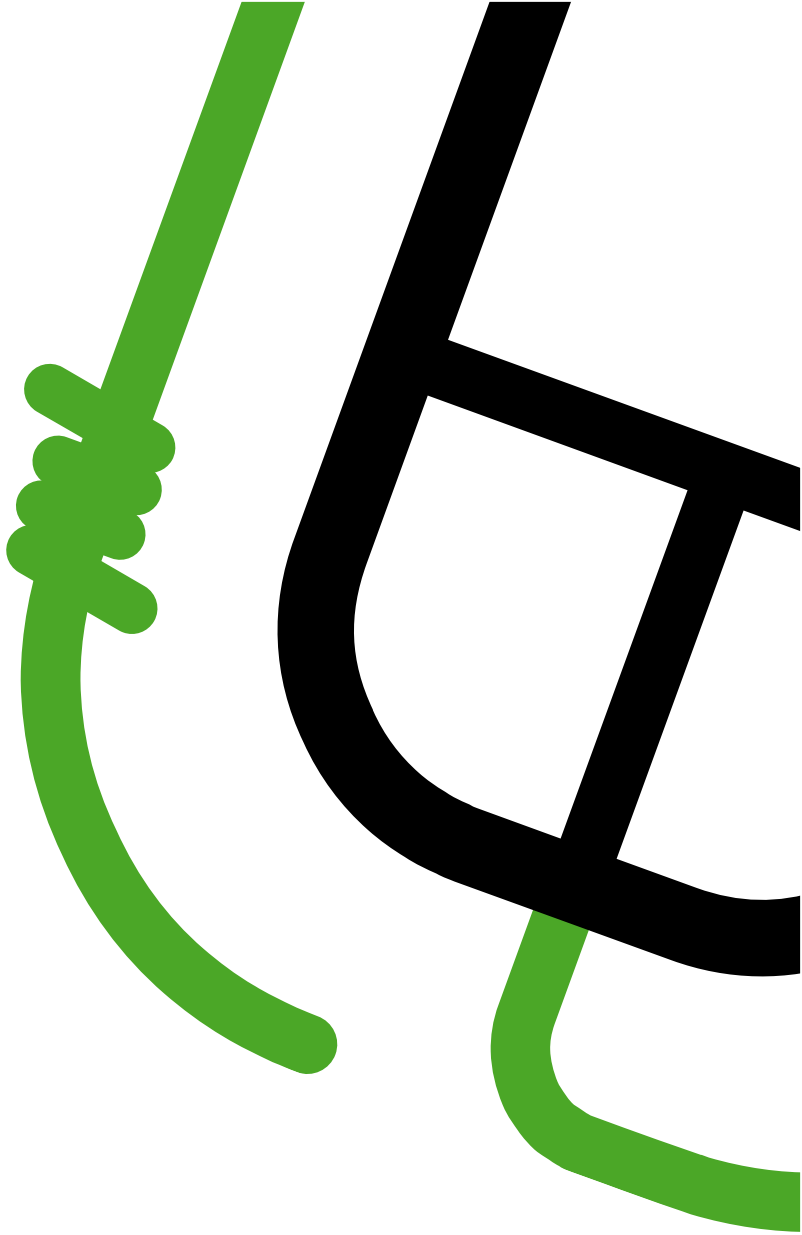
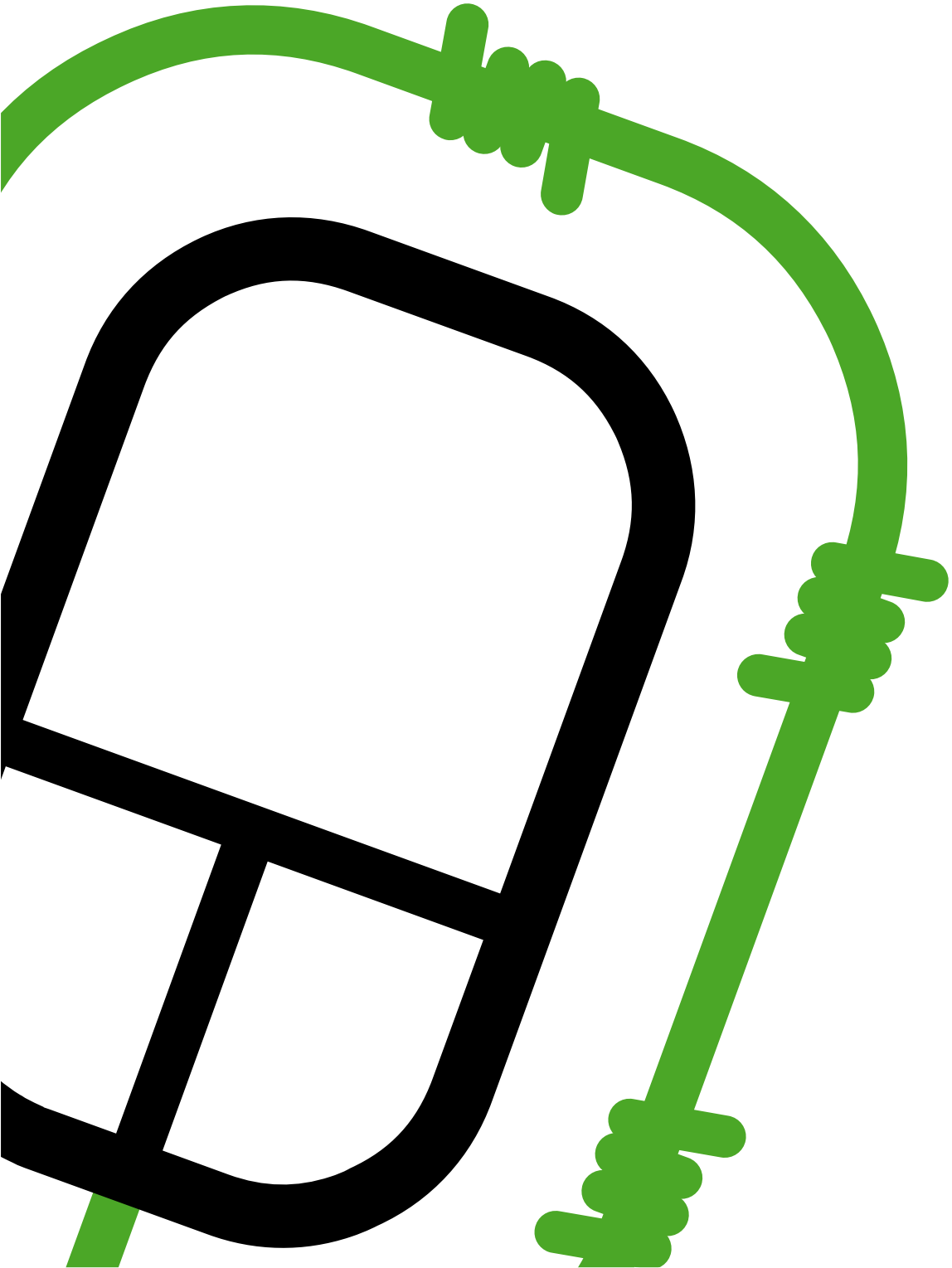
TRANSPARENT CONSUMERS



Data brokers and profiling in the Netherlands - **Floris Kreiken**

4 February 2016

BITS OF FREEDOM
VERDEDIGT DIGITALE BURGERRECHTEN



CONTENTS

BITS OF FREEDOM
VERDEDIGT DIGITALE BURGERRECHTEN

Stichting Bits of Freedom
Postbus 10746
1001 ES Amsterdam
The Netherlands
Tel. +31 6 4499 5711

floris.kreiken@bof.nl

01. EXECUTIVE SUMMARY

This research looks at data brokers and profiling for commercial marketing and credit rating in the Netherlands. It argues that most of the conduct by data brokers violates Dutch data protection law and that our current legal framework is insufficiently enforced to address profiling and conduct by data brokers. This presents risks to our autonomy and to society in general.

Recent research shows that in some countries, companies such as data brokers are increasingly collecting personal data for profiling. They can then offer these profiles to customers who use them for commercial marketing and credit rating.

The conduct of data brokers and profiling can create ethical risks. Personal data can be processed in a way that gives companies and governments power over people. It allows them to follow someone's information trail step by step, to manipulate their economic decisions, to categorize individuals, to sort and discriminate among individuals, to impede forgetfulness (the possibility to forget as well as being forgotten), to inhibit one from changing or progressing; and to infringe or steal one's identity. '[1] In other words, in the wrong hands, or applied the wrong way, profiling technologies could be used to harm people.

This research therefore looked at the Dutch situation and aimed to map the scope of the data brokers and commercial profiling industry in the field of commercial marketing and credit rating in the Netherlands, their legality, and to evaluate how society can mitigate any risks associated with it.

First, the research team conducted a literature study. Afterwards the team gathered a comprehensive list of data brokers and approached them for interviews and with data access requests. The team also acted as a data broker to see which data it could access, and made profiles with the help of experts. An expert session was organized on ethical and legal aspects.

The research reveals that there are many data brokers in the Netherlands that collect personal information (at times sensitive). They get their information from various sources (public and commercial) and make profiles on people. For those people it is difficult to control their data flows and it is not transparent. It is also difficult to obtain information about how your data is used and to get information about your credit score The research also shows that it is fairly easy for data brokers to get access to information and to make profiles. It also shows that small changes in data can have big results and that it's not always clear why some profiles have certain outcomes.

We have reached the following conclusions:

1. **Data brokers have no ground on which to collect so much data**

   People worry about control, but often feel overwhelmed by the perceived lack of choice they have. The research reveals that in the case of data brokers, users have little control over what happens to their data. Data is processed in a take it or leave it way, and once it is processed there are few possibilities to prevent further processing.

   The research shows that none of the grounds for processing legitimize the current practices of data brokers. Consent can't be a ground as there is no direct contact between the data broker and the data subject. The processing is also not necessary for the performance of a contract. The balancing provision is the remaining ground, but it is weak, as we argue

that the privacy interests of data subjects prevail.

The research also shows that sensitive data are processed, without asking for explicit consent. This means data brokers are not just in breach of the law but also of their own code of conduct.

2. The purpose limitation is not respected

The purpose limitation is the cornerstone of data protection. Recent societal unrest surrounding the ING bank in the Netherlands and the conduct of TomTom show that denying this right is not accepted by the general public. The purpose limitation protects the most important values of data protection, like the ability to confide.

As this research reveals the increasing amount of data collected and re-used by different companies and how easy re-use and collection is, the purpose limitation is an increasingly important safeguard as data slips away from user control.

It remains questionable to what extent this is respected by data brokers. Although some commercial entities state that data is shared with list brokers, this says nothing about the purpose of those data brokers. Data subjects have no way to know how their data is further processed by those parties. Furthermore, data brokers have given us little information about with whom the data is further shared.

3. There is little transparency about data brokers and data traffic

The current practices of data brokers are not transparent. Notices provided by parties that share data with brokers are vague and unspecific.

Data brokers themselves are also not transparent about how they use their data, where they get their data and with whom they share their data. Data brokers should be more open about the type and amount of data they use, where they get this information and how they create profiles.

4. There is no way for people to object

The research shows that once data is shared, it is further shared with third parties. There should be limits to this chain and opportunities for people to object to processing. Onward sharing makes it increasingly difficult for users to exercise control over their data and to prevent further processing.

We should also critically evaluate the reuse of public data and allow people the opportunity to object to processing.

5. Data subject rights are insufficiently respected

The research reveals that some data brokers don't respond to access requests and that people don't get the information about their profile to which they are legally entitled.

People should be meaningfully informed about profiling. They should also be able to tell what that profile is and be able to ask for human intervention and due process when decisions are made that concern them on the basis of profiles.

6. There should be more enforcement

The practices of data brokers and profiling should be carefully monitored by the competent authorities. The new Dutch law that has come into effect on January 1st of 2016 and the European Data Protection Regulation both promise an enforcer 'with teeth'.

This is promising as this research reveals some shadowy practices. After the laws enter into force, the data authority should closely watch the behavior of these companies. It is also important that active monitoring and enforcement happens by other organizations, mandated by groups of people, or through class action lawsuits.

We also recommend more proactive research and activity by the anti-discrimination authority. Research reveals that some profiles have the ability to indirectly discriminate against certain groups of people. The problem is that this discrimination is difficult to spot, in particular when companies don't use sensitive data. This requires new monitoring tools for the anti-discrimination authorities.

02. INTRODUCTION

Dutch people worry about protecting their data. Recent research[2] shows that 82,5% of the Dutch population thinks privacy is very important, and that a large majority takes active steps to mask their Internet behavior and protect their computer. The research also makes apparent that people ask for more control, more transparency, and more accountability from companies. People are worried about what happens with their data once it is shared.

Current research and political debates have focused mainly on the front side: tracking by websites and the use of data by big companies like Facebook and Google. But what happens in the background? Where do the data end up? How are the data further processed? What instruments exist to prevent illegitimate processing?

In the US, there is increasing attention to data brokers[3]. An FTC report showed that some data brokers keep extensive profiles on data subjects with sensitive information, ranging from medical searches online to social security numbers and personal interests. Little is known about those practices in the Netherlands. How do data brokers buy and sell personal information about Dutch people? This research expands our knowledge in this field.

Data brokers buy data from different sources and can create profiles or sell data to other data brokers, who can use these profilers for commercial purposes. For example, marketeers can use profiles to tailor their messages to people that fit a certain profile. Little is known about these profiles. Every now and then, the media reports on new tools for analytics that may be used to analyze people's behavior. Promises are made by the vendors of these technologies: for example, they promise more innovation, more security and more efficiency, but it is unclear what these promises really mean and to what extent these promises collide with societal values.

These promises are particularly made in markets that involve risk, like employment, credit rating and insurance. In those markets vendors try to minimize their risk, by acquiring more data on people. Traditionally, there were limits to what market vendors could know about their customers. Vendors knew more about their products or services, while people knew more about themselves and their preferences. This balance is under pressure by increased data collection and analysis. For example, it is not allowed as an employer to ask a woman about a pregnancy wish during a job interview, as that could lead to discriminatory outcomes. But using data analytics, employers can make accurate predictions without having to ask for that information. This threatens the position of women on the labor market. Meanwhile, risk minimization in credit rating could for example disproportionally harm people with certain social economic or cultural backgrounds.

Actors in those risk markets have expressed interest in using big data technology to minimize their risks. This is particularly true for the US. In the case of credit rating for example, companies like Zest finance promise to optimize credit lending by using large quantities of data.[4] For employment, companies like Cornerstone promise to use big data analytics to find "talent" and make "better workforce decisions".[5] Insurance companies are increasingly using marketing data to make insurance decisions.[6] Similar practices have extended to the Netherlands. Companies like Klarna and Afterpay are active on Dutch soil. How far do these practices go?

Furthermore, little is known about the vendors that

create, sell, and use those profiles, although they might influence people's lives greatly. This research attempts to lift that veil. Apart from the collectors of personal data, this research will look at where those data ends up, how they're analyzed, and how that knowledge is used. What kind of companies do this? What are their goals?

Companies have increasingly used profiling to improve marketing and credit rating. However, profiling presents risks as well, it can lead to exclusion and unfair discrimination. Therefore, the main research question this research aims to answer is: what is the scope of the commercial profiling industry in the field of commercial marketing and credit rating in the Netherlands and how can we mitigate any risks associated with it?

This research was done by Bits of Freedom and De Correspondent. We explored the conduct of data brokers and wrote a legal analysis on the basis of this exploration.

This report first offers some background, by describing data brokers, profiling and the theoretical and societal risks connected to tracking and profiling. It then describes the findings by our researchers. Afterwards it describes the legal framework in connection to those findings. It finishes with some concluding remarks and recommendations. The appendix contains more information about our methodology.

This study expands our understanding of data brokers in the Netherlands, maps the ecosystem behind the creation and use of profiling, and explores potential new safeguards that the legislator or companies can offer users to mitigate any negative effects from the use of profiling.

The study offers some much needed transparency on the way data on Dutch people are currently used in the Netherlands, a wish expressed in surveys. This presents societal benefits, and could raise awareness of the potential risks connected to the use of big data for profiling. Furthermore, it adds to the societal debate currently held on the application of these technologies, in which Bits of Freedom and De Correspondent play a big role. These societal benefits extend beyond the Dutch context, as similar technologies might be applied in other countries.

Profiling technologies increasingly rely on big data technologies to make predictions about human behavior. This presents risks for values protected by human rights. These risks arise in the collection, transfer and application of data. Identifying these risks and the legal frameworks regulating them provides policy opportunities in the Netherlands, but also extends beyond the national context.

03. BACKGROUND

Data brokers and profiling

In the US, the Federal Trade Commission has researched the practices of data brokers. In their report they say that data brokers - "companies that collect consumers' personal information and resell or share that information with others"- are important players in the new data economy.[7]

They say that there are different types of data brokers: (1) data brokers that are subject to the fair credit reporting act, (2) brokers that maintain data for marketing purposes and (3) brokers that maintain data for non-marketing purposes (for example to find people).[8]

According to the FTC, data brokers have three types of products: (1) marketing products (data brokers offer their customers information about people and their

preferences so that their customers can send marketing message to those people), (2) risk mitigation products (lenders contact data brokers to see whether people applying for a loan will be likely to pay back or whether they have a history of fraud) and (3) people search products (to allow customers to find people).[9]

The research also revealed that data brokers get their information from on- and offline sources, exchange information with each other, operate outside of the scope of consumers' knowledge and that data brokers collect and combine this information to create profiles. These profiles can include sensitive inferences.[10]

Profiling is a key tension area

Big data, the collection, storage, and analysis of data on an incredible scale, challenges existing notions of data protection, human rights and societal values. Although the technology promises benefits, human rights organizations like the Electronic Frontier Foundation[11] and EPIC[12] and academics like Nissenbaum and Barocas[13] warn us for the adverse consequences implementations of this technology might have on society.

One of the key areas where this tension surfaces is the area of profiling. Profiling is a new form of knowledge generation that makes visible patterns that are "invisible to the naked human eye."[14] Profiling adds new forms of knowledge: "profiles do not describe reality, but are detected by the aggregation, mining and cleansing of data. They are based on correlations that cannot be equated with causes or reasons without further inquiry; they are probabilistic knowledge"[15]

Profiling is the use of algorithms to discover correlations and patterns in data representing people. This technique is referred to knowledge discovery in databases and is also associated with 'machine learning.' It is defined as "[T]he nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data."[16] Because of the possibilities of big data, companies are using larger datasets and have more advanced analytical tools for profiling.

To effectively profile, data are collected on people from different sources and then used to create profiles. People can then be treated on the basis of predictions made in those profiles. The data can reveal an enormous amount of information. Based on websites visitors' clickstream data, predictions can be made about gender, age, level of education and occupation.[17] Data from social media is even more telling. Based on just Facebook 'likes', researchers were able to predict with relative accuracy characteristics like sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender.[18]

There is a difference between static profiling and dynamic profiling. The latter is connected to machine learning and not the subject of this research.

The observations and patterns are interesting for new business models or for new, more efficient and effective types of governance.[19] Profiles are increasingly used: for marketing, credit analysis and for risk determination.

In the case of risk determination for example, the Dutch SyRI law allows the creation of risk scores for children on the basis of which authorities can decide to preventively act in "problematic" families.[20] Border control makes risk scores that determine whether you require additional checks at the airport.[21] The national tax service determines the risk that you will not pay

your debt.[22] Insurance companies determine your premium insurance in exchange for private information,[23] or they will determine your premium for your car and house insurance on your postal code, home number and sometimes even home number addition.[24]

In Singapore, the government has a program called 'Total information awareness' (TIA). This program collects all kinds of electronic data: email, telephone logs, Internet searches, reservations, hotel bookings, credit card transactions, medical reports, everything. On the basis of that information, they scan for problems. At first instance this was aimed at defense and anti-terrorism but it now has been expanded to economic planning as well.

The Chinese government launched a social credit system, that allocates a score to every citizen based on their everyday behavior, ranging from the things they buy, the books they read to even what other people think of them. The score rewards behavior in line with the wishes of the ruling party: this score is then societally relevant: a higher score will allow them to get access to important jobs, loans and for example discounts on products.[25]

This research will not focus on governmental use of profiling.

For marketing purposes, profiles are used to target specific groups to increase the chance that people buy products or services.

For credit rating, companies increasingly exchange financial information and credit scoring. Credit data is combined with other data sources to recognize different groups in a population. This classification is used to predict the financial capabilities of people. Mathematical algorithms or statistical programs determine the probable debt repayments by consumers and assign a score to people according to risk classes.[26] These credit rating agencies aim for the stability of the financial system, the fight against consumer overindebtedness, and risk-management balancing in the interest of the profitability of the retail-credit industry. They can be public or private agents, of which the latter offer the market risk-management tools to improve economic efficiency and the profitability of credit providers. [27]

Profiling and the conduct of data brokers are risky

Information technology allows companies and governments to follow someone's information trail step by step, to manipulate their economic decisions, to categorize individuals, to sort and discriminate among individuals, to impede forgetfulness (the possibility to forget as well as being forgotten), to inhibit one from changing or progressing; and to infringe or steal one's identity. '[28] In other words, in the wrong hands, or applied the wrong way, profiling technologies could be used to harm people.

One of the risks associated with profiling comes from the way conclusions are reached. The predictions made on the basis of large quantities of data are not absolute and have biases and error rates. It is impossible to catch all data that are relevant to the social reality. Quantifying reality already presupposes a certain qualification. How can you turn reality into bits?[29]

These error rates and biases mean one should be careful with allocating to much credibility to technology and be wary of 'bad science.' Statistics can be misused or interpreted in the wrong way.[30] The larger the amount of observations, the larger the chance we can find correlations that aren't causal per se. A great illustration of this fact is the website

'Spurious Correlations,'[31] that shows the amount of questionable outcomes we can create from correlations. For example, like the connection between cheese consumption and the number of people who die by being entangled in their bedsheets.

Another risk associated with profiling is that it disturbs the balance between companies and government on the one hand and people on the other. Companies for example can expand their knowledge to a point where they know more about certain aspects of people than people know about themselves. A search engine can for example distill an incredibly personal profile based on past Internet searches leading to what has been referred to as a database of intentions.[32]

This shifting balance has societal consequences. On the basis of human dignity and autonomy, privacy and data protection rights protect us against the unjustified meddling with our private lives by others (but mostly the state), unless there are very good reasons to do so. These rights extend to our family, our home, our property, our communication, reputation and honor. New technologies have allowed for the easier infringement of privacy rights.

Harm to our security and autonomy could follow from information gathering, processing and spreading, as well as from intrusion in our private space. Real harm rarely follows from one intrusion, but rather follows from a series of intrusions.

For people privacy rights are important, because they protect against social pressure and offer the opportunity for concentration and rest. It grants people moral autonomy and freedom of choice and protects them from self-censorship and conformity. It protects against harmful categorization and against being judged out of context and allows for physical space where someone can be themselves. It also allows for a new start and enables people to play different social roles.

For society, privacy rights are important because trust is important in society. For instance, People share information with their doctor, and this is good for public health.

Especially important is that privacy rights protect the balance between individuals and companies/the government. This is good for democracy.[33] Profiling coupled with big data puts pressure on this information symmetry. Companies can share our behavior with institutions that have different interests than people.

People unknowingly generate input for analyses on an aggregated level and are then confronted with the results of such analyses on a personal level. In those cases, it is often uncertain what those decisions are based on.

The use of these profiles is not always visible. This is because a lot of data is collected without consent. This is difficult when people are then confronted with information used against them. For example, the ING bank in the Netherlands wanted to use information of its customers for different commercial goals. It created a lot of societal unrest.[34]

Unrest like this can be simple annoyance about receiving unsolicited advertisement but can be based on more serious disturbances: for example, being offended at being profiled in a way different to how people view themselves, or for being treated differently, by paying another price for products.

This treatment can lead to exclusion or can disrupt the balance of power between companies and people. For example, marketeers could abuse sensitive

information to influence people. Certain psychological triggers could be activated to increase conversion.

These risks are even higher in markets for credit rating, insurance and employment. For example, an employer might be able to accurately predict pregnancy, and refrain from hiring a woman. That poses risks of unfair discrimination. Insurance companies might ask higher fees for insurance in low income families as they are more prone to health problems. This could lead to exclusion of certain groups in our society from certain products or services (such as online facilities, credit, mortgages, or renting a home).

As Ferretti notes: "In short, information processing and technologies have a clear potential to dramatically influence the lives of people, and this influence puts an exceptional power in the hands of those who use information processing and technologies; this is a risk only recently perceived by business and consumer associations alike."[35]

This can be troubling in particular for the marginalized in society. Solon Barocas writes that profiling and machine learning confirm pre-existing links. Barocas and Andrew Selbst write that "Often, the "patterns" it discovers are preexisting societal patterns of inequality and exclusion. Unthinking reliance on data mining can deny members of vulnerable groups full participation in society. Worse still, because the resulting discrimination is almost always an unintentional emergent property of the algorithm's use rather than a conscious choice by its programmers, it can be unusually hard to identify the source of the problem or to explain it to a court."[36] The article shows that people with a societally worse position will be confronted with the effects of profiling earlier, because companies will not want to deal with people that have a deduced bad status like that. These potentially excluding effects call for more diligence when devising new policies. It is for example well known that people in lower income groups live less healthily.

Profiling can thus create unfair discriminatory outcomes by sharing benefits only with "good" people, not the other people. For both the credit industry and the advertisement industry it is economically wise to penalize vulnerable people.

04. FINDINGS

Part 1: Mapping and researching Data Brokers

Our researcher interviewed various data brokers, experts from the field and academics. She also send out data requests asking 25 data brokers if they had any information about her. The 25 companies were all registered in the Netherlands, and are: Experian, Graydon, Focum, 4Orange, Autoriteit Consument & Markt, Bureau Krediet Registratie, Cardatapool, Cardec, CDDN, Cendris, Company.info, Creditsafe, DAT.Mobility, Dun & Bradstreet, EDM, EDR, Facebook, Geodan, Geoscape, Google, Mastercard, Mint Marketing, Omniprofiles, PostNL, Rabobank, Sandd & T-Mobile.

These 25 were selected because they were either well known for dealing with consumer data, were companies that delivered services to our researcher (like her bank and credit card firm), or were suggested to us by readers of De Correspondent.

To answer the question on where they get their data from, our researcher conducted interviews with data brokers. Six companies were selected to visit and interview. A potential difficulty the team faced was that

profilers were not willing to share all information.

- There are many data brokers operating in the Netherlands.

The research reveals that there are a large number of data brokers active in the Netherlands. The number ranges to at least 180 companies.[37] They collect, analyze and sell data on people. For example, they help marketeers find new audiences for their products. When they do, they aim for better conversion (to increase the effectiveness of advertisement and the chance that their product is bought). To increase this conversion rate, they collect data on people. For example, they collect data on living areas. People that live in richer neighborhoods may be more easily enticed to buy luxury products.

Some data brokers use data for other purposes, and for example specialize in risk management. They use data to create credit scores. This score is an advice that tells companies how credit-worthy people are. Is someone going to pay back their loan? A bank will ask for this score when someone applies for a mortgage or a loan, and a web shop will use this to see whether someone can order products without paying in advance.

In both markets (commercial advertisement and credit rating) the same rule applies: data brokers collect enormous amounts of data from a range of sources and little is known about this collection.

- They collect personal and sensitive information and create group profiles.

The research reveals that data brokers process very sensitive information. For example, data broker 'NAW plus'[38] has lists of preachers, pastors and 'active Christians', which they collected from a publisher that sells mostly religious literature. They also sell names and addresses of visitors of a Christian camping and have email addresses of visitors of the young online Christian community Refoweb and Christian dating website 'Christianmatch.'

Another data broker, 'WIJ special media'[39] claims to have the addresses of all pregnant women in the Netherlands. They collected these because they offer free pregnancy kits to people in exchange for personal information. They not only collect address data but also the predicted birth date, name and sex of the child.

Other brokers, like 'Elite-miljonairs'[40] and 'Vip leads'[41] have the contact details of "wealthy" Dutch people and famous dutch people. The more money the data subjects have, the more expensive it is to get access to the data as a customer of those data brokers.

Any company can buy or rent a list of these addresses and names of people. At 'NAW plus' one would pay around 7.000 Euros for the data of 20.000 readers of the religious newspaper 'Reformatorisch dagblad.' 'Elite-miljonairs' sells the private home addresses on 1.500 "wealthy heirs" for 720 Euros.

Three data brokers had data on our principal researcher. For example, 'Experian'[42] knew that she was a woman, where she lived and that she had a land line. 'Sandd'[43] knew things about her home and living environment. 'Graydon'[44] had a copy of her company profile (she is registered freelancer), including a credit score.

- They get information from various sources.

When asked where they got the information on our researcher, 'Sandd' and 'Graydon' mentioned that they had their information from public records, without mentioning which ones. An 'Experian' employee says that the company "has a rich source of information."

Data brokers get a lot of information from public records, like the Dutch chamber of commerce, the Central bureau for statistics, and the Kadaster (a public register for data on real estate, housing value, property, borders and other geographic data). Although this public information may seem general, that is not the case: this information can be used to predict levels of income, levels of crime, other financial information and even predicted death rates.

Data brokers constantly update their information. They don't just use recent information from public records. Companies like '4Orange', 'Cendris' and 'EDM' also send out research surveys, on interests, living style and living situations. This personal information is connected to group profiles.

Apart from this, data brokers get information from Internet sources, like social media and from other commercial parties. These commercial parties (in retail and commerce) have agreements with data brokers. If someone buys shoes in a web shop, some of that data is shared with selected partners. Data brokers claimed that some of this information is obtained because people agree to this in the general terms and conditions of these web shops. Terms & conditions will mention that information is shared with "selected partners." However, they don't mention who their partners are. And those partners share information as well.

The data brokers we spoke to refused to mention what commercial sources they used for their data, because they claim this information is of competitive value. They also claim they don't have to mention this because it is a company secret. Some of them revealed that the sources are publishers, webshops, telecommunications companies and retailers, but they did not get specific.

- Data brokers use different data to create profiles.

The data brokers use the data to make profiles, but those aren't very accurate yet. The profiles our researcher fell into were wrong ('Experian' said she was a 45 year old Volvo driver and read certain magazines – she is significantly younger, does not even have a car and does not read those magazines).

Because most of the information came from the chamber of commerce, she was profiled in a certain way. The profile in this case was connected to her living situation.

Interestingly, some companies that don't have data on her can still profile her, because it is coupled to her living address. This means that it is possible for companies to create profiles on someone and treat someone on the basis of that profile without actually collecting data on that person.

The information collected is used to generate profiles and for example credit scores. Credit scores are important, interviewed parties claim, because they decrease risks for commerce. For example, in the case of phone subscriptions, they allow customers to get a phone on credit.

However, those scores aren't always correct. In 2012, Dutch TV program 'Kassa' investigated instances where people were labeled as a payment risk.[45] The credit rater in this case, 'EDR,' recognized however, that in some cases this rating was based on previous home owners.

- People have few ways to exercise control.

All data brokers assert that the data collection from these sources happens in a lawful way. Most of it happens through opt-in, according to data brokers in our interviews. When data comes from some

commercial sources, when buying something, people agree to terms and conditions that allow data processing to take place.

This makes it difficult for people to decide whether they want to share their information with third parties or to know in advance with who their data is shared. Resisting data processing has to happen after the fact, and even then is difficult.

Our research shows that it takes a lot of time to reach those companies. Also, some companies still don't react appropriately to requests for data. For example, four companies of the twenty five didn't respond at all to our researcher's requests. Another four spuriously claimed to need more information.

An additional problem was that it was difficult to get access to the creditscore or the way the score was made. 'Experian' didn't mention her score. At 'Graydon' she had access to the score, but not why she had this score.

Data brokers also claim a lot of advantages for commercial advertisement. The costs of the product go down, some claim, because companies make money by renting out their databases.

Part 2: 'Heel Holland Transparant'

'Heel Holland transparant'[46] was made with Atelier Yuri Veerman (an artist and designer) and is an ironical way to show how easy it is to collate data and use it for profiling. For this project, the research team collected (mostly) publicly available information from a group of famous and non-famous Dutch people and linked it with social media information.

Public sources were the central bureau for statistics, the Cadastre, the chamber of commerce, the police, the insolvency register, the public register for energy labels of housing, the provincial risk map, and social media companies like LinkedIn, Twitter and Coosto. Most of these data were free. Some social media data and data of the chamber of commerce required a modest fee.

As a part of the research, the project reveals that it is easy to get access to personal information. For example, for one person it became clear that there had been nine successful recent burglaries in the neighborhood, which for example could be interesting for home insurers. It is also easy to see whether early deaths are common in a neighborhood (relevant for life insurance) or whether someone lives in a high income neighborhood. Other information included: the use of medicine in a certain postal code, whether there are a lot of motor vehicles there, and information about bankruptcy and debt rescheduling.

It was also easy to create risk profiles with the help of experts. The research shows that small adjustments in the algorithm led to large consequences. Another expert explained that an important element of machine learning is that it may end up being unclear why the computer generates certain outcomes, making it hard to justify a particular decision.

05. LEGAL FRAMEWORK

The research shows that there are many data brokers that collect and sell information on Dutch people. It shows that this information includes sensitive data relating to religion and health. Information is obtained from various sources, including public and commercial. These data are then used to create profiles and apply those profiles to people for the purposes of marketing and credit rating. The actions are regulated by different sets of laws. Those laws range from human rights to EU legislation. EU legislation has been transposed into specific Dutch laws. Therefore, we will not describe the EU

legislation in detail. The following chapter will describe how the actions of data brokers are regulated by legislation.

Data brokers process personal data. This means data protection provisions are relevant. This right to data protection was enshrined in the constitutions and legislation of continental European countries.[47] It was intended to "minimize the threats posed by free and unregulated use and manipulation of personal information"[48] When the EU adopted the Treaty of Lisbon, the Charter of Fundamental Rights of the EU became binding. Article 8 of this charter created a right to the protection of personal data as an autonomous right distinguished from privacy.[49]

This article and later legislation specifying it lay out the specific rules that those who process personal data (data controllers) must follow when processing personal data. These rules aim to protect people against the unjustified collection, storage, use and dissemination of their data. While privacy rights protect "the legitimate opacity of individuals through prohibitive measures" data protection rules lay down the conditions under which data processing is legitimate, by creating transparency and allowing some user control when data processing is not authorized by the law.[50] Like privacy, data protection rules have at their center that democratic societies should not be based on control, surveillance, actual or predictive profiling, classification, social sorting, and discrimination."[51]

Data protection rules are necessary to protect "the collective social good and the fundamental values of a [..] democratic order where a citizen freely develops her personality and autonomy."[52]These data protection rules are relevant to the processing of data by data brokers. They are also relevant when data brokers apply their profiles to certain people.

Provisions that prevent unfair discrimination are relevant to the application of profiles. Data protection rules contain special provisions for the processing of sensitive data, like race or ethnic origin, political opinions, religious beliefs, trade-union membership, health or sex life, or criminal convictions. Meanwhile, the Dutch equal treatment act (called the "Algeme Wet Gelijke Behandeling") prohibits discrimination on grounds of "religion/belief, political beliefs, race, parentage, sex (man, woman, or transgender), pregnancy, nationality, sexual orientation (including bisexuality) and marital status". This discrimination may not be direct or indirect.[53] Indirect discrimination means that even though there is no explicit discrimination, differential treatment has discriminatory effects on the basis of one the above stated grounds (race, etc.). There is also an equal treatment law that focuses on handicaps and chronic illness.[54] The law states that this discrimination can not take place by the providers of services or goods,[55] unless "this differential treatment is justified by a legitimate interest and the measures to fulfill that interest are suitable and proportionate."[56]

The Data Protection directive regulates the processing of personal data in the EU. This includes the collection and use of personal data. It calls on Member states to set conditions for data processing. This directive has been implemented in to the Dutch data protection law (called the 'Wet Bescherming Persoonsgegevens," (Wbp)) the Dutch data protection law.

The EU also includes sector specific legislation on data protection. The directive on privacy and telecommunications translates the rules of the data protection directive to the telecommunications sector. The law states that getting access to the information

in terminal equipment of the end users is only allowed if the data subject is informed by clear and complete information and has consented to this. Providers of electronic communications services can for example place cookies on the terminal equipment of end users and these cookies can be used to follow their actions online. The directive has been implemented in Dutch legislation in the 'Telecommunicatiewet (TW),' the Dutch telecommunications law. As this research doesn't focus on tracking through cookies, its provisions have been left out the analysis.

On the basis of article 25 of the dutch data protection law, organizations that sufficiently represent a certain sector can draft a code of conduct which can be approved by the Dutch data protection authority. Such codes specify the interpretation of Dutch data protection rules in a specific sector. There are codes of conduct for private detectives, for financial services, the pharmaceutical industry, research, smart meters, network operators, health insurance, and data brokers. The code of conduct for data brokers is relevant to this research. However, these codes of conduct do not mean that specific conduct has been approved.

There is also sector specific legislation relating to consumer credit. This has been left out of this analysis.[57] Also, some legislation creates duties for entities to publish certain information,[58] like the Chamber of Commerce, which can provide information about people relating to their professional life. The sectoral legislation has been left out of this analysis, although the theme is discussed broadly in relation to the Dutch data protection law.

Data brokers

Data brokers collect, store and disseminate personal data. This means the provisions of the Dutch data protection law apply. These provisions require that processing is fair and lawful.[59] To be lawful, data may only be processed for a specific, explicit and legitimate purpose and may not exceed the specified purpose[60]. It requires a lawful ground for processing.[61] Users have to be informed (for example through Privacy statements) about the use, purpose, and recipients of personal data processing and the ground and details of processing. Data has to be accurate and up-to-date, and controllers have to take the reasonable steps to ensure the rectification and erasure of inaccurate data.[62] The data may be stored no longer than necessary based on the purpose for which they were processed.[63] The profiles data brokers can make, only fall under the data protection rules when they are applied to specific people.

According to some, "the cornerstone of the legislation is its requirement of individual consent for the processing of data, unless the processing is: necessary; subject to notice; for the performance of a contract to which the data subject is a party; for compliance with a legal obligation of the data controller; to protect a vital interest of the data subject; for the public interest; or for overriding rights of the data controller or third parties."[64]

The law also creates specific user rights. The data subject can ask the data controller to grant access to his data.[65] This includes the possibility to know where the data was collected.[66] He can also obtain the logic behind certain automated data processing.[67]

Enforcement of these provisions is handled by the Dutch data protection authority. Since January 2016, the authority also has the competence to hand out fines.

Data brokers and data processing

The collection and further processing of those data is

subject to specific requirements. Our research focused on data brokers themselves, and from what we found we cannot be certain how data brokers get their data, except for what has been said in interviews.

Data brokers process personal data. In the code of conduct on data brokers, the types of data that are specified to be collected range from broad to specific. From contact details, employment history, education history, income, debt and property data, legal history (related to debt), any data relevant to credit worthiness, and "other for the purpose of the processing relevant data." These data may be stored for 12 months after which they should be checked for accuracy.[68] In total, data may be stored for 8 years.[69]

First of all, these forms of data processing require a legitimate ground for processing.[70] We can rule out three grounds. There is no legislative obligation for data brokers to process these data (C of that article), it is not for the protection of a vital interest of the data subject (D of that article, this would be the case of a medical emergency) and there is no public task that requires data brokers to process data (E of that article). This means three possible grounds remain: consent (A), necessary for the performance of a contract (B) and the balancing provision (F).

Article 8A of the Wbp says the ground for processing can be consent. Consent has to be freely given, specific and on the basis of the information provided by the controller. In the information provided to the data subject, it should be made specifically clear with whom the data is shared and for what purposes.

The research reveals that data is obtained from commercial sources like publishers, web shops, telecommunications companies and retail companies. They also get data from surveys. As the data brokers haven't revealed the commercial sources from which they have received their information, it is difficult for us to verify whether the information provided to people is sufficient. Our research does suggest that in some cases, commercial parties simply say that information is shared with 'selected partners' or 'listbrokers.' That is insufficiently precise and would imply a free for all. One also wonders if consent is freely given or specific. Is there a choice offered when buying products that allows people to buy a product without sharing data with data brokers? More importantly, consent for processing requires direct contact between the processor (in this case the data broker) and the data subject, in which the data subject knows the purpose of the processing and the consequences of processing. The data subject can then give consent on the basis of that information. Because there has been no contact between the data broker and the people whose data is actually processed, the above would rule out consent as a valid ground for processing.

Article 8B of the Wbp could be another ground for data processing. This would mean that the data processing would be necessary for the performance of a contract. This requires that people are subject to an agreement, consciously participate in the agreement and that the processing is truly necessary. In the case of WIJ Media, filling out a survey allows people to get a particular package for pregnant people. That however, does not make this processing necessary for the agreement. In the other cases, data is generated apart from the transaction: buying something from a publisher, web shop, telecommunications company or retail doesn't make disseminating your data with third parties necessary for the conclusion of the contract. Apart from the data people share with the commercial party itself, which is necessary for billing, further processing is not necessary for the performance of this contract, so this ground would also be ruled out

as a valid ground for processing.

Article 8F of the Wbp allows for the processing of personal data if it is in the legitimate interest of the controller or a third party. This provision is also referred to as the balancing provision because the legitimate interest has to be weighed with the fundamental rights and freedoms of the data subject. In the code of conduct, the provisions reveal that data brokers can process data, either in a credit score or not, if it supports their decisions or the decisions of a third party on: selecting trade partners, maintaining trade relations, credit transactions, finishing trade relations, credit worthiness and entering work relations. All these goals are considered the legitimate interest of the data broker or a third party, according to the code.[71]

However, the balancing provision stipulates that processing for a legitimate interest is allowed unless the interests and fundamental freedoms of the data subject, in particular the rights to privacy, prevail. This balance requires the weighing of a number of factors.

First of all, it requires weighing the legitimate interest itself. For advertisers, this is the commercial interest. However, as a right that has to be balanced with other values, the commercial interest is not that strong. In the Google Spain case, the European Court of Justice has said that commercial interests do not necessarily prevail over human rights and fundamental values. For credit rating this is a different story. Preventing indebtedness is a stronger legitimate interest. However, the data broker in this case is not the party giving credit, so this would not qualify as a legitimate interest in this case.

Also, there are many reasons for overindebtedness, which mostly relate to misfortune[72] How far should people sacrifice their rights in the interest of the credit industry?[73]

Is this sharing necessary for that legitimate interest? There is still debate on this. Some authors suggest that it is possible to advertise without the use of extensive profiles on people. The same applies to credit rating: is its goal to prevent people from buying on credit if it leads to overindebtedness? Also, it is not the goal of the data broker to give out credit and prevent overindebtedness. The data broker's goal is to provide credit scores to other parties.

If we weigh this with the privacy interest of the data subject, a troubling image appears. Those interests are heavily infringed, especially because of the spill over effects. Data is not just shared with one list broker, but easily shared with others as well. This can create a detailed picture of people's lives, only exacerbated by new technological developments like Big Data.

The article 29 working party issued an influential opinion saying that in this balancing test, one should take into consideration protective measures taken by the controller.[74] This could for example be an opt out offered by the company, or more transparency. That does not appear to be the case. Taken together, this would make the balancing provision a very weak ground for processing by data brokers.

The code also names all the sources of data. These can be data they get from the subject themselves, from others about the data subject (although the broker has to be sure that those others have gotten the data lawfully) or from employers. For address and phone numbers they can also approach sources like neighbors, associations, etc. They can also access public sources like the chamber of commerce and the Kadaster.[75]

Art 9 of the Wbp lays down the purpose limitation. It

should be evaluated if the data is processed for a compatible purpose. To verify this, we need to look at the goal of processing and goal for getting the data (a), nature of data (b), consequences (c), way of getting data (d), guarantees to subject (e). In the code it also says data can be processed or transferred to others for one of the goals mentioned earlier, and can't be further processed for other goals.[76]

It remains questionable whether data brokers can legally operate within the limits of the purpose limitation. Although some commercial entities state that data is shared with list brokers, this says nothing about the purpose of those data brokers. Data subjects have no way to know how their data is further processed by those parties. Furthermore, data brokers have given us little information about with whom the data is further shared.

Art 33 Wbp provides that information should be provided to the subject on the data processing. A limitation to this article is that they don't have to do this, when this is impossible or an unreasonable burden. In that case, they should record the origin of the data (art 34-4 Wbp). In the code it says data brokers should notify data subjects about the processing before the first processing (when they receive the data from the data subject), when they get the data in another way, on the moment of processing or when the data is first handed to other parties (unless they suspect the subject knows this already).They don't have to notify the subject anymore when other interests prevail.[77] This research suggests that commercial data sources aren't specific about with whom they share their data. Data brokers also give limited information about where they get their data and with whom they share the data.

Data brokers and sensitive information

The research shows that sensitive data has been processed. 'NAW plus' gets sensitive information relating to religion from publishers, and a holiday park. They also get email addresses from a religious Youth community forum and dating website 'Christianmatch.' The same could be said about 'WIJ special media,' which processes information about health (pregnant women).

Art 16 Wbp says that sensitive information may not be processed. An exception to this rule is that there has been explicit consent of the data subject (art 23 Wbp). The code of conduct also specifies that special categories of data (like religion, race, political and sexual orientation) can't be processed unless there has been explicit consent by the data subject, the data subject has publicly disclosed the data, or when they are necessary for identification (in terms of race) or in other cases specialized by the law.[78]

The research would suggest that in the case of religious data it is unlikely people have explicitly consented to the reuse of their data. As we have previously mentioned, it is unlikely that people consent to the further processing of sensitive data, as there is no direct contact between the data broker and the subject.

Data brokers and public sources

The research shows that data brokers also get their data from public sources, like the Kamer van Koophandel (Chamber of Commerce), CBS (Central Bureau of Statistics) and the Kadaster (Cadastre). There are public registries that require publicity by legal obligation. With regard to personal data from public sources, the Dutch legislator has decided that the owner of such registries can provide personal data without inquiring the motives of the applicant. However, the purpose of the registry has to be taken

into account. Article 13 of the Wbp does require the registries to secure the provision of data against unjustified processing, which can be when data is used for another purpose.

The Dutch data protection authority has said that data brokers may process information from public sources, unless there is a specific legislative obligation to keep those data secret.[79] Presumably, this processing happens on the basis of a legitimate interest. However, data subjects should still retain the opportunity to resist this processing (on the basis of article 40 Wbp).

Data brokers and profiling

The law says that people can't be subjected to decisions that have significant or legal effects, if those decisions are made on the basis of automated processing only, and aim to get an image of specific aspects of someone's personality.[80] This article also specifically aims to regulate profiling, although there is no case law yet to clarify it. It hasn't been used in practice in the Netherlands. This provision applies in particular to credit rating.

Other provisions relevant to profiling relate to discrimination. In research and judgments by the 'College voor de rechten van de mens' (The Dutch human rights commission, formally called the commission for equal treatment, the Dutch authority that handles complaints of discrimination) it is made clear how important the prevention of unfair indirect discrimination is in the provisions of goods and services. Providers of goods and services may (either on purpose or not) find ways to offer goods and services that disproportionately affect people with certain characteristics. For example, mortgage providers can't just use location data as a relevant variable for credit provision, if that disproportionately affects people with a certain background (born outside of the Netherlands in this case).[81]

This legislation unfortunately does not address "the concealed forms of discrimination that credit scoring may generate, especially indirectly by not using such sensitive data." Some authors argue that "accordingly, it is impossible for a consumer to demonstrate a cause-and-effect relationship on an individual basis, between the data used, the data-mining technique employed, and the discriminatory decision affecting an entire group."[82]

Data brokers and data subject rights

The Dutch data protection law also grants a number of user rights. People have the right to access their data (art 35 Wbp) and to know the logic behind data processing (art 35-4 Wbp). They also have the right to correct incorrect data and delete data that is no longer appropriate for the purposes of collection (art 36 Wbp).

This means that data subjects have the right to access their data and have the right to get an explanation about their credit score. This includes (taking company secrets into consideration) the logic of the system, what certain numbers and symbols mean, and a description of how the score came to be. Controllers can ask for a small financial compensation in exchange for this information.[83]

The research shows that out of the 25 companies addresses, four did not react, and four others request additional information. Two companies claim not to own databases but to use the databases of others. However, that still entails processing of data and requires a legal ground. Twelve companies claim not to have any information on our researcher, but later one of them has a profile on her, coupled to her living address. Three brokers claim to have information on

her. Experian knows where she lives, that she is a woman, and that she has a land line. Sandd knows her home and living environment. Graydon has a copy of her professional information,and her credit score. Sandd and Graydon write that they have her information from public sources without specifying which ones. Regarding her credit score, Experian refuses to share the score. That is illegal, as it is personal data. Graydon shares the score, but does not mention why she has that score as it is a company secret. Graydon will have to share the reasons for the score if there are decisions taken on the basis of it.

Article 40 Wbp allows people to resist data processing, It reveals that if data is processed on the basis of the balancing provision, data subjects may at all times resist the processing and it states that they should be notified of this possibility when the broker turns to them for marketing purposes, when this is based on the balancing provision.[84] This does not apply to data from public sources – 40-4 Wbp. Article 41 Wbp however, allows people to resist direct marketing. This means that data brokers should offer people the opportunity to resist processing of their data when it is used for direct marketing.

Data brokers and the general data protection regulation

The European Union has finished negotiating the General Data protection regulation. This new European law should replace the European and Dutch national legislation in the field of data protection. The regulation has stricter rules for consent and transparency. It also has a specific provision on profiling. For the next part, we have relied on the final text as agreed in the negotiations. This same text was sent to the Dutch Parliament, but may be subject to minor changes (like the article numbering). This publication is dated on the 7th of january 2016.[85]

The regulation introduces some notable changes. First of all, The regulation will also allow data protection authorities more capabilities to fine companies in breach of the regulation.

Secondly, the conditions for consent have been strengthened. The burden of proof is on the company to prove that consent was given by the data subject (7-1). The article also lays down that if the request for consent is part of a written declaration that also includes other matters, "the request for consent must be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language." (7-2). 7-3 says that people can withdraw their consent at any time, which should be as easy as obtaining consent. Consent has to be freely given, and to determine this it shall be taken into account if the provision of a service is made conditional "on the consent to the processing of data that is not necessary for the performance of this contract. " (7-4).

Recital 25 says that "Consent should be given by a clear affirmative action establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to personal data relating to him or her being processed, such as by a written, including electronic, or oral statement." There cannot be an imbalance between the data subject and the company processing the data (recital 34). That same recital says that "consent is presumed not to be freely given, if it does not allow separate consent to be given to different data processing operations despite it is appropriate in the individual case, or if the performance of a contract, including the provision of a service is made dependent on the consent despite this is not necessary for such performance."

This would mean that commercial parties looking to share information with third parties on the basis of consent would have more obligations to offer a choice, separate from the sale of their goods or services. They would also have to offer people the opportunity to withdraw their consent. For data brokers we have already established that consent is not a valid ground for processing.

Where information is obtained on the basis of a legitimate interest, provisions in the regulation have watered down the protection of individuals. Recital 38 says that "The interests and fundamental rights of the data subject could in particular override the interest of the data controller where personal data are processed in circumstances where data subjects do not reasonably expect further processing." However, the same recital says that "The processing of personal data strictly necessary for the purposes of preventing fraud also constitutes a legitimate interest of the data controller concerned. But because data brokers don't aim to prevent fraud themselves, but sell information, this interest would not apply. "The processing of personal data for direct marketing purposes may be regarded as carried out for a legitimate interest."However, as the data brokers don't themselves perform direct marketing, this concerns the legitimate interest of the customer of the data brokers, and then, data subjects would have the ability to resist this direct marketing on the basis of the new article 19. Unfortunately the regulation says in article 6(3A) that data may be processed for ¨another purpose than the one for which the data have been collected."

Pseudonymous data, which means personal data that have been made more difficult to identify, have a lighter regime, which could have consequences for user rights, transparency and the grounds for processing.

The regulation also creates a new right to data portability (art 18), which allows people to receive all the personal data concerning them and allows them the right to transmit those data to another company.

The regulation contains provisions that allow people to object to processing based on a legitimate interest, including profiling on that basis (art 19). It creates the requirement that people should be allowed this opportunity clearly and at the first communication with them. It also contains a specific article on automated individual decision making, including profiling (article 20). This article again allows people the right not to be subject to decisions based on automated processing. Unfortunately, it only applies when decisions are based "solely" on automated processing, if those decisions produce legal effects or significantly affect him or her. The European parliament had wanted to include "or predominantly", but this did not make it. Member states are allowed to create new rules to allow certain kinds of profiling (b) if there are also suitable measures to safeguard people's rights. The article also says that there should be safeguards like "at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision."

06. CONCLUSIONS

This research shows that it remains questionable to what extent the conduct of data brokers is legal and that many risks remain to fundamental values in our society that are not mitigated right now.

1. Data brokers have no ground on which to collect so much data

People worry about control, but often feel overwhelmed by the perceived lack of choice they have. The research reveals that in the case of data

brokers, users have little control over what happens to their data. Data is processed in a take it or leave it way, and once it is processed there is little way to prevent further processing.

The research shows that none of the grounds for processing legitimize the current practices of data brokers. Consent can't be a ground as there is no direct contact between the data broker and the data subject. The processing is also not necessary for the performance of a contract. The balancing provision is the remaining ground, but it is weak, as we argue that the privacy interests of data subjects prevail.

The research also shows that sensitive data are processed, without asking for explicit consent. This means data brokers are not just in breach of the law but also of their own code of conduct.

2. The purpose limitation is not respected

The purpose limitation is the cornerstone of data protection. Recent societal unrest surrounding the ING bank in the Netherlands and the conduct of TomTom show that denying this right is not accepted by the general public. The purpose limitation protects the most important values of data protection, like the ability to confide.

As this research reveals the increasing amount of data collected and re-used by different companies and how easy re-use and collection is, the purpose limitation is an increasingly important safeguard as data slips away from user control.

It remains questionable to what extent this is respected by data brokers. Although some commercial entities state that data is shared with list brokers, this says nothing about the purpose of those data brokers. Data subjects have no way to know how their data is further processed by those parties. Furthermore, data brokers have given us little information about with whom the data is further shared.

3. There is little transparency about data brokers and data traffic

The current practices of data brokers are not transparent. Notices provided by parties that share data with brokers are vague and unspecific.

Data brokers themselves are also not transparent about how they use their data, where they get their data and with whom they share their data. Data brokers should be more open about the type and amount of data they use, where they get this information and how they create profiles.

4. There is no way for people to object

The research shows that once data is shared, it is further shared with third parties. There should be limits to this chain and opportunities for people to object to processing. Onward sharing makes it increasingly difficult for users to exercise control over their data and to prevent further processing.

We should also critically evaluate the reuse of public data and allow people the opportunity to object to processing.

5. Data subject rights are insufficiently respected

The research reveals that some data brokers don't respond to access requests and that people don't get more information about their profile, to which they are legally entitled.

People should be meaningfully informed about profiling. They should also be able to tell what that profile is and be able to ask for human intervention and due process when decisions are made that concern them on the basis of profiles.

6. There should be more enforcement

The practices of data brokers and profiling should be carefully monitored by the competent authorities. The new Dutch law that has come into effect on January 1st of 2016 and the European Data Protection Regulation both promise an enforcer 'with teeth'.

This is promising as the research reveals some shadowy practices. After the laws enter into force, the data authority should closely watch the behavior of these companies. It is also important that active monitoring and enforcement happens by other organizations, mandated by groups of people, or through class action lawsuits.

We also recommend more proactive research and activity by the anti-discrimination authority. Research reveals that some profiles have the ability to indirectly discriminate against certain groups of people. The problem is that this discrimination is difficult to spot, in particular when companies don't use sensitive data. This requires new monitoring tools for the anti-discrimination authorities.

07. METHODOLOGY

This is a case study on the market for commercial profiling in the Netherlands. We focused our research on marketing for commercial marketing purposes and for credit rating.

This case study has the following research questions:

1. What organizations offer profiling tools in the Netherlands? What are their interests, values, and goals? How are these organizations connected?
2. Where do profiling organizations get their data and what kind of data do they get?
3. What ethical risks are connected to these profiling technologies?
4. To what extent can these risks come true: Using similar technologies, what kind of information (for purposes of marketing, credit analysis, insurance, and employment) can we derive from open sources, marketing data, and social media vaults?
5. What legal frameworks currently govern these profiling technologies and applications, and prevent these risks from materializing?
6. What policy recommendations can we make on the basis of this analysis?

Case study

This research is exploratory, empirical and journalistic: it will navigate the relatively unknown area of commercial profiling in the Netherlands.

This case study combines interviews, experiments, and ethical and legal analyses using the input of experts.

The exploratory part first required a literature review, and legal and ethical analyses. The study is empirical in the sense that it gathers new information through interviews and by doing empirical tests online. The 'Privacy Insights Machine' of Bits of Freedom was used to request access to user data. This approach combines insights from both empirical study and interviews. Bits of Freedom and De Correspondent have a large network of academics, legal specialists and hackers that helped us with technological analyses and offered substantive feedback. Bits of Freedom regularly performs empirical studies on the status of digital rights online. For example, they checked whether companies accurately responded to notice and takedown requests in the field of copyright law or they had volunteers look into price discrimination online. De Correspondent did research

on trackers on popular websites and on the hidden ecosystem of smartphone apps in the Netherlands.

To answer our first question on profilers, our research team used its network and internet research to identify the most important profiling technology vendors in the fields of marketing and credit rating. To identify their values, interests and goals, we conducted a series of interviews with them. In these interviews we asked them about the goals of their organization, the sources of data they have, the types of technologies they use, and the precautions they take to guarantee the quality of their data.

This part was carried out by one of the researchers of the research team. Our researcher initially reached out to the rest of the team and to supporters of De Correspondent to explore and map the market for profilers.

Our researcher interviewed various data brokers, experts from the field and academics. She also send out data requests asking 25 data brokers if they had any information about her. The 25 companies were all registered in the Netherlands, and are: Experian, Graydon, Focum, 4Orange, Autoriteit Consument & Markt, Bureau Krediet Registratie, Cardatapool, Cardec, CDDN, Cendris, Company.info, Creditsafe, DAT.Mobility, Dun & Bradstreet, EDM, EDR, Facebook, Geodan, Geoscape, Google, Mastercard, Mint Marketing, Omniprofiles, PostNL, Rabobank, Sandd & T-Mobile.

These 25 were selected because they were either well known for dealing with consumer data, were companies that delivered services to our researcher (like her bank and credit card firm), or were suggested by readers of De Correspondent.

To answer the second question on where they get their data from, our researcher conducted interviews with data brokers). Six companies were selected to vist and interview. A potential difficulty the team faced was that profilers were not willing to share all information.

To identify the ethical risks connected to these profiling technologies, we used a literature study, and organized an expert session. The risks are outlined in the background section.

For the fourth question on the materialization of ethical risks, we gathered a group of people in an experimental session, and made analyses on the basis of their data. For this part of the research, the research team created 'Heel Holland Transparant' ('All of Holland Transparent'), a fake data broker that would enable us to see how much data the team could collect and what insights the data could generate using methods used by profilers themselves.

The research team collected (mostly) publicly available information from a group of famous and non-famous Dutch people and linked it with social media information.

Public sources were the central bureau for statistics, the Cadastre, the chamber of commerce, the police, the insolvency register, the public register for energy labels of housing, the provincial risk map, and social media companies like LinkedIn, Twitter and Coosto. Most of these data were free. Some social media data and data of the chamber of commerce required a modest fee.

The research team then approached experts to help them generate similar data analysis techniques. They made a risk score by seeing to what extent someone corresponds to an ideal type of the riskfree citizen: in this case: well educated, living in a postal code with few diseases or burglaries, and with a positive attitude on social media. The team allocated a risk score

between 0 and 100 and gave extra points for transparency.

For the fifth question, we did a legal analysis with the help of legal experts in our network. We invited numerous experts from the field for a session on profiling and data brokers in the Netherlands to speak under the Chatham House Rule. They came from an NGO, from academia and from the business world. During this session we discussed numerous propositions related to our research. We focused on transparency and control. The insights derived from this session are spread around the different chapters and have helped answering the sixth question about policy recommendations.

08. APPENDICES

As a result of this research two long form articles were published in De Correspondent. "Zo houden datahandelaren ons in de gaten"[86] by Maaike Goslinga as appendix 'a' and "Heel Holland Transparant: Zo bepalen bedrijven en overheden of je een risicoburger bent"[87] by Maurits Martijn and Dimitri Tokmetzis as appendix 'b'.

09. ACKNOWLEDGMENTS

The following people worked in some shape or form on this project:

Dimitri Tokmetzis, Floris Kreiken, Hans de Zwart, Maaike Goslinga, Maurits Martijn, Rico Disco, Sanne Blauw and Yuri Veerman.

We'd like to thank Media Democracy Fund, Ford Foundation and Open Society Foundations for their financial support (grant number: NVF MDF BOF GA#06092015); and all the interviewed experts for their insights, advice and commentary on the text.

010. CERTIFICATION

All activities by Bits of Freedom were and are consistent under the Internal Revenue Code Sections 501(c)(3) and 509(a)(1), (2) or (3). If any lobbying was conducted by Bits of Freedom (whether or not discussed in this report), Bits of Freedom complied with the applicable limits of Internal Revenue Code Sections 501(c)(3) and/or 501(h) and 4911. Bits of Freedom warrants that it is in full compliance with its Grant Agreement with the New Venture Fund, dated June 9, 2015, and that, if the grant was subject to any restrictions, all such restrictions were observed.

EINDNOTEN

1. Federico Ferretti, "Legal Framework of Consumer Credit Bureaus and Credit Scoring in the European Union: Pitfalls and Challenges-Overindebtedness, Responsible Lending, Market Integration, and Fundamental Rights, The." *Suffolk UL Rev.* 46 (2013): 791. P.810
2. TNO, *Privacybeleving op het internet* (2015), http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2015/02/01/privacybeleving-op-het-internet-in-nederland.html
3. Federal Trade Commission, "FTC recommends congress require data broker industry be more transparent and give users greater control over information", *FTC Website* (2014), https://www.ftc.gov/news-events/press-releases/2014/05/ftc-recommends-congress-require-data-broker-industry-be-more
4. "Zest finance company website," http://www.zestfinance.com/
5. "Cornerstone demand company website," http://www.cornerstoneondemand.com/evolv
6. Leslie Scism and Mark Maremont, "Insurers test data profiles to identify risky clients," *Wall Street Journal* (Nov. 19, 2010), http://www.wsj.com/articles/SB10001424052748704648604575620750998072986
7. Federal Trade Commission, "FTC recommends congress require data broker industry be more transparent and give users greater control over information", *FTC Website* (2014), https://www.ftc.gov/news-events/press-releases/2014/05/ftc-recommends-congress-require-data-broker-industry-be-more
8. Federal Trade Commission, "FTC recommends congress require data broker industry be more transparent and give users greater control over information", *FTC Website* (2014), https://www.ftc.gov/news-events/press-releases/2014/05/ftc-recommends-congress-require-data-broker-industry-be-more, P.III
9. Federal Trade Commission, "FTC recommends congress require data broker industry be more transparent and give users greater control over information", *FTC Website* (2014), https://www.ftc.gov/news-events/press-releases/2014/05/ftc-recommends-congress-require-data-broker-industry-be-more, P.III
10. Federal Trade Commission, "FTC recommends congress require data broker industry be more transparent and give users greater control over information", *FTC Website* (2014), https://www.ftc.gov/news-events/press-releases/2014/05/ftc-recommends-congress-require-data-broker-industry-be-more, P.IV
11. Electronic Frontier Foundation, "EFF's comments to the White House Office of Science and Technology Policy on Big Data," through *EFF Website* (RFI OSTP-2014-0003-0001), https://www.eff.org/nl/document/effs-comments-white-house-big-data
12. EPIC, "Big Data and the Future of Privacy," https://epic.org/privacy/big-data/
13. Solon Barocas and Helen Nissenbaum, "Big Data's End Run Around Procedural Privacy Protections," *Communications of the ACM*, Vol. 57 No. 11, P. 31-33.
14. Mireille Hildebrandt, "Who is Profiling Who? Invisible Visibility," in Gutwirth S., Poullet, Y., De Hert, P., De Terwangne C., Nouwt S. (Eds), *Reinventing Data Protection?* (2009 Dordrecht, Springer), P.. 239-252.
15. Fuster G., Gutwirth S., Erika E. (June 2010), "Profiling in the European Union: A high- risk practice," *INEX Policy Brief*, no. 10. (June 2010), P.2.
16. Mireille Hildebrandt, "Slaves of Big Data, or are we?" (October 2013). P.5-6.
17. Koen de Bock and Dirk van den Poel, "Predicting website audience demographics for web advertising targeting using multi-website clickstream data" (2010) *Fundamenta informaticae*. 98(1). p.49-70, https://biblio.ugent.be/record/967442
18. Michal Kosinski, "Private traits and attributes are predictable from digital records of human behavior" (2013), *PNAS* vol 110 no. 15, 5802-5808, http://www.pnas.org/content/110/15/5802
19. Mireille Hildebrandt, "Slaves of Big Data, or are we?" (October 2013). P.11-12.
20. Michael Persson, "Burger wordt straks doorgelicht zoals profiel van crimineel wordt opgesteld," *de Volkskrant* (October 1, 2014), http://www.volkskrant.nl/politiek/burger-wordt-straks-doorgelicht-zoals-profiel-van-crimineel-wordt-opgesteld~a3759563/
21. Dimitri Tokmetzis, "Dit gebeurt er allemaal achter de schermen als je naar de VS vliegt," *De Correspondent* (May 12, 2015), https://decorrespondent.nl/2675/Dit-gebeurt-er-allemaal-achter-de-schermen-als-je-naar-de-VS-vliegt/82272300-ee6f12b0

EINDNOTEN

22. Maurits Martijn, "Vergeet de politiestaat. Welkom in de belastingstaat," *De Correspondent* (September 30, 2014), https://decorrespondent.nl/1766/Vergeet-de-politiestaat-Welkom-in-de-belastingstaat/54315096-f35e98af
23. Laura Klompenhouwer, "Achmea wil lagere premie bieden als klanten privégegevens delen," *NRC* (Oktober 1, 2015), http://www.nrc.nl/nieuws/2015/10/01/achmea-wil-lagere-premie-bieden-als-klanten-privegegevens-delen
24. "Premies verzekeringen verschillen tot op huisnummer," *De Consumentenbond* (August 26, 2015), https://www.consumentenbond.nl/actueel/nieuws/2015/verzekeringspremies-verschillen-tot-op-huisnummer/
25. Michael Persson et al, "China kent elke burger score toe - ook voor internetgedrag," *de Volkskrant* (April 25, 2015), http://www.volkskrant.nl/buitenland/china-kent-elke-burger-score-toe-ook-voor-internetgedrag~a3980289/
26. Federico Ferretti, at (1), P. 796-797
27. Federico Ferretti, at (1), P. 810
28. Federico Ferretti, at (1), P. 810
29. Mireille Hildebrandt, "Slaves of Big Data, or are we?" (October 2013). P.13.
30. Darell Huff, *Lying with statistics*, 1954.
31. "Spurious Correlations," http://tylervigen.com/
32. John Battelle, "The Database of Intentions," *Searchblog* (November 13, 2003), see: http://battellemedia.com/archives/2003/11/the_database_of_intentions.php
33. Trina Magi, "Fourteen Reasons Privacy Matters: A Multidisciplinary Review of Scholarly Literature." *The Library* 81.2 (2011).
34. Janneke Sloetjes, "Drie vragen over big data privacy en de ING," *Bits of Freedom* (March 10, 2014), https://www.bof.nl/2014/03/10/drie-vragen-over-big-data-privacy-en-de-ing/
35. Federico Ferretti, at (1), P. 810-811.
36. Solon Barocas and Andrew Selbst, "Big Data's Disparate Impact", 104 *Calif. L. Rev* (forthcoming 2016). p.3
37. Maaike Goslinga, "Zo houden datahandelaren ons in de gaten, maar wie controleert hen", *De Correspondent* (Oktober 13, 2015), https://decorrespondent.nl/3472/Zo-houden-datahandelaren-ons-in-de-gaten-maar-wie-controleert-hen-/318414571552-e7b47e38
38. "NAW Plus company website," http://www.nawplus.nl/
39. "WIJ special media company website," http://www.wijspecialmedia.nl/
40. "Elite Miljonairs company website," http://www.elite-miljonairs.nl/
41. "VIP Leads company website," http://www.vip-leads.nl/
42. "Experian company website," http://www.experian.nl/
43. "Sandd company website," http://www.sandd.nl/
44. "Graydon company website," https://www.graydon.nl/
45. "Onterecht wanbetaler door foute postcode", *Kassa* (September 21, 2013), http://kassa.vara.nl/tv/afspeelpagina/fragment/onterecht-wanbetaler-door-foute-postcode/speel/1/
46. The website can be found at https://www.heelhollandtransparant.nl However, all the data has been taken offline.
47. Federico Ferretti, at (1), P. 807
48. Federico Ferretti, at (1), P. 808
49. Federico Ferretti, at (1), P. 809
50. Federico Ferretti, at (1), P. 809
51. Federico Ferretti, at (1), P. 809
52. Federico Ferretti, at (1), P. 811..
53. Art 1 (1 c) Algemen Wet Gelijke Behandeling (AWGB)
54. Wet Gelijke Behandeling Chronisch Zieken en Gehandicapten.
55. Art 7 (1) AWGB
56. Art 7 (3c) AWGB
57. Federico Ferretti, at (1).

EINDNOTEN

58. See for example: Wet van 22 maart 2007, Regels omtrent een basisregister van ondernemingen en rechtspersonen (Handelsregisterwet 2007), Staatsblad 153, 1 mei 2007.
59. Art 6 Wet bescherming persoonsgegevens (Wbp)
60. Art 7 Wbp
61. Art 8 Wbp
62. Art 36 Wbp
63. Art 11 Wbp
64. Federico Ferretti, at (1), P. 812
65. Art 35 (1) Wbp
66. Art 35(2) Wbp
67. Art 35 (4) Wbp
68. Art 5 of the code of conduct on data brokers.
69. Art 8 of the code of conduct
70. Art 8 Wbp
71. Art 3 of the code of conduct
72. Federico Ferretti, at (1), P. 815
73. Federico Ferretti, at (1), P. 823-824
74. Article 29 Data Protection Working Party, "Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC", WP 217 (9 April 2014), see: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf
75. Art 4 of the code of conduct
76. Art 9 and 10 of the code of conduct
77. Art 11 of the code of conduct
78. Art 6 of the code of conduct
79. See the decision of the Dutch data protection authority at CBP 7 march 2003, z2002-0499, Uitsprakenbundel Wet Bescherming Persoonsgegevens 2009, 8.26.
80. Art 42 Wbp
81. "Onderzoek uit eigen beweging naar risicoselectie op grond van postcode en verblijfsstatus," *College voor de Rechten van de Mens* (August 3, 2006), http://www.mensenrechten.nl/publicaties/detail/9993
82. Federico Ferretti, at (1), P. 814.
83. Art 12 of the code of conduct
84. Art 3.3 of the code of conduct
85. See: https://zoek.officielebekendmakingen.nl/blg-657102
86. Maaike Goslinga, "Zo houden datahandelaren ons in de gaten, maar wie controleert hen", *De Correspondent* (Oktober 13, 2015), https://decorrespondent.nl/3472/Zo-houden-datahandelaren-ons-in-de-gaten-maar-wie-controleert-hen-/318414571552-e7b47e38
87. Maurits Martijn, "Heel Holland Transparant: Zo bepalen bedrijven en overheden of je een risicoburger bent," *De Correspondent* (Oktober 12, 2015), https://decorrespondent.nl/3478/Heel-Holland-Transparant-Zo-bepalen-bedrijven-en-overheden-of-je-een-risicoburger-bent/191359285238-05385ad5

de Correspondent

Maaike Goslinga is een talentvolle journaliste die zich vastbijt in verhalen over data en privacy. Dit onderzoek naar datahandel leverde een erg goed verhaal op.

Dimitri Tokmetzis
Correspondent Hacken

13.10.2015 • *Leestijd 12 - 17 minuten*

Zonder dat je het doorhebt, worden jouw persoonlijke data elke dag verhandeld. Een wereld waar dagelijks miljoenen in omgaan. Toch weten we in Nederland weinig van deze handel af. Wat gebeurt er precies met onze gegevens? Een inzicht in een wereld waarin je constant wordt geobserveerd, geregistreerd en geïnterpreteerd.

Zo houden datahandelaren ons in de gaten (maar wie controleert hen?)

Gastcorrespondent Datastromen & Privacy

Maaike **GOSLINGA**

Illustratie: Maus Bullhorst (voor De Correspondent)

Ik ben 45 jaar oud en dol op beleggen. Mijn buren zijn hier geboren, net als ik. Ik rijd een Volvo, maar ik heb aanschafplannen voor een Hummer. Dat is handig voor mijn gezin met drie kinderen.

O, en ik lees de *Linda*.

Dit is wat datahandelaar Experian over mij weet. In 2015 zette de Nederlandse tak van dat Amerikaanse bedrijf 16,2 miljoen euro om met het verzamelen, analyseren en verhandelen van persoonsgegevens van burgers.

Experian staat hier niet alleen in. In Nederland zijn zo'n 180 datahandelaren actief die de gedragingen van burgers constant in de gaten houden.

De gegevensverzameling van deze bedrijven voltrekt zich volledig onder de radar. Denk er maar eens over na: heb jij ooit data over jezelf aan 4Orange, Cendris of Experian gegeven? Grote kans van niet. Toch hebben ze die in handen en verdienen ze er flink wat geld aan.

Maar waarom is die wereld met zoveel schimmigheid omgeven? Welke gegevens hebben bedrijven in handen? Waar halen ze die vandaan en wat doen ze daarmee?

Dat wilde ik weten. Eerst vroeg ik jullie mij te helpen met dit onderzoek. Het leverde veel bruikbare tips op. Vervolgens sprak ik met verscheidene datahandelaren, experts uit het bedrijfsleven en academici die mij een inzicht boden in deze wereld. Ook was ik benieuwd naar mijn eigen dataspoor, dus vroeg ik aan een selectie bedrijven wat zij over mij weten.

Wat blijkt: datahandelaren analyseren je constant. En daar kun je maar weinig tegen doen.

De wereld van de datahandel

Datahandelaren verzamelen, analyseren en verhandelen jouw gegevens. Die data kunnen ze voor verschillende doelen inzetten, op persoonlijk en doelgroepniveau.

Zo helpen datahandelaren marketeers om doelgroepen te selecteren. Een bedrijf dat een duur espressoapparaat wil verkopen, adverteert niet in de wijk Woensel-Noord in Eindhoven maar in een hippe, rijke buurt als het Amsterdamse Oud-West. De kans op een 'conversie,' het omzetten van een advertentie in een bestelling, is daar namelijk veel hoger.

Omdat datahandelaren over veel gegevens van burgers en woonwijken beschikken, weten zij precies waar de juiste doelgroep

Een bedrijf dat een duur espressoapparaat wil verkopen, adverteert niet in de wijk Woensel-Noord

woonwijken beschikken, weten zij precies waar de juiste doelgroep zit voor een bepaald product. Zo biedt handelaar 4Orange, die gegevens van ‘alle Nederlandse consumenten’ bezit, een zogenoemde ‘lifestyle scan’ aan. Hiermee krijgen bedrijven een ‘gedetailleerd inzicht in de karakteristieken van de doelgroep gegeven aan de hand van een groot aantal kenmerken op het gebied van socio-demografie, koopgedrag, media en lifestyle.’

Persoonsgegevens kunnen ook voor risicobeheer gebruikt worden. Handelsinformatiebureaus gebruiken data over jou (zoals je betaalgeschiedenis, maar ook je postcode) om een kredietscore van je op te stellen met een bijbehorend advies. Zo’n score geeft aan hoe kredietwaardig je bent en of je dus een rekening of lening zult (terug)betalen. Handig voor banken en webshops.

Datahandelaren verzamelen informatie over jou (zonder dat je dat doorhebt)

Op 13 augustus 2015 doe ik een grote stapel enveloppen op de post. In die enveloppen zitten inzageverzoeken aan 25 verschillende bedrijven. Met deze verzoeken hoop ik bij bedrijven en datahandelaren te ontfutselen of zij persoonsgegevens van mij hebben en, zo ja, welke. Ook wil ik weten wat ze daarmee doen.

Terwijl ik op reacties wacht, struin ik het internet af op zoek naar datahandelaren en hun waar.

Al snel stuit ik op datahandelaar N.A.W.plus. Het bedrijf heeft complete lijsten beschikbaar van predikanten, pastores en ‘actieve christenen,' verkregen van een uitgeverij die zich bezighoudt met ‘materiaal voor dagelijkse bezinning.' N.A.W.plus verkoopt ook namen en adresgegevens van de bezoekers van een christelijk vakantiepark op de Veluwe. Het beschikt over e-mailadressen van bezoekers van jongerencommunity Refoweb én datingwebsite Christianmatch. Stel dat je een nieuw christelijk magazine start, dan kun je deze data goed gebruiken.

Het is niet de enige bijzondere lijst die ik tegenkom.

Het bedrijf WIJ Special Media, onderdeel van Prénatal, claimt ‘nagenoeg alle’ adressen van zwangere vrouwen in Nederland te hebben. Het bedrijf biedt gratis zwangerschapspakketten aan in ruil voor persoonlijke informatie. Het bedrijf vraagt niet alleen om adresgegevens, maar ook naar de uitgerekende datum en zelfs de geboortedatum, de naam en het geslacht van het kind. Deze data kunnen interessant zijn voor bedrijven die baby- en kinderspullen willen verkopen. Zij weten precies wanneer een vrouw de juiste aanbieding moet ontvangen.

Er zijn ook datahandelaren, zoals Elite-Miljonairs en VIP-Leads, die contactgegevens van Nederlandse miljonairs, welgestelden en bekende Nederlanders verhandelen. Welvaart drijft de prijs op: zo betaal je bij Elite-Miljonairs voor bekende Nederlanders 70 eurocent per adres en voor een multimiljonair 1,25 euro.

Ieder bedrijf kan zo’n lijst met namen en adressen van burgers huren of kopen. Bij N.A.W.plus betaal je rond de 7.000 euro voor ongeveer 20.000 gegevens van lezers van het *Reformatorisch Dagblad*. Elite-Miljonairs verkoopt adressen van 1.500 rijke Nederlanders voor 720 euro.

Ik vraag me af: in welke categorie val ík eigenlijk?

De inzageverzoeken die ik terugkrijg helpen mij niet verder. Maar drie datahandelaren zeggen gegevens van mij te hebben. Experian weet waar ik woon, dat ik een vrouw ben en dat ik een vastetelefoonaansluiting heb. Sandd kan mij alles vertellen over mijn huis en woonomgeving; van Graydon krijg ik een kopietje van mijn bedrijfsgegevens mee, inclusief het aantal werknemers dat ik in dienst heb en mijn kredietscore.

Het is niet veel. Toch heb ik die data nooit bewust afgegeven.

Het is me bovendien niet helemaal duidelijk waar ze hun informatie precies vandaan hebben. In hun brieven schrijven Sandd en Graydon dat ze mijn informatie uit openbare bronnen hebben, zonder verder uit te leggen welke dat precies zijn. Een medewerkster van Experian laat mij schriftelijk weten dat het bedrijf rekening houdt ‘met een schat aan informatie van bronnen met registratiedata – zoals uitgebreide en gedetailleerde vastgoedgegevens en diverse personenbestanden.’

Hoe een profiel gemaakt wordt

Het blijft vaag. Dus besluit ik bij een zestal bedrijven langs te gaan. Wellicht kunnen ze mij meer vertellen over hun vergaringstactieken.

De woorden ‘openbare bronnen’ komen in elk gesprek terug. Handelaren raadplegen bijvoorbeeld de Kamer van Koophandel (KvK), het Centraal Bureau voor de Statistiek (CBS) en Kadaster, veelal voor informatie over wijken en straten.

Hoe algemeen en onpersoonlijk deze informatie ook lijkt, in de praktijk is het dat niet. Want nee, ik ben geen 45-jarige Volvo-rijder. De *Linda* lees ik hooguit bij de kapper.

Nee, ik ben geen 45-jarige Volvo-rijder. De *Linda* lees ik hooguit bij de kapper

Dit zijn gewoon kenmerken van de wijk waarin ik woon. Vergaard door databases van het CBS en Kadaster aan elkaar te koppelen. Experian deelde me op basis daarvan in bij het groepsprofiel ‘Gouden Rand.' Anders dan mensen met de profielen ‘Vergrijsde Eenvoud’ en ‘Sociale Huurders’ ben ik, als je naar mijn wijk kijkt, waarschijnlijk rijk en een harde werker.

Ook vind ik een groepsprofiel van Experian genaamd 'Minder Geslaagden.' Dit zijn arme mensen die 'veel thuis zitten' en daardoor 'kunnen nadenken over hoe het had kunnen zijn.' Ze wonen in 'vervallen wijken' met een 'ratjetoe' aan huizen en mensen. In hun huis hangen 'gordijnen voor de ramen' en staan 'vetplanten' op de vensterbank. Het profiel is inmiddels uit het assortiment gehaald. Destijds is het onder andere door brandweer Amsterdam-Amstelland gebruikt.

Uit gesprekken met handelaren blijkt dat zij dit soort groepsprofielen constant updaten – en niet alleen met actuele informatie van het CBS en andere openbare bronnen. Zo sturen 4Orange en EDM ‘onderzoeksenquêtes’ uit waarin mensen worden bevraagd over hun interesses, leefstijl en woonomgeving. Dit is persoonlijke informatie die aan individuen wordt gekoppeld. Vul je een keer zo’n enquête in, dan zullen handelaren en hun klanten jouw mening over je buurt en favoriete shampoomerk jarenlang bewaren.

Frank de Beun van EDM legt uit hoe dat gaat: ‘Als we enquêtes rondsturen, krijgen we bijvoorbeeld gegevens terug van burgers die hockeyen. We zien dat die gemiddeld een hoger inkomen hebben, tussen de 30 en 40 jaar zijn en kinderen hebben. Deze resultaten passen we toe op heel Nederland. Het is een kans dat jij een bepaalde hobby of interesse hebt. Misschien klopt het niet voor elk huishouden, maar die foutmarge heb je altijd.’

En het aantal openbare informatiebronnen neemt alleen maar toe. Cookies en andere online data vormen steeds vaker onderdeel van een profiel. ‘Als jij je hele ziel en zaligheid op LinkedIn zet, moet je je er altijd bewust van zijn dat iedereen daarnaar kan kijken. Sociale media en andere online bronnen, zoals websites van bedrijven, zijn publiek. Handelsinformatiebureaus kunnen die informatie dus ook inzien,’ aldus Gertjan Kaart, voorzitter van de Nederlandse Vereniging van Handelsinformatiebureaus (NVH).

Ook commerciële partijen blijken een goudmijn. Koop jij een paar schoenen in een webshop, dan kun je ervan uitgaan dat sommige bedrijven jouw gegevens delen met ‘geselecteerde partners.’ Op het moment dat jij besluit de algemene voorwaarden aan te vinken, ga je hier namelijk mee akkoord.

Handelaren weten veel over ons, maar wij weinig over hen

Er is zo een onbekend aantal commerciële partijen dat klantgegevens verhandelt aan datahandelaren. Wie die partijen precies zijn, dat willen datahandelaren niet delen. Zijn dat telecomproviders? Webshops? Specifieker: Etos? Albert Heijn? ‘Ik kan een openbare bron zo noemen,’ zegt Jan-Hendrik Fleury, de Director Data Management van Cendris, ‘maar de commerciële bron delen wij niet, omdat het concurrentiegevoelig is.’

‘Concurrentiegevoelig.’ Het is een woord dat ik nog vaker zal horen. Stellig delen datahandelaren mij mee dat commerciële bronnen bedrijfsgeheim zijn. ‘Uiteindelijk wil ieder bedrijf een grote, complete database,’ legt Roland Pluut van Maxidelta, ook een datahandelaar, uit. ‘Exclusieve en kwalitatieve bronnen zijn een *competitive advantage,*dat voordeel wil je als bedrijf niet kwijt. Wie zijn geheime recept verklapt, is morgen failliet.’

Uiteindelijk vertellen enkele handelaren dat het onder andere gaat om uitgeverijen, webwinkels, telecombedrijven en retailers die klantendata registreren. Ze geven verder geen commentaar, máár, stellen ze me gerust, persoonsgegevens zijn in alle gevallen wettelijk verkregen.

Het toverwoord is hier de ‘*opt-in*.' In theorie betekent dat: jij geeft expliciet toestemming dat jouw gegevens mogen worden verwerkt en gedeeld door een bedrijf waarmee je zaken doet.

Volgens Jitty van Doodewaerd, Compliance Officer van de Nederlandse brancheverenigingen

Data-Driven Marketing Association (DDMA), zijn bedrijven verplicht consumenten duidelijk te informeren over wat ze met hun data doen. ‘Het moet niet zo zijn dat je dat op pagina 15 van de algemene voorwaarden zet. Het moet voor een consument meteen duidelijk zijn welke gegevens je gebruikt en met wie je die deelt. ‘Zorgvuldig geselecteerde partners’ volstaat niet.’

Toestemming geven, daar kun je niet onderuit

Bij veel bedrijven staan dit soort verklaringen inderdaad diep begraven in de algemene voorwaarden. Je kunt je er bovendien vooraf nooit tegen verzetten: je moet eerst accepteren dat jouw gegevens worden gedeeld voordat je daar tegenin kunt gaan.

Uit mijn inzageverzoeken blijkt dat het zeer veel tijd kost om je naderhand te verzetten tegen dataverwerking. Het is namelijk onmogelijk om na te gaan waar jouw informatie ooit is beland. Bedrijven maken niet duidelijk wie hun partners zijn – dat is concurrentiegevoelig. Wellicht hebben die partners jouw data op dat moment ook al met derden gedeeld.

En zo kan het dus zijn dat jouw data uiteindelijk in een database van christenen of miljonairs terechtkomen.

Datahandelaren nemen beslissingen over ons

Het blijft onduidelijk wat handelaren precies verzamelen en waar ze die data vandaan

hebben. Kunnen ze me wel meer vertellen over waar de data voor worden ingezet?

Gertjan Kaart van de NVH doet een poging. Want aan een kredietscore en het bijbehorende advies kleeft een groot voordeel. Klanten met goede scores kunnen makkelijker spullen op rekening betalen, terwijl een bedrijf minder risico loopt. ‘Bij abonnementen voor mobiele telefonie geldt dat de provider jou een service levert en aan het einde van de rit een factuur stuurt. Die provider wil van tevoren weten of er netjes aan de betaling gaat worden voldaan. Het bedrijf schakelt een handelsinformatiebureau in dat allerlei informatie over personen verzamelt en op basis daarvan een advies uitbrengt. Dit beschermt consumenten ook tegen zichzelf, door ze niet met rekeningen te confronteren die ze niet kunnen betalen.’

Mijn kredietscore bij Graydon kan ik wel inzien, maar ik heb geen idee door welke factoren mijn score wordt beïnvloed

Maar scores blijken niet altijd te kloppen. In 2012 onderzocht *Kassa* een aantal gevallen waarin burgers ten onrechte als ‘betalingsrisico’ werden aangemerkt. Het handelsinformatiebureau in kwestie, EDR, erkende dat dit stempel in sommige gevallen te wijten was aan een eerdere bewoner op dat adres die een rekening niet betaalde, of zelfs aan het postcodegebied waarin mensen wonen. Een bijkomend probleem: de gedupeerden kregen beperkte inzage in de totstandkoming van de score.

Bij mijn eigen inzageverzoeken is dat niet anders. Zo deelt Experian niet welke score het van mij heeft. Mijn kredietscore bij Graydon kan ik wel inzien, maar ik heb geen idee door welke factoren mijn score wordt beïnvloed – wederom: bedrijfsgeheim. Buiten de meest basale bedrijfsinformatie, zoals het aantal werknemers dat ik in dienst heb en mijn bedrijfsadres, weten ze namelijk niks over mij.

Toch zullen de voor mij (en de datahandelaar zelf) onduidelijke scores een belangrijke rol spelen in toekomstige transacties met kredietverstrekkers. Als ik een hypotheek aanvraag bijvoorbeeld, of een nieuw telefoonabonnement afsluit.

Op het gebied van marketing beweren datahandelaren dat gegevensverwerking een hoop gemak met zich meebrengt. Volgens Roland Pluut van MaxiDelta drijft handel in persoonsgegevens ook de kosten van producten omlaag. Bedrijven verdienen namelijk aan de verhuur van hun databases.

Wat als ik dit niet wil?

Maar het wordt een ander verhaal als jij dit niet wilt. Het is namelijk onmogelijk je tegen deze gegevensverwerking te verzetten. Je zou ervoor kunnen kiezen niets meer online te delen, maar dat haalt uiteindelijk weinig uit. Bedrijven schuilen zich namelijk achter het feit dat zij openbare databases gebruiken, zoals die van het CBS, de KvK en Kadaster.

De informatie komt weliswaar niet direct van jou, maar koppel de databases en je krijgt

nauwkeurige inschattingen over jouw leefomgeving. Dit mag ook volgens de wet: een profiel is geen persoonsgegeven en mag vrij worden toegepast op huishoudens.

Als het om kredietscores gaat, kun je je informatie als burger vaak pas achteraf corrigeren. Zo blijkt uit mijn inzageverzoeken dat ik momenteel niet in de database van datahandelaar Dun & Bradstreet sta, maar mocht een bedrijf bij hen aankloppen, dan zullen zij automatisch een advies over mij opstellen. Dit advies is gebaseerd op allerlei databronnen waar ik geen weet van heb.

Handelsinformatiebureaus hóéven bovendien geen inzage te geven in de modellen die zij gebruiken bij het opstellen van een score. Dit is concurrentiegevoelige informatie, waardoor het voor burgers onduidelijk blijft waarom zij een bepaalde score toegewezen kregen.

Dit zeggen de experts over ongeremde datahandel

Bedrijven mogen persoonsgegevens verhandelen als je daar toestemming voor hebt gegeven. Toch is het voor burgers onduidelijk waar zij precies toestemming voor geven. Zij kunnen zich bovendien niet onttrekken aan datavergaring van bedrijven en hun partners. Je wordt hoe dan ook in een profiel geplaatst – is het niet op basis van je eigen data, dan wel op die van anderen. Uit mijn onderzoek blijkt dat het ook onduidelijk is waar data uiteindelijk belanden en waar die voor worden gebruikt.

Wat zeggen de experts over deze ontwikkelingen?

Het delen van data brengt gemakken met zich mee, vinden datahandelaren. Gerichte advertenties online en in je postbus sluiten wellicht beter bij jouw interesses aan dan willekeurige reclameboodschappen. Dit kan een hoop irritatie voorkomen. ‘Er zitten in de Bijlmer minder mensen op een Porscheaanbieding te wachten dan in Laren,’ aldus Frank de Beun van EDM.

Een goede kredietscore maakt het betalen op rekening ook een stuk makkelijker. Gertjan Kaart van de NVH vindt het vervelend dat handelsinformatiebureaus vaak negatief in het nieuws komen. ‘Je wordt pas met het systeem geconfronteerd als je er tegenaan loopt. Maar het gemak waarmee je iets krijgt in ruil voor informatie, daar wordt aan voorbijgegaan. Terwijl alle businessmodellen ook zijn gebaseerd op informatie. Jij geeft je informatie door, omdat je een bedrijf vertrouwt en er iets voor terugkrijgt, zoals op rekening geleverd krijgen.’

Illustraties: Maus Bullhorst

De Nederlandse brancheverenigingData-Driven Marketing Association (DDMA) ziet erop toe dat bedrijven zich aan de regels houden. Het gaat weleens mis, erkent Compliance Officer Jitty van Doodewaerd. ‘DDMA controleert of organisaties privacyprincipes borgen. Bij datahandelaren controleren we bijvoorbeeld of zij de burger goed informeren dat zijn gegevens verkocht worden. Bedrijven die goed uit deze tests komen, mogen het Privacy Waarborg voeren.’

Volgens Van Doodewaerd moet het wel afgelopen zijn met de ‘sneaky’ marketing. ‘DDMA heeft in het verleden leden geroyeerd die privacyprincipes niet in hun oren geknoopt kregen. Een daarvan ging begin dit jaar om die reden failliet.’

Datahandel heeft dus zeker niet alleen maar voordelen. Onderzoeker Floris Kreiken van burgerrechtenorganisatie Bits of Freedom, die zich inzet voor de bescherming van persoonsgegevens: ‘In de toekomst zullen we nog véél meer data achterlaten bij wat we doen en laten. Dat betekent dat burgers nu al veel duidelijker moeten worden geïnformeerd over wat er met onze gegevens gebeurt. Met wie worden je gegevens gedeeld? En wat zijn de effecten daarvan voor jou? Je zou ervoor moeten kunnen kiezen dat andere partijen jouw gegevens helemaal niet in handen krijgen.’

Kreiken refereert daarnaast aan de nieuwe boetebevoegdheid van het College Bescherming Persoonsgegevens (CBP). Het CBP kan bedrijven en organisaties per 1 januari 2016 forse boetes opleggen als zij slordig met persoonsgegevens omgaan. ‘Deze nieuwe bevoegdheid kan op een fiks boetefestijn uitlopen als bedrijven over de rand van de wet gaan. Ze zullen zich tweemaal bedenken voordat ze die grens overschrijden.’

Dit artikel kwam tot stand met hulp van een fonds van de Open Society Foundations. Bits of Freedom heeft met behulp van hetzelfde fonds een adviesrapport geschreven, mede op basis van de bevindingen van dit onderzoek. Dat rapport is aanstaande donderdag beschikbaar.

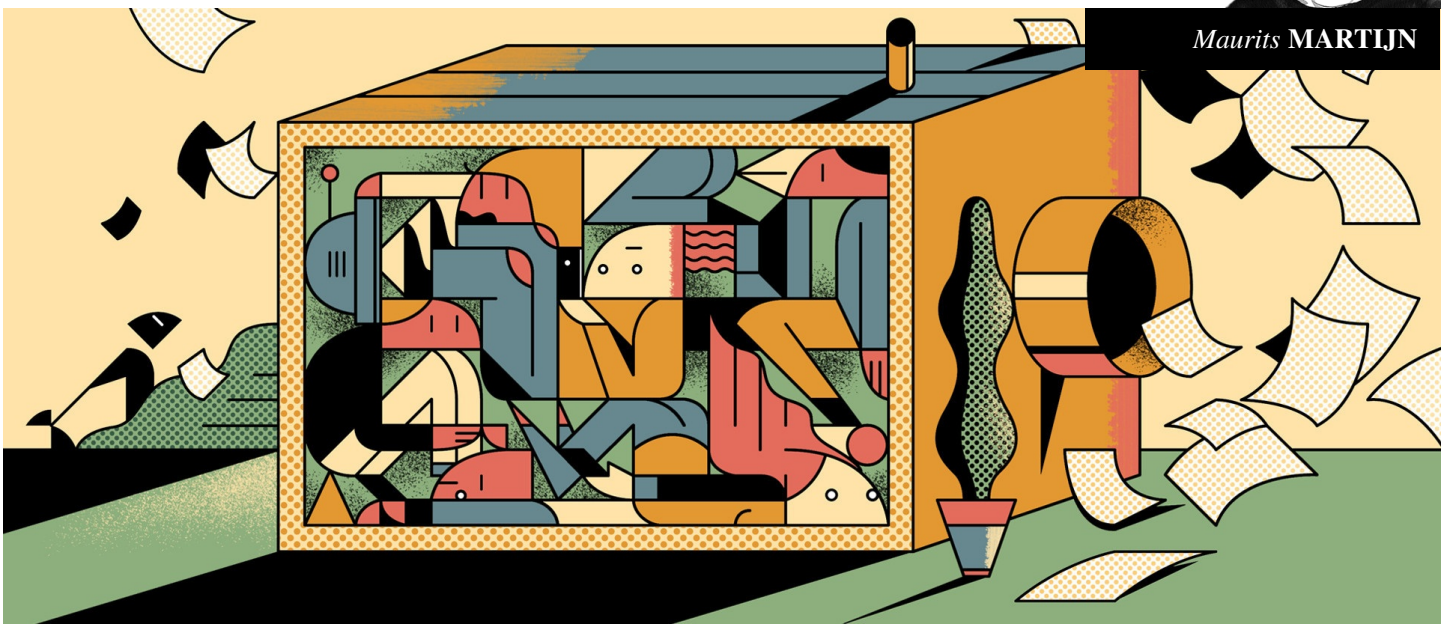
12.10.2015 • Leestijd 10 - 13 minuten

Alle Nederlanders krijgen scores toegekend door overheden, bedrijven en werkgevers. Die bepalen of je een lening, een huurauto of een baan kunt krijgen. Of: misschien wel op een fraudeur of terrorist lijkt. Hoe werkt deze scorebordsamenleving precies? Dat blijft vaak ondoorzichtig. Daarom presenteren we vandaag de website Heel Holland Transparant.

Heel Holland Transparant: Zo bepalen bedrijven en overheden of je een risicoburger bent

Correspondent Technologie & Surveillance

Maurits **MARTIJN**



Illustratie: Maus Bullhorst (voor De Correspondent)

Laten we het eens over Rob Wijnberg hebben.

Op het eerste gezicht is hij succesvol. Hoogopgeleid. Oprichter van een aantal bedrijven en bezitter van een koophuis in Amsterdam. Zijn naam staat geregeld in de kranten, zijn hoofd verschijnt weleens op tv, hij is populair op sociale media.

Er lijkt weinig mis te zijn met onze hoofdredacteur. Maar de harde data vertellen een ander verhaal.

Zou je Rob Wijnberg bijvoorbeeld een inboedelverzekering verstrekken als je weet dat er in zijn buurt negen succesvolle inbraken zijn gepleegd sinds juni?

Neem je een column van Rob Wijnberg over klimaatverandering serieus als je weet dat zijn woning energielabel 'G' heeft, de laagst mogelijke score?

En hoe sociaal is hij eigenlijk? Op Twitter heeft hij maar liefst 83.000 volgers, maar hij volgt er zelf iets meer dan 300 - waaronder ook nog eens al zijn werknemers bij De Correspondent. Wijnberg lijkt iemand die meer praat dan luistert, liever zendt dan ontvangt.

Nee, de data vellen een hard oordeel over Rob Wijnberg: zijn socialerisicoscore is 29 op een schaal van 0 (geen risico) tot 100 (extreem veel).

Het kan slechter: Bram Moszkowicz zit op 100. Maar het kan ook beter. Zo is columnist Jan Dijkgraaf met een score van 11 Wijnberg de baas.

Heel Holland Transparant

Deze score komt uit het project Heel Holland Transparant, dat we vandaag lanceren. Heel Holland Transparant doet publiek wat talloze instanties en bedrijven achter gesloten deuren doen: burgers en consumenten scoren.

Overheidsinstanties, bedrijven en werkgevers kunnen deze scores gebruiken om te bepalen of ze met de gescoorden in zee willen gaan, of ze die scherper in de gaten moeten houden, of juist allerlei aanbiedingen moeten doen.

Aan de hand van een analyse van openbare gegevens wijzen we binnen Heel Holland Transparant 35 bekende en 36 onbekende Nederlanders een risicoscore toe. VVD-Kamerlid Joost Taverne heeft bijvoorbeeld een score van 14, zangeres Marianne Weber scoort 49, de onbekende Gerda Sikkema zit op 64 en tv-presentator Matthijs van Nieuwkerk scoort met 88 heel slecht.

Wat je met die scores zou kunnen?

Ook zanger Gordon staat op de lijst. Sinds juni zijn er twintig inbraken geweest in zijn buurt. Handige informatie voor een inboedelverzekeraar

Wij zien bijvoorbeeld dat de onbekende H. van Norden in een wijk (Kruiskamp, Amersfoort) woont waar de kans op vroegtijdige sterfte hoog is. Handig voor een levensverzekeraar om te weten.

In het woonblok van oud-bestuursvoorzitter van de Universiteit van Amsterdam Louise Gunning wonen veertig mensen die een gemiddeld bruto maandinkomen van 9.200 euro verdienen. Handig om te weten als je kopers van luxeproducten zoekt.

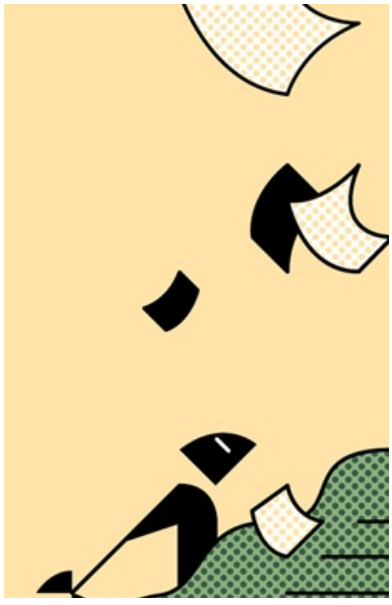
Oud-topadvocaat Bram Moszkowicz is mogelijk niet de beste persoon om een krediet aan te verstrekken. Hij staat in het openbare insolventieregister, een verzameling gerechtelijke uitspraken over faillissementen en schuldsaneringen. Veel gebruikt door bedrijven die kredietscores samenstellen.

Ook zanger Gordon staat op de lijst. Sinds juni zijn er twintig inbraken geweest in zijn buurt. Handige informatie voor een inboedelverzekeraar.

Journalisten zijn niet altijd de vrolijkste mensen. GeenStijlhoofdredacteur Marck Burema en onze eigen economiecorrespondent Jesse Frederik zijn weinig optimistische twitteraars. Van alle personen uit Heel Holland Transparant tweeten zij het negatiefst. Dat is interessante informatie voor, pak 'm beet, een potentiële toekomstige werkgever.

Sociaal kredietsysteem

Heel Holland Transparant bestaat niet écht, maar dat had je waarschijnlijk al door. Het is een project van De Correspondent, Bits of Freedom en ontwerpstudio Yuri Veerman. Wij willen hiermee de aandacht vestigen op wat wij ‘de scorebordsamenleving’ noemen.



De scorebordsamenleving is een samenleving waarin burgers scores krijgen toegekend door overheden en instanties, bedrijven en werkgevers. Die scores zijn berekeningen op basis van heel veel data. Ze voorspellen of iemand in de toekomst bepaald gedrag gaat vertonen. En die scores bepalen of je van een bepaald recht of dienst gebruik mag maken en tegen welke prijs.

Cruciaal is dat dit vaak gebeurt zonder dat burgers het doorhebben. Welke persoons- of gedragsgegevens gebruikt worden voor de scores, hoe de scores worden berekend én waarvoor ze worden gebruikt, blijft meestal in nevelen gehuld.

Een halfjaar geleden liet een artikel in *de Volkskrant* de extreemste vorm van de scorebordsamenleving zien. China heeft in 2014 het Sociaal Kredietsysteem geïntroduceerd waarbinnen iedere Chinees een score krijgt toegekend voor zijn 'gedrag.' Verschillende data bepalen die score: iemands schulden, iemands uitingen op sociale media én de scores van de mensen met wie iemand contact heeft. De score wordt voor tal van toepassingen gebruikt. Wie slecht scoort, zou kunnen worden uitgesloten van bepaalde banen, huisvesting of kredietverlening.

Schokkend. Het punt is: het Sociaal Kredietsysteem verschilt niet zo veel van wat wij in het vrije Westen al jaren aan het doen zijn. Alleen is dat niet verpakt in zo'n eenduidige sociale score en gaat het hier niet om het behoud van 'socialistische kernwaarden,' maar om het minimaliseren van allerlei risico's - van wanbetaling tot terroristische aanslagen.

Een paar voorbeelden.

1. Bijna alle Nederlanders hebben een kredietscore. Die score wordt berekend op basis van kredietverleden, faillissementen, data van de Kamer van Koophandel en, steeds vaker, data van sociale media en buurtgegevens. Die score bepaalt of jij een lening kunt krijgen en tegen welke rente.

2. Met de invoering van het Elektronisch Kinddossier krijgen alle nieuwe gezinnen een risicoscore toegewezen. Aan de hand van een lange vragenlijst wordt bepaald welke risicofactoren een gezonde ontwikkeling van het kind kunnen bedreigen. Als er een opeenstapeling van risicofactoren is, kan worden ingegrepen.

3. Van iedere passagier die naar de Verenigde Staten vliegt, wordt een score berekend. Die komt tot stand aan de hand van ongeveer dertig verschillende databronnen - waaronder

bronnen van commerciële datahandelaren en sociale media, biometrische gegevens en gegevens over eerdere reizen. Wie hoog scoort, wordt aan extra controles onderworpen of, in het extreemste geval, geweigerd.

4. Syri is een overheidssysteem dat als doel heeft om uitkerings- en belastingfraude te voorkomen. Uit een grote bak gegevens - over onder meer zorgverzekering, schulden, huisvesting en pensioenen - tovert een algoritme een risicoscore voor iedere burger. Zo weet Syri, bijvoorbeeld, dat laag watergebruik op fraude kan duiden en neemt dat mee in de score. 'Alle burgers worden onderworpen aan een integriteitstoets,' zei emeritus hoogleraar Staats- en Bestuursrecht Margriet Overkleeft-Verburg daarover.
'In feite krijgt iedere burger een rapportcijfer.'

5. Ook de Belastingdienst geeft prioriteit aan het opstellen van profielen en scores op basis van de enorme hoeveelheden data die de fiscus in huis heeft. Zo bepaalt de Belastingdienst wat de kansen zijn dat belastingplichtigen hun belastingen betalen, om vervolgens aan de hand daarvan ‘iedere belastingbetaler de behandeling te geven die hij verdient.’

We hebben dan misschien geen Chinees Sociaal Kredietsysteem, maar ook wij, Nederlanders, worden continu beoordeeld, in rankings geplaatst, doorgemeten en geanalyseerd. De taal en de beweegredenen van het Chinese systeem mogen anders zijn, de logica is hetzelfde: de score die wij krijgen toebedeeld doet een voorspelling over ons toekomstige gedrag.

Slimme machines

Bij Heel Holland Transparant zijn alle data handmatig ingevoerd en is de score per persoon berekend. Een simpele én archaïsche vorm van scoren, want de meeste scores komen tegenwoordig geautomatiseerd tot stand, aan de hand van de analyse van grote

hoeveelheden data.

Daar zitten positieve aspecten aan. Eenvoudige beslissingen voor én over mensen zijn prima te automatiseren. Wat voor advertentie je te zien krijgt. Welke zoekresultaten relevant zijn. LinkedIn, Spotify en Netflix kunnen allerlei scores berekenen om je goede aanbevelingen te doen waardoor je net die juiste persoon bevriendt, dat prachtige liedje vindt of die bijzondere film ontdekt die je anders over het hoofd zou hebben gezien.

Ook voor meer complexe beslissingen zijn geautomatiseerde beslissingen vaak heel nuttig. Mensen zijn goed in het vinden van patronen, maar computers kunnen dat doorgaans nog veel beter. Als de data goed zijn en de rekenprocedure ook, dan kunnen computers mensen helpen betere beslissingen te nemen, bijvoorbeeld door prioriteiten te stellen als veel mensen beoordeeld of gecontroleerd moeten worden. Schaarse tijd wordt nuttiger besteed. De mens, met al zijn vooroordelen en bagage, wordt bijgestaan door een computer die het niets kan schelen of je arm of rijk, blank of zwart, ongezond of fit, atheïst of moslim, hoogopgeleid bent of nooit een opleiding hebt afgemaakt.

Deze automatische benadering lijkt eerlijker en minder willekeurig.

Maar die claim is niet waar te maken, zegt Solon Barocas, die aan Princeton onderzoek doet naar geautomatiseerde besluitvorming. Vorig jaar publiceerde hij met jurist Andrew Selbst een invloedrijk artikel over de impact van Big Data. De kern: de manier waarop computers grote datasets verwerken, leidt geregeld tot onbedoelde discriminatie. Dit kan bestaande ongelijkheden in de samenleving juist vergroten.

Om dit te begrijpen, legt Barocas uit, moeten we weten hoe dit algoritmische proces werkt. Hierbij draait alles om het zogenoemde machine learning. Barocas geeft leningen als voorbeeld. Er zijn bedrijven die telefoondata gebruiken om de kans op terugbetaling te berekenen en zo beslissen of het verstandig is iemand een lening te verstrekken. Zij hebben een dataset van de belgeschiedenis van 10.000 telefoonabonnees én ze hebben een dataset met het kredietverleden van die mensen. Die voegen ze samen, waarop ze de computer vragen: als je nu kijkt naar hoe telefoons worden gebruikt, wat valt je dan op bij de mensen die hun schulden niet afbetalen? De computer zoekt en vindt patronen en komt met een antwoord: de machine heeft geleerd.

Illustraties: Maus Bullhorst

Hypothetisch voorbeeld: bij mensen die ‘s avonds laat bellen, is de kans groot dat ze hun lening niet op tijd af kunnen betalen. Als er dan een aanvraag komt voor een lening, kijk je naar iemands belgeschiedenis om te zien in hoeverre die aan dat negatieve profiel voldoet. Op basis daarvan wijs je de lening toe of af, of bereken je een hogere of lagere rente.

Dit lijkt eerlijk, maar is het niet altijd.

Barocas legt uit dat de patronen die een algoritme ontdekt vaak bestaande maatschappelijke patronen zijn. Neem het gebruik van *machine learning* in het veiligheidsdomein. Steeds meer politiekorpsen gebruiken dat om misdaadvoorspellingen te doen. Ze gaan dan preventief surveilleren in bepaalde wijken waar een hogere kans lijkt op crimineel gedrag. Maar juist doordat de politie daar extra surveilleert, zal zij misdaad vinden. Die misdaad vindt elders ook plaats, maar blijft daar onopgemerkt. De volgende keer zal de politie naar dezelfde wijk gaan. Op deze manier kan een ogenschijnlijk valide statistisch model bestaande ongelijkheden en discriminatie ‘herontdekken' en zo onbedoeld versterken.

'Als de makers van de scores en algoritmen ze al niet snappen, dan heeft het niet zoveel zin ze te openbaren'

Maurits Kaptein, docent Kunstmatige Intelligentie aan de Radboud Universiteit benoemt nog een andere eigenschap van *machine learning*: de bouwers van de algoritmen snappen de uitkomsten ook niet altijd. Kaptein beschrijft een onderzoek waar hij aan werkt in samenwerking met een bank. ‘De uitdaging van dat onderzoek is: wat is de optimale prijs voor een lening? Welke rente kan de bank aan een individu vragen zodat de bank er de meeste winst op maakt? Wij bedenken dan allerlei formules en berekeningen, maar uiteindelijk gaat de machine zelf leren en komt er een prijs uit. Ik kan dan ook niet meer exact terughalen waarom die prijs op dat moment naar voren komt.’

Dit legt misschien wel het grootste probleem van de geautomatiseerde scorebordsamenleving bloot, zegt Hans de Zwart van Bits of Freedom: bij wie moet je als individu aankloppen als er een fout is gemaakt? ‘Wat gebeurt er als je aan de verzekeringsmaatschappij vraagt: hoe zijn jullie tot deze beslissing gekomen? en zij zeggen: die is gebaseerd op de totale hoeveelheid data die wij hebben en onze rekenmodellen. En, nee, wij weten ook niet exact hoe dit komt?’

Rob Wijnberg - of waarschijnlijker: Bram Moszkowicz - kan ervoor kiezen om Heel Holland Transparant voor de rechter te slepen. Als hij het niet eens is met de score en last heeft van de gevolgen bijvoorbeeld. Maar hoe zit dat met al die geautomatiseerde systemen, die wij vaak niet zien en daardoor niet kunnen adresseren? Wat als de gevolgen van een score niet meer tot een oorzaak zijn te herleiden? Een algoritme kun je niet aanklagen.

Volgens sommige critici is totale transparantie de oplossing: bedrijven zouden hun

algoritmen en scores moeten openbaren. ‘Dat zie ik niet gebeuren,’ zegt De Zwart. ‘Je zit er als individu niet op te wachten om continu op zoek te moeten gaan naar alle profielen en scores die van jou zijn. Bovendien, als de makers van de scores en algoritmen ze al niet snappen, dan heeft het niet zoveel zin ze te openbaren. Hoewel het natuurlijk belangrijk blijft om de gegevens over jezelf in te kunnen zien, geloof ik veel meer in transparantie in de zin van: kan ik als maker het model nog uitleggen? Ik vind dat je als bedrijf bij een rechter moet kunnen verantwoorden met een aannemelijk verhaal waarom jij bepaalde beslissingen hebt genomen.’

Energielabel G?

Rob Wijnberg las op Heel Holland Transparant dat zijn woning energielabel G heeft. Dit is het slechtst mogelijke label dat aangeeft dat zijn huis zeer energieonzuinig is.

Vreemd, vond onze hoofdredacteur. Zijn huis is twee jaar geleden verbouwd en opnieuw geïsoleerd. Waar is dat label dan op gebaseerd? En welke consequenties heeft dat?

Vragen waar hij niet zomaar een antwoord op kan krijgen. Om zijn energiescore te verbeteren, is hij in zijn administratie gedoken. Op zoek naar zijn nieuwe, naar eigen zeggen, veel betere score. Tevergeefs, tot nu toe.

Heeft Rob Wijnberg geen belangrijker dingen te doen? Ja, natuurlijk. Maar niemand wil achtervolgd worden door een score die niet klopt.

Dit artikel schreef ik samen met Dimitri Tokmetzis.

Het project Heel Holland Transparant is een samenwerking van De Correspondent, Bits of Freedom en Atelier Yuri Veerman. Het project is deels gefinancierd door Open Society Foundations. De score van Heel Holland Transparant kwam tot stand met behulp van Maurits Kaptein van de Radboud Universiteit en correspondent Sanne Blauw. Wij bedanken ook Maarten de Rijke, Anne Schuth en Christoph van Gysel van de Universiteit van Amsterdam voor hun adviezen over de analyse.

Speciale dank aan Rico Disco voor hulp bij het onderzoek en de analyse daarvan.

Morgen publiceren wij een artikel over een onderzoek van gastcorrespondent Maaike Goslinga naar datahandel in Nederland.