and acts towards the devices.[14]

This non-comprehensive overview supports the ultimate argument of this thesis: the feeling of privacy violation is not a contemporary problem in the age of new media but has been a widely debated issue each time a new media form comes into existence. Previous sources are almost exclusively from the field of law and only in recent years has the privacy debate entered studies with prominent media perspectives.

Modern Conceptualizations of Privacy

In modern times, social media and the Internet have sparked a lot of debate on privacy. In 2011, law scholar Daniel Solove argues that former privacy concepts suffer from the 'secrecy paradigm', a wrongful assumption that when one appears, or poses information, in the public (online) space, privacy is automatically suspended and information no longer private. This assumption no longer holds up in the new Internet era, and Solove thus argues that privacy concepts should be rethought in relation to information accessibility, information control and a balance between the struggle between privacy rights and the freedom of speech online, of which the latter is often unchallengedly preferred.[15]

One reconceptualization of privacy in relation to technology, is information scientist Helen Nissenbaum's 'contextual integrity.' Nissenbaum argues that *every* aspect of our lives is influenced by flows of information. To maintain privacy within these flows, two norms should be upheld: appropriateness and distribution. Referring to the former norm, privacy is maintained when the amount or type of information shared, is appropriate to the specific situation. For example, it might be appropriate to share medical data of patients amongst different hospitals, but it might be inappropriate if this information is shared between hospitals and financial institutions. Referring to the latter norm, the way information is distributed might maintain or harm an individual's privacy: within certain contexts, confidentiality and discretion are important values, while in others the obligation to share information for the benefit of others might be of greater interest. Contextual integrity thus refers to the specific context of information flows and whether these flows do or do not comply with norms of appropriateness and distribution.[16] For Nissenbaum, the Internet is such a context and thus requires values and principles similar to offline scenarios.[17]

Social media scholars Alice E. Marwick and danah boyd both critique and build upon Nissenbaum's contextual integrity, mainly in relation to social media. According to them, the notion of contextual integrity relies too heavily on the ability of an individual to perceive a context correctly, as not everyone understands and agrees on the ways data should be shared in different situations. Marwick and boyd coin the notion of 'network privacy': social media are networked, made up of individuals, companies, social norms, technical infrastructures, etc. Privacy, then, is not an individualistic matter, but

[14] Timothy B. Jay, "Computerphobia: What to Do About It," *Educational Technology* 21, no. 1 (January 1981): 47–48.
[15] Daniel Solove, "Speech, Privacy, and Reputation on the Internet," in *The Offensive Internet*, ed. Martha Craven Nussbaum (Cambridge: Harvard University Press, 2011), 15 - 30.
[16] Helen Nissenbaum, "Privacy as Contextual Integrity," *Washington Law Review Association*, no. 79 (2004): 136–142.
[17] Helen Nissenbaum, "A Contextual Approach to Privacy Online," *Daedalus* 140, no. 4 (2011): 33–34.

encompasses many factors. Any factor (an individual, a company or anything else) can easily violate the privacy of others.[18]

These conceptualizations of privacy are very focused on the present. Outside the academia privacy is often times discussed in relation to new media, social media and the increasing digitization of our everyday surroundings. As the first part of this theoretical framework has shown, privacy has a history embedded in the emergence of new technologies and media forms. By mapping the discourse of the 1971 census, and the applicability of modern conceptualizations to it, this thesis offers a dissection of a socially and temporally specific case study.

[18] Alice E. Marwick and danah boyd, "Networked Privacy: How Teenagers Negotiate Context in Social Media," *New Media & Society* 16, no. 7 (2014): 1054, 1062-1064.

Chapter 2: Exploring the 1971 Census through the Method of Distant Reading

This chapter describes the method employed in this thesis. It first describes how the analyzed dataset of newspapers was created, followed by a description of (the relevance of) the chosen approaches: distant reading and topic modeling.

NEWSPAPER DATASET

The current dataset is downloaded from Delpher, an online newspaper database.[19] Delpher allows for searching their database through queries, which can consider a time span, a certain type of newspapers (local, national, etc.) or type of article (advertisement, images, articles, obituaries, etc.). Each article in Delpher is protected through copyright, which prohibits reproduction and therefore cannot be fully included in this thesis. Delpher does note that the right of citation allows (partial) citations from articles for academic purposes.[20] The copyright protection did not obstruct the creation of the dataset, as Delpher provided all its data as long as it is for private use.

Through a search scraper in the Python-environment Jupyter Notebook (see Appendix IV, page 36 – 40), it was possible to download newspaper articles from Delpher based on one query. *Volkstelling* was the query used, and the scraper exported the newspaper articles including metadata (parent newspaper, title, date) into a CSV file. Unfortunately, due to the way the scraper was built and the Delpher API, it was not possible to fill in queries consisting of multiple words. Hence it limited the creation of a dataset built on queries of multiple words. However, as *volkstelling* is a specific word in a timeframe in which the word itself was highly popular, it generated sufficient data to gain insight in the discussion of the time.

Twelve newspapers were selected: these were published on March 1st, 1971, one day after the census was held (the day of the census was a Sunday and no newspapers were published then). For each of these twelve newspapers, articles were selected that contained the query from the time period of 01-08-1970 up until and including 31-3-1971. The selection consists of *Leeuwarder Courant, Limburgs Dagblad, Nederlands Dagblad, Nieuwsblad van het Noorden, NRC, Het Parool, Telegraaf, De Tijd, Trouw, Volkskrant, Vrije Volk* and *De Waarheid*. As there were also censuses in other countries, not all articles within the dataset were covering the Dutch census. Hence, I manually sorted out all the articles which did not belong to the Dutch census. 345 articles (including duplicates) were removed, leaving a dataset of 1,434 articles to be analyzed.

A few notions on the quality of the dataset. The number of articles mentioned is not the exact number of relevant articles within the newspapers published then. Due to Delpher's interface, articles were sometimes split up into multiple files in the dataset, often when the article would have different column lengths in the newspaper. Thus the number of 1,434 is higher than the actual number of articles published in the newspaper. However, as quantities are not the main research interest of this thesis,

[19] Delpher is an online newspaper archive, containing millions of newspapers and magazines and hundred thousands of books, ranging from the 1500s to the end of the 20th century. Delpher is part of the *Koninklijke Bibliotheek* (Royal Library), accessible through www.delpher.com.
[20] Gebruiksvoorwaarden, Delpher.com, last modified January 4, 2019, https://www.delpher.nl/nl/platform/pages/helpitems?title=gebruiksvoorwaarden

this does not invalidate the results. Secondly, the newspaper articles within Delpher are digitized copies from the actual pages, they are not digitally created. The text derived from those pages was transformed through an OCR (optical character recognition) software, that turns the copies in Delpher into digital .txt documents. However, due to the limits of the OCR software and/or the quality of the scanned documents by Delpher, the text files differ a lot in quality. Some articles were easily readable, others had to be read in their digitized form on the website of Delpher. Furthermore, as computational textual analysis was employed on the created dataset, the computational analysis also suffered from the faulty OCR, limiting its results. This is compensated through the oscillation between computational text analysis (a form of distant reading) and single, individual articles (or citations from them), so that they both can support the claims made.

Distant Reading as Method

This thesis analyses the discourse on the national census of 1971. In short, discourse is the way people talk about and understand an aspect of the world. Media scholars Jorgensen and Phillips note several premises of discourses: they are historically and culturally specific, they take a critical approach to taken-for-granted knowledge, and they link knowledge to social processes and social action.[21] This thesis bases its definition of discourses indeed on these premises: the national census debate of 1971 is seen as a particularity, both in time and place. The census debate also was a highly political and even activist discussion, relating the debate to the social processes and action as mentioned by Jorgensen and Phillips. An important nuance within the definition of discourse used within thesis is that discourse is not only 'the way people talk about…', but also the way the discourse is shaped by the 'world': discourse is both constitutive as it is constituted. This nuance is an important notion within Norman Fairclough's critical discourse analysis.[22]

To analyze the discourse, this paper focusses on newspaper articles as its research material. Jorgenson and Phillips argue the importance of linguistic utterances within (critical) discourse analysis:

> the relationship between texts and social practice is mediated by discursive practice. Hence it is only through discursive practice – whereby people use language to produce and consume texts – that text shape and are shaped by social practice.[23]

The discourse is thus the connective element between a text (for this thesis: newspaper articles) and the social practice. By analyzing the newspaper articles, this thesis finds out how the Dutch newspapers constitute a discourse around the census.

The approach taken within this research is partially that of distant reading. Literary scholar Franco Moretti is known for his different researches with the distant reading approach. In *Conjectures on World Literature*, Moretti opposes distant reading to close reading: close reading focusses on a small corpus – within literature often a very selective and small literary canon – while distant reading focuses on a

[21] Marianne Jorgensen and Louise J. Philips, *Discourse Analysis as Theory and Method* (London: SAGE Publications, 2002), 1-6.
[22] Idem, 65.
[23] Idem, 69.

large dataset (the larger the better).[24] In *Bankspeak,* Moretti, together with history of science scholar Dominique Pestre, also take a distant reading approach. “Quantitative linguistic analysis” is the name of the method in *Bankspeak*, through which they analyze financial reports of the World Bank. The actual conclusions and findings are, in the context of this thesis, not as relevant as the methodological foundations their work has. While the authors do not explicitly reflect on their methodology, their focus is two-part: mapping both semantic transformations and grammatical patterns. The former focusses on the occurrences and (dis)appearances of word clusters, the latter on patterns within word types and transitions.[25]

While traditional humanities often focus on close reading smaller datasets, analyzed mostly by a human researcher, using the distant reading approach (a large dataset analyzed with help of computational tools) offers new perspectives. For example, this thesis computationally maps the general themes and stakeholders in a comprehensive dataset. General themes in a large dataset cannot be found by analyzing just a few articles, and reading everything *manually* would take too much time. The computer can do this in a relative short time span. It has to be noted that I am not invalidating traditional humanities work: close reading still is essential in understanding texts, as interpretative work is primarily a human endeavor. Both approaches have their own benefits and neither one is intrinsically qualitatively better – both approaches just ask different questions and thus generate their own answers. I personally believe an oscillation between both forms is the most productive way of generating knowledge. Moretti and Pestre also oscillate between a close and a distant approach to the text: the authors use citations of single texts to explain the recognized changing patterns in the reports, an approach also taken within this thesis.[26]

Topic Modeling

One manifestation of distant reading is topic modeling, which is the prime method of this thesis. Topic modeling knows different models, of which the LDA model (latent Dirichlet allocation) is used here.[27] Within topic modeling, algorithms are used to ‘discover and annotate’ large datasets. It investigates what topics run through the texts of the dataset, how they are related and possibly how they evolve over time. The distinct character of the LDA model, is that it assumes that all the documents within the dataset share the same topics, but to a different degree. Topic modeling can thus not only find what topics can be found within a dataset, but also in what way an article relates to all these topics.[28]

The process of creating topics was executed in the Python environment Jupyter Notebook. All articles were loaded into Python and needed cleaning before the topic modeling could be executed. First of all, stop words were removed from the dataset. Stop words, within textual analysis, are often occurring words which do not carry a lot of meaning, such as *the*, *a*, *to do*, *she*, *you*, etc. The Dutch

[24] Franco Moretti, “Conjectures on World Literature,” *New Left Review*, February 2000, 57.
[25] Franco Moretti and Dominique Pestre, “Bankspeak: The Language of World Bank Reports,” *New Left Review* 92 (2015): 76–96.
[26] Idem, 78.
[27] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research* 3 (2003): 933–1022.
[28] David M. Blei, “Probablistic Topic Models,” *Communications of the ACM* 5, no. 4 (April 2012): 77–84.

equivalents of such words were removed from the dataset. Also, at the same time, punctuation marks and symbols have been removed from the dataset.

Furthermore, the words within the dataset were lemmatized through Frog.[29] Lemmatization is a form of text normalization, in which different derivatives of a certain word are all brought back to one term. For example, *to vote*, *voting*, *voted*, *votes*, can all be brought back to the verb *vote*. The processes of removing stop words and lemmatization improves the quality of the data: words are brought back to their roots and meaningless, often occurring words, are removed to show what unique and meaningful words can be found.

Topic modeling was accomplished with scikit-learn, a machine learning package in Python.[30] Within topic modeling, two important points need to be highlighted. First of all, the topics that are returned do not have a label and are not always coherent: it is up to the research to find out why the algorithm placed the word together and then (if the researcher wishes to do so) label it accordingly. Secondly, LDA does not define the number of topics, that is the researcher's task as well. The number of topics the algorithm will return has to be defined before the topic modeling takes place. It is a process of trial-and-error to see what number of topics gives the best results. For this research, 9 topics were chosen. As the dataset is relatively small (compared to other examples using LDA) the number of topics is fairly low. With a number of topics higher or lower than 9, insights were lost as topics became too general or not coherent enough. The results of the topic modeling – in combination with the close reading – can be found within the next chapter.

[29] *Van den Bosch, A., Busser, G.J., Daelemans, W., and Canisius, S*, "An efficient memory-based morphosyntactic tagger and parser for Dutch," in F. van Eynde, P. Dirix, I. Schuurman, and V. Vandeghinste (Eds.), *Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting*, Leuven, Belgium (2007) pp. 99-114.
[30] Skicit Learn, https://scikit-learn.org/stable/, last visited 31 – 05 – 2019.

Chapter 3: An Analysis of the 1971 Census

This chapter is the analysis of the dataset covering the 1971 census. It starts with a look into a timeline of the captured data. It continues by analyzing the word frequencies and the topic models. The main part of this chapter is the in-depth discussion of the labeled topics through a close reading of newspaper articles. This chapter answers the sub questions of this thesis: *What are the privacy concerns expressed in Dutch newspaper, covering the Volkstelling of 1971? What role does technology play in the privacy debate?*

Timeline

1,434 newspaper articles were analyzed from the period of August 1970 up until March 1971. As Figure 1 (Appendix I, page 26) shows, from October 1970 onwards there was a rapid increase in the appearance of the word *volkstelling* in the Dutch newspapers, with a peak of 728 articles in February 1971 – the month in which the actual census was held, which is 26 articles per day. Afterwards, there is a big decrease in the appearance of the word. When plotted separately per newspaper as seen in Figure 2 (Appendix I, page 26), each newspaper seems to follow a similar pattern: there is a slight peak in November 1970, follow by a minor drop in December of the same year, followed by a peak in February 1971.

There is a difference in the number of articles per newspaper, with *De Tijd* having the most, 180 articles, and *Nederlands Dagblad* containing the least amount, 39 articles. It is hard to tell if this is because of their lack of interest in the topic, Delpher not having a complete dataset or more articles in *De Tijd* being spread over multiple pages, in which case each part of the article is counted as one specific instance. Therefore, no conclusions can be drawn from the frequency differences between the newspapers. However, based on the similarities in frequencies of each newspaper, the timeline shows that the national census debate had a short time span of about four months. As all newspapers in the dataset have the same patterns, the census debate can be regarded as highly relevant on national level for its short-lived existence.

Word Frequencies

For the cleaned corpus, a frequency list was created showing the most occurring words in the dataset. In Table 1 (Appendix II, page 28 – 31), the 50 most occurring words are shown, with their English equivalent, and the number of occurrences of the word within the cleaned dataset.

Volkstelling was the word with the highest number of occurrences (3,292 times), obviously because the dataset was created based on articles containing this word. The second highest ranking word is *gegeven* (1,308 times), which is, besides the past participle of the Dutch equivalent for 'to give', the lemmatized version of *gegevens*, which means 'data'. The high occurrence of this word shows that, indeed, the national census debate was a debate primarily centered around data.

Besides data being an important aspect, the debate seems to be political as well. The 8th most common word is *minister* (occurring 714 times), the 22nd word is *gemeente* (municipality, occurring 508 times) and the 31st and 32nd word are *regering* and *overheid* (government, respectively occurring 376

and 363 times). Another political actor which cooperated with the government, is the Central Bureau for Statistics (CBS), organizer of the census. *Statisitiek* (statistics, 323 times) and *bureau* (*bureau,* 298 times) can respectively be found at the 46th and 50th place in the list.

Other words in the frequency list make sense in an obvious way, as they relate to the practice of the census. Words such as *vraag* (question), *vragen* (to ask / questions), *teller* (person who counts), *stellen* (to argue or to pose a question), *tellen* (to count), *invullen* (to fill in), *februari* (February*)* all have to do with the actual practice of the census and therefore seem obvious to appear in the frequency list.

One last word I would like to highlight is *computer*, which ranks 48th with 313 occurrences. Especially for this thesis, which is written in the field of media studies, the presence of the computer within the debate, adds a technological aspect to it. Briefly concluding, based on the frequency list, the census debate seems a highly political, slightly technological and partially practical discussion about data. The rest of this chapter explores the debate in the newspapers in depth, guided by algorithmically created topic models.

TOPIC MODELS

To gain further insights in how the words relate to each other, topics were identified through LDA. As researcher, I want to re-emphasize two points. Firstly, the number of topics identified is determined by me, not by the algorithm. LDA requires a fixed number of topics it needs to identify, and through a process of trial-and-error, nine topics seemed to give the best results. Secondly, the labels put on the topics is an interpretative act by the researcher, and hence should not be seen as an objective category name. A third remark is that the topics are numbered 0 – 8, as Python (the programming language used) starts counting at 0.

The LDA topics identified from the cleaned dataset can be found in Table 2 (Appendix II, page 28 – 31). In the table, the Dutch words, their English translations and the weights can be found, with weight being a measure of importance for a word in that specific topic. The topics in the table are unlabeled, to emphasize the interpretative act of naming the topics. The nine topics, their labels and their top 15 most important words (translated) are shown in Table 3 (see page 14).

Each topic will be explored in-depth in the next part of this chapter, except for topic 1 and topic 3. Topic 1 does not have a strong coherence. There seem to be a slightly politically themed cluster of words, but as topic 6 has a stronger political coherence, the theme of politics will be discussed within that topic. Topic 3 seems about research and inquiries. As the census itself is an inquiry, and the different sides of the debate explain their attitudes and objections towards the census, topic 6 seems too general to discuss in its own specificity for mapping the debate.

Based on a selection of these seven topics, the debate on the census will be discussed. It is important to note that within LDA, topics are distributed over articles, but articles are also distributed over topics. In other words, an article always connects to all nine topics but not every topic is as prevalent as the other. Because of the relatively short length of each article and the OCR not working perfectly, it was hard to find the most important articles within each topic. Therefore, a choice was made to base the close reading on the identified labels through the topic modeling, so cited articles discussed within

one topic might also be relevant for other topics. Furthermore, the order in which the topics are discussed is not from 0 – 8, but in an order to create the best narrative for this thesis.

Table 3: Topic Models with their Labels

No:	Label	Words
0	Practicalities and punishments of the census	*municipality, counter, question, to ask, to refuse, guilders, civil servants, hague, bureau, data, to receive, census, Rotterdam, doctor, statistics*
1	No strong coherence, slightly political	*good, to know, human, to see, to receive, politics, to think, to say, to stand, big, to sit, country, to give, year, to hold*
2	Census form and questions	*question, child, job, to ask, to fill in, card, residence, woman, how much, to follow, year, human, to know, head, to pose*
3	Research, inquiry and explanation	*scientific, explanation, letter, research, social, data, right, Amsterdam, human, big, association, objection, council, Dutch, public*
4	Resisting and critiquing factors	*Amsterdam, to say, leprechaun, room, comity, to pose, to hold, to ask, human, party, politics, to let, big, good*
5	Counters and counting	*counter, human, to say, Amsterdam, big, number of, census counter, percentage, to know, to let, to hold, sir, to count, census, to give*
6	Politicians, political parties and data	*minister, sir, PvdA, room, census, to say, data, Nelissen, motion, delay, possible, Beernink, case, party, to pose*
7	Philosophical and religious beliefs	*church, to say, minister, census, to hold, human, amount of, association, to give, to let, year, word, humanistic, objection, young*
8	Data	*data, name, year, question, big, census, computer, to say, human, to pose, to stand, to give, to know, government, possible*

TOPIC 2: CENSUS FORM AND QUESTIONS

Within topic 2, many words can be associated with the census form: *vraag* (question), *invullen* (to fill in), *kaart (*card). The words also seem to hint at the content of those questions: *beroep (job)*, *woning* (residence), *hoeveel* (how much, how many) and *kind* (child). As mentioned earlier in this thesis, the 1971 Census was organized by the Central Bureau for Statistics to capture data which would lead to insights to better organize living conditions, traffic and employment.[31] But what questions were actually asked? In Appendix III (page 32 – 35 of this thesis), a translation of the national census can be found as presented by legal scholar Jan Holvast in his book on the census.[32] The census inquired into the

[31] "De burger in kaart," *Andere Tijden* [2:56 - 4:26].
[32] Holvast, *De Volkstelling van 1971,* 401 - 416, translated by the researcher.

areas of marital and employment status, education, religion, living conditions and transportation. While some questions are understandable for a national survey, other questions seem more peculiar:

> 8.2: *Does he/she have a home telephone?*
> 10.6: *In case one has a personal car, where does he/she park the car*
> 10.11a: *Where is the toilet located? Indoors, Outside the living area, Outdoors or No toilet present*.

It's understandable these questions raise a feeling of reluctance: to share the location of your bathroom with the government does seem a little unnecessary. However, I will not critique the census myself. In Appendix III, a reader can judge for herself on the intrusiveness of the questionnaire. The rest of this thesis will focus on the way newspapers have written about the census and how, according to them, the census was (or was not) a danger for privacy and data security.

TOPIC 0 AND TOPIC 5: PRACTICALITIES, PUNISHMENTS AND COUNTING

It seems that topic 0 is mostly about the actual practice of the census: *gemeente* (municipality), *teller* (person who counts), *vraag* (question), *vragen* (to ask/questions) *ambtenaar* (civil servant), *gegeven* (data, singular), *telling* (census). Furthermore, it also seems about the punishments: *weigeren* (to refuse) and *gulden* (guilders). Topic 5 seems to share the focus on practicalities: *Tellers* and *volkstellers* (the civil servants who count) are prominent words, just as *tellen* (to count) and *telling* (census). This paragraph will focus not on practicalities, but on the punishments, as that was one of the main concerns people had with the census.

The census was obligatory: if people refused to fill in the census, they would either be fined for 500 Dutch guilders or could face imprisonment for fourteen days.[33] This raised, understandably, concerns amongst the Dutch population, as two citations from sent in letters in *De Telegraaf* show:

> When one does not want to risk the fine of 500 guilders or an imprisonment of 14 days, one can always fill in the census opposite to the truth (Necessity know no law!)[34]

> I wonder if the ill, the mad and the people abroad are also part of the census or if these people, by not obeying to their legal duty, will be dragged to court and convicted to the minimal or maximal punishment? [35]

In the documentary of *Andere Tijden*, these sentiments are reflected as well. One of the interviewees replies that even though she does not agree with the national census being held, she will fill it in as the punishments are too severe to refuse. Another interview refuses to fill in the census, as she thinks it is not okay that the census was made obligatory through the punishments.[36]

Within topic 0, the words *statistiek* and *bureau* can also be found, which refer to the Central Bureau for Statistics, the organizer of the census. Head of the census department, L.J.S. De Jonge responded to the sentiments against the obligatory nature. He stated the census would only work if the

[33] *De Tijd,* "Volkstelling hangt van potlood af," August 1, 1970 translated by the researcher.
[34] *De Telegraaf, "*Boycotten," November 13, 1970, translated by the researcher.
[35] *De Telegraaf*, "Voor wie?" November 13, 1970, translated by the researcher.
[36] "De burger in kaart," *Andere Tijden,* [11:37 - 12:00].

least amount of people refuse, hence it was made obligatory. To create something obligatory, refusing had to be punishable. De Jonge also notes that it is the population's duty to provide information.[37]

TOPIC 4: RESISTING AND CRITIQUING FACTORS

The obligatory nature was not the only aspect of the census that raised concerns. In topic 4 different anti-census movements can be recognized. *Kabouter* (the Dutch word for leprechaun) is derived from the Kabouter movement, a political activist group of that time. *Comit,* the lemmatized version of *committee* (comity) refers to *Committee Waakzaamheid*, an organization which voiced its concerns on the national census. *Oranje* (orange) in its turn refers to the Oranje Vrijstaat, another activist group. This paragraph discusses these groups, their concerns and their acts of resistance.

The biggest activist group in the census debate was the *Committee Waakzaamheid Volkstelling* (CWV, national census vigilance comity). The first appearance of the CWV in the dataset is in *NRC Handelsblad* on the 27th of October 1970) when the artist Peter Muijlwijk, one of the founders of the comity, is interviewed. Muilwijk's concerns are twofold. Firstly, Muijlwijk is not convinced the personal data will be kept anonymously. Secondly he is afraid of the possibility that the data would get into hands of malicious institutions.[38] In another interview Muijlwijk claims that the census will only provide data for corporations, and that it is merely a tool for measuring labor potential. Also, Muijlwijk believes the census data can be transferred from the CBS to other institutions as well, for example the Dutch intelligence service. Furthermore, the comity also published a documentation folder with the purpose of informing Dutch citizens.[39] The CWV also published a pamphlet and distributed it amongst churchgoers throughout the Netherlands to warn against the census. It read:

> Please think about it a thousand times, before you fill in something of which the consequences can be incalculable (…). No one know who will rule our people, through oppression, with deportation or extermination.[40]

On January 9th, 1971, Muijlwijk is quoted in *De Tijd*, in which he says that the comity's purpose is not solely to prevent a national census, but mainly to wake up the people and warn them for the dangers of automatization. According to Muijlwijk the census serves just as a concrete example to show these dangers.[41]

Holvast notes that the comity, in January 1971, gets an increasing number of members and attention. A total of 88 local subcommittees have been founded, all inspired by the CWV. Local committees can be found in, for example, Nijmegen, Den Bosch, Haarlemmermeer, Zaandam and Zeist.[42] The local committees organized their own protests, as for example the *Nijmeegse comité Waakzaamheid Volkstelling*, which wanted to publicly burn census forms on the 26th of February.[43] The Amsterdam committee, in cooperation with the *Kabouterpartij*, distributed pamphlets which called to

[37] *Het Vrije Volk*, "De Volkstelling," November 13, 1970, translated by the researcher.
[38] *NRC Handelsblad*, "Anonimiteit is in gevaar," October 27, 1970, translated by the researcher.
[39] *Het Vrije Volk*, "De Volkstelling," November 13, 1970, translated by the researcher.
[40] *Het Vrije Volk*, "Registers," December 28, 1970, translated by the researcher.
[41] *De Tijd*, "Alternatief volkstelling," January 9, 1971, translated by the researcher.
[42] Holvast, *De Volkstelling van 1971,* 204 - 207.
[43] *NRC Handelsblad*, "Verbranding formulieren volkstelling," January 20, 1971, translated by the researcher.

sabotage the census by folding the census forms, making the forms unreadable for the computer.[44] Also, the national CWV officially called to boycott the census and refuse participation. They furthermore asked people to come to the Dokwerker in Amsterdam on the 26th February of to protest against the census.[45] The Dokwerker is a statue of to commemorate the February Strike of 1941, which was the first large-scale act of resistance against the persecution of the Jewish population by the German occupier. The 26th of February was not only one day before the census was held, but also the date of the February Strike in 1941.

A second important activist group showing up as relevant through the topic modeling, is the Kabouter movement. In a short article, *Het Parool* quotes the movement on sabotaging the census:

> The group thinks the census is a "tumor that changes people in numbers, in well-oiled products, in robots. When the census will continue, we are a toy for the ones in power." [46]

On 19th of November 1970, *Het Parool* writes another article on the *Kabouter* movement. According to *Kabouters*, the national census could easily be held through a sample instead of inquiring the entire population. It would give the same trustworthy results. The movement then concludes :

> That it is not just about statistical research, but about centralization on national scale of very private data of Dutch citizens.[47]

Within different articles the movement was mentioned bursting into the parliamentary discussion room, on two different occasions, to protest against the census.[48]

Other activist groups also wanted to sabotage the census. Even though the CBS found out and prevented them from succeeding, the *Aktiegroep Anoniem* was planning on sending fake civil servants out on the street. These would collect the filled-in forms on the day of the census to make sure the data of the people would not get to the CBS.[49] The group *Oranje Vrijstaat* published a post stamp as protest against the census (Appendix I, page 27). The imagery on the post stamp is inspired by the punch card system. Another act of protest, presumably by the *Kabouters*, was to send in children into a parliamentary discussion. The children handed out soap. A similar protest was held on the Dam in Amsterdam where a person was washing himself with soap. Both protests argue for legal privilege (*verschoningsrecht,* the right to 'clean' yourself): the right to refuse to participate [50].

The protest groups protested not only against the obligatory nature but feared for a lack of anonymity and for their data to be distributed to other parties without their knowledge. The centralization of data, especially combined with the technology of automatization, would make them numbers, or toys, in the

[44] *Het Vrije Volk*, "Actie tegen volkstelling," January 26, 1971, translated by the researcher.
[45] *De Volkskrant*, "Oproep tot bezinning: Comité voert bezwaren aan tegen volkstelling," January 27, 1971, translated by the researcher.
[46] *Het Parool*, "Oproep sabotage volkstelling," October 26, 1970, translated by the researcher.
[47] *Het Parool*, "Kabouters contra volkstelling '71," November 19, 1970, translated by the researcher.
[48] *De Leeuwarder Courant*, "Jonge man zette Kamer op stelten," February 12, 1971, translated by the researcher.
[49] *Nieuwblad van het Noorden*, "Actiegroep wil kaarten van volkstelling gaan ophalen," February 23, 1971, translated by the researcher.
[50] Holvast, *De Volkstelling van 1971*, 292.

hands of evil people in power. Their way of protesting was through claiming they had to right to refuse, destroying the materiality of the census forms and creating distrust amongst the population.

TOPIC 7: PHILOSOPHICAL AND RELIGIOUS BELIEFS

Topic 7 is partially referring to philosophical and religious beliefs: *kerk* (church), *humanistisch* (humanistic), *verbond* (covenant). Secondly, the words in this topic are quite similar to the topics already mentioned. Regarding the 1971 census, there seem to be different religious sentiments. One of the questions (question 9, card 2) of the census inquired into the religious denomination of the participants. The pre-made answers were different forms of Christianity. Other denominations could be filled in on the back of the census punch card. In a letter to *Het Parool*, one reader notes that this is a topic far too private to inquire on this scale. In England, according to the reader, the national census does not ask question on religious beliefs, because it's a violation of privacy.[51] Another letter, in *Nieuwsblad van het Noorden*, also disapproves of the religious question:

> Is this [the question of religious denomination] not contrary to the human rights as dictated by Genève? (…) In this way, we return to the Medieval times, the times before the French Revolution, in which the ecclesiastical authority ruled, and the people had to say amen. No, we need to be free if we want to share our religious denomination.[52]

Another letter in the same newspaper sees the census as an operative tool for the Roman Catholic Church in Rome to increase its power. According to the letter, the census provides the Church statistics to increase their power, eliminate its opponents and rule over all who consider themselves in any way connected to the Church.[53]

The *Raad van Kerken* (Council of Churches, a cooperation between Dutch churches) issued different statements regarding the census. On the 4th of February 1971, they issued a statement in which they protest against the obligatory nature of the census. The Council of Churches argues that the importance of a national inquiry is less of value than the importance of being free from total registration.[54] However, on the 26th of February, the Council of Churches – while still having their former concerns – also stated that they did hope that a majority of the Dutch population will fill in the national census, as it does make economic and societal policies easier to make.[55]

Words related to the Jewish community do not seem to take up an important spot within the topic modeling. However, a close look in the dataset, exposes the sentiments expressed by the Jewish community are related to the administration of the Jewish population in the Second World War, which led to their deportation. In the year 1971, the Jewish church community advised to not fill in to which Jewish community their members belonged, fearing a repetition of the horrors of the Second World War.[56]

The *Humanistisch Verbond* (HV, the Dutch Humanistic Association) had its own concerns regarding the census. Besides the re-occurring concerns about anonymity, the humanists felt

[51] *Het Parool*, "Vragen staat vrij – maar met mate," November 20, 1970, translated by the researcher.
[52] *Nieuwsblad van het Noorden*, "Volk moet weer amen zeggen," November 21, 1970, translated by the researcher.
[53] *Nieuwblad van het Noorden*, "Nieuwe wet vertrouwen wij niet," January 30, 1971, translated by the researcher.
[54] *De Telegraaf*, "Amsterdamse raad vraagt uitstel van volkstelling," February 4, 1971, translated by the researcher.
[55] *Trouw*, "Raad van kerken geeft volkstellers steun in de rug," February 26, 1971, translated by the researcher.
[56] *Trouw*, "Weerstand bij joden tegen volkstelling," February 2, 1971, translated by the researcher.

discriminated, as for the question on religious belief, 'humanistic' was not one of the pre-given options. The HV asked their members to write 'under protest' at the back of the punch cards. The CBS approved this protest form and said they would count the instances this would be filled in.[57] Furthermore, there was critique amongst the humanists that the census was only filled in by the head of the household, and not the individual members, resulting in the head of the household choosing whether and how the census would be filled in.[58]

For religious and philosophical groups, their concerns were fairly similar to the protest grouped but are centered around the question of religious denomination. Again, there was fear the data would be used for wrong ends. Partially fueled by the horrors of the Second World War, people were afraid of the dangers a central administration would bring.

TOPIC 8: DATA

Topic 8 did not have a very strong coherence but has one interesting word not seen in other topics: computer. While not the most prevalent word within this topic, it is the 48th most frequent word within the dataset and appears a total of 313 times. Hence, this paragraph focuses on the concerns people had with the use of the computer within the administration of citizens.

One article stated that 'distrustful, pessimistic or chase sensitive people' think that the evil government can, within only one push of the button, gain insights into the population.[59] Another article also warns for the computer as war machine:

> When an occupier [in times of war] can mobilize all men younger than 50 years, with one push on the button, something is thoroughly wrong.[60]

An interview with professor Guus Zoutendijk ,'one of the main computer experts', reflects on the national census, privacy and computers. The only concern Zoutendijk has with the national census it its questions on religious beliefs, but furthermore he sees no privacy dangers in relation to the census. However, Zoutendijk does see a future in which the computer threatens privacy:

> Computers shall also contain almost all information over a person. This can be either demographic data, but also results of exams, medical, fiscal, financial, police and judicial data. (…) The main question is, who will get access over this information, what guarantees do we get so misuse and mistakes are prevented.[61]

In an elaborate article of October 10th, 1970 in *NRC Handelsblad*, legal expert Frank Kuitenbrouwer analyzes the issues with the rise of the increasing computerization of personal data, and refers to the discussion on the national census as case study. Kuitenbrouwer notes that there is no central institution responsible for the quality of the privacy and the security of the population registers, and neither is there systematic research into these matters – and this is problematic. The administration of the population

[57] *Nieuwsblad van het Noorden*, "Volkstelling," February 10, 1971, translated by the researcher.
[58] *NRC Handelsblad*, "Brieven: Volkstelling," February 24, 1971, translated by the researcher.
[59] *De Tijd*, "Volkstelling ontmoet plotselinge tegenwind,' November 21, 1970, translated by the researcher.
[60] *De Leeuwarder Courant*, "Verzet zonder zin," January 25, 1971, translated by the researcher.
[61] *Het Parool*, "Prof. Dr. G. Zoutendijk over: Computer als bedreiging," February 24, 1971, translated by the researcher.

into the computer is inevitable, says Kuitenbrouwer, but there has to be a balance of power, parliamentary control and official secrecy.[62]

In relation to the computer, as already argued, there was a fear it would make data easily accessible for harmful institutions and governments. While privacy experts note that the use of the computer is inevitable, they do pose questions about access, security and privacy – and they see the census as exemplary case study of these issues. The computer was seen as a danger, but the arguments for it were extremely superficial: it was mentioned as a machine which made wrongdoing accessible, but there was little elaboration on the precise notion of it. These sentiments correspond with the two fold attitude towards computers, which Lee identified in the 60s: the computer was seen as beneficial tool (an inevitable machine which made central administration easer) and as something that felt threatening for the population.[63] The fear for the computer is also a form of Jay's *computerphobia*.[64] The census adheres to the privacy discourse on computers in the 60s and 70s, and as Chapter 1 has shown, is thus part of a genealogy of different privacy and technology discourses.

TOPIC 6: POLITICIAN, POLITICAL PARTIES AND DATA

Within topic 6, names of individuals are present: *Beernink, Nelissen* and *Goudsmit*. All three of them were ministers in the government. There are more politically themed words in this topic: *kamer* (chamber*,* probably reference to the Dutch House of Representatives), *motie* (motion), *partij* (party), *PvdA* (Dutch political party). The Dutch government was divided on the census, with many arguments similar to the sentiments discussed earlier. This paragraph does not map the entire discussion of the census in parliament, but highlights a few key moments, based on the above.

While most of this chapter has discussed the arguments against the thesis, within the political sphere there are also many arguments defending the census. For example, Minister Nelissen of Economic Affairs wrote a letter to the House of Representatives, in which he guarantees that the personal census data are secure and very hard to trace back to individuals.[65] Furthermore, Nelissen noted that the questions are not inappropriate or intimate according to general Dutch attitudes.[66] This was in line with the argumentation of the CBS. In an interview, head of the census department De Jonge notes that the data are secure, as the Dutch intelligence services and the tax inspection would have no access to the data. He also argues that the questions are – even though they may seem intrusive – necessary for future planning. Anonymity is secured as well.[67] On the first page of the census people had to fill in their name, address, date of birth and sex. After the punch card were collected, this front page was separated from the rest of the data – and the questionnaire thus anonymized. However, both the front page as the rest of the punch card booklet had the same serial number printed upon them. The CBS could look up the name and address that belonged to a certain filled-in questionnaire, through the serial number. The other way around was harder: as the name and address was stored by serial number,

[62] *NRC Handelsblad*, "Big Brother is een bureaucraat," October 10, 1970, translated by the researcher.
[63] Lee, "Social Attitudes,", 53 – 59.
[64] Jay, "Computerphobia," 47 – 48.
[65] *Trouw,* "Beernink licht uitlating over volkstelling toe," December 24, 1970, translated by the researcher.
[66] Holvast, *De Volkstelling van 1971*, 194.
[67] *Het Vrije Volk*, "De Volkstelling," November 13, 1970, translated by the researcher.

it was immensely difficult to look up a certain individual, as one had to go through all the cards and find it by accident. Furthermore, after administered into the computer, 10% of all the questionnaires would be stored, for statistical reasons.[68]

The promised anonymity of the census was met with disbelief. Minister Goudsmit, of D66, requested an interpellation on the census. Her primary goal was to delay it to get full anonymity.[69] The delay did not happen, but minister Nelissen promised two concessions: the 10% of the census punch cards that would have been stored, were now going to be destroyed. Furthermore, the punch cards with the address and name on it would from now on be stored at municipalities, while the rest of the questionnaires would be stored at the CBS. This increased the complexity of matching the census answers with specific individuals.[70]

Not all politicians were as successful in convincing the population of the benefits of the census. Minister Beernink, minister of the Interior, defended the census, by arguing the information would not be shared with the Dutch intelligence services or the municipal governments. However, Beernink also said that data leaks were possible. Even though the chances it would happen were small, Beernink acknowledged that the *tellers* could make use of the personal data they gathered.[71] This interview caused a lot of fear amongst the population and lead to the founding of different organizations against the census, according to *Nederlands Dagblad.*[72]

AFTER THE CENSUS

Some closing remarks on the results of the census. In the end, the number of people refusing to participate in the national census was only minimal. According to Holvast, only 22,400 out of 13 million refused to fill in the census forms. The low number of people who refused might be because of the fear of the punishments. Another reason could be that people, when the actual census forms were presented to them, thought the questions seemed not that bad. A third reason could be that after the changes made after Goudsmit's interpellation, people were assured their privacy was secured [73].

In the end, after some negotiations, none of the 22,400 people who refused were prosecuted. Holvast also notes that 246,000 people were not at home during the census, hence they were counted administratively through the data already known about them. There were more shortcoming: an unknown part of the population had provided (consciously or accidentally) wrong or unbelievable data. One million (out of the 80 million) punch cards had to be manually copied by CBS employees, because a printing mistake in the color of the original cards made them unreadable for the computer.[74]

Taken all together, the census was a debate in which the nature of the data collection and computerization of the central administration by the Dutch government were questioned. Already based on the frequency of words, it became apparent that the discourse on the census involved many political

[68] Holvast, *De Volkstelling van 1971*, 73 - 75.
[69] *Nederlands Dagblad*, "Uitstel van volkstelling is mogelijk," February 9, 1971, translated by the researcher.
[70] *Trouw*, "Geen uitstel volkstelling," February 11, 1971, translated by the researcher.
[71] *De Tijd,* "Beernink: gegevens volkstelling kunnen uitlekken," December 1, 1970, translated by the researcher.
[72] *Nederlands Dagblad*, "Uitstel volkstelling zal geld kosten," February 10, 1971, translated by the researcher.
[73] Holvast, *De Volkstelling van 1971,* 346 - 348.
[74] Idem, 395 - 396.

and practical factors, and was linked to technology. The seven topics explored in-depth show the exact nature of the objections towards the census. Fear of harmful use of the data, made accessible by the computer, was one of the biggest concerns. Nissenbaum's breach of contextual integrity, as described in Chapter 1, can be recognized here: people fear their data would be used in inappropriate ways because it would be distributed to evil institutions.[75] Marwick and boyds notion of networked privacy is also present: Dutch citizens felt their privacy and security of their personal data not only threatened by the government, but were that anyone could know anything about them.[76] The applicability of these modern concepts on a historic case study does not only strengthen the concepts themselves, but shows that – indeed – the modern conceptualizations of privacy should not only be regarded in their own contemporality.

The next, final chapter of this thesis reflects on the parallels and differences between the discourse on the 1971 census and our contemporary privacy, data and technology debates.

[75] Nissenbaum, "Privacy as Contextual Integrity," 136 - 142
[76] Marwick and boyd, "Networked Privacy," 1062 – 1064.

Chapter 4: Conclusions

The main question of this thesis is: *How does the (data) privacy debate within Dutch newspapers on the Volkstelling of 1970 relate to our contemporary privacy discussion?* While this thesis mainly focused on the debate of 1971, this conclusion provides a short, somewhat explorative elaboration on the relation between the census debate and our contemporary discussions. Last chapter has shown that the 1971 census was a highly political and technological debate. While the CBS and most parliamentarians defended the census, many others feared it. A large concern was the obligatory nature which forced people to either participate or face an excessive monetary sanction or even imprisonment. Protest movements also critiqued the centralization of the data into computers: people still remembered the deportations of the Jewish people during the Second World War based on administrated data, and the computer would – as they believed – only simplify the possibility to exert power through the administration. Through protests, the destruction of the punch cards, and giving false data, individuals threatened to sabotage the census on large-scale, but in the end, almost everyone participated. After the census, the CW (the anti-census comity) focused themselves on computerization in general, and set up five principles in relation to the privacy of our data:

> Everyone has the right to know what data is stored about them
> Everyone has the right to contest wrong or irrelevant data about themselves
> Everyone has the right to know and to control who has access to their data
> Everyone has the right to claim compensation for unjust use of confidential or wrong information
> Everyone has the right to demand removal of information on personal ideals or intimate habits[77]

In our current society, data debates often cover similar themes of control over, access to and agency with data(bases). The five principles published in 1971 even make sense in regards to our own data debates. New within the data debates is the presence of the private sector. The debates now also involve big media conglomerates, social media platforms, data brokers and everyday technological devices that monitor people's behavior. And the monitoring of behavior is the second main difference with the 1971 debate: the data gathered in our contemporary society is not just names, addresses and employment status, but much, much more. Our Internet habits, our movements, who we befriend, our purchases, our location – all are monitored. We are no longer just citizens who provide data for government planning, we are customers, clients, patients and users whose data is used, sometimes for our own benefits, sometimes for the sake of the data collector. The 1971 concerns were justified: datafication would only increase. And while it's not per se a bad thing, it's not per se a good thing either. To question the motives and practices of the data capturing was, and always will be, an endeavor in which we protect our rights to privacy from powerful institutions.

While not delegitimizing the contemporary concerns with privacy and personal data – as there are of course huge differences in the data the census gathered and contemporary data collections – this thesis has shown that with each new media form, people fear for their privacy. Whether it's newspapers that report on daily events, statistical research through computers or data collection by commercial companies, the fear that technology knows us and takes what is ours, is present.

[77] *Het Vrije Volk,* "Natellen," March 16, 1971, translated by the researcher.

This thesis solely focused on the sentiments in the debate in the newspaper dataset of one specific case study. By oscillating between a distant reading of a text and in-depth research into specific articles, it mapped the debate in the newspapers, structured by different themes (protest groups, religious beliefs, politics). The research has mostly focused on the opinions express against the census, with a little space for opinions which argued in favor of the census. It should be emphasized that this thesis' objective was not to criticize the census, but to map its debate. Other research could investigate in what way the census indeed *does* violate privacy, but as this requires a clear definite definition of privacy, this did not seem fruitful to me. Other research could also investigate how governmental personal databanks in the Netherlands used data: was it really only statistical research? As Frank Kuitenbrouwer points out in one of the newspaper articles, before the law was changed in 1968, the Dutch government sold the data they had of their civilians to third parties, so these could advertise directly to certain groups of people.[78] This example is similar to the discussion on personalized advertisement online as it is based on the same question: to what extent can data be used and for what purposes?

The objective of this thesis was met: to map the debate on privacy and personal data of an pre-Internet era, and argue that the privacy concerns expressed nowadays are not something unique of the 21st century: there has always been, and always will be be, a fear that technology knows us better than we do.

[78] *NRC Handelsblad*, "Big Brother is een bureaucraat," October 10, 1970, translated by the researcher.

Bibliography:

Blei, David M. “Probablistic Topic Models.” *Communications of the ACM* 5, no. 4 (April 2012): 77–84.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3 (2003): 933–1022.

Elson, Sam. “Recent Developments in the Right of Privacy.” *St. Louis Law Review* 14, no. 3 (1929): 306–14.

Gajda, Amy. “What If Samuel D. Warren Hadn’t Married a Senator’s Daughter?: Uncovering The Press Coverage That Led To ‘The Right to Privacy.’” *Michigan State Law Review* 35 (2008): 35–60.

Godkin, E. L. “The Rights of the Citizen: To His Own Reputation.” *Scribner’s Magazine* 8, no. 1 (1890): 58–67.

Holvast, Jan. *De Volkstelling van 1971: Verslag van de Eerste Brede Maatschappelijke Discussie over Aantasting van Privacy*. Zutphen: Uitgeverij Paris, 2013.

Jay, Timothy B. “Computerphobia: What to Do About It.” *Educational Technology* 21, no. 1 (January 1981): 47–48.

Jorgensen, Marianne, and Louise J. Philips. “Critical Discourse Analysis.” In *Discourse Analysis as Theory and Method*, 60–95. London: SAGE Publications, 2002.

Lee, Robert S. “Social Attitudes and the Computer Revolution.” *The Public Opinion Quarterly* 34, no. 1 (1970): 53–59.

Marwick, Alice E., and danah boyd. “Networked Privacy: How Teenagers Negotiate Context in Social Media.” *New Media & Society* 16, no. 7 (2014): 1051–67.

Miller, Arthur R. “Personal Privacy in the Computer Age: The Challenge of a New Technology in an Information-Oriented Society.” *Michigan Law Review* 67, no. 6 (April 1969): 1089–1246.

Moretti, Franco. “Conjectures on World Literature.” *New Left Review*, February 2000, 54–68.

Moretti, Franco, and Dominique Pestre. “Bankspeak: The Language of World Bank Reports.” *New Left Review* 92 (2015): 75–99.

Nissenbaum, Helen. “A Contextual Approach to Privacy Online.” *Daedalus* 140, no. 4 (2011): 32–48.

———. “Privacy as Contextual Integrity.” *Washington Law Review Association*, no. 79 (2004): 119–57.

Nizer, Louis. “Right of Privacy - A Half Century’s Developments.” *Michigan Law Review* 39 (1941): 526–60.

Prosser, William L. “Privacy.” *California Law Review* 4, no. 3 (1960): 383–423.

Solove, Daniel. “Speech, Privacy, and Reputation on the Internet.” In *The Offensive Internet*, edited by Martha Craven Nussbaum. Cambridge: Harvard University Press, 2011.

Van den Bosch, A., Busser, G.J., Daelemans, W., and Canisius, S. (2007). "An efficient memory-based morphosyntactic tagger and parser for Dutch." In F. van Eynde, P. Dirix, I. Schuurman, and V. Vandeghinste (Eds.), *Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting*, Leuven, Belgium, pp. 99-114.

Warren, Samuel D., and Louis D. Brandeis. “The Right To Privacy.” *Harvard Law Review* 4, no. 5 (December 1890): 193–220.

Non-literary sources, packages and data tools:

Delpher, https://www.delpher.nl/.

“De burger in kaart: de Volkstelling in 1971.” *Andere Tijden*, 10 2011. https://anderetijden.nl/aflevering/156/De-burger-in-kaart.

Jupyter Notebook, https://jupyter.org/.

Skicit Learn, https://scikit-learn.org/stable/, last visited 31 – 05 – 2019

Stamp Oranje Vrijstaat, Internationaal Instituut voor Sociale Geschiedenis, http://hdl.handle.net/10622/30051002228002?locatt=view:level3

Tableau, https://www.tableau.com/nl-nl.

Appendix I: Figures & Images

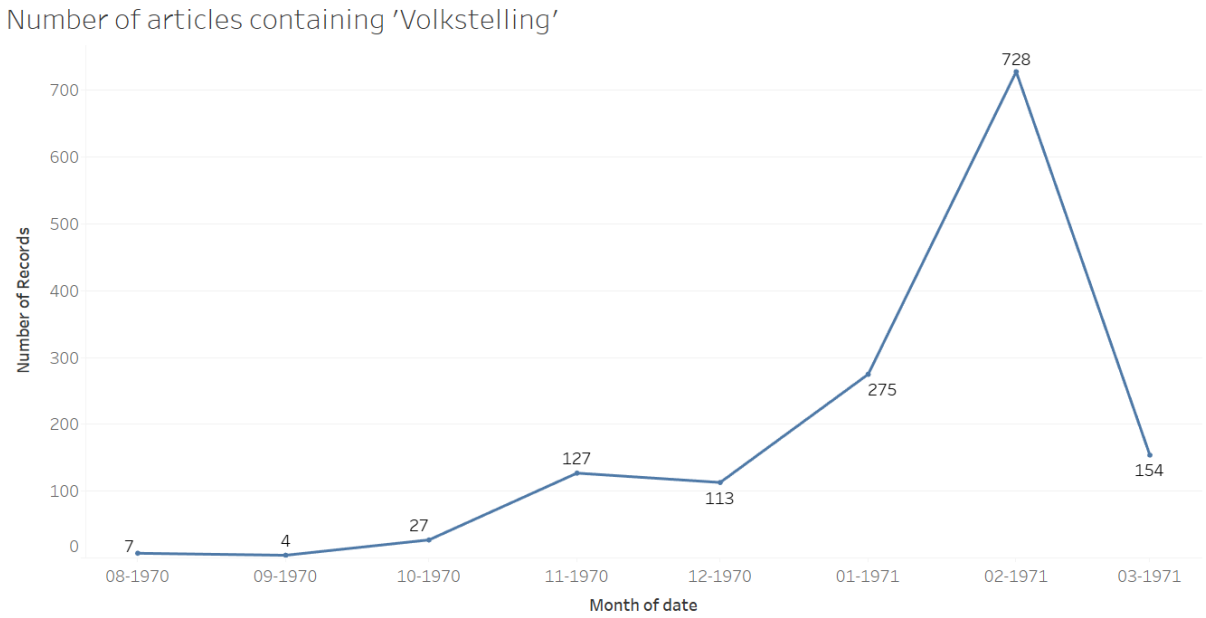

Figure 1: Number of articles containing 'Volkstelling'

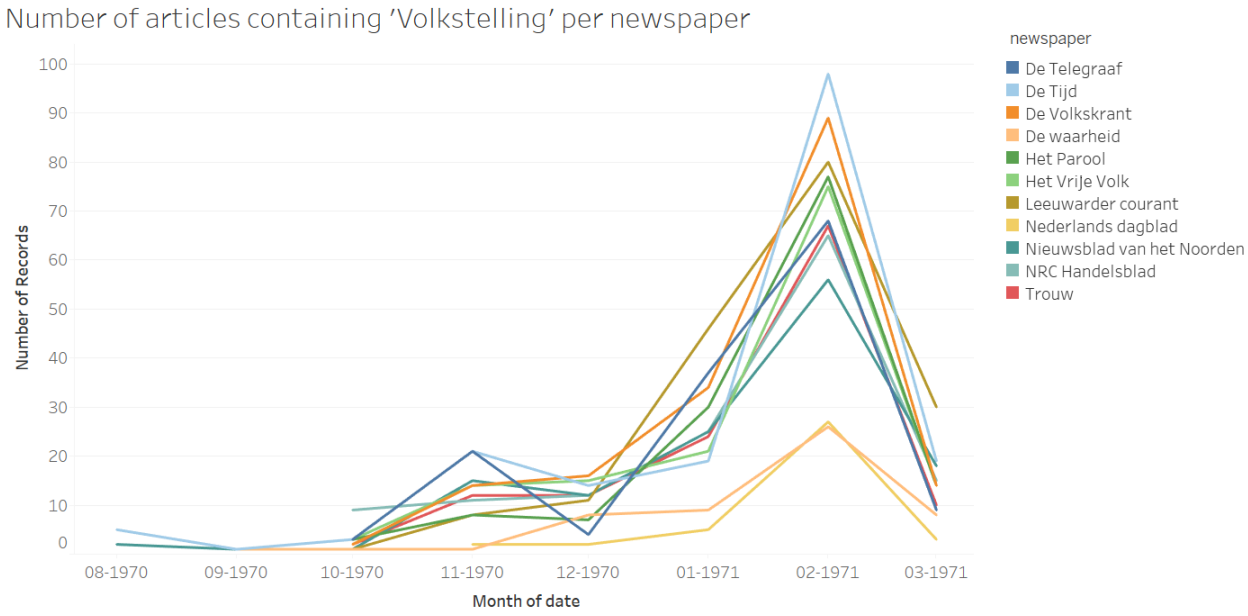

Figure 2: Number of articles containing 'Volkstelling' per newspaper

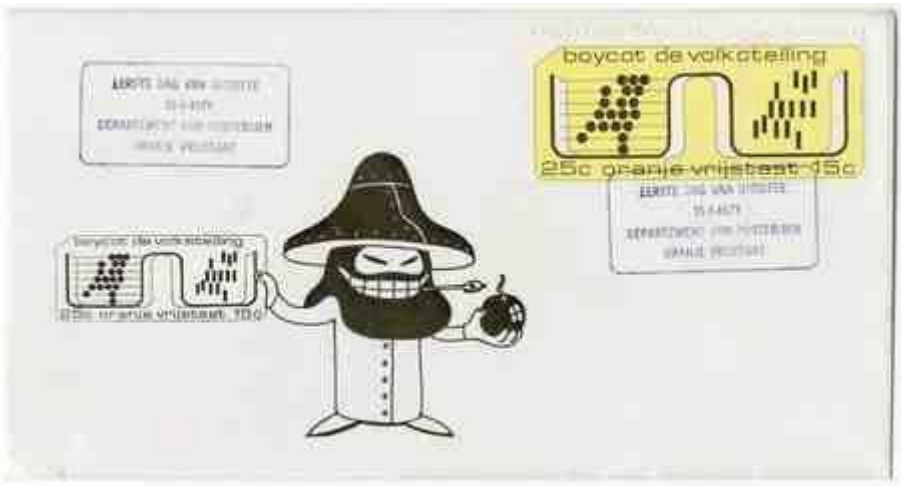

Figure 3: Stamp of Oranje Vrijstaat to protest against census
Source: Internationaal Instituut voor Sociale Geschiedenis, http://hdl.handle.net/10622/30051002228002?locatt=view:level3

Appendix II: Topic Models

Table 1: Word Frequencies

Rank	Dutch word	English equivalent	No. of occurrences
1	volkstelling	*National census*	3292
2	gegeven	*Data (singular)*	1308
3	zeggen	*To say*	973
4	mens	*Human / Person*	882
5	vraag	*Question*	881
6	groot	*Big / Large*	782
7	telling	*Census*	748
8	minister	*Minister*	714
9	goed	*Good*	705
10	vragen	*To ask / Questions*	686
11	teller	*Counter (person who counts)*	680
12	houden	*To hold*	669
13	geven	*To give*	666
14	jaar	*Year*	658
15	krijgen	*To get / To receive*	647
16	stellen	*to argue or to pose a question*	641
17	weten	*To know*	627
18	heer	*Sir / Man*	625
19	laten	*To let*	616
20	staan	*To stand*	590
21	zaak	*Case*	527
22	gemeente	*Municipality*	508
23	zien	*To see*	490
24	naam	*Name*	476
25	amsterdam	*Amsterdam*	466
26	aantal	*Number of / Amount of*	463
27	mogelijk	*Possible*	427
28	nederland	*The Netherlands*	409
29	laat	*To Let / Late*	401
30	land	*Country*	379
31	regering	*Government*	376
32	overheid	*Government*	363
33	volgen	*To follow*	362
34	kamer	*Room / Chamber*	359
35	nieuw	*New*	357
36	blijven	*To stay*	356
37	tellen	*To count*	352
38	kaart	*Card / Map*	349
39	invullen	*To fill in*	349
40	bezwaar	*Objection*	346
41	februari	*February*	344
42	week	*Week*	337

43	nederlands	*Dutch*	331
44	tijd	*Time*	326
45	werken	*To work*	325
46	statistiek	*Statistics*	323
47	zitten	*To sit*	313
48	computer	*Computer*	313
49	bevolking	*Population*	311
50	bureau	*Bureau / Office*	298

Table 2: Identified Topics within Dataset

Topic 0	**Translated**	**weights**	**Topic 1**	**Translated**	**weights**
gemeente	*Municipality*	290.1	goed	*Good*	234.4
teller	*Counter*	265.2	weten	*To know*	136.3
vraag	*Question*	218.9	mens	*Human*	116.9
vragen	*To ask*	148.2	zien	*To see*	113.6
weigeren	*To Refuse*	135.1	krijgen	*To receive / To get*	111.1
gulden	*Guilders*	126.4	politiek	*Politics*	100.6
ambtenaar	*Civil servants*	118.8	denken	*To think*	99.8
haag	*Hague*	117.9	zeggen	*To say*	99.8
bureau	*Bureau*	114.3	staan	*To stand*	99.4
gegeven	*Data*	110.9	groot	*Big / Large*	96.7
krijgen	*To receive*	107.2	zitten	*To sit*	89.7
telling	*Census*	106.7	land	*Country*	85.8
rotterdam	*Rotterdam*	101.0	geven	*To give*	81.6
arts	*Doctor*	99.0	jaar	*Year*	79.3
statistiek	*Statistics*	98.1	houden	*To hold*	78.6
invullen	*To fill in*	97.1	vrijheid	*Freedom*	75.7
heer	*Sir / man*	92.7	laten	*To let*	72.3
zaak	*Case*	90.3	nederland	*The Netherlands*	70.1
beantwoorden	*To answer*	90.1	overheid	*Government*	68.1
houden	*To hold*	85.5	nieuw	*New*	66.4

Topic 2	**Translated**	**Weights**	**Topic 3**	**Translated**	**weights**
vraag	*Question*	227.6	wetenschappelijk	*Scientific*	67.2
kind	*Child*	147.5	verklaring	*Explanation*	58.6
beroep	*Job*	102.2	brief	*Letter*	50.8
vragen	*To ask*	101.9	onderzoek	*Research*	49.5
invullen	*To fill in*	89.4	sociaal	*Social*	46.0
kaart	*Card / Map*	89.2	gegeven	*Data*	44.2
woning	*Residence*	85.4	recht	*Right*	43.7
vrouw	*Woman*	82.8	amsterdam	*Amsterdam*	41.1
hoeveel	*How much*	81.1	mens	*Human*	40.5
volgen	*To follow*	78.9	groot	*Big / Large*	40.2

jaar	*Year*	78.6	vereniging	*Association*	37.0
mens	*Human*	75.3	bezwaar	*Objection*	37.0
weten	*To know*	73.3	raad	*Council*	36.7
hoofd	*Head*	70.9	nederlands	*Dutch*	35.8
stellen	*to argue or to pose a question*	68.8	openbaar	*Public*	35.4
werken	*To work*	66.6	werkgroep	*Work group*	35.1
leven	*To live*	66.2	telling	*Census*	34.9
krijgen	*To watch*	60.8	artikel	*Article*	33.8
werk	*To work*	60.6	krommenie	*Krommenie*	33.1
huishouden	*Household*	59.1	organisatie	*Organization*	32.3

Topic 4	**Translated**	**weights**	**Topic 5**	**Translated**	**weights**
amsterdam	*Amsterdam*	133.6	teller	*Counter*	328.7
zeggen	*To say*	107.0	mens	*Human*	279.4
kabouter	*Leprechaun*	106.0	zeggen	*To say*	225.3
kamer	*Room / Chamber*	96.9	amsterdam	*Amsterdam*	176.8
comit	*Comity*	91.9	groot	*Big / Large*	128.5
stellen	*to argue or to pose a question*	89.2	aantal	*Number of*	126.4
houden	*To hold*	87.5	volksteller	*Census counter*	125.5
vragen	*To ask / Questions*	79.4	procent	*Percentage*	123.2
mens	*Human*	74.4	weten	*To know*	121.9
partij	*Party*	70.4	laten	*To let*	111.4
politiek	*Politics*	70.1	houden	*To hold*	109.4
laten	*To let*	69.1	heer	*Sir / man*	109.4
oranje	*Orange*	59.9	tellen	*To count*	106.6
groot	*Big / Large*	58.9	telling	*Census*	105.0
goed	*Good*	56.7	geven	*To give*	104.7
schmidt	*Schmidt*	53.8	week	*Week*	89.0
geven	*To give*	52.8	goed	*Good*	88.5
amsterdams	*Amsterdam*	51.6	comit	*Comity*	85.7
krijgen	*To receive*	51.0	thuis	*Home*	85.5
nieuw	*New*	50.9	avond	*Evening*	85.1

Topic 6	**Translated**	**weights**	**Topic 7**	**Translated**	**weights**
minister	*Minister*	451.8	kerk	*Church*	120.7
heer	*Sir / man*	229.7	zeggen	*To say*	80.8
pvda	*PvdA*	189.2	minister	*Minister*	65.8
kamer	*Room / Chamber*	185.3	telling	*Census*	64.7
telling	*Census*	154.2	houden	*To hold*	58.2
zeggen	*To say*	142.7	mens	*Human*	56.8

gegeven	*Data*	137.6	aantal	*Amount of*	56.6
nelissen	*Nelissen*	123.9	verbond	*Association*	56.3
motie	*Motion*	120.8	geven	*To give*	55.5
uitstel	*Delay*	116.2	laten	*To let*	55.2
mogelijk	*Possible*	115.7	jaar	*Year*	55.1
beernink	*Beernink*	113.4	woord	*Word*	52.8
zaak	*Case*	111.9	humanistisch	*Humanistic*	50.1
partij	*Party*	93.9	bezwaar	*Objection*	49.9
stellen	*to argue or to pose a question*	91.9	jong	*Young*	49.6
regering	*Government*	88.5	gegeven	*Data*	49.5
enqu	*Survey*	78.4	volk	*Population*	46.0
congres	*Congress*	78.4	heer	*Sir / man*	44.5
goudsmit	*Goudsmit*	70.8	huis	*Home*	44.2
teur	*Teur*	69.1	televisie	*Television*	43.7

Topic 8	**Translated**	**weights**
gegeven	*Data*	876.1
naam	*Name*	303.3
jaar	*Year*	278.1
vraag	*Question*	264.4
groot	*Big / Large*	249.6
telling	*Census*	239.3
computer	*Computer*	239.3
zeggen	*To say*	229.1
mens	*Human*	224.1
stellen	*to argue or to pose a question*	200.4
staan	*To stand*	197.5
geven	*To give*	193.5
weten	*To know*	191.4
overheid	*Government*	188.0
mogelijk	*Possible*	186.4
privacy	*Privacy*	185.7
vragen	*To ask / Questions*	185.6
goed	*Good*	183.6
krijgen	*To receive*	155.5
persoonlijk	*Personal*	153.5

Appendix III: National Census 1971 – Translated

Disclaimer: The original census was a punch card system, in which the participant had to color in a circle in front of the corresponding answer. This made the census easily readable for the computer. For this translation, only the questions and the answers are translated – not the style of the original census form.

The census has been translated from the national census as presented by Jan Holvast in his book *De Volkstelling van 1971: Verslag van de Eerste Brede Maatschappelijke Discussie over Aantasting van Privacy*, 2013.

Card 1 (front side):
1. Sex: *Male, Female*
2. I am: Unmarried, *Married, Divorced ('from table and bed'), Divorced ('for real'), Widowed*
3a. Birth month: *Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec*
3b. First numbers birth year: *186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197*

3c. Last number birth year: *1, 2, 3, 4, 5, 6, 7, 8, 9, 0*
4a. Was he/she born in his/her current municipality of residency? *Yes, No*
4b. When is the most recent time he/she moved into this municipality? *At birth, Before 1930, 1930 – 1939, 1940 – 1944, 1945 – 1949, 1950 – 1954, 1960 – 1964, 1965 – 1969, After 1969*
5. Was he/she not born in his/her current municipality of residency, in which province was he/she born? *Gr, Fr, Dr, Ov, Gld, Utr, N-H, Z-H, N-B, L*
In case he/she was born abroad, fill in question 5 on the back of this page.

Card 1 (back side):
Only fill in if none of the pre-printed answers was applicable.
5. In which country and which municipality was he/she born?
Birth municipality _____
Birth country _____

Card 2 (front side):
6a. Does he/she have a job? *Yes, No*
6b. If not, is he/she: *Employed in their own household or that of their parents, Retired, Pupil or student ,Unemployed, looking for a job Not employed due to other reasons*
7a. Is he/she because of illness, accident, old age, congenital defect or similar issues dependent on: *Help of others, Special aids*
7b. If yes, this help incorporates: *Own care, Household chores, Tasks or placements out of the house*
7c. Is he/she constantly bedridden? *Yes , No*
8. Is he/she: *Head of the household, Married to the head, Unmarried child of the head Single*
In case he/she has another task within the household, fill in question 8 on the backside of this page.
9. What is his/her religious denomination? *Nederl. Herv., Rooms-Kath., Geref. Kerken, Geref. Kerken (vrijgemaakt), None*
In case he/she is of other religious denomination, fill in question 9 on the backside of this page.
10. What is his/her nationality? *Dutch, Belgian, German, Italian, None (stateless)*
In case he/she is of other nationality, fill in question 10 on the backside of this page.

Card 2 (back side):
Only fill in if none of the pre-printed answers was applicable.

8a. The relation to the head of the household of which he/she is part (*example: father, father-in-law, granddaughter, resident maid, aunt, friend, etc.*) _____
8b. Does he/she live with unmarried, own and/or stepchildren? *Yes, No*
8c. In case of marriage: does he/she live with their spouse? *Yes, No*
8d. In case of not married: Does he/she live with their own father and/or mother? *Yes, No*
9. Fill in the religious denomination as specific and complete as possible (*see also the separate accompanying explanation*) _____
10. Fill in the nationality as specific and complete as possible (if more than one nationality, fill in all nationalities) _____

Card 3
11. What was his/her primary source of income (life support) in 1970? *Income through labor, Property income, Pensions, Social benefits or student loans, Support of partner or parents or caretakers*
12. When receiving pensions, does he/she also have: *Social benefit, Other income, No other income*
13. What class is his/her income? *A, B, C, D, E* (*see attachment*)

ONLY for MARRIED WOMEN, WIDOWS and DIVORCED WOMEN
1. How many children has he she brought to life (also count children who passed away)?
In case married:
2a. How many of these children are born from the current marriage?
2b. Do all of these children still live with her?
2c. When was the current date of marriage?
2d. Been married before? *Herself: Yes, No Her husband: Yes, No*

Card 4 (front side)
ONLY for PEOPLE of 12 YEARS and OLDER

1a. Does he/she have daily education? *Yes, No*
1b. If yes, what type of education is this?
Education type: _____
Field of study: _____
Year: _____
1c. If he/she does not have daily education, did he/she after primary school follow at least one year of education? *Yes, No*
2a. Did he/she have one of the following educations: *VGLO, LAVO, (M)ULO, MAVO or VHMO? Yes, No*
2b. Did he/she successfully pas the third year of VHMO? *Yes, No*
2c. Which of the following diploma's does he/she have? *ULO or MULO, HBS or MAVO, Trade school, MMS or HAVO, HBS 5 or 6 year, Gymnasium*
3a. Did he/she have other education of at least one year? *Yes, No*
3b. If so, please fill in question 4 on the back of this card

5a. Did he/she follow an study in the field of education? *Yes, No*
6a. Did he/she study at a university, college, theological college or *grootseminarie*? *Yes, No*
6b. If yes, which one? _____
6c. Field of study? _____
6d. Exams successfully participated in: *Candidate exam, Doctorate, Promotion*

Card 5
1a. What job/function does he/she have? _____
1b. Describe the activities of this job? _____
2. Does he/she execute this job as: *Employee, Self-employed, Family company*

3a.If he/she is self-employed or in charge of people, how many people is that? *0, 1 – 4, 5 – 9, 10 – 19, 20 – 49, 50+*
3b. What is the nature of this managerial function? _____
4. If self-employed and not solely the manager, what is his/her function? _____
5a. Where does he/she work (*name of the company, possibly name of the director*) _____
5b. What sort of company, office, school, practice, shop or other institution is this? _____
5c. For which department does he/she work?
5d. Is the company part of the government, municipality, church, a particular person, etc.?
_____
6. What is the address where he/she has to go daily? (*If no fixed address, fill in address of employer*)

Card 6
7. Is the address given at question 6 his/her: *Fixed working address, Address where to report to daily, Not the fixed address or address to report to*
8a. In case of a fixed address, how much time does it take to get to work? *He/she works at home, Less than 15 minutes, 15 – 29 min., 30 – 44 min., 45 – 59 min., 60 – 89 min., 90 – 119 min., 2 hours or more*
8b. With what type of transportation does he/she travel? *Bicycle, Moped, Motor or scooter, Train, Tram or subway, ,Public bus, Corporate bus, Personal car/van as main driver, Personal car/van as fellow traveler, Other, No vehicle (by foot)*
8c. Does he/she usually go to the working address at least 4 times a week?
9. Where does he/she mainly work? *In the municipality of residence, In another municipality, In different municipalities or at sea*
10. In case one has a personal car, where does he/she park the car? *In the open air on a public road, In the open air somewhere else, Indoors in a commercial garage, Indoors in a personal garage, Indoors somewhere else*

Card 7
11. How many hours does he/she work at the primary job as filled in on card 5, on average per week? *Less than 10 hours, 10 – 14 hours, 15 – 19 hours, 20 – 24 hours, 25 – 29 hours, 30 – 34 hours, 35 – 39 hours, 40 – 44 hours, 45 or more hours per week*
12a. Does he/she partake in payed secondary jobs or functions? *Yes, No*
12b. If yes, how many hours per week does he/she work there on average? *Less than 5 hours, 5 – 9 hours, 10 – 14 hours, 15 – 19 hours, 20 – 24 hours, 25 – 29 hours, 30 or more hours per week*
2c. Does he/she partake in the secondary job: *As employee, Self-employed, As part of a family business*

ONLY if SEEKING FOR A JOB, *WACHTGELDERS* and *TEWERKGESTELDEN*
1a. Is he/she looking for a job? *Yes, No*
1b. If yes, is this the first time he/she tries to find a job? *Yes, No*
2. Is he/she registered at the employment office? *Yes, No*
3. What is the job he/she tries to find? Or, if not looking, what is the last job he/she had? Answer this question at card 5, question 1a.
4. Is he/she placed at a job: *At a social workplace, Complementary labour, He/she was not placed at a job*

Card 8
ONLY for HEAD OF THE HOUSEHOLD and for SINGLE PEOPLE
1. Is he/she: *Main occupant without co-inhabitants, Main occupant with co-inhabitants, Co-inhabiting and looking for own residence, Co-inhabiting not looking for own residence*
2. Does he/she have a home telephone? *Yes and one subscription number, Yes and two subscription numbers, No*

3. How many rooms does he/she at the current address of residency use for commercial practices?
4a. If co-inhabiting, how many rooms does he/she rent from the main occupant (*Also count the kitchen, if shared*).
4b. Does he/she have: *A private kitchen, A shared kitchen, No kitchen*
4c. How many more rooms in the residency are shared with the main occupant or other inhabitants?

END of the census to be filled in by the POPULATION, FURTHER to be filled in by the CIVIL SERVANT
1. Is the house of residency: *Inhabited by one household or single person, Inhabited by more than one household or single person, A residency without main occupant, A second residency, An empty residency*
2. Is the residency: *a normal home, A home with a shop or working place, a farm or gardener's house, Other (fill in on the back of this card)*

Card 9
3a. Is the residency inhabited by the owner? *Yes, No*
3b. If no, is it property of: *The municipality, A housing association, The province or the State, A private person, A private institution*
3c. Is the residence*: A woningnet-home, Official residence (dienstwoning), Charity home (liefdadigheidswoning)*
4a and 4b. How much is the rent (value) of the home?
4c. In this amount, the following is included: *Water, Fuel, Other costs*
5. Is the residence part of a housing complex of at least four residencies for: *Elderly people (Yes, No), Students or working women and such (Yes, No)*
6a. Is the home a single family home which is: *Detached, Connected to neighbors on one side, Connected to neighbors on both sides, A ground floor or upstairs apartment, Part of a corporate building*
6b. Where is the main living area? *In souterrain, On ground level, On the ___ floor*

Card 10
7. Does the residence have its own door of access? *Yes, No*
8. When was the residence build? *before 1906, 1906 – 1918, 1919 – 1930, 1931 – 1944, 1945 – 1949, 1950 – 1954, 1955 – 1959, 1960 – 1964, 1965 – 1969, after 1969.*
9. Does the residence have connection to: *Water (Yes, No), Electricity (Yes, No), Gas (Yes, No)*
10a. Does the residence have a (living) kitchen of: Less than 4 m^2, 4 – 12 m^2, 12 m^2 or more, There is no (living) kitchen
10b. And above all this number of rooms:
11a. Where is the toilet located? *Indoors, Outside the living area (hallway, portal, other), Outdoors, No toilet present*
11b. In case a toilet is present, is there*: Drainage (Yes, No), Running water (Yes, No)*
12a. Does the residence have: *A sink (possibly with shower), A bathtub (possibly with shower), A shower (no bathtub), No bathing spot*
12b. If a sink, shower or bathtub is present, is this placed in a separate room? *(Yes, No)*
13a. Does the residence have: *A private heating system, Warm water heating, City heating, No central heating system*
13b. What is the primary source of heating? *Coals, Oil, Gas, Other*

Appendix IV: Python Codes

For the codes used within this thesis, I have to thank my research colleague J. Veerbeek for providing me with help, advise and exemplary scripts on which I eventually based my thesis. Both codes were provided to me, and the second code was altered by me to fit my research.

Delpher Search I:

```
import requests
import json
import datetime
from lxml import etree
from urllib.request import urlopen
from xml.etree import ElementTree

class DelpherAPI:
    page = 0
    records_processed = 0
    number_of_records = None

    def __init__(self, ppn, from_date, until_date, queryterm, collection='ddd',
record_type='artikel'):
        self.ppn = ppn
        items_per_page = 100
        self.query_template =
'https://delpher.nl/nl/api/results/coll/{collection}/query/%22{queryterm}%22/' \
'facets[type]/{record_type}/cql/%28date+_gte_+{from_date}%29/cql/%28date+_lte_+{until_da
te}%29/' \

'cql/%28ppn+any+%28%22{ppn}%22%29%29/maxperpage/{items_per_page}/sortfield/date/
page/'
        self.query = self.query_template.format(**locals())

    def result_pages(self):
        while self.number_of_records is None or self.records_processed <
self.number_of_records:
            self.page += 1
            articles = self.list_next_articles()
            self.records_processed += len(articles)
            yield articles

    def results(self):
        for result in (result_page for result_page in self.result_pages()):
            for r in result:
                root = etree.parse(urlopen(r['identifier']))
                r['text'] = ' '.join(root.xpath('//p/text()'))
                yield r

    def results_url(self):
        return self.query + str(self.page)

    def list_next_articles(self):
        url = self.results_url()
        try:
            r = requests.get(url)
```

```
            response = r.json()
        except:
            return []
        self.number_of_records = response['numberOfRecords']
        return response['records']

    def get(url):
        expected_status = 200
        print(url)
        r = requests.get(url)
        return ElementTree.fromstring(r.content)
```

Delpher Search II:

```
import delpher_search as dh
import pprint
import csv

pp = pprint.PrettyPrinter(indent=4)

OUTFILE = 'datasets/set_waakzaamheid.csv'
field_names = ['id', 'title', 'newspaper', 'page', 'date', 'url', 'text', 'query name']
writer = csv.DictWriter(open(OUTFILE, 'w', encoding='utf-8'), fieldnames=field_names,
                        lineterminator='\n', delimiter=';')
writer.writeheader()

volkskrant = '412869594'
trouw = '412789353'
waarheid = '832737666'
leeuwarder = '865061483'
nrc = '400367629'
parool = '412869543'
limburg = '83245351X'
telegraaf = '832675288'
tijd = '842127143'
nieuwsblad = '833013246'
vrijevolk = '832737143'
neddagblad = '810209039'
niw = '831178310'

numbers = {volkskrant, trouw, waarheid, leeuwarder, nrc, parool, limburg, telegraaf, tijd,
nieuwsblad, vrijevolk, neddagblad, niw}
words = [‘volkstelling’]
for number in numbers:
    ppn = number
    from_date = '01-01-1970'
    until_date = '31-12-1971'
    for word in words:
        query = word
        uri = dh.DelpherAPI(ppn, from_date, until_date, query)
        for article in uri.results():
            article_dict = {'id':article['metadataKey'],
                        'title':article['title'],
                        'newspaper':article['papertitle'],
```

```
            'page': article['page'],
            'date':article['date'],
            'url':article['identifier'],
            'text':article['text'],
            'query name':query,
        }
        writer.writerow(article_dict)
```

Script for LDA Topic Modeling for Dataset:

```
import nltk
from nltk.tokenize import RegexpTokenizer
from nltk.corpus import stopwords
import pandas as pd
import numpy as np
import re
import os
import pickle
import gensim
import scipy as sp;
import sklearn;
import sys;
import nltk;
from gensim.models import ldamodel
import gensim.corpora;
from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer;
from sklearn.decomposition import LatentDirichletAllocation;
from sklearn.preprocessing import normalize;
import pickle;

#import lemmatized files and create string of second column of each file
filelist = os.listdir(r"C:\Users\Ruudd\Desktop\Thesis_dataset\topic
modelling\texts_lemmatized\texts_lemmatized")
path = r'C:\Users\Ruudd\Desktop\Thesis_dataset\topic
modelling\texts_lemmatized\texts_lemmatized\\'
art_list = []
df = pd.DataFrame()

for file in filelist:
    article = pd.read_csv(path + file, header=None, sep='\t', error_bad_lines=False)
    s = article[2]
    string = s.str.cat(sep=' ')
    txt = string.lower()
    art_list.append(txt)

df_articles = pd.DataFrame(art_list)


#function to remove stopwords
f = open('stopwords-nl.txt', 'r')
lines = list(f)
my_stopwords = []
for i in lines:
    my_stopwords.append(i.rstrip('\n'))
```

```
word_rooter = nltk.stem.snowball.PorterStemmer(ignore_stopwords=False).stem
my_punctuation = '!"$%&\'()*+,-./:;<=>?[\\]^_`{|}~•@>>»„'

# cleaning master function
def clean_articles(article, bigrams=False):
    article = article.lower() # lower case
    article = re.sub('['+my_punctuation + ']+', ' ', article) # strip punctuation
    article = re.sub('\s+', ' ', article) #remove double spacing
#     article = re.sub('([0-9]+)', '', article) # remove numbers
    article_token_list = [word for word in article.split(' ')
                          if (word not in my_stopwords and len(word) > 3)] # remove stopwords

    if bigrams:
        article_token_list = article_token_list+[article_token_list[i]+'_'+article_token_list[i+1]
                                    for i in range(len(article_token_list)-1)]
    article = ' '.join(article_token_list)
    return article

#remove stopwords and create dataframe with clean articles and save as pickle
df_articles['clean'] = df_articles[0].apply(clean_articles)
df_articles['clean']
df_clean = df_articles['clean']
pickle.dump(df_clean, open('clean_data.p', 'wb'))


#load data
data = pickle.load(open('clean_data.p', 'rb'))

#LDA Model

# the vectorizer object will be used to transform text to vector form
vectorizer = CountVectorizer(max_df=0.95, min_df=5, token_pattern='\w+|\$[\d\.]+|\S+')
# apply transformation
tf = vectorizer.fit_transform(data).toarray()
# tf_feature_names tells us what word each column in the matric represents
tf_feature_names = vectorizer.get_feature_names()
number_of_topics = 9
model = LatentDirichletAllocation(n_components=number_of_topics, random_state=0)
model.fit(tf)

def display_topics(model, feature_names, no_top_words):
    topic_dict = {}
    for topic_idx, topic in enumerate(model.components_):
        topic_dict["Topic %d words" % (topic_idx)]= ['{}'.format(feature_names[i])
                    for i in topic.argsort()[:-no_top_words - 1:-1]]
        topic_dict["Topic %d weights" % (topic_idx)]= ['{:.1f}'.format(topic[i])
                    for i in topic.argsort()[:-no_top_words - 1:-1]]
    return pd.DataFrame(topic_dict)

no_top_words = 15
df = display_topics(model, tf_feature_names, no_top_words)
display_topics(model, tf_feature_names, no_top_words)

pickle.dump(df, open('topic_words.p', 'wb'))
```

```
doc_topic = model.transform(tf)

for n in range(doc_topic.shape[0]):
    topic_most_pr = doc_topic[n].argmax()
    print("doc: {} topic: {}\n".format(n,topic_most_pr))

topic_distr = pd.read_csv(r'high_topic_per_doc.csv', sep=';')
topic_distr.head()
pickle.dump(topic_distr, open('topic_distr.p', 'wb'))

topic_no_of_docs = topic_distr['topic'].value_counts()
topic_no_of_docs
pickle.dump(topic_no_of_docs, open('topic_no_of_docs.p', 'wb'))

series = data.str.split(expand=True).stack().value_counts()
pickle.dump(series, open('word_frequencies_total.p','wb'))

df = pickle.load(open('topic_words.p', 'rb'))

df.to_csv(r'topic_table.csv')

pickle.load(open('topic_distr.p', 'rb'))


pickle.load(open('topic_no_of_docs.p', 'rb'))


word_freq = pickle.load(open('word_frequencies_total.p', 'rb'))
word_freq[:50]
```