Wolfie Christl, Sarah Spiekermann
Networks of Control

Wolfie Christl, Sarah Spiekermann

Networks of Control

A Report on Corporate Surveillance, Digital Tracking,
Big Data & Privacy

Wien 2016

facultas

Bibliografische Information Der Deutschen Nationalbibliothek

Satz und Druck: Facultas Verlags- und Buchhandels AG
Printed in Austria
ISBN 978-3-7089-1473-2

Contents

Preface **7**

1. Introduction **9**

2. Analyzing Personal Data **11**

2.1 Big Data and predicting behavior with statistics and data mining 11

2.2 Predictive analytics based on personal data: selected examples 13

2.2.1 The "Target" example: predicting pregnancy from purchase behavior 14

2.2.2 Predicting sensitive personal attributes from Facebook Likes 14

2.2.3 Judging personality from phone logs and Facebook data 16

2.2.4 Analyzing anonymous website visitors and their web searches 19

2.2.5 Recognizing emotions from keyboard typing patterns 20

2.2.6 Forecasting future movements based on phone data 20

2.2.7 Predicting romantic relations and job success from Facebook data 21

2.3 De-anonymization and re-identification 21

3. Analyzing Personal Data in Marketing, Finance, Insurance and Work **24**

3.1 Practical examples of predicting personality from digital records 25

3.2 Credit scoring and personal finance 28

3.3 Employee monitoring, hiring and workforce analytics 31

3.4 Insurance and healthcare 35

3.5 Fraud prevention and risk management 38

3.6 Personalized price discrimination in e-commerce 41

4. Recording Personal Data – Devices and Platforms **45**

4.1 Smartphones, mobile devices and apps – spies in your pocket? 46

4.2.1 Data abuse by apps 48

4.2 Car telematics, tracking-based insurance and the Connected Car 52

4.3 Wearables, fitness trackers and health apps – measuring the self 58

4.3.1 A step aside – gamification, surveillance and influence on behavior 60

4.3.2 Example: Fitbit's devices and apps 62

4.3.3 Transmitting data to third parties 64

4.3.4 Health data for insurances and corporate wellness 65

4.4 Ubiquitous surveillance in an Internet of Things? 69

4.4.1 Examples – from body and home to work and public space 72

5. Data Brokers and the Business of Personal Data **76**

5.1 The marketing data economy and the value of personal data 76

5.2 Thoughts on a 'Customers' Lifetime Risk' – an excursus 80

5.3 From marketing data to credit scoring and fraud detection 82

5.4 Observing, inferring, modeling and scoring people 84

5.5 Data brokers and online data management platforms 87

5.6 Cross-device tracking and linking user profiles with hidden identifiers 90

5.7 Case studies and example companies 94

5.7.1 Acxiom – the world's largest commercial database on consumers 94

5.7.2 Oracle and their consumer data brokers Bluekai and Datalogix 97

5.7.3 Experian – expanding from credit scoring to consumer data 101

5.7.4 arvato Bertelsmann – credit scoring and consumer data in Germany 104

5.7.5 LexisNexis and ID Analytics – scoring, identity, fraud and credit risks ... 106
5.7.6 Palantir – data analytics for national security, banks and insurers ... 108
5.7.7 Alliant Data and Analytics IQ – payment data and consumer scores ... 109
5.7.8 Lotame – an online data management platform (DMP) ... 110
5.7.9 Drawbridge – tracking and recognizing people across devices ... 111
5.7.10 Flurry, InMobi and Sense Networks – mobile and location data ... 112
5.7.11 Adyen, PAY.ON and others – payment and fraud detection ... 115
5.7.12 MasterCard – fraud scoring and marketing data ... 116
6. Summary of Findings and Discussion of its Societal Implications ... 118
6.1 Ubiquitous data collection ... 119
6.2 A loss of contextual integrity ... 120
6.3 The transparency issue ... 121
6.4 Power imbalances ... 123
6.5 Power imbalances abused: systematic discrimination and sorting ... 124
6.6 Companies hurt consumers *and* themselves ... 126
6.7 Long term effects: the end of dignity? ... 127
6.8. Final reflection: From voluntary to mandatory surveillance? ... 129
7. Ethical Reflections on Personal Data Markets (by Sarah Spiekermann) ... 131
7.1 A short Utilitarian reflection on personal data markets ... 131
7.2 A short deontological reflection on personal data markets ... 133
7.3 A short virtue ethical reflection on personal data markets ... 136
7.4 Conclusion on ethical reflections ... 138
8. Recommended Action ... 139
8.1 Short- and medium term aspects of regulation ... 140
8.2 Enforcing transparency from outside the "black boxes" ... 144
8.3 Knowledge, awareness and education on a broad scale ... 145
8.4 A technical and legal model for a privacy-friendly digital economy ... 147
List of tables ... 151
List of figures ... 152
References ... 155

Preface

In his book "How Our Days Became Numbered" historian Dan Bouk looks into how life insurers started to predict people's lives and their relative risk of death at the end of the nineteenth century. A few companies started to quantify, sort and rate people, based on statistical models and rough demographic information. Today, **a vast landscape of partially interlinked databases has emerged** which serve to characterize each one of us. Whenever we use our smartphone, a laptop, an ATM or credit card, or our 'smart' TV sets detailed information is transmitted about our behaviors and movements to servers, which might be located at the other end of the world. A rapidly growing number of our interactions is monitored, analyzed and assessed by a network of machines and software algorithms that are operated by companies we have rarely ever heard of. Without our knowledge and hardly with our *effectively informed* consent, our individual strengths and weaknesses, interests, preferences, miseries, fortunes, illnesses, successes, secrets and – most importantly – purchasing power are surveyed. **If we don't score well, we are not treated as equal to our better peers.** We are categorized, excluded and sometimes invisibly observed by an obscure network of machines for potential misconduct and without having any control over such practices.

While the media and special interest groups are aware of these developments for a while now, we believe that the full degree and scale of personal data collection, use and – in particular – *abuse* has not been scrutinized closely enough. This is the gap we want to close with the study presented in this book.

Our investigation is published at an important moment in time. A time, where a new scale of corporate surveillance is becoming effective, amplified by the rising use of smartphones, apps, social networks and ambient intelligence devices. Many of today's devices and services are deeply embedded in our private lives. In the early 2000s, we could believe that turning the computer off or not using a mobile phone would protect our privacy. Many people believed that if they did not have a share in the digital world their lives would not be affected by it. But, as this report shows in detail, old players in fields such as direct marketing, loyalty programs, credit reporting, insurance and fraud prevention are increasingly teaming up with the new online players and their pervasive data ecosystems. They make use of our clicks and swipes and link them with our "offline" purchases. Specialized data companies help others to recognize us across devices and platforms and provide access to behavioral data. Each of our interactions contributes to an ongoing evaluation of how "valuable" or potentially "risky" we might be for companies. Algorithmic decisions based on our personal data play an increasingly important role for our **options, opportunities and life-chances**. Those of us presumed unworthy by the invisible network of personal data market players and their machines can expect to face serious disadvantages. They have been categorized as "waste" by data brokers.[1]

While we were writing this report and analyzing all the facts for it, we became increasingly appalled. While both of us have been working on privacy for a while and are aware of what is happening, the pure scale of it has overwhelmed us. We are wondering whether the modern ubiquitous data-driven IT world makes us sacrifice our dignity. The readers of this book shall decide for themselves.

The title "Networks of Control" is justified by the fact that there is not one single corporate entity that by itself controls today's data flows. Many companies co-operate at a large scale to complete their profiles about us through various networks they have built up. The profiles they trade are filled with thousands of attributes per person. These networked databases are not only abused to discriminate against people with specific profile attributes, but also **attempt to make us change our behavior at scale**. Data richness is increasingly used to correct us or incentivize us to correct ourselves. It is used to "nudge" us to act differently. As a result of this continued nudging, influencing and incentivation, our autonomy suffers. Very swiftly we lose control of

[1] Singer, Natasha (2012): Mapping, and Sharing, the Consumer Genome. New York Times, June 16, 2012. Online: http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html

many aspects in our life. The idea and trust that humans are very well capable of acting responsibly is slowly evaporating.

A few words on how this report was created and on its structure: Our main goal was to investigate and summarize today's personal data ecosystem. For this purpose, the report thereafter first accumulates the facts we were able to discover. Based on an extensive range of examples from different areas and industries we aim to create a better understanding of what is happening. Some of these corporate practices have already been discussed by others, but many of them have been rarely investigated up to now, or not at all. However, this selection of examples is needed to understand the full potential and scope of corporate surveillance, digital tracking and of the business models in place today. Therefore a large part of our investigation is descriptive. This shall enable others to use our findings for their research, conclusions and ongoing initiatives. In later sections we provide a discussion of the societal and ethical implications, and recommended actions to challenge these developments.

A few words on the history of this report. A shorter first version of this report was a single-authored piece in German by Wolfie Christl who accumulated a lot of material in a study he conducted on behalf of the consumer protection department of the Austrian Chamber of Labour (Österreichische Arbeiterkammer). This study was published in November 2014[2]. This original piece was translated by the *Vienna University of Economics and Business (WU)*, while keeping only its most important parts. A master student of Sarah Spiekermann, Isabella Garraway, helped with this translation and provided some additional research. Between January and August 2016, Wolfie Christl extended and updated the investigation with extra research. Sarah Spiekermann overhauled, enriched and amended all sections, adding in particular an ethical reflection on personal data markets. Esther Görnemann, a Ph.D. student of Sarah, added reflections on a "Customer' Lifetime Risk" index. The final editing and shaping of the report was done by Wolfie Christl, Esther Görnemann, Sarah Spiekermann and Sabrina Kirrane before the publishing house *Facultas* took over.

Wolfie Christl & Sarah Spiekermann

[2] Christl, Wolfie (2014): Kommerzielle digitale Überwachung im Alltag. Studie von Cracked Labs im Auftrag der Bundesarbeitskammer. Wien, November 2014. Online: http://crackedlabs.org/dl/Studie_Digitale_Ueberwachung.pdf

1. Introduction

Classifying and sorting people

In 1994 David Lyon, a Canadian sociologist, published a book called "The Electronic Eye: The Rise of Surveillance Society". In this book Lyon foresaw the rise of a **surveillance society**, in which databases belonging to corporations and governments routinely collect, store, retrieve and process precise details of the personal lives of individuals (Lyon 1994, p.3). Lyon also introduced the concept of **social sorting**. Building on the work of Oscar Gandy, he described how electronic surveillance would lead to the constant classification and sorting of populations according to varying criteria, based on software algorithms using personal data and group data (Lyon 2003, p. 13 et seq.). As the individual groups generated by the algorithms are treated differently, this sorting would be discriminatory per se and thus may affect **choices and life-chances of individuals**.

Corporate surveillance

David Lyon's predictions of a surveillance society were made in the mid 1990s and many probably doubted the realism of his predictions at the time or put the raised threats far off for future generations to care about. Today, many of the aspects Lyon described have already become reality. The digital collection of personal data is invading everyday life more and more. The clicks, locations, paths, networks, likes and dislikes of billions of online users are stored, processed and utilized to an extent that was unthinkable only a few years ago. By now, thousands of companies are in the business of tracking and analyzing every step in the lives of citizens that live in countries with a well-developed digital infrastructure. Whether shopping in a store, using a smartphone or surfing the web, digital traces are systematically collected everywhere. Moreover, an increasing number of devices are now equipped with sensors that can broadcast information beyond the private domain of the phone. These sensors increase the amount of profiling that is being done on individuals and their behavior. The information is collected and shared across services, platforms and devices. Then, behaviors and movements are evaluated. Individuals' personality and interests are analyzed in detail. Comprehensive personal profiles are created and updated automatically. And finally digital communication and advertisements as well as offerings in the physical world are individually tailored; mostly according to their estimated profit potential for the company.

Against this background, we argue that **the surveillance society has effectively materialized.** This is not only the result of the extent of governmental surveillance, which was brought to public attention by Edward Snowden, but it is also caused by the systematic surveillance corporations have started to engage in.

What is surveillance?

S**urveillance** is defined as „the focused, systematic and routine attention to personal details for purposes of influence, management, protection or direction" *(Lyon 2007, p. 14).* Surveillance is **focused**, when it is oriented toward the individual, even though aggregate data may be used in the process. It is **systematic** when it is intentional, deliberate, and depending on certain protocols and techniques; when it doesn't happen randomly or spontaneously. In addition, surveillance happens when data collection becomes a **routine**. In "societies that depend on bureaucratic administration" based on information technology it occurs as a "normal" part of everyday life. Usually, surveillance results in power relations, in which the "watchers are privileged" (ibid).

The facts presented in this book give an account of how these three criteria are evolving, the "smarter" our cities, infrastructures and devices become.

The questions investigated in this report

Networks of control?

The objective of this report is to give a comprehensive overview of the practices in today's personal data ecosystems and their implications for individuals and society. The report addresses the following questions:

- **Data networks:** Who are the players in today's networks of digital tracking and personal data business? How do tech companies, data brokers, online data management platforms and many other businesses actually collect, collate, share and make use of personal information? How is information recorded by smartphones and other devices linked with customer records in companies?
- **Data network's sources:** Which kinds of information are recorded and shared by smartphones, fitness trackers, e-readers, smart TVs, connected thermostats and cars, and many other devices and platforms? Will the Internet of Things lead to ubiquitous surveillance of everyday life?
- **The scope of data networks:** Where is information being used in other contexts or for other purposes than it was initially collected for? To what extent is today's marketing data ecosystem merging with applications of risk management such as fraud prevention, identity verification, credit scoring, insurance analytics, background checks for employers and landlords, or even law enforcement?
- **How data networks observe the population:** How is personal data analyzed in times of Big Data? What is inferred from purchases, calls, messages, website visits, app usage, web searches and likes? How can analytics be used to predict sensitive personal attributes and to judge personality? Where are methods of data mining and Big Data analytics used today in fields such as marketing, retail, insurance, banking, healthcare and work? To what extent are consumers profiled, categorized, rated and ranked by businesses?
- **How data networks exercise control:** Do the fundamental principles of advertising that have been in effect for decades still hold? Or did advertising perhaps turn to something different through real-time targeting and personalization? How are people nudged and influenced using personalized content, rewards and other incentives based on digital tracking?

Structure of the report

These questions are addressed in four main chapters that focus on: the analysis of personal data (**chapter 2**), the use of analytics by businesses (**chapter 3**), devices and platforms (**chapter 3**) and the business of personal data (**chapter 4**). This structure was chosen as a reasonable functional differentiation, but it is still a compromise. In practice these fields are highly interconnected. Subsequently - based on the findings - the implications of corporate surveillance on individuals and society are summarized and discussed (**chapter 6**). This includes issues such as how automated decisions based on digital profiling may affect the lives of consumers and how this may this lead to unfair discrimination, social exclusion and other harms. After an ethical reflection on personal data markets by Sarah Spiekermann (**chapter 7**) an overview about recommended action is provided (**chapter 8**).

Methodology

Networks of corporate surveillance remain largely obscure. Their services, apps, platforms and algorithms are sometimes comprehensible on the surface, but the deeper functionalities are opaque and still poorly understood by the majority of users. It is therefore not surprising that the information presented hereafter is grounded in many years of research by the authors. The report is based on a systematic literature review and analysis of hundreds of documents and builds on previous research by scholars in various disciplines such as computer science, information technology, data security, economics, marketing, law, media studies, sociology and surveillance studies. Existing academic research was utilized where applicable and available. Sources also include reports by international organizations, regulators, data protection authorities, privacy advocates, civil rights organizations, industry associations, market research and consulting firms. In addition, systematic searches in online archives of newspapers, online media and blogs were conducted.

Corporate sources

As comprehensive information on corporate practices is often missing, incomplete or outdated, we selected some services and companies as examples to illustrate wider practices. We did so with the help of various corporate websites, marketing materials, brochures, data catalogs, case studies, corporate videos, developer guides, API docs etc. On occasion we also used historical versions of corporate resources. Information published by trade magazines in online marketing turned out to be particularly revealing. We also included talks of company representatives at conferences. That said, many corporate practices are kept as secret as possible. The fact that this report is only based on publicly available information is, therefore, a limitation.

Data-intensive companies communicate in a vague and ambiguous way, however they are more open when it comes to selling their services and in this context they reveal internal practices through public statements. Such statements have to be treated with caution though. Some of the sources, which cite corporate representatives may have cited them out of context (and without us being able to know this). Some sources may be altered or vanish from the Internet soon. Companies constantly change the products and services they offer. Some companies are acquired by others. Some of the sources that we found a few months ago when this study was uptaken are no longer available online, however we have still included them along with the date when they were accessed. Especially in chapters 3, 4 and 5 we often cite and document corporate statements at length for the purpose of evidence. Nevertheless, due to the ambiguity and incompleteness of these corporate sources the information in this report must be read with caution and when citing it, please make sure that you don't present our findings as a scientific fact.

2. Analyzing Personal Data

"We feel like all data is credit data, we just don't know how to use it yet"
Douglas Merrill, former Chief Information Officer at Google, 2012[3]

"Big data is the new plutonium. In its natural state it leaks, contaminates, harms. Safely contained & harnessed it can power a city"
Robert Kirkpatrick, Director UN Global Pulse, 2014[4]

2.1 Big Data and predicting behavior with statistics and data mining

In the course of digitalization, storage and computing power has multiplied tremendously. Since the turn of the millennium, data is stored, processed and analyzed on a much higher level than ever before. In public debate, the term **Big Data** often refers to the processing of these large amounts of data, sometimes it also refers to methods of analysis and prediction, and sometimes even to areas of application. There is no established definition, it has been branded as a vague[5] term that is often used as a buzzword.

---

3 Hardy, Quentin (2012): Just the Facts. Yes, All of Them. New York Times, 24.03.2012. Online: http://www.nytimes.com/2012/03/25/business/factuals-gil-elbaz-wants-to-gather-the-data-universe.html [27.07.2016]

4 Tweet: https://twitter.com/rgkirkpatrick/status/535830741247344641 [27.07.2016]

5 Harford, Tim (2014): Big data: are we making a big mistake? Financial Times, 28.03.2014. Online: http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html [27.07.2016]

A vague term

According to a definition dating back to a report from the *META Institute* (2001), which became popular during the last years, the term "Big" refers to the three dimensions **volume** (the increasing size of data), **velocity** (the increasing rate at which it is produced and transmitted) and **variety** (the increasing range of formats and representations employed).[6] The consulting company *McKinsey* uses an "intentionally subjective" definition, stating that Big Data "refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze". The size of datasets that could be referred to as Big Data could "vary by sector, depending on what kinds of software tools are commonly available and what sizes of datasets are common in a particular industry".[7]

In many fields…

The processing of large amounts of digital data has become common in many fields – from scientific fields such as meteorology, genomics, physics and astronomy to many sectors of business, financial markets, industry and government. Massive data are generated and processed in financial reporting, telecommunication, web search, social media and government surveillance as well as by sensor networks in manufacturing plants or airplanes. Every second, every device from smartphones to machines in industry are generating sensor data, software applications are generating log files and Internet users are generating clickstreams (see Krishnan 2013).

Probabilities instead of precise numbers

But Big Data is not only about volume, velocity and variety. According to Mayer-Schönberger and Cukier (2013, p. 2 et seq.) it is about "applying math to huge quantities of data in order to infer probabilities", it turns exact numbers into "something more probabilistic than precise", and it causes **three major shifts**:

- Today it is possible to "analyze vast amounts of data about a topic rather than be forced to settle for smaller sets"
- The "willingness to embrace data's real-world messiness rather than privilege exactitude"
- A "growing respect for correlations rather than a continuing quest for elusive causality"

Statistical correlations describe the "relation existing between phenomena or things or between mathematical or statistical variables which tend to vary, be associated, or occur together in a way not expected on the basis of chance alone"[8]. But "correlation does not imply causation".[9] If a statistical correlation is found between two variables and it is assumed to be a causal relationship by mistake it is called a **spurious correlation**.[10]

Analyzing personal information

Society can benefit from the technologies and practices known as Big Data in many fields, often without the use of personal data. However, it has also become common for companies to use statistical methods to analyze large amounts of very personal information – **to recognize patterns and relations, to profile, rate and judge people**

---

6 Ward, Jonathan Stuart and Adam Barker (2013): Undefined By Data: A Survey of Big Data Definitions. arXiv:1309.5821, 20.09.2013. Online: http://arxiv.org/pdf/1309.5821v1.pdf [27.07.2016]

7 Manyika, James; Chui, Michael; Brown, Brad; Bughin, Jacques; Dobbs, Richard; Roxburgh, Charles; Hung Byers, Angela (2011): Big data: The next frontier for innovation, competition, and productivity, McKinsey&Company, McKinsey Global Institute. Online: http://www.mckinsey.com/~/media/McKinsey/Business Functions/Business Technology/Our Insights/Big data The next frontier for innovation/MGI_big_data_full_report.ashx [27.07.2016]

8 http://www.merriam-webster.com/dictionary/correlation [27.07.2016]

9 Helen Beebee, Christopher Hitchcock, Peter Menzies (2012): The Oxford Handbook of Causation. OUP Oxford.

10 Many examples can be found on: http://www.tylervigen.com/spurious-correlations [28.07.2016]

and to predict their future behavior. The technologies used are summarized under the term "data mining". Their outcomes and results don't have to be completely accurate in every case. A certain amount of fuzziness is accepted. It is all about probabilities.

Identify valueable customers, avoid risk

In the context of corporate surveillance, **data mining** is, according to surveillance studies scholar Oscar H. Gandy (2006, p. 364), a process to transform "raw data into information that can be utilized as strategic intelligence" for an organization's goals. It is "directed towards the identification of behavior and status markers that serve as reliable indicators of a probable future". Companies analyzing customer data focus on identifying the most valuable customers, the best prospects, and on minimizing risk. Similarly, from a business perspective, data mining has been defined as the "process of analyzing data from different perspectives and summarizing it into useful information – information that can be used in order to increase revenue, reduce the costs, or both".[11]

In a technical sense data mining is the task of "discovering interesting patterns from large amounts of data", based on methods from statistics, pattern recognition and machine learning – for example, cluster analysis, classification, association analysis and social network analysis (see Han et al 2011). Although the terms **data mining and predictive analytics** are often used synonymously in media and public discussions, a structured classification of data mining methods has been suggested by Koh Hian and Chan Kin Leong (2011, p. 4). According to them, data mining methods are classified according to the purpose they serve:

- Methods for description and visualization
- Methods for association and clustering
- Methods for classification and estimation (prediction)

2.2 Predictive analytics based on personal data: selected examples

The following section will explore the possibilities of deriving sensitive information about people's lives from digital records that on the surface do not seem to carry a lot of information and shed light on the information that can be inferred from transactional data such as purchases, calls, messages, likes and searches.

The selection of analysis methods summarized in the following chapters show that today's digitally tracked data allows companies to **predict many aspects of a person's personality as well as sensitive personal attributes**. Although these methods are based on statistical correlations and probabilities their outcomes and conclusions are considered good enough to automatically **sort, rate and categorize people**.

A summary of academic research

After a brief summary of the often cited predictive analysis conducted by the U.S. supermarket chain Target **several academic studies on predictive analytics** are reviewed. Some of these studies were partly conducted in collaboration with companies like *Nokia*, *Microsoft*, and *Facebook*. However, the majority of such analyses and their practical applications are realized by companies that don't publish details about their practical application of predictive analytics.

[11] Information Resources Management Association (2012): Data Mining: Concepts, Methodologies, Tools, and Applications. IGI Global, 2012.

2.2.1 The "Target" example: predicting pregnancy from purchase behavior

One of the most cited examples about the prediction of sensitive information based on the analysis of everyday digital data is the case of the **U.S. supermarket chain *Target*** and its attempt to identify pregnant customers based on their shopping behavior. As Charles Duhigg reported in the New York Times[12] and in his book "The Power of Habit" (Duhigg 2012), *Target* assigns a unique code to all of its customers. All purchases and interactions are recorded – regardless of whether people are paying by credit card, using a coupon, filling out a survey, mailing in a refund, calling the customer help line, opening an email from them or visiting their website. Additionally, *Target* buys additional information on customers from data brokers.

Identifying unique moments in people's lives

Duhigg spoke extensively with a statistician from *Target*, whose marketing analytics department was tasked with **analyzing the behavior of customers** and finding ways to increase revenue. The statistician reported that one of the simpler tasks was to identify parents with children and send them catalogues with toys before Christmas. Another example he gave was the identification of customers who bought swimsuits in April and to send them coupons for sunscreen in July and weight-loss books in December. But the main challenge was to identify those major moments in consumers' lives when their shopping behavior becomes "flexible" and the right advertisement or coupon would be effective in causing them to start shopping in new ways – for example college graduation, marriage, divorce or moving house. According to a researcher cited by Duhigg, specific advertisements sent exactly at the right time, could change a customer's shopping behavior for years.

Estimating birth dates

One of the most lucrative moments would be the birth of a child. The shopping habits of exhausted, new parents would be the more flexible than at any other point in their lives. According to *Target*'s statistician, they identified 25 products which were significant to create a so called "**pregnancy prediction" score** and could even estimate the birth date. It is important to understand that they didn't simply look at purchases of baby clothes or buggies, which would be obvious. Instead, they analyzed **statistical patterns** about people purchasing certain quantities of specific lotions, soaps, hand sanitizers, cotton balls, washcloths or nutritional supplements at precise points in time.

Influencing behavior

When pregnant women were identified they received different kinds of personalized advertisements, coupons or other incentives at specific stages of their pregnancy. Duhigg also reported that a father reached out to *Target* and accused them of encouraging his daughter to get pregnant, because they sent coupons for baby clothes to her. To her father's surprise it turned out that the girl was indeed pregnant and did not tell him about it.

Regardless of whether this anecdote is true, Duhigg's research about *Target* became one of the most prominent examples of how today's companies are **collecting and analyzing personal data** to influence their customer's behavior on an individual level.

2.2.2 Predicting sensitive personal attributes from Facebook Likes

Just 170 Facebook Likes

A study conducted at the University of Cambridge showed that it is possible to accurately predict **ethnicity, religious and political views, relationship status, gender, sexual orientation** as well as a person's consumption of **alcohol, cigarettes and drugs** based on the analysis of *Facebook* Likes (see Kosinski et al 2013). The analysis was based on data of

[12] Charles Duhigg: How Companies Learn Your Secrets. New York Times, 16.02.2012. cited am 14.09.2014 von http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

58,466 users from the United States, who participated in surveys and voluntarily provided demographic information through a specific *Facebook* app called *myPersonality*[13]. This app also analyzed what they "liked" on *Facebook*, i.e. their positive associations with popular websites or other content in areas such as products, sports, musicians and books. Researchers were able to automatically predict sensitive personal attributes quite accurately, solely based on an average of 170 Likes per *Facebook* user:

Predicted attribute	Prediction accuracy
Ethnicity – "Caucasian vs. African American"	95%
Gender	93%
Gay?	88%
Political views – "Democrat vs. Republican"	85%
Religious views – "Christianity vs. Islam"	82%
Lesbian?	75%
Smokes cigarettes?	73%
Drinks alcohol?	70%
Uses drugs	65%
Single or in a relationship?	67%
Were the parents still together at 21?	60%

Table 1: Predicting personal attributes from Facebook Likes. Source: Kosinski et al 2013.

Not obvious Likes, but correlations

This shows that, for example, 88% of participants who declared themselves as gay when providing their demographic data were correctly classified as gay by the analysis based on *Facebook* Likes only. Researchers used the statistical method of logistic regression[14] to predict these dichotomous variables (e.g. yes/no) above. In addition, they also used linear regression[15] to predict numeric variables like **age, which was predicted correctly for 75%** of participants. As the researchers explain, only a "few users were associated with Likes explicitly revealing their attributes". For example, "less than 5% of users labeled as gay were connected with explicitly gay groups" such as "Being Gay", "Gay Marriage" or "I love Being Gay". Predictions rely on **less obvious, but more popular Likes** such as "Britney Spears" or "Desperate Housewives" – which proved to be weak indicators of being gay. It's remarkable that even the question whether user's parents have stayed together after this user was 21 years old was correctly predicted with an accuracy of 60%.

Likes are a generic type of data

This study shows that sensible personal attributes, which are usually considered as rather private, can be automatically and accurately inferred from rather basic information about online behavior. According to Kosinski et al, **Facebook** Likes represent a very generic type of digital records about users, similar to **web searches, browsing histories and credit card transactions**. For example, *Facebook* Likes related to music and artists are very similar to data about songs listened to or artists searched for online. Yet, in comparison to web searches and purchases the Likes of *Facebook* users are publicly accessible by default.

[13] http://www.mypersonality.org/wiki
[14] See e.g. http://www.biostathandbook.com/simplelogistic.html
[15] See e.g. http://www.biostathandbook.com/linearregression.html

2.2.3 Judging personality from phone logs and Facebook data

The five-factor model of personality, also known as the **Big Five** model, is one of the leading models of personality psychology.[16] It has been the subject of nearly 2,000 publications alone between 1999 and 2006.[17] Many studies have proven its reproducibility and consistency among different groups of age and culture.[18] The model is regularly used in the context of predicting user characteristics based on digital data.

„Big Five" personality model

According to the "Big Five" model, every person can be rated along five dimensions:[19]

Personality Dimension	People who are rated as high in this dimension could be
Extraversion	Active, assertive, energetic, enthusiastic, outgoing, talkative
Agreeableness	Appreciative, forgiving, generous, kind, sympathetic, trusting
Conscientiousness	Efficient, organized, planful, reliable, responsible, thorough
Neuroticism	Anxious, self-pitying, tense, touchy, unstable, worrying
Openness	Artistic, curious, imaginative, insightful, original, wide interests

Table 1: The five dimensions of the "Big Five" personality model. Source: McCrae and Joh 1992.

Recording smartphone usage

A Swiss study in collaboration with *Nokia Research* showed that these "Big Five" personality traits can be predicted based on smartphone metadata with an accuracy of up to 75,9% (see Chittaranjan et al 2011). At first 83 persons were asked to assess themselves using a questionnaire. Second, their communication behavior was tracked using special software installed on their phones for 8 months. For example, the following data was recorded:

Category	Which data was recorded and analyzed?
App usage	Number of times the following apps were used: Office, Internet, Maps, Mail, Video/Audio/Music, YouTube, Calendar, Camera, Chat, SMS, Games
Call logs	Number of incoming/outgoing/missed calls, number of unique contacts called and unique contacts who called, average duration of incoming/outgoing calls, …
SMS logs	Number of received/sent text messages, number of recipients/senders, Ø word length,…
Bluetooth	Number of unique Bluetooth IDs, times most common Bluetooth ID is seen, …

Table 2: Recorded mobile phone data to predict personality traits. Source: Chittaranja et al 2011

Phone usage and personality

Chittaranjan et al. recorded "data that provides information about other data", also known as **metadata**[20] **–** not the contents of the communication.[21] Applying multiple regression

---

[16] McCrae, R. R.; John, O. P. (1992): An introduction to the five-factor model and its Applications. Journal of Personality, 60, pp.175-215. Online: http://www.workplacebullying.org/multi/pdf/5factor-theory.pdf

[17] John, Oliver P.; Naumann, Laura P.; Soto, Christopher J. (2008): Paradigm Shift to the Integrative Big Five Trait Taxonomy. Handbook of Personality Theory and Research. 3. Edition, pp. 114-117. Online: http://www.ocf.berkeley.edu/~johnlab/2008chapter.pdf

[18] There are also assessments doubting the significance and accuracy of its theoretical basis. For example, its explicit focusing on the statistic method of factor analysis is criticized, see e.g. Block, Jack (2010): "The five-factor framing of personality and beyond: Some ruminations". Psychological Inquiry 21 (1): 2–25. Online: http://psychology.okstate.edu/faculty/jgrice/psyc4333/Block_Jack_2010.pdf

[19] McCrae, R. R.; John, O. P. (1992): An introduction to the five-factor model and its Applications. Journal of Personality, 60:175-215, 1992. Online: http://www.workplacebullying.org/multi/pdf/5factor-theory.pdf

[20] http://www.merriam-webster.com/dictionary/metadata

[21] To be precise, due to different definitions of "metadata" one could also argue, that information such as the „average word length" of text messages is not metadata.

analysis[22], the following significant statistical correlations between smartphone metadata and personality traits were detected (instead of "neuroticism" the inverted variant "emotional stability" was used):

Smartphone usage		Emotional Stability	Extra-version	Open-ness	Conscientious-ness	Agreeable-ness
Apps most frequently used:	Office	- 0.23		- 0.26		- 0.18
	Calender	- 0.16		- 0.18		- 0.18
	Internet		- 0.26	- 0.15		
	Camera		- 0.15			
	Video/Music				-0.18	
Calls received		0.15	0.13			0.20
Ø duration of incoming calls			0.18	0.12		
Missed calls				- 0.12		
Unique contacts called						0.17
Unique contacts SMS sent to					-0.13	- 0.13
Ø word length (sent)		0.14	- 0.15			

Table 3: Pairwise correlations between features and traits having p<0.01, ranked by absolute value of r Source: Chittaranjan et al 2011

A lack of emotional stability?

The table above shows the probability of certain personality traits based on data about smartphone usage. For example, participants who received a higher number of calls, were more likely to be agreeable (r = 0.20) and emotionally stable (r = 0.15). In contrast, participants who used the Office app more, were less likely to be open for new experience (r=-0.26). Relationships with a correlation coefficient < 0.5 are weak but still exist.[23]

Rating users

Furthermore, a machine learning model was developed to **automatically classify users** based on their smartphone metadata.

Do participants score a) low or b) high in these personality traits?	Prediction accuracy
Emotional Stability	71.5 %
Extraversion	75.9 %
Openness for Experience	69.3 %
Conscientiousness	74.5 %
Agreeableness	69.6 %

Table 4: Accuracy of predicting personality traits from phone data. Source: Chittaranjan et al 2011

Significantly above chance

Although a binary classification scheme was used, which only allows individuals to be rated as either low or high in one of the five dimensions, this shows that it is possible to infer the personality type of users based on phone usage with up to 75.9% accuracy, which is significantly above chance.

Another study based on phone logs

Researchers of *MIT*, *Harvard* and *ENS Lyon* limited themselves even more and only used so-called **Call Data Records** (CDR),[24] which all carriers keep about their customers – the same records that governments are accessing for "data retention"[25]. Their study (see Montjoye et al 2013) was based on both questionnaires and mobile phone logs of 69 participants in the United States. Data was recorded over 14 months with software

[22] See e.g. http://www.biostathandbook.com/multipleregression.html
[23] See e.g. http://www.statstutor.ac.uk/resources/uploaded/pearsons.pdf
[24] See e.g. https://www.privacyinternational.org/node/76
[25] See e.g. https://www.epic.org/privacy/intl/data_retention.html

installed on smartphones. The raw data recorded was divided into groups of indicators, for example:

Category	Evaluated Data
Regularity	E.g. Average time interval between calls and text messages, variance
Diversity	E.g. Entropy of contacts, contacts to interactions ratio, number of contacts
Movement	E.g. Daily distance traveled, number and entropy of visited places
Active Behaviour	Eg. Percent of self-initiated communication, response rates

Table 5: Evaluated mobile phone data. Source: Montjoye et al 2013

After applying a machine learning model Montjoye et al. were able to classify users along three grades of each of the "Big Five" dimensions. For example, they were able to rate participants as low, average or high in neuroticism. A comparison of the automated predictions with the personality traits measured by questionnaires lead to the following results:

Do participants score a) low b) average c) high in these personality traits?	Prediction accuracy
Neuroticism	63%
Extraversion	61%
Openness	49%
Conscientiousness	51%
Agreeableness	51%

Table 6: Accuracy of predicting personality traits from phone data. Source: Montjoye et al 2013

According to the authors, their study "provides the first evidence that personality can be reliably predicted from standard mobile phone logs". On average, the results were 42% better than random.

Judging personality better than humans?

A newer study from 2015 suggests that **computer-based personality judgments could be even more accurate than those made by humans** (see Youyou et al 2015). Again, analysis was based on data obtained through the "myPersonality" *Facebook* app. And, again, the researchers Michal Kosinski and David Stillwell were involved. They compared the "accuracy of human and computer-based personality judgments" using the results of questionnaires from 17,622 participants and **data about Facebook Likes** from 86,220 participants. Their automated predictions on personality based on Facebook Likes (r = 0.56) were more accurate than those of people, who are the participant's Facebook friends and filled out a questionnaire (r = 0.49). While the judgements of individuals considered as "spouse" (r = 0.58) were more exact than the computer models, the answers of participants considered as "family" (r = 0.50) were also less accurate than the predictions of the machines.

In addition to the "Big Five" personality traits, Montjoye et al further examined "13 life outcomes and traits previously shown to be related to personality" such as **life satisfaction, impulsivity, depression, sensationalist interest, political orientation, substance use and physical health**. As a result the "validity of the computer judgments" was again "higher than that of human judges in 12 of the 13 criteria". They state that *Facebook* Likes "represent one of the most generic kinds of digital footprint" and that their results present "significant opportunities and challenges in the areas of psychological assessment, marketing, and privacy".

2.2.4 Analyzing anonymous website visitors and their web searches

Several studies focus on how to infer personality from anonymous users doing web searches or visiting websites.

Personality profiles based on website visits

At the University of Cambridge, a study in cooperation with *Microsoft Research* about "Personality and Website Choice" was conducted, which determined correlations between visited websites and, again, the "Big Five" (see Kosinski et al 2012). More than 160 000 users were evaluated, data was provided by the previously mentioned *Facebook* app "myPersonality". Results included "Big Five" profiles of thousands of websites, based on the personality of their average visitors. The following table shows three websites in the context of arts and "do it yourself". The predicted personality traits of the average visitors of those websites are quite similar:

Domain	Open-ness	Conscienti-ousness	Extra-version	Agreeable-ness	Neuro-ticism	Frequ-ency	Default deviation
deviantART.com	0.40	- 0.19	- 0.42	- 0.05	0.16	3,154	0.01 – 0.02
Tumblr.com	0.23	- 0.23	-0.16	- 0.10	0.22	639	0.03
Etsy.com	0.41	0.14	-0.26	0.07	0.1	612	0.03

Table 7: "Big Five" profiles of average visitors of three websites. Source: Kosinski et al, 2012

When "Big Five" website profiles are known for many websites, they can be used to estimate the character of unknown, anonymous users who also visited those websites – without the need for additional information.

Age and gender

Another study by *Microsoft Research,* also based on data from the *myPersonality* app, analyzed 133 million search queries from 3.3 million unique users of the search engine Bing (see Bi et al 2013). Based on anonymous search queries it was possible to predict the age of users and the gender with 74% and 80% accuracy respectively. Religious and political views were also inferred rather accurately from web searches.

Education level and occupation

A Belgian study examined the automatic prediction of demographic attributes like gender, age, level of education and occupation from anonymous website visitors (See De Bock and Van den Poel 2010). More than 4,000 users participated in an online survey indicating their demographic information, while in parallel their clickstream data was extracted out of log files of 260 associated Belgian websites. Their surfing behavior with regard to visited websites was evaluated based on frequency, duration, the time of the day and the day of the week. After a training and scoring phase, rather reliable predictions about the demographic attributes of anonymous visitors of websites were derived:

Attribute	Possible values	Error rate
Gender	Male; female	4.94 – 6.23 %
Age	Age 12-17; age 18-24; age 25-34; age 35-44; age 45-54; age 55 and older	2.92 – 4.05 %
Occupation	Top management; middle management; farmer, craftsman, small business owner; white collar worker; blue collar worker; housewife / houseman; retired; unemployed; student; other	1.99 – 3.01 %
Education level	None or primary/elementary; lower/junior high school; high school; college; university or higher	2.56 – 4.03 %

Table 8: Predicting gender, age, level of education and occupation from website visits. Source: De Bock and Van den Poel 2010

The indicated error rates represent the average absolute error of the estimations in percentage.

2.2.5 Recognizing emotions from keyboard typing patterns

A Canadian study dealt with the recognition of user emotion by analyzing the rhythm of their typing patterns on a standard keyboard (see Epp et al 2011). 12 participants were monitored for 4 weeks using specific software, **which recorded every keystroke**, and showed a dialog with a short questionnaire about their emotional states throughout their day.

Typing patterns on a standard keyboard

Recorded data included all key press and release events of participants. The researchers then analyzed the timing of single keystroke events, but also grouped keystrokes into two-letter (e.g. "ab", "cd") and three-letter (e.g. "asd", "sdf") combinations, and prepared it as follows:

Two-letter combinations	Three-letter combinations
Duration between key 1 pressed & key 2 pressed	Duration between key 1 pressed & key 2 pressed
Duration between key 1 pressed & key 2 released	Duration between key 2 pressed & key 3 pressed
Duration between key 1 released & key 2 pressed	Duration between key 1 pressed & key 3 released
...	...

Table 9: Keyboard input evaluated. Source: Epp et al, 2011

Many mistakes in typing?

Additionally, they prepared variables like the **number of mistakes** (backspace and delete keys) and the **number of special characters** (e.g. punctuation, numbers). Longer pauses in typing were excluded. After applying machine learning models and classification algorithms, they achieved rather impressive results:

Confidence	Hesitancy	Nervousness	Relaxation	Sadness	Tired
83%	82%	83%	77%	88%	84%

Table 10: Accuracy of predicting emotional states from keystroke dynamics. Source: Epp et al, 2011

Up to 88% accuracy

Although those predictions are dichotomous (e.g. more or less "nervous"), they were able to automatically identify emotional states of users based on their keystroke dynamics with an accuracy of up to 88%, which is clearly above chance (50%).

The researchers suggest that the "ability to recognize emotions is an important part of building intelligent computers" and see their work in the context of "affective computing", which refers to "computing that relates to, arises from, or deliberately influences emotions".[26] In their related work section, Epp et al state that in prior approaches, computers successfully identified emotional states based on "facial expressions, gestures, vocal intonation, and language". Keystroke dynamics have also been successfully used to identify and authenticate users.

2.2.6 Forecasting future movements based on phone data

I know where you will be tomorrow...

Based on the analysis of smartphone data from 25 participants, researchers in the U.K. were able to predict what the participants' probable geographic position would be 24 hours later. In their study from 2012, De Domenico et al were able to exploit the **correlation between movement data and social interactions** in order to improve the accuracy of forecasting of the future geographic position of a user.

Mobility data from friends

Using data logs from 25 phones, including "GPS traces, telephone numbers, call and SMS history, Bluetooth and WLAN history", the scientists forecasted the future GPS coordinates of the users based on their movement. This resulted in an average error of 1,000 meters.

---

[26] Picard, R.W. Affective Computing. MIT Press, Cambridge, 1997, p.3

When the prediction model was subsequently extended to include the mobility data from user's friends, the **average error of the prediction could be reduced to less than 20 meters**.[27] The friendship relation between two users was, for example, derived based on one of them appearing in the address book of others.

Use cases?

The researchers outline that previous work has already shown that "human movement is predictable to a certain extent at different geographic scales" (De Domenico et al 2012, p. 1). In their study, they point to the fact that their "dataset contains a small number of users, so it is difficult to make claims about the general validity of this finding" (ibid., p. 4). However, the authors show that knowledge about a user's **social contacts** can increase the accuracy of predictions about that user considerably. Forecasting movements of people based on digital records could be used in several fields from marketing to governments. For example, **law enforcement authorities** could keep a special eye on people whose movements don't conform to the predicted ones.

2.2.7 Predicting romantic relations and job success from Facebook data

A study, which was conducted in direct collaboration with *Facebook* in 2013, analyzed data from **1.3 million** randomly chosen users who had between 50 and 2,000 friends, and who list a "relationship status" in their user profile (see Backstrom et al 2013).

Identifying partners and predicting breakups

The focus of the analysis was to examine relationships amongst users. The basic question under consideration was: "given all the connections among a person's friends, can you recognize his or her romantic partner from the network structure alone?" To recognize romantic relationships between two users, not only the **number of mutual friends** was examined but also how deeply those friends were **interconnected**. Using machine learning algorithms, the researchers were able to identify the true partner from the user's friends list in **60% of cases**. To a limited extent they were even able to predict if couples will separate in near future. Couples, who declared a relationship status in their profile, but were **not recognized as couples** by the algorithm, had a 50% higher probability of separation within 2 months.

Experiments on users

As this study reveals, the **analysis of social networks** between individuals offers a large potential for predictive analytics. Other digital records such as phone and email contacts between people offer similar options.

Facebook regularly conducts experiments on users.[28] During a very controversial[29] experiment leading to a study published in 2014, not only the behavior of users was analyzed without their knowledge, but also the user's newsfeed was manipulated (see Kramer et al 2014).

2.3 De-anonymization and re-identification

In many fields from scientific research to digital communication technology data sets, which include information on individuals, are anonymized or pseudonymized to protect individuals.

---

27 See also: Talbot, David (2012): A Phone that Knows Where You're Going. MIT Technology Review, 09.07.2012. Online: http://www.technologyreview.com/news/428441/a-phone-that-knows-where-youre-going/ [06.06.2016]

28 Hill, Kashmir (2014): 10 Other Facebook Experiments On Users, Rated On A Highly-Scientific WTF Scale. Forbes, 10.07.2014. http://www.forbes.com/sites/kashmirhill/2014/07/10/facebook-experiments-on-users [27.07.2016]

29 See e.g. Tufekci (2014)

Anonymized and de-identified

Pseudonymization involves the replacement of names and other identifying attributes with pseudonyms, for example by combinations of letters and digits. The EU General Data Protection Regulation defines it as the "processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information".[30] When additional information, for example how names relate to pseudonyms, is known, pseudonymity can be easily reverted. In contrast, the purpose of **anonymization** is to get rid of any information that would allow the re-identification of individuals. There are many challenging aspects and concepts around pseudonymity and anonymity (see Pfitzmann and Hansen 2010).

Besides the fact that different assessments about which attributes should be considered as "personally-identifiable", many of today's companies are using terms such as "anonymized" or "de-identified" in ambiguous or even wrong ways.[31] There are also fundamental problems concerning anonymization today, as for example Paul Ohm (2009) showed.

Re-identify

Depending on the kind and quantity of anonymized or pseudonymized data records it may still be possible to identify a person. If, for example, a small data set doesn't contain names, but instead **initials and birthdates**, it is often possible to identify a person by means of additional databases or publicly available information, for example because the combination of initials and birthdates is often unique.[32] A study from 1990 discovered that the combination of **zip code, gender and birth date** was unique for 216 of 248 million U.S. citizens (87%) and therefore makes identification possible. Consequently, data records with names removed but zip codes, gender and birth dates still included cannot be seen as anonymized. Therefore, it is not sufficient to only remove obviously identifying information such as name, social insurance number or IP address to anonymize data records.

„Anonymous" searches

The more detailed a data record is, the more potential links to other sources. In addition, the better the technologies use are the easier it is to identify a person, even if data seems to be anonymized. Since more and more various data about individuals is stored, this issue became increasingly severe. When, for example, *AOL* published detailed "anonymous" log files about web searches of 675,000 users in 2006, some of them could be identified just based on their search history (see Ohm 2009).

„Anonymous" movie ratings

In recent years, elaborate statistical methods for de-anonymization were developed. When Netflix published an "anonymized" data set containing movie ratings of 500,000 subscribers in 2006, a study showed that a subscriber could be easily identified, when a bit of background knowledge about this person was available. To achieve this, researchers compared and linked the "anonymized" movie ratings of the *Netflix* subscribers with publicly available reviews on the website *imdb.com*, where users often used their real names. On average between two and eight reviews from *imdb.com* were needed to identify persons in the Netflix dataset (see Narayanan and Shmatikov 2008).

---

30 Full definition: pseudonymization means the "processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person" (see EU 2016).

31 See chapter 5.6

32 Pelleter, Jörg (2011): Organisatorische und institutionelle Herausforderungen bei der Implementierung von Integrierten Versorgungskonzepten am Beispiel der Telemedizin. Schriften zur Gesundheitsökonomie, Universität Erlangen Lehrstuhl für Gesundheitsmanagement, S. 296ff

4 data points are enough

A study from 2013 analyzed the mobility data of 1.5 million mobile phone users and proved that just four spatio-temporal data points were enough to uniquely identify 95% of the users. The combination of **four times and locations where users made or received calls** is highly unique amongst different people (see Montjoye et al 2013b). According to another study, a combination of just **four apps installed on a users' smartphone** was sufficient to re-identify 95% of the users amongst a data set with lists of installed apps of 54,893 smartphone users (Achara et al 2015). It might be reasonably assumed that other types of similar data such as **purchases, search terms, visited websites** and **Facebook Likes** provide similar results.

Academic studies aside, such technologies are already used in practice to re-identify users. For example, online marketers and data brokers use **browser fingerprints** or **device fingerprints** to re-identify users based on the specific characteristics of their web browsers and devices (seeBujlow et al 2015). Also biometric data from iris, voice and face recognition as well as analyses of **keystrokes and mouse dynamics** (see Mudholkar 2012) can be used to re-identify people – akin to traditional fingerprints or DNA profiles.

3. Analyzing Personal Data in Marketing, Finance, Insurance and Work

"The privileged, we'll see time and again, are processed more by people, the masses by machines"
Cathy O'Neill , 2016

"Data scientists created the means to predict how voters will vote, or how patients will follow treatment protocols, or how borrowers will pay off debts. It wasn't long before HR realized the same technologies and approaches could be applied to predicting how employees will behave around key metrics like attrition and performance"
Greta Roberts, CEO of Human resources consulting firm Talent Analytics, 2014 [33]

The following chapter depicts how Big Data and data mining methods applied to information about human beings are already being used in the fields of marketing, retail, insurance, finance and at work. A significant focus is put on areas where these methods are applied in ways that could impact or harm individuals.

Examples in different fields

This section introduces examples in several business fields – starting with an overview about how the predictive models on personality examined in the previous chapters are already used in marketing, credit scoring and voter targeting. In addition, **five other areas** were chosen for further exploration, ranging from personalized pricing based on digital tracking to work, insurance, finance and risk management. Some fields of application are not covered in this chapter (e.g. education) or lack completeness (e.g. marketing).

Marketing is one of the areas where the analysis and exploitation of personal data is already very common at a large scale. Customer analytics try to precisely understand consumers' behaviors and preferences down to the individual level – to attract, avoid, persuade, retain or to get rid of them. Further examples of common practices can be found in chapters 4 and 5 about data-gathering devices and data brokers.

It is often **difficult to draw the line** between different areas of application. An app like *BagIQ*, which offers consumers to calculate a health score from automatically logged online and offline food purchases, is related to marketing and loyalty as well as to health.[34] While digital marketing technology is more and more incorporating aspects of consumer scoring and risk management, insurers and credit rating companies are increasingly using data about individuals, which were collected in the context of social media, marketing and online advertising. *Facebook* has already registered a patent about credit scoring.[35]

Marketing, fraud and employment

Predictive technologies such as face recognition are used on social network platforms, on consumer devices as well as for marketing, identity verification and law enforcement.[36] Fraud analytics based on vast amounts of data from different sources is used by intelligence agencies as well as by insurance companies – and also to prevent benefits fraud and social program abuse (see chapter 3.5). When *UPS* tracks and analyzes package

---

33 Roberts, Greta (2014): Making The Business Case For Predictive Talent Analytics. SAP Business Innovation, 12.05.2014. Online: http://blogs.sap.com/innovation/human-resources/making-business-case-predictive-talent-analytics-01250921 [01.08.2016]

34 https://bagiq.com [01.08.2016]

35 http://www.theatlantic.com/technology/archive/2015/09/facebooks-new-patent-and-digital-redlining/407287/ [25.01.2016]

36 See e.g. Wadhwa, Tarun (2016): How Facial Recognition Will Soon End Anonymity. Heatstreet, June 2, 2016. Online: http://heatst.com/tech/how-facial-recognition-will-soon-end-anonymity [01.08.2016]

movements and transactions[37], this is not only about improving logistical business processes, but also about monitoring and controlling employees. Similarly, when *United Healthcare* records and analyzes customer calls to call centers, to automatically detect dissatisfaction[38], this data could potentially be also used to sort and rate call center agents. Possibly, the same audio analysis technology from the same technology providers, which in this case is used to identify unsatisfied customers and improve service, can be used by fraud prevention companies and intelligence agencies to discover suspicious behavior

3.1 Practical examples of predicting personality from digital records

The analysis of personal traits based on digital records, often applying the "Big Five" model, was discussed in numerous academic papers and has gained popularity in many different areas. Several websites were launched, letting users automatically calculate their "Big Five" profile based on Facebook likes or texts written, for example, by the *Psychometrics Centre* of the *University of Cambridge*.[39]

Even the British intelligence agency **GCHQ** has used it, as was pointed out by one of the documents leaked by Edward Snowden. One of the slides shows that they had investigated correlations between the five personality traits and web browsers used – such as Chrome, Firefox, Safari and Internet Explorer.[40]

Analyzing personality for marketing

IBM predicted the "Big Five" personality traits by analyzing what users posted on *Twitter*. Michelle Zhou, the leader of *IBM*'s "User Systems and Experience Research Group", explained to *Technology Review* that extroverted persons had more desire for rewards and attention – for example, as bonus miles within a frequent flyer program. Call center agents could react differently, depending on consumers' predicted personality. She also believes that customer conversion rates could become higher if this kind of knowledge is taken into consideration– for example, in order to identify the customers that are susceptible to marketing emails or phone calls.[41]

Online quizzes

VisualDNA goes beyond testing. They use online quizzes and "psychometric" personality tests to gather data from consumers[42]. Up to now the tests were taken by more than **40 million people** "without any paid incentive".[43] Based on the collected data and analytics *VisualDNA* has created personality profiles, which could be used to predict a wide range of personal attributes, including the "Big Five", for 500 million people[44] across the globe: [45]

---

37 Davenport, Thomas H., Jill Dyché (2013): Big Data in Big Companies. SAS Institute, May 2013. Online: http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf [01.08.2016]

38 Ibid.

39 For example: http://applymagicsauce.com [01.08.2016]

40 NBC News Investigations: GCHQ PowerPoint Slideshow Presentation 2012. Online: http://www.statewatch.org/news/2014/apr/snowden_youtube_nbc_document.pdf [01.08.2014]

41 Simonite, Tom (2013): Ads Could Soon Know If You're an Introvert (on Twitter). MIT Technology Review, 08.11.2013. Online: http://www.technologyreview.com/news/520671/ads-could-soon-know-if-youre-an-introvert-on-twitter [01.08.2016]

42 http://www.visualdna.com/press-and-news/?p_id=7601 [01.08.2016]

43 http://www.visualdna.com/profiling [01.08.2016]

44 http://www.visualdna.com/press-and-news/?p_id=7601 [01.08.2016]

45 http://www.visualdna.com/profiling [01.08.2016]

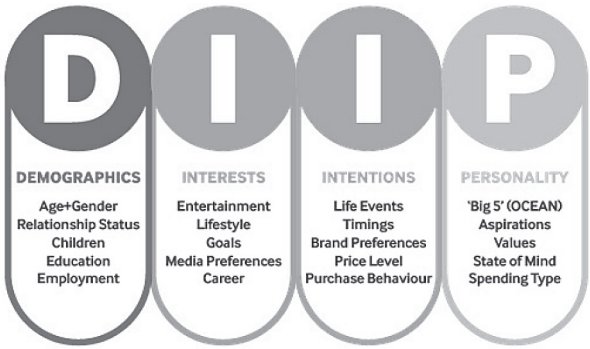

Figure 1: Types of data offered by VisualDNA. Source: Screenshot VisualDNA website

Credit risk and insurance

The company offers its data for marketing and online targeting purposes, but also for customer data management and even to **predict credit risk**.[46] According to their website they started to work with "Experian, Callcredit and MasterCard across four continents to find banking solutions for millions and even billions of people".[47] ***MasterCard*** states in a report that firms "like VisualDNA and EFL have predicted willingness to repay and other risk factors and generate a personal credit-risk score lenders can use in assessing applicants". In 2016 the company started to collaborate with *Admiral*, **a leading UK insurer**, to "explore the impact of personality on motor insurance risk assessment".[48] The *Psychometrics Centre* of the *University of Cambridge*, where much of the academic research on the prediction of "Big Five" personality traits from digital records was conducted, lists *VisualDNA* as a partner.[49]

Voter Targeting

Similarly, the consulting firm ***Cambridge Analytica*** used predictive models based on the "Big Five" personality traits[50] for Ted Cruz's U.S. presidential candidate campaign. The firm is not affiliated with the *University of Cambridge*, but a subsidiary of UK-based *SCL Group*. The company states that it helped the campaign to "identify likely pro-Cruz caucus voters and reach out to them with messages tailored to resonate specifically with their personality types" by "combining advanced data analytics with psychological research".[51] In a promotion video, their CEO explains that the "more you know about someone, the more you can align a campaign with their requirements or their wants and needs". Subsequently, it would be possible to "take one specific issue and communicate it in multiple ways to different audiences depending on their personalities".[52]

Data on 220 million citizens

Cambridge Analytica uses a "database of over 220 million Americans", which enables them to sort and categorize people along different **segments**.[53] According to their website, their analytics is based on data such as age, gender, ethnicity, income, relationship status,

---

46 http://www.visualdna.com/creditandrisk [01.08.2016]
47 Ibid.
48 http://www.visualdna.com/press-and-news/?p_id=7601 [01.08.2016]
49 http://www.psychometrics.cam.ac.uk/client-showcase [01.08.2016]
50 Davies, Harry (2015): Ted Cruz using firm that harvested data on millions of unwitting Facebook users. The Guardian, 11.12.2015. Online: https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data [01.08.2016]
51 https://cambridgeanalytica.org/news/cambridge-analytica-congratulates-senator-ted-cruz-on-iowa-caucus-win [01.08.2016]
52 Cambridge Analytica (2015): Applying Data Science to Political Campaigns. YouTube video, published on Aug 13, 2015. Online: https://www.youtube.com/watch?v=c_SlD7D_xug [01.08.2016]
53 https://cambridgeanalytica.org/datamodels [01.08.2016]

children as well as information about voting registration and many details about earlier voting behavior. From this raw data, the company predicts "swing" voters and estimates people's political views, for example:[54]

Ideology of Voters	Description
Moderate conservative	People who are likely moderate conservatives
Very conservative	People who are likely very conservative
Establishment conservative	People who are likely establishment conservatives
Liberal	People who are likely liberal leaning
Libertarian	People who are likely libertarian leaning
Tea party	People who likely support the Tea Party

Table 11: Data models to predict political ideology of voters. Source: Cambridge Analytica

Also the likely opinions about specific political issues are predicted:[55]

Specific Issues	Description
Fiscally Responsible	People who are likely to oppose government spending
Pro life	People who have a high likelihood of being pro-life
Pro environment	People who have a high likelihood of prioritizing the environment
Pro gun rights	People who have a high likelihood of prioritizing gun rights as an important issue
Pro National Security	People who have a high likelihood of prioritizing national security as an important issue
Anti Obamacare	People who are likely to oppose the Affordable Care Act
Anti immigration	People who are likely to oppose Immigration

Table 12: Data models to predict political opinions. Source: Cambridge Analytica

Personalized messages

Based on this data, voters can be targeted with specific messages and ads. People who are categorized as moderate-conservative and anti-immigration could be addressed differently than people who are categorized as libertarian and pro-environment. *Cambridge Analytica* states it helped the campaign "devise messages for a variety of direct-mail pieces, digital ads including video spots, and customized scripts for volunteers to use while contacting voters". According to the Guardian, the company has "harvested data on millions of unwitting Facebook users".[56]

Defense

Cambridge Analytica's parent company ***SCL Group*** sees itself as "[w]orking at the forefront of behavioural change" and not only as a "global election management agency", but also as a "leading practitioner of psychological approaches to conflict resolution, including population messaging and information operations"[57], providing "governments and militaries with defence and homeland security solutions"[58]

54 Ibid.
55 Ibid.
56 Davies, Harry (2015): Ted Cruz using firm that harvested data on millions of unwitting Facebook users. The Guardian, 11.12.2015. Online: https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data [01.08.2016]
57 https://sclgroup.cc [01.08.2016]
58 http://web.archive.org/web/20160109175653/http://scldefence.com/ [08.08.2016]

3.2 Credit scoring and personal finance

In recent years, several companies around the globe started to predict the creditworthiness of individuals based on data from many sources. Some of them purchase data from a wide range of third parties, some use mobile and location data, social network profiles or even carefully watch how many mistakes the applicants make when filling out an online form. While most of these companies can still be considered as "startups", some of them already received hundreds of millions in funding or started to partner with major players in finance. For example, *VisualDNA*, which was previously mentioned in the chapter on analyzing personality, started to work with *MasterCard*.

Credit scoring for social good

Some of the companies are providing their credit scoring technology to other companies. Others are also running online platforms offering payday loans, usually with rather high interest rates. Nearly all of them are constantly emphasizing, that their products will help the underbanked and unbanked – people without a credit history, who don't have access to traditional financial institutions. This is especially a problem for people in many countries in South America, Asia or Africa. However, requiring people to expose their most private details to Big Data algorithms in order to get a loan raises serious ethical concerns.

All data is credit data

One example is the U.S.-based company **ZestFinance,** which sees itself as "tech platform that applies Google-like math to credit decisions"[59]. Its founder Douglas Merrill, former Chief Information Officer at *Google*, said in 2012: "We feel like all data is credit data, we just don't know how to use it yet". And he added: "Data matters. More data is always better".[60] *ZestFinance* offers its credit scoring technology to lenders and to collectors in "auto financing, student lending, legal and healthcare",[61] but also runs an own online platform to provide loans to consumers.[62]

Data from smartphones and social networks?

ZestFinance explains that its scoring models are based on "thousands of raw data elements including **third-party data** and data collected from borrowers".[63] For example, people who "made a number of small housing moves since they graduated from college repay less than those who have moved fewer times".[64] *ZestFinance* stated that it "analyzes thousands of potential credit variables—everything from **financial information to technology usage**—to better assess factors like the potential for fraud, the risk of default, and the viability of a long-term customer relationship".[65] According to Fortune, the company looks at "**how people use smartphones and social network**".[66] According to Cathy O'Neill (2016), *ZestFinance* also uses "observations, such as whether applicants use proper spelling and capitalization on their applications forms, how long it takes them to read it, and whether they bother to look at the terms and conditions".

Credit scoring based on web searches

In 2015, *ZestFinance* started to partner with ***JD.com***, China's second largest e-commerce business. According to Fortune, they will "use data from consumers' past and present **online shopping habits**" to predict credit risk to customers and Chinese lenders based on

---

59 https://www.zestfinance.com/our-story.html

60 Hardy, Quentin (2012): Just the Facts. Yes, All of Them. New York Times, 24.03.2012. Online: http://www.nytimes.com/2012/03/25/business/factuals-gil-elbaz-wants-to-gather-the-data-universe.html [27.07.2016]

61 https://www.zestfinance.com/pdf/ZestFinance_Collections_Model.pdf [28.07.2016]

62 https://www.zestfinance.com/pdf/BaiduZestFinancePressRelease_ABSOLUTEFINALFINAL.pdf [28.07.2016]

63 https://www.zestfinance.com/pdf/ZestCashHollerREV.pdf [28.07.2016]

64 https://www.zestfinance.com/pdf/ZestFinance_Collections_Model.pdf [28.07.2016]

65 http://fortune.com/2015/12/01/tech-loans-credit-affirm-zest [28.07.2016]

66 Ibid.

data such as which items customers are purchasing, at what time of day, taking into consideration "their history of buying expensive items".[67] *JD.com* reported to have 155 million customers.[68] In 2016, a partnership with ***Baidu*** – China's dominant web search provider – was announced, to "apply *ZestFinance*'s underwriting technology to Baidu's search, location, and payment data in order to improve credit scoring decisions in China".[69] They state that Baidu's "rich **user search data** will be valuable for loan underwriting and assessing credit risk".

Valueable friends

Lenddo, a company focused on credit scoring and identity verification based in Hong Kong and operating in 20 countries such as India, South Korea, Mexico and Philippines[70], uses a wide range of data sources. According to its chairman Jeff Stewart, Lenddo helps dozens of banks analyze data from millions of smartphones globally.[71] Their *LenddoScore* is "derived from the customer's social data and online behavior" – including mobile data, browser data, application data, transactional data from telecom companies, as well as data from web publishers and social networks. In its factsheet, *Lenddo* also mentions "mouse data", "biometrics", "digital footprints", "personality analysis", "spending patterns" and "form filling".[72] A CNN article[73] points out that Lenddo's analyses include "everything from a smartphone user's messaging and browsing activity, to the apps and Wi-Fi network they use […] Elements such as foreign language used and text length reveal behavioural patterns". Even the battery level could impact the calculated credit score for a user: "the company looks at how that changes over a specific duration -- that can convey how consistent someone is and how much they plan ahead." *Lenddo*'s **identity verification** technology is called "Social Verification". It is based on similar data, and can additionally "also be configured to include document and/or face capture".[74] On an earlier version of its website, *Lenddo*'s FAQ stated that the users' credit score is based on their "character" and their "connections" to their "community" who would impact their score "both positively and negatively". Therefore, customers ought to be "selective when adding members" to their community.[75]

GPS data, social media and shopping behavior

The German company **Kreditech** has developed a "credit scoring technology which uses artificial intelligence and machine learning to process up to 20,000 data points per application".[76] They offer loans and digital banking products to consumers in Poland, Spain, Czech Republic, Russia, Mexico – but not in Germany.[77] On their platform ***Mondeo,*** they claim to have 2,000,000 "consumers scored" and additionally offer a "digital wallet", which also serves as a "Prepaid MasterCard".[78] Unlike earlier[79], *Kreditech* doesn't publish

---

[67] Rao, Leena (2015): This partnership wants to bring credit scores to China. Fortune, June 25, 2015. Online: http://fortune.com/2015/06/25/zestfinance-jd-credit-china [28.07.2016]
[68] http://ir.jd.com/phoenix.zhtml?c=253315&p=irol-homeProfile [21.08.2016]
[69] https://www.zestfinance.com/pdf/BaiduZestFinancePressRelease_ABSOLUTEFINALFINAL.pdf [28.07.2016]
[70] https://lenddo.com/about.html [28.07.2016]
[71] Hope King, L. (2016): This startup uses battery life to determine credit scores. CNN Money, Aug. 24, 2016. Online: http://money.cnn.com/2016/08/24/technology/lenddo-smartphone-battery-loan/index.html
[72] https://lenddo.com/pdfs/Lenddo-Scoring-Factsheet-2015.pdf [28.07.2016]
[73] Hope King, L. (2016): This startup uses battery life to determine credit scores. CNN Money, Aug. 24, 2016. Online: http://money.cnn.com/2016/08/24/technology/lenddo-smartphone-battery-loan/index.html
[74] https://lenddo.com/pdfs/Lenddo-Verification-Factsheet-2015.pdf [28.07.2016]
[75] https://web.archive.org/web/20140629080959/https://www.lenddo.com/pages/faq [28.07.2016]
[76] https://www.kreditech.com/what-we-do [28.07.2016]
[77] Ibid.
[78] https://www.monedo.com [28.07.2016]

much information about the data sources they are using for scoring anymore today. According to the Financial Times, the company asks loan applicants to share information on their browsing history and shopping habits as well as data from their social media accounts.[80] The applicant's interactions with *Kreditech*'s websites are also being analyzed,[81] and even the kind of fonts installed on computer can play a role.[82] In 2012, according to an earlier press release, the company used **"[l]ocation data** (GPS, micro-geographical), **social graph** (likes, friends, locations and posts), **behavioral analytics** (movement and duration on the webpage), people's e-commerce **shopping behavior** and **device data** (apps installed, operating systems)".[83]

Accessing bank data

Kreditech's subsidiary[84] ***Kontomatik*** offers a "Banking API"[85] (application programming interface) for banks and lenders[86] that "allows financial organisations to perform KYC[87], credit scoring and contextual offers online".[88] They explain that their product lets companies "access banking data" of their users with "95 supported banks" in "8 available countries".[89] *Kontomatik*'s credit scoring product "**Financial Health Indicator**" promises to help online lending companies to benefit from "detailed financial assessment of their clients".[90] On their developer website they explain that end users are asked for bank credentials in order to be able to use "screen scraping to mimic a human using a web browser" to access bank data. Therefore, they would not need agreements with supported banks – which they call "permissionless innovation".[91]

Credit scoring based on phone data

The U.S. Company ***Cignifi*** uses the previously mentioned mobile phone data to calculate "credit risk and marketing scores".[92] According to a promotional video, they "partner with mobile operators and analyze patterns from users call data records" such as "call duration, time calls are made, who initiates a call or text, numbers frequently called, and the timing, frequency and amount that uses top up their prepaid phones" to "help predict people willingness and ability to repay a loan or propensity to respond to a marketing offer".[93]

Mobile providers and credit rating agencies

According to its website, *Cignifi* partners with large mobile phone network providers such as **Telefonica**, **Airtel** (India) and **Globe Telecom** (Philippines) – they see themselves as

---

[79] https://web.archive.org/web/20140701082523/http://www.kreditech.com/#kreditechnology [28.07.2016]
[80] Vasagar, Jeevan (2016): Kreditech: A credit check by social media. Financial Times, 19.01.2016. Online: http://www.ft.com/cms/s/0/12dc4cda-ae59-11e5-b955-1a1d298b6250.html [28.07.2016]
[81] Ibid.
[82] Seibel, Karsten (2015): Gegen Kreditech ist die Schufa ein Schuljunge. Welt, 17.04.2015. Online: http://www.welt.de/finanzen/verbraucher/article139671014/Gegen-Kreditech-ist-die-Schufa-ein-Schuljunge.html [28.07.2016]
[83] Kreditech (2012): Kreditech raises 4m USD for international expansion of B2C microloans and roll-out of B2B 'Scoring as a Service' products. Press release, 17.12.2012. Available on an earlier version of their website: https://web.archive.org/web/20140117000645/http://www.kreditech.com/kreditech-raises-4m-usd-for-international-expansion-of-b2c-microloans-and-roll-out-of-b2b-scoring-as-a-service-products , and on other platforms: http://www.dgap.de/dgap/News/dgap_media/kreditech-raises-usd-for-international-expansion-microloans-and-rollout-scoring-service-products/?newsID=743379 [28.07.2016]
[84] https://www.kreditech.com/what-we-do [28.07.2016]
[85] http://kontomatik.com/ [28.07.2016]
[86] http://kontomatik.com/press [28.07.2016]
[87] "KYC" means "know your customer".
[88] http://kontomatik.com/ [28.07.2016]
[89] Ibid.
[90] http://kontomatik.com/post/kontomatik-announces-financial-health-indicator [28.07.2016]
[91] http://developer.kontomatik.com [29.07.2016]
[92] http://cignifi.com/company [29.07.2016]
[93] Cignifi (2016): How does Cignifi work? Promotional video on YouTube, published on Jun 22, 2016. Online: https://www.youtube.com/watch?v=AEJs0gw6PKw [29.07.2016]

the "ultimate data monetarization platform for mobile network operators".[94] Apart from "credit scoring models" for use in "traditional loan application processing", they also offer "Credit & Risk Models for Retailers", which predict the "likelihood of the default based on Telco data for on-line and off-line retailers".[95] In 2016, they announced a "multi-year partnership" with ***Equifax***, one of the largest consumer credit reporting agencies in the U.S., to "help Equifax expand its credit scoring capabilities in Latin America". In "partnership with local telecommunications companies in each country of operation" *Cignifi*'s scoring technology should help to "assess creditworthiness, propensity, and risk based on mobile phone usage data" for "banks, retailers, and insurers".[96]

3.3 Employee monitoring, hiring and workforce analytics

As Frank Pasquale summarized[97], in today's world of work, vast amounts of information about employees are collected and analyzed – from traditional time tracking and data from devices and machines used by the workers[98] to monitoring keystrokes and tones of voice. More and more companies are trying to measure "workers' performance, levels of concentration, attentiveness, and physical condition". In warehouses, employees are asked to wear connected handheld scanners, electronic armbands or even GPS tags.[99] One employer was even found forcing employees to use an app on their smartphone, which monitored their location 24/7.[100]

Combining available data

While in many European countries, ongoing monitoring of employees is more restricted by regulation, many companies throughout the world are making considerable efforts to enhance workforce tracking and combine available data on employees. More and more specialized service providers are developing technologies to apply predictive analytics to **workforce data** as well as to **recruiting**. Consultants and technology providers often emphasize the opportunities for both business needs and employees.

Data on 3 million employees

For example, ***Evolv***, the self-declared "leader in big data workforce optimization", claimed to have access to "500 million points of employment data"[101] on "over 3 million employees in a variety of industries and job types"[102]. As part of the hiring assessment process, according to media reports, the company utilized criteria, such as the web browsers used when sending a job application, as performance predictors.[103] It also included criteria such

---

[94] http://cignifi.com [29.07.2016]
[95] http://cignifi.com/creditfinance [29.07.2016]
[96] Cignifi and Equifax (2016): Cignifi and Equifax Partner to Bring Next-Generation Credit Scores to Unbanked Population in Latin America. Press release, March 30, 2016. Online: http://www.businesswire.com/news/home/20160330005361/en/Cignifi-Equifax-Partner-Bring-Next-Generation-Credit-Scores [29.07.2016]
[97] Pasquale, Frank (2015): The Other Big Brother. The Atlantic, Sep 21, 2015. Online: http://www.theatlantic.com/business/archive/2015/09/corporate-surveillance-activists/406201 [31.07.2016]
[98] See also chapter 4.5 on the Internet of Things
[99] Solon, Olivia (2015): Wearable Technology Creeps Into The Workplace. Bloomberg, August 7, 2015. Online: http://www.bloomberg.com/news/articles/2015-08-07/wearable-technology-creeps-into-the-workplace [31.07.2016]
[100] Ibid.
[101] https://web.archive.org/web/20141009143203/http://www.evolv.net/company/news-and-events/press-releases/evolv-achieves-triple-digit-booking-growth-big-data-workforce-solutions
[102] http://www.cnbc.com/2014/02/12/inside-the-wacky-world-of-weird-data-whats-getting-crunched.html
[103] Ibid.

as how many social networks somebody uses "to evaluate candidates for hourly work".[104] Aside from data from questionnaires and employment histories *Evolv*, reportedly also collected data about "various measures of job performance such as customer satisfaction surveys".[105] In 2015 *Evolv* was acquired by *Cornerstone*, a NASDAQ listed company[106], which offer cloud-based software for human resources.

Measuring employee performance

Cornerstone promises companies to "recruit, train and manage their people" and serves "over 25 million people in 191 countries and in 42 languages", including employees from "hundreds of the world's largest companies" such as *Walgreens*, *Xerox* and *Deutsche Post DHL*.[107] They offer several products from recruiting, training and performance management to analytics and "unified talent management" as online services[108]. In this case, data about employees is managed by *Cornerstone*. The company's **Performance Management** product offers to "measure individual employee performance". While managers are able to "continuously encourage goal achievement, productivity, and development", employees can receive "continuous feedback and coaching" – for example "social feedback and badges".[109] For example, managers in retail can "observe employees performing service competencies directly from the field, in real-time" and provide "ratings".[110]

Monitor, compare and filter workers

Cornerstone's **Insights** product promises to apply "sophisticated data science to workforce data" and uses "machine learning technology to collect and analyze data from every segment of the employee lifecycle" to obtain "actionable insights down to the individual employee, including immediate risks, opportunities and recommendations"[111] With their **View** product companies can "[i]dentify performance & compensation gaps", "[c]ompare selected employees' succession metrics & performance reviews" and use "filters to create short lists of people to solve key talent challenges".[112] In a press release *Cornerstone* explains that they offer the "**largest network of shared talent data**" from recruiting to performance management, representing "nearly 16 years of talent management activity across more than 19 million users" from "more than 2,200 organizations".[113] According to a *Cornerstone* representative cited in Fortune, other companies can compare their "historical internal data" with *Cornerstone*'s "large data sets".[114]

Scores on employees

A whitepaper about long-term unemployed provides details on the **kind of data and the type of analytics** *Cornerstone* uses. In this case they compiled "performance data on entry level frontline sales and service workers" with "nearly 500,000 performance data points from nearly 20,000 employees" from 6 employers. Data about employees used for analysis included **key performance indicators** such as the average time employees needed to

---

[104] Ito, Aki (2013): Hiring in the Age of Big Data. Bloomberg, October 25, 2013. Online: http://www.bloomberg.com/news/articles/2013-10-24/new-way-to-assess-job-applicants-online-games-and-quizzes [30.07.2016]
[105] Ibid.
[106] http://www.nasdaq.com/symbol/csod [30.07.2016]
[107] https://www.cornerstoneondemand.com/company [30.07.2016]
[108] Ibid.
[109] https://www.cornerstoneondemand.com/performance [30.07.2016]
[110] https://www.cornerstoneondemand.com/retail [30.07.2016]
[111] https://www.cornerstoneondemand.com/news/press-releases/cornerstone-ondemand-announces-cornerstone-insights-unlocking-big-data-potential [30.07.2016]
[112] https://www.cornerstoneondemand.com/resources/datasheet/cornerstone-view/4860 [30.07.2016]
[113] https://www.cornerstoneondemand.com/news/press-releases/cornerstone-ondemand-announces-cornerstone-insights-unlocking-big-data-potential [30.07.2016]
[114] Vanian, Jonathan (2015): Cornerstone OnDemand thinks big data can tell you who to hire. Fortune, May 12, 2015. Online: http://fortune.com/2015/05/12/cornerstone-recruiting-data [30.07.2016]

"complete a transaction on any given day" and a score that indicates "how satisfied the customer was with the service they received".[115]

Analyzing the tone of voice

Another company providing similar technologies is ***Workday***, which "incorporates people, business, and talent data in a single system" and enables companies to "gain detailed insight" into "employee data such as behavior, productivity, skills, and aspirations" on a wide scale.[116] In contrast, ***Humanyze*** is a U.S. startup focusing on specific kinds of data. It offers to analyze the communication patterns of team members and provides a wearable "badge" device to record behavioral data of employees.[117] According to TechCrunch, the device contains a microphone, a motion sensor and a Bluetooth connection and measures aspects such as "how people moved through the day, who they interacted with" and "what their tone of voice was like".[118]

Algorithmic Hiring

In the context of recruiting and hiring, companies also increasingly use predictive analytics to automatically rank and score applicants. In a paper on "Networked Employment Discrimination", *Data & Society* summarized how large companies in the U.S. use **Applicant Tracking Systems (ATS)** to automatically "score and sort resumes" and to "rank applicants" (see Rosenblat 2014). Only the applications and resumes with top scores are considered. The sorting algorithms are not only based on traditional criteria such as education certificates, but sometimes also incorporate online tests to assess personality and cognitive skills – or even use third-party data from blacklists to non-work related data such as social media profiles.

Predicting poor performers and risks

One example of a company providing predictive analytics for recruitment is ***HireIQ***, which offers web-based interview technology for call center recruitment and is used by "dozens" of Fortune 500 companies[119] in more than one million interviews. The company's "hiring analytics" technology[120] "automatically identifies applicants who exhibit the characteristics of long-tenured, well-performing employees".[121] Reversely, the service promises to "identify those who are likely to be poor performers and early-tenure flight risks".[122] The company's "virtual interviewing" technology supports several interviewing techniques such as quizzes and math assessments, surveys and the evaluation of typing skills.[123] But it centers on analyzing **the applicant's voice**.[124]

Voice scoring of call center candidates

HireIQ's **Audiolytics** product promises to "analyze key audio and speech characteristics to predict likely performance". It calculates "candidate scores for **vocal energy, answer length, and pace**" to "identify applicants who exhibit energy and personality".[125]

---

[115] Cornerstone (2014): The Truth About the Long Term Unemployed. Whitepaper. Online: https://www.cornerstoneondemand.com/sites/default/files/whitepaper/csod-wp-long-term-unemployed-102014.pdf [30.07.2016]
[116] Workday (2016): Workday Talent Management. Online: https://forms.workday.com/Documents/pdf/datasheets/datasheet-workday-talent-management.pdf [31.07.2016]
[117] http://www.humanyze.com [31.07.2016]
[118] Miller, Ron (2015): New Firm Combines Wearables And Data To Improve Decision Making. TechCrunch, Feb 24, 2015. Online: https://techcrunch.com/2015/02/24/new-firm-combines-wearables-and-data-to-improve-decision-making [31.07.2016]
[119] http://www.hireiqinc.com/company [31.07.2016]
[120] http://www.hireiqinc.com/solutions [31.07.2016]
[121] HireIQ (2015): Emotional Assessments: A Disruptive Innovation in Candidate Selection. A Hiring Optimization White Paper. Online: http://www2.hireiqinc.com/l/6892/2015-04-29/39jpt9/6892/155518/HireIQ___Audiolytics_White_Paper_FINAL.pdf [31.07.2016]
[122] Ibid.
[123] http://www.hireiqinc.com/solutions [31.07.2016]
[124] Ibid.
[125] Ibid.

Subsequently, a "proprietary algorithm automatically scores each candidate's interview".[126] In addition, the system is able to provide "[a]utomatic scores" to assess "**language proficiency, fluency, critical thinking, and active listening**".[127]

Gamified assessment

Other examples include ***Knack***, which offers a mobile game to measure "abilities, competencies, and interpersonal and work skills" and to "identify the right people" in hiring. It promises to "predict potential in real time" as people play, and calculates a "Knack Score" for every player.[128] ***pymetrics*** is another startup also offering games for recruitment. Their games are based on "neuroscience research" and promise to "assess 90 key cognitive and personality traits", which should result in a "snapshot of a person's unique characteristics".[129]

People analytics

Today, many human resource departments of large companies have established analytics departments or are buying technology from external companies. These technologies are often labeled as "talent management", "workforce analytics" or "people analytics". ***Oracle***, one of the largest providers of business software, criticizes companies that are still storing and managing data on their employees in separate data "silos", which, according to *Oracle*, need to be broken down.[130] Companies are encouraged to focus on employee data such as demographics, skills, rewards, engagement, attendance, adoption, key projects, assignments, goal attainment, performance ratings and data captured from the use of instruments.

Always-on surveillance

According to Josh Bersin, the founder of *Bersin by Deloitte*, **people analytics** means "bringing together all the people data in the company" – from workforce productivity to customer satisfaction.[131] He states that companies are now "opening up the floodgates" to "'always-on' listening tools" and recommends to "pull data from many different systems", for example:

- business data
- human resources data such as "tenure, salary history, job mobility, location, training history, performance rating"
- "data about individual people at work" such as "patterns of communication, location, feedback (ie. from pulse surveys), testing and assessment data, and soon heartbeat and other biometrics"
- "organizational network data" such as structures, locations, team sizes and "who reports to whom"
- external data or "data collected during recruitment" like job history, schooling, experience, and educational history
- and "new sources of data like location, travel schedule, commute time, and now even fitness, heartbeat"

---

[126] http://www.hireiqinc.com/voice-matters [31.07.2016]
[127] http://www.hireiqinc.com/solutions [31.07.2016]
[128] https://www.knack.it [31.07.2016]
[129] https://pymetrics.com/the-science [31.07.2016]
[130] CIP/Oracle (2013): Talent analytics and big data – the challenge for HR. Research Report, November 2013. Online http://www.oracle.com/us/products/applications/human-capital-management/talent-analytics-and-big-data-2063584.pdf [31.07.2016]
[131] Bersin, Josh (2016): People Analytics Market Growth: Ten Things You Need to Know. July 1, 2016. Online: http://joshbersin.com/2016/07/people-analytics-market-growth-ten-things-you-need-to-know [31.07.2016]

3.4 Insurance and healthcare

As Dan Bouk showed in his book "How Our Days Became Numbered", the early origins of the phenomenon we call "Big Data" date back to the end of the nineteenth century, when life insurance companies started to predict people's lives and relative risk of death, to quantify, sort and to rate them – they started to make them "statistical individuals".

Big Data in insurance

Insurance companies have used statistical and predictive methods for a long time. More than a century later, insurers seem to be slower, employing more data sources and advanced predictive technology. Possibly, because it is common understanding that many of these technologies are unreliable. Certainly, because the insurance sector is by far more regulated than, for example, the general digital economy, including social media platforms, online marketing and consumer data brokers. The *Boston Consulting Group* also outlines that insurance companies do not have the "rich transactional data" that bank have, because insurers have less frequent interactions with customers. In the context of Big Data, they see the "highest potential" in the following areas of the insurance value chain (see Brat et al, 2013):

- Risk assessment and pricing
- Marketing and sales (e.g. cross-selling and chum prevention)
- Fraud detection
- Claims prevention and mitigation

Risk assessment based on digital tracking

Risk assessment and pricing based on digital records of consumers' everyday behavior are already well-established. **Car insurance rates** based on actual, digitally monitored driving behavior were started more than 10 years ago. By now, **life, health and dental insurance programs** including data from wearables and activity trackers are also on the rise. The latter sometimes also incorporates data from purchases into risk assessments and pricing, for example, by asking consumers to provide access to data about the kind of food that they buy. These programs are mostly focused on rewarding – or, sometimes punishing – customers depending on how much their recorded behavior conforms to the system's rules. These programs are further investigated in chapters 4.3, 4.4 and 4.5.

Credit scores and consumer data

There are other ways how insurers and healthcare providers use predictive analytics for risk assessment. Consumer data from the wide range of online and offline sources available today can be useful for insurance companies in many different ways. In the United States, it is already common for car insurers to use data from credit reports to create their own risk scores for drivers. According to Cathy O'Neill (2016) these scores, which include all kinds of demographic data about consumers, are often more relevant to pricing than driving records. One major insurance company derived its pricing from sorting the population into more than 100,000 micro segments, which are "based on how much each group can be expected to pay". This resulted in discounts of up to 90% and increases of up to 800% for individual consumers.

Predicting health from purchases

In the field of life and health insurance, there is less evidence of practices like this. Already back in 2010, the U.S. branch of the large British insurer ***Aviva*** conducted a test to estimate individual health risks of 60,000 insurance applicants based on consumer data purchased from **data brokers**, which is traditionally used for marketing. According to the Wall Street Journal, the predictive model was developed together with the consulting firm ***Deloitte*** and aimed to examine whether *Aviva*'s traditional methods of health assessment based on blood and urine tests could be replaced by analyzing purchases, lifestyle choices and information about financial status. Consequently, they compared the traditional

methods to predict health risks such as diabetes, high blood pressure or depression with the results obtained from predictions based on consumer data. According to *Aviva*, the results were "closely aligned with those of purely traditional underwriting decisions".[132]

Diabetes, depression or cancer?

A presentation by a *Deloitte* representative explains that third-party data was acquired from *Equifax*, a consumer data broker. It included "**over 3,400 fields of data**" about occupation, education, income level and sports activities. Using this consumer data the company had built models to "predict if individuals are afflicted with any of 17 diseases (e.g. diabetes, female cancer, tobacco related cancer, cardiovascular, depression, etc.)".[133] In another report by *Deloitte* they state that they "do not propose predictive models as replacements for underwriters", medical tests would still be important. But these models could be used to "identify the higher risk applicants early". Nevertheless, they conclude that predictive analytics in life insurance "may raise ethical and legal questions".[134]

Social isolation

Similarly, the consulting firm *McKinsey* confirmed in 2012 that it helped to predict the hospital costs of patients from consumer data of a "large US payor". They used information about demographics, family structure, purchases, car ownership and other data to "construct a social isolation index" and found that **hospital costs were 24% higher** for "socially isolated individuals than for socially connected individuals". *McKinsey* concluded that such insights could help "identify key patient subgroups before high-cost episodes occur".[135]

Predicting health risks

Another U.S. company goes beyond this. ***GNS Healthcare*** sees itself as a "big data analytics company"[136] that "applies causal machine learning technology to match health interventions to individual patients". It promises to "unlock value from increasingly rich streams of patient data, including data from electronic medical records, mobile health devices, medical and pharmacy claims, genomics, consumer behavior, and more".[137] *GNS Healthcare* offers to predict individual health risks, progression of illnesses, medication adherence or intervention outcomes from a wide range of data:[138]

---

[132] Scism, Leslie and Mark Maremont (2010): Insurers Test Data Profiles to Identify Risky Clients. Wall Street Journal. Updated Nov. 19, 2010. Online: http://www.wsj.com/articles/SB10001424052748704648604575620750998072986 [31.07.2016]
[133] Kroll, Alice and Ernest A. Testa (2010): Predictive Modeling for Life Insurance Seminar. Society of Actuaries, May 19, 2010. Online: https://www.soa.org/files/pd/2010-tampa-pred-mod-4.pdf [31.07.2016]
[134] Deloitte (2010): Predictive Modeling for Life Insurance. April 2010. Online: https://www.soa.org/files/pdf/research-pred-mod-life-batty.pdf [31.07.2016]
[135] McKinsey (2012): Changing patient behavior: the next frontier in healthcare value. Online: http://healthcare.mckinsey.com/sites/default/files/791750_Changing_Patient_Behavior_the_Next_Frontier_in_Healthcare_Value.pdf [31.07.2016]
[136] http://www.gnshealthcare.com/about [31.07.2016]
[137] http://www.businesswire.com/news/home/20151208005172/en/GNS-Healthcare-Secures-10M-Series-Financing-Accelerate [31.07.2016]
[138] http://www.gnshealthcare.com/technology-overview/technology [31.07.2016]

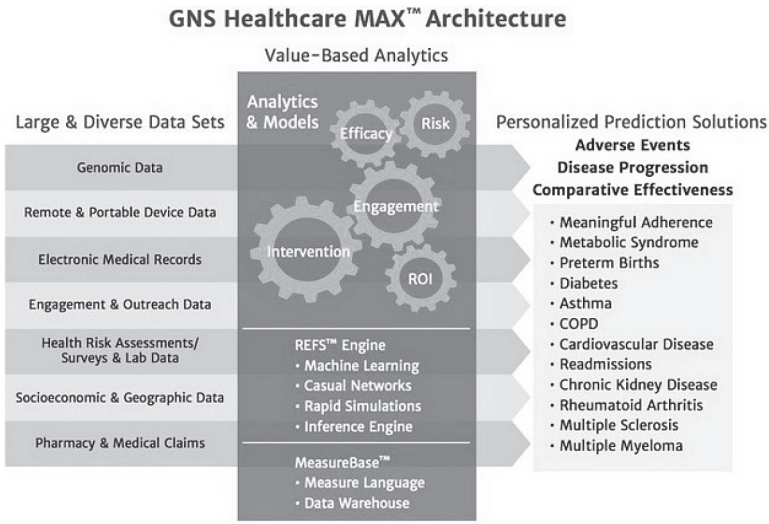

Figure 2: Data sets and models to predict health. Source: GNS Healthcare website, screenshot 31.07.2016

GNS Healthcare's ***MAX*** product promises to quantify "how interventions drive changes in behavior" and to predict the "risk of negative outcomes, including adverse events, sub-optimal outcomes and progression to disease states". Their ***IEScore*** identifies "people likely to participate in interventions".[139]

Individual risk profiles

In 2014, *GNS Healthcare* partnered[140] with the large insurer ***Aetna*** to predict the "future risk of metabolic syndrome on both a population level and an individual level" for 36,944 policyholders, based on a wide range of data – from insurance eligibility, medical and pharmacy claims records, screenings and lab results to **demographic variables** such as age, body mass index, ethnicity, cigarette usage and sleep.[141] The results were published as a study, which concludes that the predictive ability of their models was "good to excellent". The researchers created "individual risk profiles", for example, they mention a "46-year-old male", who had "92% predicted probability of developing metabolic syndrome within 12 months, and a 73% probability of developing abnormal blood glucose".

Exclude hopeless cases?

They emphasize that they achieved good results based on predictive analytics within only three months, as "opposed to the years that clinical trial and longitudinal studies take". Their methodology would allow "personalized risk predictions" and "targeted interventions for individuals with or at risk of metabolic syndrome" or, as they state, "individualized targeting based on personalized data". This could help to improve "intervention program design, impact, and returns" and also "reduce costs". According to a

[139] Ibid.
[140] Steinberg, Greg (2014): Using "Big Data" to Predict – and Improve – Your Health. Aetna Website, June 2014. Online: https://news.aetna.com/2014/06/big-data-can-predict-and-improve-health [31.07.2016]
[141] Steinberg, Gregory B.; Bruce W. Church, Carol J. McCall, Adam B. Scott, Brian P. Kalis (2014): Novel Predictive Models for Metabolic Syndrome Risk: A "Big Data" Analytic Approach. The American Journal of managed care, Vol. 20, No. 6, June 2014. Online: http://www.ajmc.com/journals/issue/2014/2014-vol20-n6/novel-predictive-models-for-metabolic-syndrome-risk-a-big-data-analytic-approach

report[142] by *Stat*, *GNS Healthcare* has also **ranked patients** "by how much return on investment the insurer can expect if it targets them with particular interventions" for other customers. As indirectly cited by *Stat*, *GNS Healthcare*'s CEO stated that the "algorithm" could "tell the insurer not to waste time and money trying to get certain patients to take their pills", who won't.

Analyzing claims

While the use of large-scale analytics in risk assessment and pricing seems to be on the rise, it has already arrived in **fraud detection and claims investigation**. Many big software vendors offer analytics products in this field. For example, the *SAS* product offers to process "all data through advanced analytics models", to apply "risk- and value-based scoring models to prioritize output for investigators" and to identify "linkages among seemingly unrelated data and uncover previously unknown relationships". It promises to "prevent substantial losses early using social network diagrams and sophisticated data mining techniques" and to create scores for claims in "near-real time with an online scoring engine that combines business rules, anomaly detection and advanced analytic techniques".[143] Often, the same software platforms, which were developed for intelligence services or military, are also being used for fraud detection in insurance companies. Examples include *IBM*'s "i2" products and *Sentinel Visualizer* (see chapter 3.5).

Social media and online data

Social Intelligence, which sees itself as a "social analytics platform built for risk"[144], offers insurers a means to "leverage social media and online data" for claims and fraud investigation. Its "real-time predictive fraud scores" promise to "assess risk during claims filing, determining the likelihood of a fraudulent claim based on a claimant's online presence".[145]

3.5 Fraud prevention and risk management

It is certainly important to protect individuals and businesses from fraud, especially in online environments where not only individual fraud is a threat but also technology-based, automated fraud. At the same time, companies in fraud prevention are often closely monitoring everyday online behavior and processing vast amounts of data on individuals. They connect to other realms of personal data collection such as marketing, credit scoring, law enforcement, intelligence services and military. Fraud prevention companies increasingly use predictive analytics based on rich data sources to classify persons or behaviors as "suspicious". For individuals, it is usually not transparent why certain interactions or options such as registering for a service or a specific purchase method get denied.

Thousands of data points

Trustev, for example, is an online fraud prevention company headquartered in Ireland, which was sold to ***TransUnion*** in 2015.[146] They offer to evaluate "online transactions in real time" for customers in financial services, government, healthcare and insurance, based on "profiling devices, analyzing digital behaviors and verifying online identities".[147]

---

[142] Robbins, Rebecca (2015): Insurers want to nudge you to better health. So they're data mining your shopping lists. Stat, 15.12.2015. Online: https://www.statnews.com/2015/12/15/insurance-big-data [31.07.2016]
[143] http://www.sas.com/en_sg/industry/insurance/fraud-framework.html [01.08.2016]
[144] http://socialintel.com [01.08.2016]
[145] http://socialintel.com/claims [01.08.2016]
[146] TransUnion (2015): TransUnion Expands Fraud and Identity Management Solutions with Acquisition of Trustev. Press release, December 10, 2015. Online: http://newsroom.transunion.com/transunion-expands-fraud-and-identity-management-solutions-with-acquisition-of-trustev [29.07.2016]
[147] Ibid.

Trustev states that it "fuses identity data with digital data in real time"[148], and promises to "examine all data, in context, for every transaction".[149] They offer customers tools that can be used to analyze "how visitors move through [their] site, click, and interact" and to employ "1000s of data points like device, IP, phone, and email" to "make a realtime decision on whether it is fraudulent or not", powered by "behavioral analysis" and machine learning.[150]

Blacklisting and analyzing friends

Trustev lists a wide range of methods and data they employ to feed their fraud detection algorithms, including phone numbers, email addresses, postal addresses, browser and device fingerprints, credit and ID checks, "cross-merchant" transaction history, IP analysis, carrier details and cell location, "smart blacklisting" and even "friend list analysis".[151] This is how they advertise their "full-spectrum transaction analysis" on their website:[152]

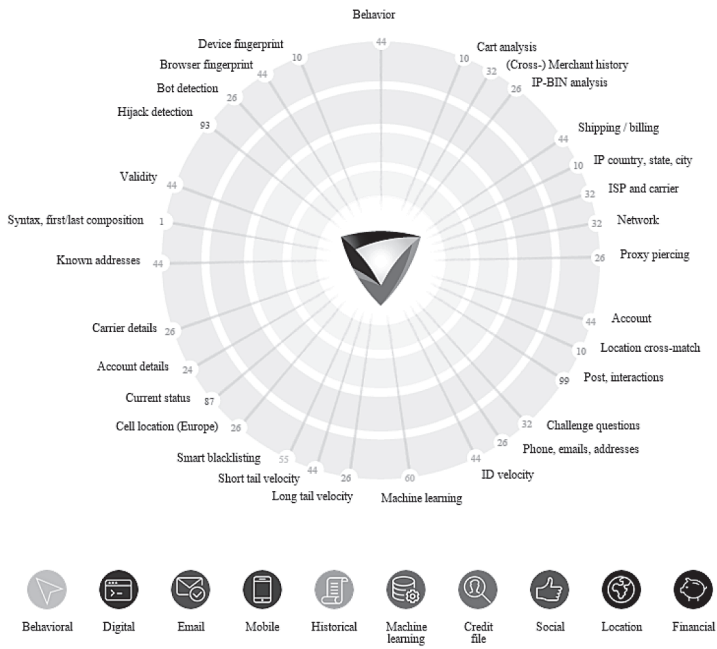

Figure 3: "Full-spectrum transaction analysis" for online fraud prevention. Source: Trustev website, screenshot from 29.07.2016.

An earlier version of the *Trustev* website explained in 2015 that their "**Social Fingerprinting** technology analyzes the information behind transactions via pattern analysis of social network information" and offers features such as "Validate personal details", "Friend list analysis" and "Pattern identification for types of content, content

[148] http://www.trustev.com [29.07.2016]
[149] http://www.trustev.com/how-it-works [29.07.2016]
[150] Ibid.
[151] Ibid.
[152] Screenshot from http://www.trustev.com/how-it-works [29.07.2016]

repetition, time stamps and differentials, interactivity".[153] *Trustev* also offers services for "RETAIL/INSTORE" fraud detection.[154]

Risk services and marketing

TransUnion states that it has "already integrated Trustev technology into its ID Manager product"[155], a suite of services for identity verification, fraud detection and authentication, which combines "insights on consumers and the devices they use in digital channels".[156] *TransUnion* sees itself as a "global risk and information solutions provider" and claims to have data on **one billion consumers** globally, obtained from 90,000 data sources. Besides risk services the company also provides marketing solutions to help businesses to "cross-sell to existing customers", "monitor and manage risk in their existing portfolios" and "display personalized messages", including "offline-to-online matching".[157]

A similar fraud prevention company, *41st Parameter*, has been acquired by the major data broker *Experian* in 2013. Their fraud detection technology allows *Experian* to link profiles in marketing.[158] *Experian* and other companies in the fields of fraud prevention, risk management and payment such as *LexisNexis*, *ID Analytics*, *Adyen*, *PAY.ON* and *MasterCard* are covered in chapter 5.7.

Data from governments and businesses

Another example of a technology to analyze fraud and risk, which is used by intelligence services as well as by insurance companies, is *IBM*'s "i2" platform. **IBM** offers a wide range of Big Data and analytics products, many of them under the label "i2".[159] For example, **i2 Analyst's Notebook** is a "visual intelligence analysis environment" to "help identify, predict, prevent and disrupt criminal, terrorist and fraudulent activities", based on "massive amounts of information collected by government agencies and businesses". A wide range of "structured and unstructured data from a variety of sources" can be imported, including "telephone call records, financial transactions, computer IP logs and mobile forensics data". This data can then be used to "Identify key people, events, connections and patterns" and to "highlight key individuals and relationships and their connections to key events", based on "integrated social network analysis capabilities". The software is used by intelligence agencies, police departments, prisons and military, and also by insurance companies.[160] For example, a "**major US insurance company**" uses IBM's i2 software to "prevent and detect auto and medical insurance fraud", to "ingest data from multiple sources, including policy, claims and medical billing data" and to gain "insight into customers before granting them insurance".[161]

Analyzing phone call data

Analyst's Notebook can be extended by other products such as the **i2 Pattern Tracer**, which is a tool for the "analysis of telephone call data", which can be used to "rapidly analyze[s] large volumes of call detail records to identify call clusters and uncover key participants" to

---

[153] Trustev website from January 2015 on archive.org: https://web.archive.org/web/20150111133910/http://www.trustev.com/how-it-works [29.07.2016]
[154] http://www.trustev.com/pricing [22.08.2016]
[155] TransUnion (2015): TransUnion Expands Fraud and Identity Management Solutions with Acquisition of Trustev. Press release, December 10, 2015. Online: http://newsroom.transunion.com/transunion-expands-fraud-and-identity-management-solutions-with-acquisition-of-trustev [29.07.2016]
[156] https://www.transunion.com/product/id-manager [22.08.2016]
[157] TransUnion (2016): 2015 Annual Report. Online: http://s21.q4cdn.com/588148537/files/doc_financials/2015/YE/TRU-2015-Annual-Report-FINAL.PDF [22.08.2016]
[158] See chapter 5.7.3
[159] http://www.ibm.com/software/industry/i2software [29.07.2016]
[160] http://www-03.ibm.com/software/products/en/analysts-notebook [29.07.2016]
[161] IBM (2013): A major US insurance company improves online insurance fraud detection. IBM Case Study. Online: https://public.dhe.ibm.com/common/ssi/ecm/bi/en/bic03034usen/BIC03034USEN.PDF [29.07.2016]

"[u]ncover and visualize clusters and patterns of interest hidden within telephone data".[162] IBM's i2 products were originally developed by a company called *i2 Inc.*, which was acquired by *ChoicePoint* in 2005[163] and by IBM in 2011.[164] According to a presentation of IBM, their i2 products are used by "80% of National Security agencies worldwide" and "25 of the 28 NATO member countries", but also by "8 of the top 10 largest companies, 12 of top 20 banks".[165]

Fraudulent unemployment claims

IBM offers many other products for analytics, such as software to prevent "**social program waste and abuse**", which is supposed to help governments to "reduce improper payments through better matching of eligibility information, gain insight into familial relationships, enhance in-take and eligibility determination, and reduce fraudulent claims through identity resolution".[166] According to an article[167] by Natasha Singer in the New York Times on Big Data and "**benefits fraud**", *IBM*'s software is used by U.S. state agencies to "identify patterns that could indicate benefit abuse". Other business intelligence companies like *SAS* and *LexisNexis* also provide analytics or data about U.S. citizens to "mitigate fraud, waste and abuse", for example, to reveal "fraudulent unemployment claims". Of course, *IBM* also offers several products for marketing and customer analytics to "uncover consumer insights with predictive analytics".[168]

CIA technology for banks and insurers

Similarly to IBM's software products, **Sentinel Visualizer** also offers features such as link and social network analysis to "discover hidden relationships, connections, and patterns among people, places, and events" in all kinds of data.[169] Its analysis software for telephone call records allows customers to "gain insight into the massive number of phone calls" by discovering relationships between phone numbers and people through multiple levels.[170] According to the company's own statement "In-Q-Tel, the CIA's venture capital arm" has invested in this technologies[171]. It is "in use by several agencies within the U.S. Federal Intelligence, Defense and Law Enforcement".[172] But it is also used for fraud detection or customer relationship mining by **banks, insurance companies and healthcare organizations** – for example, they list *Capital One* and *CIGNA Insurance* as customers.[173]

3.6 Personalized price discrimination in e-commerce

Dynamic pricing has been a common practice for a long time – from traveling (e.g. flights, hotels), entertainment (e.g. tickets for events) to retail (e.g. food). Prices vary depending on the time of a purchase or booking, inventory, available seats, popularity of a product, or prices of competitors. It is also usual to customize pricing depending on the number of units bought – and based on a specific attribute of consumers, for example discounts for children, families or elders (see Borgesius 2015). What is new today is that it is now possible to personalize pricing in real-time, based on digital records containing attributes or behaviors of consumers.

---

[162] http://www-03.ibm.com/software/products/en/pattern-tracer [29.07.2016]
[163] http://web.archive.org/web/20080703182253/http://www.i2inc.com/company/factsheet.php [29.07.2016]
[164] http://www-03.ibm.com/press/us/en/pressrelease/35255.wss [29.07.2016]
[165] IBM (2014): IBM – Analytics Solutions. Online: https://www-356.ibm.com/partnerworld/wps/servlet/download/DownloadServlet?id=6Pgu3HVIdfKiPCA$cnt&attachmentName=Introducing_IBM_i2_Enterprise_Insight_Analysis_for_Cyber_Crime_Threat_and_European_Legislation_Webinar.pdf [29.07.2016]
[166] http://www.ibm.com/analytics/us/en/industry/government/social-programs [29.07.2016]
[167] Singer, Natasha (2015): Bringing Big Data to the Fight Against Benefits Fraud. New York Times, Feb 20, 2015. Online: http://www.nytimes.com/2015/02/22/technology/bringing-big-data-to-the-fight-against-benefits-fraud.html [29.07.2016]
[168] http://www.ibm.com/analytics/us/en/business/customer-analytics [29.07.2016]
[169] http://www.fmsasg.com/
[170] http://www.fmsasg.com/LinkAnalysis/telephone_logs/call_data_records.htm
[171] http://www.fmsasg.com/LinkAnalysis/Partners/Solutions.asp
[172] http://www.fmsasg.com/AboutUs/
[173] http://www.fmsasg.com/LinkAnalysis/Commercial/Solutions.asp

Different prices or products

Jakub Mikians et al (2012) differentiate between price and search discrimination:

- **Price discrimination** is defined as the "practice of pricing the same product differently to different buyers", depending on an assumed maximum price, which a particular customer possibly would pay. It is clearly distinguished from different pricing across different stores, which may want to reduce their stock or have better deals with manufacturers than other stores.
- In the case of **search discrimination,** different users see different products, when browsing an online shop or certain product categories. For example, some users may see more expensive hotels than others on the top of the list. As most users do not view more than the first page of a search result or category listing, they are steered towards specific offers. Consequently, this practice is also called **price steering** (see Borgesius 2015).

Mac vs. PC

According to the Wall Street Journal, the travel and booking platform Orbitz was found in 2012 to "steer" users of *Apple*'s *Mac* computers to pricier hotels, because they "spend $20 to $30 more a night on hotels" than PC users on average.[174] During their tests, the hotels listed on the first page of results were up to **13% more expensive** when using a Mac than when using a PC. According to the article, *Orbitz* conformed that they were "experimenting with showing different hotel offers to Mac and PC visitors", and also that other factors such as the location of the user and their previous behavior on the website could have an influence on the offers shown. But they were not "showing the same room to different users at different prices". In 2014, Orbitz emphasized that their "experiment lasted approximately one month" and "was discontinued".[175]

Based on location and web browsing

Another investigation, also conducted by the *Wall Street Journal* later in 2012, found that the large U.S. office supply company *Staples* offered "different prices to people after estimating their locations".[176] Testing suggested that the company could have inferred the ZIP codes of online shoppers by analyzing their **IP addresses**. When they simulated visits to *Staple*'s website from 29,000 different ZIP codes in the U.S. and tested 1,000 randomly selected products offered, they found **price differences of 8% on averag**e. In addition, it was discovered that other companies like *Discover Financial Services, Rosetta Stone* and *Home Depot* were also "adjusting prices and displaying different product offers based on a range of characteristics that could be discovered about the user". The office supplier *Office Depot* confirmed using "customers' **browsing history** and geolocation" to show different offers and products to shoppers. *Home Depot* confirmed using IP addresses of users in order to "match users to the closest store and align online prices".

Refering site and financial status

A Spanish study used sophisticated methods of measurement to automatically monitor the prices of 600 products in 35 categories of 200 large online shops (see Mikians et al 2012). It was found that **prices differed up to 166%** depending on the geographical location of users. In some cases prices varied also depending on the referring URL. For example, prices in some product categories were 23% lower, when the online shop's website was

---

[174] Mattioli, Dana (2012): On Orbitz, Mac Users Steered to Pricier Hotels. Wall Street Journal, 23.08.2012. Online: http://online.wsj.com/news/articles/SB10001424052702304458604577488822667325882 [29.07.2016]

[175] Letter to Christo Wilson, Assistant Professor, College of Computer and Information Science, Northeastern University. August 13, 2014. Online: http://personalization.ccs.neu.edu/PriceDiscrimination/Responses/Orbitz/OWW to Christo Wilson.pdf [29.07.2016]

[176] Valentino-Devries, Jennifer; Singer-Vine, Jeremy; Soltani, Ashkan (2012): Websites Vary Prices, Deals Based on Users' Information. Wall Street Journal. Online: http://online.wsj.com/news/articles/SB10001424127887323777204578189391813881534 [29.07.2016]

not visited directly, but was instead accessed via a "discount aggregator site". Furthermore, Mikians et al found evidence of search discrimination depending on whether a customer was simulated as **"affluent" or as "budget conscious"**. Prices of products shown to customers were "up to 4 times higher for affluent than for budget conscious customers". The researchers created an elaborate technical framework for measurement. For example, to analyze whether a user's financial status had an influence on pricing, the measurement framework automatically simulated previous visits of hundreds of specific websites from discount sites to shops selling luxury products.

Different prices on mobile

A similar study from 2014 examined 16 major e-commerce sites, including 10 general retailers and 6 travel sites (see Hannak et al 2014) to investigate price steering and price discrimination. They used both a technical measurement framework and 300 real-world users, and found "evidence of personalization on four general retailers and five travel sites, including cases where **sites altered prices by hundreds of dollars**". Through technical measurement they discovered differences in the products shown to users based on the history of clicked or purchased products, and their operating system or browser – for example, when using a mobile device. Two travel sites conducted A/B tests[177] that "steer users towards more expensive hotel reservations". In some cases different prices where shown depending on whether the site was browsed while logged in or not. The authors emphasize that they were "only able to identify positive instances of price discrimination and steering" and "**cannot claim the absence of personalization**". They received several responses from investigated companies, which are documented on an additional website.[178]

Hard to reveal

As the studies described above show, it is very challenging – if not impossible – to accurately investigate and prove **price or search discrimination** based on individual attributes or user behavior. Even when differences in pricing are obvious, it is still unknown, whether these differences depend on individual characteristics or on other criteria – prices may vary for other reasons. It also **remains unknown** how user attributes or behaviors were identified, which personal data was used, whether additional data was purchased and which algorithms were used. Under these circumstances **consumers have no chance** to understand what their individual offers and prices are based on. It is even very difficult for them to recognize, whether they receive individual offers and prices at all or not. The situation could aggravate when more information about users and better analytics are added. Apart from transparency issues companies could, for example, "overcharge [people] when the data collected indicates that the buyer is **indifferent, uninformed or in a hurry**".[179]

Really personal offers

Some scholars note that it "seems that companies rarely personalise prices" today (Borgesius 2015). However, dynamic pricing is on the rise. According to a price monitoring company, ***Amazon* varies its prices 2.5 million times** on an average day. Sometimes prices of specific products are changed more than 10 times a day.[180] While *Amazon* still doesn't seem to use information about personal characteristics for pricing, it could be difficult to prove, when they would someday decide to introduce it. Personalized

---

[177] A/B tests are experiments by website providers, where different groups of users see different versions of a website's functionality.
[178] Hannak, Aniko; Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson (2014): Paper Overview. Online: http://personalization.ccs.neu.edu/PriceDiscrimination/Press [29.07.2016]
[179] Zarsky T (2004): Desperately Seeking Solutions: Using Implementation-Based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society. 56(1) Maine Law Review 13, p. 52. See also p. 30-31. Quoted from: Borgesius (2015)
[180] Callard, Abby (2014): The right price, not the lowest price. internet RETAILER, December 30, 2014. Online: https://www.internetretailer.com/2014/12/30/right-price-not-lowest-price [29.07.2016]

pricing can also present itself in other ways than simply during an everyday visit of a **centralized shopping or travel website**. For example, via web and mobile advertisements, email and other channels, Internet users increasingly often directly receive individual offers and discounts based on digital tracking.

Based on a "customer value score"

TellApart, a "predictive marketing platform" provides companies ways to send personalized "unique-to-one messages that are consistent across channels and devices", based on analytics and a wide range of data.[181] According to their website, for "each shopper and product combination" a "Customer Value Score" is created – a "compilation of **likelihood to purchase, predicted order size, and lifetime value**".[182] Companies can then deliver personalized messages to consumers "through channels and devices", for example "through advertising in display, social sites, mobile, email and even dynamic on-site promotions".[183] By using "Dynamic Promotions" such as **personalized "discounts and offers"**, companies can send the "right offer to the right person at the right time".[184]

Online and offline data

In order to do so *TellApart* creates a "**customer data profile**", which is constantly updated "via feeds and tags", and "represents a merging of 100s of online and in-store signals about a particular anonymous customer across the channels and devices they use". Because "[p]ersonalization begins with a person" their "Identity Network" service "incorporates anonymous data – from both online and offline sources – to **create an ID for shoppers**".[185] *TellApart* often emphases that it uses "anonymous data" and creates "anonymous profiles" about "anonymous identities" and "anonymous customers", but nonetheless still creates an "ID for shoppers".[186] In 2015, *TellApart* has been acquired by ***Twitter*** for $479 million.[187]

Personalized terms and conditions

Many companies from advertising to data brokers are working on technologies and products with the objective of sending personalized offers and discounts to **valuable consumers**, and excluding others. Not least, personalized pricing based on digital tracking is already present on an even more problematic level. Since a few years **insurance companies** offer programs, for which pricing depends on steps, activity and car driving behavior (see chapter 4.3.4). Other companies offer **loans**, whose conditions are linked to extensive digital records on individuals (see chapter 3.2). Not only "prices, but also terms and conditions can be personalised" (Helberger 2016).

---

[181] https://www.tellapart.com/solutions [29.07.2016]
[182] https://www.tellapart.com/platform [29.07.2016]
[183] Ibid.
[184] https://www.tellapart.com/solutions [29.07.2016]
[185] https://www.tellapart.com/platform [29.07.2016]
[186] Ibid.
[187] Lunden, Ingrid (2016): Twitter Ended Up Paying $479M For Adtech Startup TellApart, 10-K Reveals. TechCrunch, Feb 29, 2016. Online: https://techcrunch.com/2016/02/29/twitter-479m-tellapart/ [29.07.2016]

4. Recording Personal Data – Devices and Platforms

"He had won the victory over himself. He loved Big Brother"
Last sentence of George Orwell's novel „1984" [188]

In the early days of digital tracking the only way to recognize users across multiple website visits was to either require them to register and authenticate or to pass unique identifiers from one to another page while users are interacting. When HTTP cookies became available in 1994 they "fundamentally altered the nature of surfing the Web", from "being a relatively anonymous activity, like wandering the streets of a large city, to the kind of environment where records of one's transactions, movements and even desires could be stored, sorted, mined and sold".[189]

Cookies

Cookies are "small pieces of data", which are "placed in a browser storage by the web server" (Bujlow et al 2015, p. 5). When a website is visited the first time, a unique identification code can be stored in the cookie file on the user's computer. Subsequently, the website can recognize the user across further page visits by accessing this identifier again and again. While **session cookies** expire when the web browser is closed, **persistent cookies** can be stored for hours, days or years (see Bujlow et al 2015). Both types can be used for authentication purposes or to remember information entered by the user, such as items in an online shopping cart, but also to track which pages were visited and how a user interacted with the website in the past. While **first-party cookies** are directly set by the domain the user visited, **third-party cookies** are stored by other domains embedded in the website initially visited (see Roesner et al 2012, p. 2). This way, third parties can track users of multiple websites.

Tracking website visits

Today, most websites contain small code fragments from third-party companies, which record every click, track users across websites and transmit information to those companies. These code fragments are referred to as **web beacons** or **web bugs**[190]. They can be visualized with browser extensions such as *Lightbeam*[191]. After installation, *Lightbeam* shows all third parties, which information is transferred when a website is visited. Many services still use persistent third-party cookies to recognize users throughout different website visits. But also other sophisticated technologies are used, from other storage-based and cache-based mechanisms to browser fingerprinting, which use specific attributes of computers and web browsers to recognize users again at their next website visit. So-called **evercookies** try to rebuild themselves after they have been deleted by users (see Bujlow et al 2015, p. 4).

Every click being transferred to 234 third parties

In an elaborate study of 50 of the most popular websites, the Wall Street Journal noted already in 2010, that nearly all of them, except *Wikipedia*, transferred user data to third parties. 37 of the 50 most popular websites transferred information about every click to over 30 third parties, 22 of them even to more than 60 third parties. The website dictionary.com transmitted data on every page request to 234 external services (see Wall Street Journal 2010). The third parties, which information about website visits was transferred to, were often largely unknown ad networks and web analytics services, but

---

[188] Orwell, George (1949): 1984. Secker and Warburg, London.
[189] Schwartz, John (2001): Giving Web a Memory Cost Its Users Privacy. New York Times, 04.07.2001. Online: http://www.nytimes.com/2001/09/04/business/giving-web-a-memory-cost-its-users-privacy.html?pagewanted=all [18.07.2016]
[190] https://w2.eff.org/Privacy/Marketing/web_bug.html [25.01.2016]
[191] https://www.mozilla.org/en-US/lightbeam/ [25.01.2016]

also prominent companies like Google, *Facebook* or *Twitter*. *Google*'s countless services such as *Google Analytics, DoubleClick* and *AdMob* are embedded in almost every website. *Facebook* is embedded at least in every website which contains a *Facebook Like* button.

Tracking the behavior of users surfing the web, from their searches to the sites that they visited, is still one of the most common ways to obtain rich information about their preferences, interests, problems, likes and dislikes. Web tracking and the companies who are receiving the information are further examined in chapter 5 which focuses on the business of personal data.

In recent years, other devices and platforms besides web tracking have evolved into rich sources of digital information about individuals. **Smartphones** and the **apps** installed on them transmit extensive information about everyday life to a wide range of companies. **Fitness trackers** are recording in-depth body and health data. Insurance programs based on recording **car driving behavior** could become prototypes for other fields of life, and in the **Internet of Things** surveillance becomes ubiquitous. The following chapter will explore those four areas of personal data collection.

4.1 Smartphones, mobile devices and apps – spies in your pocket?

"Location data, created all day long just by having a phone in your pocket, is probably the richest source of information in the world today"
Greg Skibiski, co-founder of Sense Networks, 2009 [192]

Wireless connections and sensors

Due to the rapid evolution of mobile technology and the introduction of Apple's iPhone in 2007, smartphones and installed applications became one of the most important gateways for companies to collect data on consumers. Smartphones offer various wireless connections for data transfer, including WLAN, GSM, UMTS, HSPA/3G, LTE/4G, Bluetooth and NFC (Rothmann et al 2012). In addition, smartphones are equipped with a variety of sensors. Besides microphones, cameras, GPS receivers and fingerprint sensors the Android developer guide lists three categories of sensors[193]:

- **motion sensors** measuring "acceleration forces and rotational forces along three axes" – accelerometers, gravity sensors, gyroscopes, rotational vector sensors
- **environmental sensors** – barometers, photometers, thermometers
- **position sensors** – orientation sensors magnetometers

1.4 billion new devices a year, 82.8% Android

According to Gartner, worldwide smartphone sales reached more than 1.4 billion devices a year in 2015[194], after 1 billion in 2013[195], and 472 million devices in 2011.[196] The market for operating systems is dominated by *Android* and *iOS* (according to IDC their market shares in Q2 2015 were 82.8% and 13.9% respectively).[197] Other platforms such as *Windows Phone* or *Blackberry* are representing niche existences. While iOS is being used only on *Apple devices, Android*, which was developed by the *Open Handset Alliance*[198] led

[192] http://www.wired.co.uk/article/the-hidden-persuaders-mining-your-mobile-phone-log [18.07.2016]
[193] https://developer.android.com/guide/topics/sensors/sensors_overview.html [18.07.2016]
[194] http://www.gartner.com/newsroom/id/3215217 [18.07.2016]
[195] http://www.gartner.com/newsroom/id/2996817 [18.07.2016]
[196] http://www.gartner.com/newsroom/id/1924314 [18.07.2016]
[197] http://www.idc.com/prodserv/smartphone-os-market-share.jsp [18.07.2016]
[198] http://www.openhandsetalliance.com

by *Google*, is able to serve as an operating system and software platform on devices of several manufacturers.

App permissions

Devices of both worlds provide a basic software installation, responsible for fundamental features such as phone calls, contact management, texting, photography and video. In addition, it is possible to install software by third-party developers, referred to as **apps**. In June 2016, 2.2 million apps were available for Android, and 2 million for iOS.[199] A permission system defines which sensors and stored data can be accessed by an app. Android for example presents a list of all permissions requested by an app before it is installed (e.g. access to contacts or to the current location). Without permitting access to all these functions the app cannot be installed. Within iOS, every installed app has certain standard permissions (e.g. accessing the internet), other access permissions (e.g. location, microphone or motion sensors) are requested by iOS at runtime when an app is trying to access the resource (see Kulyk et al 2016). Since version 6, *Android* also introduced "runtime permissions", overcoming the need for an app to request all permissions at once, when users install an app.[200]

Spy in your pocket?

Typically, a smartphone is used by a single person, and carried on the body more or less permanently. It is therefore seen as very personal and private device, which one would not want to hand to someone unknown (see Urban et al 2012). The information stored on such devices, including calls, text messages, contact lists, calendar, photos, videos, visited websites, the phone's location and motion behavior, provides **detailed insights into the user's personality and everyday life**. It is not only information about friends and family that is stored on such a device, but also work, finance and health contacts. Most of the time, mobile devices are connected to the Internet. Potentially, the integrated sensors can always be activated. Many users also store passwords on their smartphone, which provide access to personal user accounts such as email, social networks and e-commerce.

Smartphones entail several specific risks regarding users' privacy:

Privacy risks

- **Data security**: Unauthorized access to the device (e.g. through loss or theft) and security flaws within the OS and apps, which can be exploited by computer viruses, malware and for targeted attacks (see Lopes et al 2013).
- **Data transfer to app provider**: Storage, processing or transfer of personal data through apps of third-party developers, which can access different types of information stored on the phone as well as sensor data and send it to other companies (see chapter 4.2.1).
- **Data transfer to platforms or app store providers**: Most *Android* users are linking their device with a *Google* account. According to *Google* worldwide 1.4 billion *Android* devices were "active" at least once in a month in 2015[201]. Most likely this describes the number of monthly active *Android* users with a *Google* Account. Also, most iOS users connect their devices to an *Apple* account ("Apple ID"), as the device cannot be used appropriately without this. *Apple* didn't publish clear numbers about monthly active iOS users.
- **Mobile networks**: Information about the users' communication behavior is also stored by wireless communication service providers.

---

[199] http://www.statista.com/statistics/276623/number-of-Apps-available-in-leading-App-stores/ [18.07.2016]
[200] https://developer.android.com/training/permissions/requesting.html [18.07.2016]
[201] Alphabet (2016): Google Annual Report 2015. Online: https://abc.xyz/investor/pdf/2015_google_annual_report.pdf [18.07.2016]

4.2.1 Data abuse by apps

Mobile apps cover many areas of life and often depend on the access to certain information in order to fulfill their purpose. For example, a camera app has to be able to access the built-in camera, a navigation app needs location data to function and an address book app needs to access the address book. But many apps require access to sensors and data, which **isn't required for their functionality**. Required or not, there is a risk that apps transfer data to **third parties** without the users' knowledge.

Wall Street Journal app survey 2010

In 2010, a famous investigation of *Android* and *iOS* apps by the *Wall Street Journal* showed that **47 of the 100 most popular apps** transferred the phone's location not only to the developer, but also to third parties (see Thurm et al 2010). 56 of the assessed apps transmitted the unique device ID to third party companies without the users' awareness or consent, mostly to advertising networks. Developers of free apps were especially guilty of integrating tracking modules that exploit data for targeted advertising and other uses.

At the time of the *Wall Street Journal's* investigation the music app ***Pandora*** transferred age, gender, location and device ID to several advertising networks. The popular gaming app ***Angry Birds*** sent the users' address book, location and device ID to a third party. The dating app ***Grindr*** transferred gender, location and device ID to three advertising networks. The gaming app ***PaperToss*** sent location and device ID to five advertising networks, the texting app ***TextPlus*** sent the device ID to eight of these.[202]

Locating users every 30 seconds

According to another a U.S. study from 2010, **15 of 30 examined *Android* apps** transferred **location data** to advertising networks, again without notifying the users (see Enck et al, 2010). In some cases, the phone's location was transferred every 30 seconds. In one case it was sent directly after the installation, before starting the app even once. A further examination on 94 iOS apps from 2011 revealed that **84% of the apps** analyzed **connected to external domains** and 74% transferred the device ID.[203]

c't magazine survey 2012

According to an analysis by the German magazine *c't* from 2012, many of the **60 apps examined transmitted data to advertising networks**.[204] For instance, the popular app *Flashlight* sent data to five of them. The magazine also reported on the company *Onavo* which operates an online service that promises to reduce the cost of expensive mobile data transfer through a proxy server and data compression. At the same time, *Onavo's* apps transferred information about the phone's location, the frequency of the usage of specific apps and the websites visited by the user, to the app maker. In 2013 *Onavo* was acquired by *Facebook*.[205]

Accessing address books without consent

In 2012, it was also revealed that several iOS apps from social networks such as *Path*, *Foursquare*, *Twitter* and *LinkedIn* were uploading the user's address book without explicit consent, sometimes including names, email addresses, phone numbers and postal

---

[202] Thurm, S.; Kane, Y. (2010): Your Apps Are Watching You, Wall Street Journal. Online: http://online.wsj.com/article/SB10001424052748704694004576020083703574602.html [18.07.2016]
[203] Cortesi, Aldo (2011): How UDIDs are used: a survey, 19.05.2011.Online: https://corte.si/posts/security/apple-udid-survey/index.html [18.07.2016]
[204] Venne, Patrick Kollaten; Eikenberg, Ronald; Schmidt, Jürgen (2012): Self service shop smartphone. c't, book 7/2012, S. 114.
[205] Goel, Vindu (2013): Facebook Buys Israeli Maker of Data Compression Software for Mobile Web Effort, 14.10.2013, http://bits.blogs.nytimes.com/2013/10/14/facebook-acquires-onavo-and-a-foothold-in-israel [18.07.2016]

addresses.[206] After scandalizing reports in the media and an inquiry from the U.S. Congress, these apps were updated to ensure that users are informed about data sharing, and Apple introduced an explicit permission for apps to access contact data.[207]

Study on risky apps from 2014

Although data abuse by apps was often debated in the media, the situation seems to have become even worse. *Appthority* examines the reputation of apps for corporate use on a regular basis. In 2014, the 200 most popular apps of *Android* and *iOS* were analyzed for risky behavior patterns (see Appthority 2014):

Risky mobile app behaviors 2014	Top 100 apps / Android		Top 100 apps / iOS	
	free	paid	free	paid
Track user's location	82%	49%	50%	24%
Track users with device ID	88%	65%	57%	28%
Access address book	30%	14%	26%	8%
Share data with ad networks	71%	38%	32%	16%
Share data with social networks	73%	43%	61%	53%
Share data with analytics & SDKs	38%	20%	31%	41%

Table 13: Risky mobile app behaviors (Source: Appthority, 2014)

Sharing with data brokers

82% of free Android apps and 50% of free iOS apps were accessing **location data**, while nearly a third of the free apps on both platforms access the **address book**. Behind the scenes, many apps transfer data to **advertising networks** and **data brokers**, who in turn sometimes share the data collected with even more companies. Those apps didn't show ads to the user in every case. In some cases the app developers get paid according to the amount of information collected about the user. Surprisingly, paid apps are also transferring data to third parties. Many developers of apps use embedded frameworks, libraries or *Software Developer Kits (SDK),* which frequently collect detailed data about users and submit this data to analytics providers like *Google Analytics* or *Flurry* (see Appthority 2014).

59% of apps raising concerns

According to another study conducted in 2014 by 26 privacy enforcement authorities in 19 countries, **31% of 1,200 popular apps** access data that is not necessary for the app's functionality (see Office of the Privacy Commissioner of Canada 2014). 59% of the apps raised privacy concerns even before they were downloaded because they do not adequately inform the user about which data is used and shared.

Free Android and iOS apps in 2015

Apps sharing data with third parties

In a recent study, 110 popular *Android* and *iOS* apps where tested to discover which of them shared personal, behavioural and location data with third parties (see Zang et al 2015). Android apps sent sensitive data to 3.1 external domains on average, iOS apps to 2.6 third parties. Overall, sensitive data was shared with **94 distinct third-party domains**. 45 % of the 110 tested apps shared **email addresses** with third parties, 40% **location data**, and 34% the user's **name**. More in detail:

---

[206] Bohn, Dieter: iOS apps and the address book: who has your data, and how they're getting it. The Verge, 24.02.2012, http://www.theverge.com/2012/2/14/2798008/ios-apps-and-the-address-book-what-you-need-to-know [19.07.2016]

[207] Paczkowski, John: Apple: App Access to Contact Data Will Require Explicit User Permission. All Things Digital, 15.02.2012, http://allthingsd.com/20120215/apple-app-access-to-contact-data-will-require-explicit-user-permission [13.08.2014]

Data sent to third parties	All apps tested	Android apps	iOS apps
Email address	45%	73%	16%
Location data	40%	33%	47%
Name	34%	49%	18%
Username	20%	25%	15%
Gender	15%	20%	9%
Search terms	10%	9%	11%
Information on friends	11%	16%	5%
Job-related information	4%	4%	4%
Medical information	3%	2%	4%

Table 14: Sensitive data free mobile apps send to third parties (Source: Zang et al 2015)

Up to 17 third parties per app

73% of free Android apps shared identifying information such as email addresses with third parties, and 46% of iOS apps shared the phone's location. The tested apps sent sensitive information, including personally identifiable, behavioral and location data, to up to 17 third-party domains. The study also shows that a significant proportion of apps **share data from user inputs** with third parties – from search terms and information on friends entered by users during app usage to employment and health related information. Some examples:

Health and job-related data

- The app *Text Free* sent sensitive data to 11 third-party domains, with 9 domains receiving personally identifiable data and 6 receiving the user's location.
- The app *Local Scope* sent location data to 17 third-party domains.
- The *Drugs.com* app shared "**medical info input by the user**" with 5 third parties (e.g. words such as "herpes" or "interferon").
- The app *Period Tracker Lite* shared the "**input into a symptom field**" with one third-party domain.
- The Android apps *Job Search* and *Snagajob* shared "**employment-related search terms**" such as "driver," "cashier," and "burger" with four third-party domains.
- The iOS apps *Indeed.com* and *Snagajob* shared "**employment-related inputs**" such as "nurse" and "car mechanic" with 4 third parties.
- In general, apps listed in the categories "Health & Fitness" and "Communication" sent sensitive data to more third-party domains than apps in other app categories.

Encrypted data sharing not included

The study design has some **limitations**. For example, Zang et al didn't look at non-TCP traffic and they just tested for data leakages in clear text. It was not tested whether apps share simply encrypted versions of sensitive data, e.g. by using **common hashes like MD5**. It is very likely that many cases where apps share sensitive data with third parties were not discovered. Therefore, the observed numbers and percentages are in fact probably **higher than found**.

A study on free and paid apps from 2015

A study from 2015 on the top 100 free and paid apps in Australia, Brazil, Germany and the U.S. showed that tracking is less invasive in paid apps, but still very common (see Seneviratne et al, 2015). They found that around 60% of paid apps were "connected to trackers that collect personal information", compared to 85-95% of free apps. About 20% of paid apps were connected to more than 3 trackers.

App usage is tracked by many companies

By combining lists of installed apps from 338 different smartphone users with their research they found that 50% of these users were exposed to more than 25 trackers and 20% of them to over 40 trackers. Trackers were categorized as: "advertising" (e.g. *Google Ads, Millennial Media, Inmobi, Mopub*), "analytics" (e.g. *Flurry, Google Analytics, Comscore,*

Amazon Insights, Localytics, Kontagent, Apsalar) and "utilities" (e.g. *Crashlytics, Bugsense*). Overall they identified 124 different trackers[208] in 509 unique apps in Australia, Brazil, Germany and the United States. Many of these trackers were present in a high percentage of users' devices:

Tracker	Users affected
Google Ads	96%
Flurry	91%
Google Analytics	87%
Millennial Media	86%
Crashlytics	85%
Mopub	81%
Inmobi	79%
Hockeyapp	71%
Comscore	70%
Crittercism	68%

Table 15: Trackers users of smartphone apps are exposed to (Source: Seneviratne et al 2015)

Taken together, the study shows that "tracking behaviors of paid apps are almost the same as those of free apps". The researchers started to build *Privmetrics*[209], a "framework to secure user privacy in smartphones".[210]

How do users think about apps and privacy?

Users often have little knowledge and awareness about what information is accessible by apps. A study from 2012 showed the **difference between expectations and reality** on the basis of the 100 most popular Android apps.[211] 95% of 179 participants were surprised that the app *Brightest Flashlight* is accessing location data. 90% were surprised that the app *Background HD Wallpaper* is accessing their address book. On the other hand, nobody was surprised that *Google Maps* is accessing information about the phone's location. Overall, participants were startled about which apps are accessing the device ID, location data or the address book. Consequently, the authors of this study interpreted a small level of surprise as a form of "informed consent".

1,173,265 apps analyzed

Based on the study mentioned above and further research about mobile app privacy and usability (see Lin et al 2014), a team of researchers from *Carnegie Mellon University* created ***PrivacyGrade***,[212] an online platform about mobile apps and privacy, which measures "the gap between people's expectations of an app's behavior and the app's actual behavior"[213]. By July 2016, the site offered information on 1,173,265 mobile apps, whose privacy-related behaviors are summarized in the form of grades, using a scale of A+ to D. Additionally, detailed information is available for each app, including the permissions

[208] http://www.privmetrics.org/wp-content/uploads/2016/04/Tracker_List-11.xlsx [19.07.2016]
[209] http://www.privmetrics.org/
[210] Seneviratne, Suranga; Aruna Seneviratne, Johan Kestenare (2014): PrivMetrics: A Framework for Quantifying User Privacy in Smartphones. W3C Workshop on Privacy and User Centric Controls, Berlin, Germany, November 2014. Online: https://www.w3.org/2014/privacyws/pp/Seneviratne.pdf
[211] Lin, Jialiu; Sadeh, Norman M.; Amini, Shahriyar; Lindqvist, Janne; Hong, Jason I.; Zhang, Joy (2012): Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In: UbiComp, 2012. Online: http://www.winlab.rutgers.edu/~janne/privacyasexpectations-ubicomp12-final.pdf [07.07.2014]
[212] http://privacygrade.org
[213] http://privacygrade.org/faq

used by the app, a short description about why an app may need access to specific data, and the third-party libraries contacted by the app.

Flashlight app accessing location data

For example, the app profile page for *Flashlight - Torch LED Light* shows that it "appears this app uses" data on the phone's location for "market/customer analysis" and recorded audio via microphone for "delivering targeted advertisement".[214] According to the last analysis from April 2015, the third parties to whom data may be transferred include *Flurry*, *Facebook*, *Twitter*, *Chartboost*, *Inmobi*, *Millenial Media* and *Mopup*. An extra page with overall statistics on third-party libraries used by all apps analyzed is available.[215] For example, *Google*'s *Admob* was found in 407,181 apps, *Flurry* in 65,515 different apps.

Summary

Taken together, the use of today's smartphones and mobile apps is **deeply invading the privacy** of a substantial part of the world's population, consumers are often not aware of how many companies receive information about their everyday lives, and our knowledge about how apps collect data and transfer it to third parties is limited, incomplete, and often outdated.

4.2 Car telematics, tracking-based insurance and the Connected Car

"You know the way that advertising turned out to be the native business model for the internet? I think that insurance is going to be the native business model for the Internet of Things"

Tim O'Reilly, 2014 [216]

"We know everyone who breaks the law, we know when you're doing it. We have GPS in your car, so we know what you're doing."

Jim Farley, Head of Marketing and Sales at Ford, 2014 [217]

Driving cars is a kind of everyday life behavior, which has been digitally tracked for many years. What initially started as a technology used in freight and fleet management to make logistics more efficient (and to control employees)[218] also became common in consumer space.

Insurance rates based on driving behavior

Today, so-called **black boxes**, which monitor the vehicle around the clock and transmit information about its position, time, velocity, braking and acceleration values to several service providers, are increasingly being built into consumer cars. Terms like the **connected car** and the **smart car** are somehow tied to these developments. Customized insurance rates based on actual driving behavior are around since the mid-2000s (see Ptolemus 2016). As this kind of tracking and employing data about everyday life behavior may be a role model for other fields of life, it is worth taking a closer look.

---

[214] http://privacygrade.org/apps/com.rvappstudios.flashlight.html [20.07.2016]
[215] http://privacygrade.org/third_party_libraries/
[216] Myslewski, Rik (2014): The Internet of Things helps insurance firms reward, punish. The Register, 24.05.2014. cited on 19.09.2014 from http://www.theregister.co.uk/2014/05/23/the_internet_of_things_helps_insurance_firms_reward_punish
[217] Jim Farley told this to an audience at the Consumer Electronics Show in Las Vegas in January 2014. A week later he said that his statement was "hypothetical", see: http://www.businessinsider.com/ford-jim-farley-retracts-statements-tracking-drivers-gps-2014-1 [24.07.2016]
[218] See: Kanngieser, 2013

Plugging into the car

According to a report by the U.S. *National Association of Insurance Commissioners* (NAIC) most devices which record data about consumers' driving behavior are plugged into special interfaces such as the "on-board diagnostics" (OBD-II) port. They record information about **dates, times, locations and distances driven**, more sophisticated ones also report data about **speed, cornering, acceleration and braking**. There are different technologies available to track the driving behavior (see Karapiperis et al 2015):

- **Dongles** are devices provided by the insurer for a certain time. They are self-installed by the driver, record data on location and driving style and could soon be technologically obsolete.
- **Black boxes** are professionally installed and provide more detailed information. When accelerometers are integrated, they can also track speed, braking or harsh cornering – and they can use the cars sensors by plugging into its electronic control unit (ECU).
- **Embedded telematics** also directly connects to the vehicle's systems and can record a wide range of data on both the car and on driving behavior.
- **Smartphones and apps** are also increasingly used for car telematics, either standalone or plugged into the car's system. They provide a range of relevant sensors from accelerometers and gyroscopes to GPS and a network connection.

12 million customers worldwide

Usage-based insurance (UBI), which takes the recorded data into consideration for pricing purposes, is on the rise. According to an extensive industry report by consulting firm *Ptolemus* (2016), more than **200 insurance programs based on telematics** are available in 34 countries on five continents, covering 12 million customers. The number of customers in **Europe** increased from 2.1 million in July 2013 to 4.4 million in November 2015. The U.S.-based insurer ***Progressive* has 2.8 million UBI customers**. *Generali* has **800,000 UBI customers in Italy** and claims that 33% of new policies in Italy include telematics. But also several telematics programs are available in countries like the UK, Canada, France, Germany, Spain, Ireland, Russia and South Africa, and currently launched in many more countries from Columbia to China. Ptolemus predicts that "nearly 100 million vehicles will be insured with telematics policies" by 2020. However, according to another industry report, today's "market penetration is lower than predicted" and still <5% in the U.S.[219]

Risk ratings about drivers

Ptolemus (2016) differentiates between **Pay-As-You-Drive (PAYD)**, where insurance premiums are based on mileage, sometimes also on time and location data, and **Pay-How-You-Drive (PHYD)**, where insurance companies receive "driving style data" and calculate "risk ratings" about drivers. In general, most of today's vehicle insurance policies use information such as age, gender, vehicle age, place of residence, occupation and the customer's historical claims profile to calculate pricing.[220] Policies based on telematics add "new, dynamic parameters", which are recorded by **motion sensors** like accelerators, **GPS devices** or the car's sensors, and are automatically transmitted to the insurance company or telematics providers. The information **recorded and analyzed** ranges from the distance travelled, the day of the week and the time of the day, the average length of the trips, the type of the road to the driving behavior, including acceleration, braking, speed and cornering.

---

[219] Novarica (2016): Telematics in Insurance: Current State of UBI and Market Challenges. July 2016. Summary online: http://novarica.com/telematics-2016/ [24.07.2016]

[220] Insurances may not be allowed to use certain parameters in certain regions, e.g. gender in the EU: http://ec.europa.eu/justice/newsroom/gender-equality/news/121220_en.htm [24.07.2016]

Progressive's telematics-based insurance offer in the U.S.

20% discount for safe drivers, 10% increase for risk drivers

Customers who decide to participate in *Progressive*'s popular *Snapshot* program in the U.S. receive a small device, which they plug into their car's OBD port. It records the vehicle's speed, time information, and "in some devices" also "G force".[221] Information about driving behavior is transmitted wirelessly "to and from Progressive", including the **Vehicle Identification Number**. *Progressive* then calculates **a score for each driver**, who can receive a personalized discount on their insurance premium based on their driving habits.[222] The details vary by state.[223] In addition to discounts they started **penalizing "bad" driving behavior** in some states in 2015.[224] In Ohio, drivers can get a maximum 20% "discount for safer driving habits", and a maximum 10% "increase for risker habits".[225]

Calculating scores on safe or risky driving behavior is based on the following parameters:[226]

Behavior	Description (according to Progressive)
Hard braking	Hard brakes are decreases in speed of seven mph per second or greater. Your Snapshot device will "beep" when you brake hard. Minimize hard braking to work toward a discount.
Amount of time driven	The number of minutes that your engine is running during a trip. To earn a discount, try to minimize your time behind the wheel by combining trips, carpooling or using public transportation.
Time and day	The number of minutes you spend driving during higher risk hours—the highest risks are between midnight and 4 a.m. on the weekends.
Fast starts	Fast starts are increases in speed of nine mph per second or greater. Also known as "jackrabbit starts" or just "putting the pedal to the metal." Use a lighter foot on the gas pedal to work toward a discount.
Trip regularity	The frequency with which you drive at the same time of day and same duration.

Avoid hard braking and late-night drives

Table 16: How Progressive is scoring safe or risky car driving behavior in Ohio. Source: Progressive (2016)

Originally, Progressive had also included location and GPS data, but later they removed them. In 2016, *Progressive* was thinking about including location data again.[227]

Discovery's telematics-based insurance in South Africa

The leader in health plans involving fitness trackers

The South African company *Discovery* offers another usage-based insurance product called *VitalityDrive*.[228] Drivers who participate can earn points based on the monitoring of their driving behavior and other parameters. When they behave according to the system's algorithmic rules, they can get discounts and rewards such as a refund of up to 50% of their "BP fuel and Gautrain spend". Data are recorded by a telematics device, but

[221] https://www.progressive.com/auto/snapshot-terms-conditions [24.07.2016]
[222] https://www.progressive.com/auto/snapshot-terms-conditions [24.07.2016]
[223] https://www.progressive.com/auto/snapshot-details [24.07.2016]
[224] Passikoff, Robert (2015): Progressive Adds 'Bad Driver' Surveillance To Snapshot Telematics. Forbes, 31.03.2015. Online: http://www.forbes.com/sites/robertpassikoff/2015/03/31/progressive-adds-bad-driver-surveillance-to-snapshot-telematics/ [24.07.2016]
[225] Progressive (2016): Everything you want to know about Snapshot. Ohio. February 20, 2016 to Present. Online: https://www.progressive.com/auto/snapshot-details/ [24.07.2016]
[226] Ibid.
[227] Scism, Leslie (2016): Car Insurers Find Tracking Devices Are a Tough Sell. The Wall Street Journal, 10.01.2016. Online: http://www.wsj.com/articles/car-insurers-find-tracking-devices-are-a-tough-sell-1452476714 [24.07.2016]
[228] https://www.discovery.co.za/portal/individual/insure-vitality-drive [24.07.2016]

customers can also opt in to a **smartphone-enabled plan** to earn "more" points.[229] In that case, *Discovery*'s mobile app additionally "uses accelerometer, gyroscope and GPS data" to measure driving behavior.

The following graphic shows how drivers participating in *VitalityDrive* can earn points:[230]

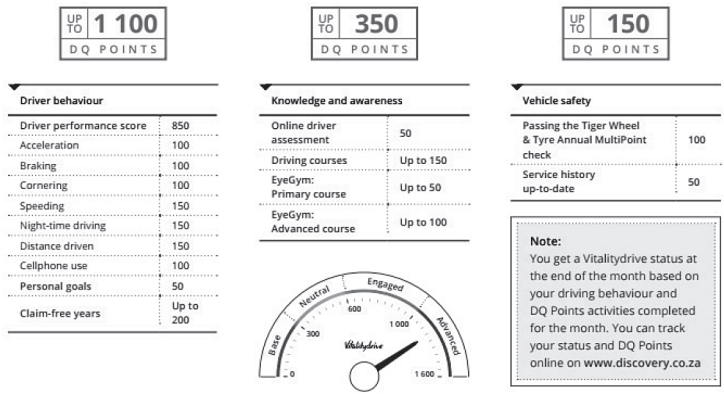

Figure 4: Earning points for desired behavior at Discovery's usage-based insurance offer. Source: Discovery.

Driver performance score

Each month, a "driver performance score" is calculated. Drivers receive less than maximum when they either drive "10km/h over the speed limit" too often, when the distance travelled is too far, when the "number of harsh accelerations, harsh brakes and harsh corners" is too high, or when they drive during nighttime too often.[231] Additionally, *Discovery* regularly appoints "**personal goals**" for each driver. When reaching these goals, drivers can earn additional points. As part of a special program, young adults aged up to 26 can choose **a more sophisticated regime**. They get a 25% refund on their insurance premium, payable "in cash" every six months. However, after six months, they may have a "**premium increase**" of 10% when they drive more than 50 kilometers during the night, and an increase of 25% when they have more than 200 "monthly average night-time kilometers".[232]

Scoring for rating and underwriting

Discovery states that it "will not use" the data recorded by their tracking device "in the event of a **claim**, other than to confirm the time and place of an incident". But, according to their privacy policy, customers consent to the use of "scoring information for rating and underwriting purposes".[233] The company is operating **similar programs in the field of health**. Its *Vitality* program is a global leader in health insurance and corporate wellness programs that integrate data from on-body trackers, point-based reward systems and provide discounts on premiums (see chapter 4.3.4).

---

[229] https://www.discovery.co.za/portal/individual/insure-faqs [24.07.2016]
[230] Discovery (2016): Vitalitydrive terms and conditions. Online: https://www.discovery.co.za/discovery_coza/web/linked_content/pdfs/insure/vitalitydrive.pdf [24.07.2016]
[231] https://www.discovery.co.za/portal/individual/insurance-news-mar15-driver-performance-score [24.07.2016]
[232] Discovery (2016): Vitalitydrive terms and conditions.
[233] Discovery Insure Limited driving/mobile application privacy policy and terms of use. Online: https://www.discovery.co.za/discovery_coza/web/linked_content/pdfs/insure/discovery_insure_driving_terms_and_conditions.pdf [24.07.2016]

Allianz's telematics-based insurance in Germany

Discounts up to 40%

German insurer *Allianz* also introduced a telematics-based program in 2016.[234] Data is recorded with a combination of a Bluetooth beacon and a smartphone app. Customers can get a discount of up to 40% on their premium when they drive according to the system's desired behavior. Scores are calculated based on the components **hard braking** (30%), **fast starts** (20%), **harsh cornering** (20%), **exceeding speed limits** (10%) as well as from **day, time and type of street** (20%). *Allianz* states that for example "driving in the city during the rush hour" would involve a higher risk than "driving on the highway on Sunday morning".[235]

Concerns about privacy, transparency and discrimination

Major concerns

It is undisputed that strengthening careful driving and reducing risky driving behaviour is very desirable for society. Vehicle telematics offer many additional opportunities, for example better "remote diagnostics, roadside assistance, emergency response and stolen vehicle location services" (Karapiperis et al 2015). However, in a report from the *U.S. National Association of Insurance Commissioners (NAIC)* major concerns are raised:

- NAIC states that "insurers have turned telematics into just **another black box rating factor**, like credit scoring but without even the limited protections afforded consumers for insurers' use of consumer credit information".

Credit scoring for car drivers

- Insurers may use and distribute the recorded data "for **purposes other than loss mitigation and pricing**, including, for example, insurers using information from telematics in claim settlements when helpful to insurers but not making the data available to consumers when helpful to consumers".
- Usage-based insurance may lead to a "[d]isproportionate impact of offer and sale" against "**low- and moderate-income and minority communities**".
- Insurers may use telematics data as "merely another data mining exercise following on insurer use of credit information – including **penalizing consumers** not because of driving behavior but because of where and when they drive as a function of work and housing segregation".

A prototype for other areas of life

The criteria of how specific driving behavior is rewarded or penalized are **arbitrary in general and nontransparent in detail for drivers**. The algorithms used to calculate the resulting scores are mostly secret. Penalizing behaviors like harsh braking could even be **dangerous**, because it is very unlikely that these systems can reliably differ between required and willful harsh braking. When driving during the night or in the city is penalized in general, the individual's freedom of action gets restricted. In the long run, concepts like this may result in **corporate governance of everyday life**, where citizens are controlled by private companies, especially when similar practices are adopted in other fields of life such as healthcare.

The „new normal"?

Participation in this "usage-based insurance surveillance", as it has been called by Robert Passikoff in *Forbes*[236], is clearly **voluntary** today. But one of the main concerns is that it could become **mandatory**, either because insurance companies could completely drop

---

[234] https://www.allianz.de/auto/kfz-versicherung/telematik-versicherung [24.07.2016]
[235] Translation by the authors, original: „Eine Fahrt während der Rushhour in der Stadt birgt zum Beispiel größere Gefahren als eine Sonntagsausfahrt vormittags auf der Autobahn.", https://www.allianz.de/auto/kfz-versicherung/telematik-versicherung [24.07.2016]
[236] Passikoff, Robert (2015): Progressive Adds 'Bad Driver' Surveillance To Snapshot Telematics. Forbes, 31.03.2015. Online: http://www.forbes.com/sites/robertpassikoff/2015/03/31/progressive-adds-bad-driver-surveillance-to-snapshot-telematics/ [24.07.2016]

offers not based on telematics, or because offers not based on telematics become non-affordable. A publication by the consulting giant *Ernst & Young* asked whether usage-based insurance could already be the "new normal".[237] Regarding Pay-As-You-Drive (PAYD) models, the author asks: "Why stop there?", and suggests to introduce "**Pay-As-You-Live** (PAYL)" for life and health insurance solutions based on surveillance and personalized pricing.

Scores are relevant data

Today's insurers, who are offering products based on tracking and scoring driving behavior, are emphasizing that the **raw data recorded is stored by separate service providers** which they don't have access to. They only receive the calculated scores about how safe or risky people drive. However, one could argue that those scores are actually the relevant information, not the raw data recorded. Why shouldn't companies in other business fields ask people to "voluntarily" consent to provide this information to them through incentives such as rewards or discounts? In addition, telematics providers are sometimes part of larger corporate players, who are active in the personal data ecosystem in several ways. For example, *LexisNexis*, a large provider of solutions in risk management and scoring based on data about 500 million consumers, owns *Wunelli*, a large telematics service provider (see chapter 5.7.5).

Privacy risks on several levels

When it comes to the **connected car,** much more aspects are relevant. Today's automobiles are full of information technology – from sensors and cameras to network connections. Many existing or upcoming features depend on data, such as brake assistance, traction control, collision avoidance and video-based obstacle and pedestrian detection to systems that monitor the drivers' attentiveness. Last but not least, the **autonomous car** is on the rise, but this is beyond the scope of this report.

The *Canadian B.C. Freedom of Information and Privacy Association* carried out an extensive study on "privacy and onboard vehicle telematics technology" (see FIPA 2015):

- They state that the "connected car" is becoming a "**major new source of data about individual drivers**". The customer data generated is "now seen as a major new source of revenue" for many parties. The privacy risks are "amplified in an industry ecosystem characterized by multiple players" who are all "vying for a piece of the data pie". At the same time, data provided by telematics and vehicle infotainment systems is "**highly revealing of personal lifestyles, habits and preferences**", especially when "tracked, combined or linked with other available data".

Many parties are interested

- The **parties interested in telematics data** include not only automakers and their partners, but also car dealers, insurance companies, lenders, telematics service providers, call center operators, third-party app developers, vehicle infotainment content providers, mobile network operators, and mobile device system providers such as *Google* and *Apple*. Also, many third parties **outside the telematics industry** are interested, including local retailers and merchants, online advertising agencies, data brokers, law enforcement agencies, debt collectors, fraud investigators, litigants and many more.
- In addition, telematics has become a "standard tool by which businesses and others manage their automotive fleets" from delivery trucks to taxis, rental cars or company cars. Thus, these technologies are also used "for **detailed monitoring of employees** using company vehicles".

[237] Walter Poetscher (2015). Usage Based Insurance. The New Normal? EY, July 2015. Online: http://www.ey.com/Publication/vwLUAssets/EY-usage-based-insurance-the-new-normal/$File/EY-usage-based-insurance-the-new-normal.pdf [24.07.2016]

Who will have access?

Finally, FIPA asks: **Who will own the data** generated by telematics? Which of the companies will have access, and to what extent will data be available to third parties?

4.3 Wearables, fitness trackers and health apps – measuring the self

"Smart devices are constantly collecting information, tracking user habits, trying to anticipate and shape their owners' behaviors and reporting back to the corporate mother ship"
Javob Silverman, 2016[238]

What had been common just for chronically ill patients and top athletes a few years ago, increasingly became a daily routine for broad sections of the population: the optimization of the self through to the continuous measuring of activity, vitality and body functions – with different tools from mobile apps and portable devices to "smart" scales. Terms like the ***Quantified Self***[239], *self-tracking* or *life-logging* (see Almalki et al 2015, Crawford et al 2015, Lupton 2016) describe a variety of approaches and products for the collection, analysis and evaluation of body and health information.

Body and health data

Most activity and fitness trackers **record the number of steps** taken while walking or running, GPS location data and pulse rate. Often, tracking of the **duration and quality of sleep** is added. Most products offer the measurement and improvement of sports activities, weight loss or eating habits – some of them also of the **menstrual cycle**[240], **alcohol and nicotine** consumption[241] or even **mood or mental well-being**[242]. Portable "wearable" devices such as wristbands and smartwatches are typically carried on the body, but also standard smartphones can be used in a similar way, for example together with armbands or other holding mechanisms. All these devices measure the body activity with several sensors, most importantly with **sensors that recognize the directions and intensity of movements** (see Su et al 2014), for example:

Sensor Type	Measuring
Accelerometer	acceleration force applied to the device
Gravity sensor	force of the gravity applied to the device, in three axes (x; y; z)
Gyroscope	orientation of a device in pitch, roll and yaw
Magnetometer	ambient geomagnetic field in three axes (x; y; z)
Barometer	ambient air pressure

Table 17: A set of sensors used for activity recognition (Source: Su et al, 2014)

Sensors, algorithms and charts

Other sensors used to track activity and fitness are **optical sensors** (e.g. optical heart rate monitoring, cameras), **temperature sensors**, wearable electrodes, chemical sensors, and

238 http://www.nytimes.com/2016/06/19/magazine/just-how-smart-do-you-want-your-blender-to-be.html [22.08.2016]
239 http://quantifiedself.com
240 Weigel, Moira (2016): 'Fitbit for your period': the rise of fertility tracking. The Guardian, 23.03.2016. Online: https://www.theguardian.com/technology/2016/mar/23/fitbit-for-your-period-the-rise-of-fertility-tracking [20.07.2016]
241 Schumacher, Florian (2014): Wearables that help cope with addiction. Wearable Technologies, 19.08.2014. Online: https://www.wearable-technologies.com/2014/08/wearables-that-help-to-cope-with-addiction/ [20.07.2016]
242 Medical Xpress (2016): New mental health app helps track moods and promotes emotional self-awareness. 20.04.2016. Online: http://medicalxpress.com/news/2016-04-mental-health-app-track-moods.html [20.07.2016]

stretch and pressure sensors.[243] The raw values measured by the devices are, more or less accurately[244], converted into steps, distances walked, calories burned or sleeping quality by **software algorithms**. Results, including statistics and graphs, are accessible by users via the provider's web platforms and mobile apps. The data measured can typically be extended with **information manually entered** by the user (see Crawford et al 2015), for example details about eating habits, gender, age, body height, weight, blood pressure or blood sugar.

Tracking many areas of life

The website quantifiedself.com[245] lists a database of more than 500 self-tracking tools which are not restricted to fitness, exercise and physical health. Many areas of life are covered, from apps for tracking and improving **personal growth, psychological wellness, meditation, relationships, sexual activity, cognitive performance, work productivity and personal finance** to the detailed analysis of one's own online or social media behavior. All this is mostly done with the goal of digitally storing and analyzing as much information about daily life as possible. In a broader sense, social media and blog platforms such as *Facebook*, *Twitter*, *Instagram* and *Tumblr* are sometimes also used under paradigms of life-logging and self-tracking.

24/7 monitoring

One important element of self-tracking apps and platforms is that they make the raw data collected **accessible for users** via personalized reports, tables, charts, diagrams and interactive infographics. Usually these platforms also try to motivate users to **utilize the devices as much as possible** in order to achieve significant results, for example in optimizing one's body in terms of beauty norms or health aspects. Munson and Consolvo (2012) summarized goal-setting, rewards, reminders and sharing as four main strategies used to motivate physical activity.

Set goals, get rewards and share

Most self-tracking apps offer possibilities to define **goals and target values**, for example a certain running distance per week or a certain number of steps to achieve. Achievements are **rewarded with activity points, trophies and virtual badges**. The road to success is visualized as a progress bar. Often, users are encouraged to **share results** and successes with others. Many apps allow users to share their successes on social media platforms or to publish it on a public profile site. In some cases this is even the default setting.

From a small community to mainstream

In 2007, only a small community of people exchanged information about self-measurement and the tools needed on platforms like quantifiedself.com. Its co-founder Gary Wolf published his manifest "The Data-Driven Life"[246] in 2010, and is considered to be a thought leader of the movement. In 2015, according to IDC, a total of 79 million wearables have been shipped. The global market for wearables is dominated by ***Fitbit*** (26.9%), ***Xiaomi*** (15.4%), ***Apple*** (14.9%), ***Garmin*** (4.2%) and ***Samsung*** (4%). While *Apple* focuses on its smartwatch, *Xiaomi* offers "inexpensive fitness trackers" with "prices

[243] Hayward, James and Guillaume Chansin (2016): Wearable Sensors 2016-2026: Market Forecasts, Technologies, Players. Online: http://www.idtechex.com/research/reports/wearable-sensors-2016-2026-market-forecasts-technologies-players-000470.asp [20.07.2016]
[244] Lee JM; Kim Y, Welk GJ (2014): Validity of consumer-based physical activity monitors. Med Sci Sports Exerc. 2016 Aug;48(8):1619-28, DOI: 10.1249/MSS.0000000000000287.
[245] http://quantifiedself.com/guide [20.07.2016]
[246] Wolf, Gary (2010): The Data-Driven Life. New York Times, 02.05.2010. Online: http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html?pagewanted=all [20.07.2016]

far below the competition".[247] IDC predicts wearables to reach 111 million units by 2020, half of it smartwatches and about a third of it wristband trackers.[248]

Health and fitness apps

Apart from wearable devices, **thousands of smartphone apps** are available which either connect to the software platforms of hardware vendors via APIs or use the sensors of today's smartphones. A report[249] by a market research company from 2014 estimates the "number of monthly active users who track at least one health & fitness parameter" is approximately 100 million people. According to Nielsen, 46 million U.S. consumers "accessed apps in the fitness and health category in January 2014". That is around one-third of U.S. smartphone owners.[250]

Health and fitness apps

By July 2016, **health and fitness apps with more than 10 million downloads** listed in *Google*'s app store include apps from companies and brands such as *MyFitnessPal, Runkeeper, Nike+, Runtastic, Pedometer*, *Endomondo* and *Azumio* – some of them offering multiple apps. Other apps with more than 10 million downloads include for example "Calorie Counter by FatSecret", "My Diet Coach - Weight Loss", "Period Tracker, My Calendar" and "Pregnancy & Baby Today". Many popular fitness and health apps have been acquired by larger companies during the last few years. *Runkeeper* was bought by *Asics* and *Runtastic* by *Adidas*. *Endomondo, MyFitnessPal and MapMyFitness* were acquired by *Under Armour*.[251] *Moves* was bought by *Facebook*.[252]

The large players in today's digital economy like *Google*, *Apple* and *Samsung* have also started to offer apps and platforms for health and fitness data.[253]

4.3.1 A step aside – gamification, surveillance and influence on behavior

Most fitness apps in recent years are based on functionality, which has been frequently discussed under the term of "Gamification", i.e. the "use of game design elements in non-game contexts" (Deterding et al 2011) **to influence user's behaviors** (see Whitson 2013)

---

[247] IDC (2016): The Worldwide Wearables Market Leaps 126.9% in the Fourth Quarter and 171.6% in 2015, According to IDC. Press Release, 23.02.2016. Online: http://www.idc.com/getdoc.jsp?containerId=prUS41037416 [20.07.2016]

[248] IDC (2016): IDC Forecasts Wearables Shipments to Reach 213.6 Million Units Worldwide in 2020 with Watches and Wristbands Driving Volume While Clothing and Eyewear Gain Traction. Press Release, 15.06.2016. Online: http://www.idc.com/getdoc.jsp?containerId=prUS41530816 [20.07.2016]

[249] research2guidance (2014): mHealth App Developer Economics 2014. The State of the Art of mHealth App Publishing. Fourth annual study on mHealth app publishing. May 6, 2014. Online: http://research2guidance.com/r2g/research2guidance-mHealth-App-Developer-Economics-2014.pdf

[250] Nielsen (2014): HACKING HEALTH: HOW CONSUMERS USE SMARTPHONES AND WEARABLE TECH TO TRACK THEIR HEALTH. 04-16-2014. Online: http://www.nielsen.com/us/en/insights/news/2014/hacking-health-how-consumers-use-smartphones-and-wearable-tech-to-track-their-health.html [21.07.2016]

[251] Goode, Lauren (2016): The age of indie fitness apps is over. The Verge. Feb 12, 2016. Online: http://www.theverge.com/2016/2/12/10978234/fitness-app-brand-takeover-runkeeper-under-armour-adidas-fitbit [21.07.2016]

[252] Dredge, Stuart; and Alex Hern (2014): Facebook buys fitness-tracking app Moves. The Guardian, 24.04.2014. Online: https://www.theguardian.com/technology/2014/apr/24/facebook-buys-moves-fitness-tracking-health-data [21.07.2016]

[253] Gibbs, Samuel (2014): Google launches Fit app to take on Apple's Heath and Samsung's S Health. Guardian, 29.12.2014. Online: https://www.theguardian.com/technology/2014/oct/29/google-launches-fit-app-apple-heath-samsungs-s-health [01.08.2016]

and **to increase participation and engagement**[254]. Therefore, more or less complex rule sets are implemented, which are complemented by mechanisms that incentivize and reward desired behavior, or more rarely, penalize non-desired behavior. An industry guide formulated by Oracle in order to help business organizations to "design and implement a successful gamification project"[255] suggests four main categories of game mechanics[256]:

Rewarding desired behaviors

- **Feedback mechanisms:**
  These mechanisms reward users for their performance, for example through **points** ("awarded for an action or a combination of actions"), **levels** ("reward those accumulating points" and "reflect that a user is improving or continuing to show the desired behavior" to "motivate users" or "unlock content"), **badges** (a "highly visible, social aspect of gamification" to "reward users for specific behaviors" and make them "show their statuses to others"), **bonuses** ("extra rewards for completing a set of actions" which "serve a similar function to bonuses awarded at work") and **notifications** (to "alert users of changes in their statuses", including "when they have earned points, badges, and bonuses").

- **Indicator mechanisms:**
  These mechanisms define a "user's relative position" in time or in relation to other users, for example **countdowns** ("give users some sense of urgency" to "increase activity" or to "trigger an action for a user who hasn't committed to an action"), **progress indicators** ("help users understand where they are in the system and how much farther they have to go" to "get users to continue interactions within the system") and **leaderboards** ("list top performers in particular areas"; better "show the user's position relative to those closest to them in scores" or "create different groups" instead of displaying the overall "top 5-10").

- **Game design mechanisms:**
  Oracle suggests the use of game design mechanisms for "larger goal and reward states" and "longer-term engagements". These mechanisms include **quests, missions and challenges** ("involve completion of a set of actions that follow a particular order or path" to "motivate users to complete particular sets of activities"), **competitions** ("events that encourage rivalry for some prize, honor, or advantage" to "motivate users by having them compete against each other to achieve some goal or objective") and **virtual economies** ("enable users to trade on their successes", to trade "something gained in the system" (e.g. points) for "goods or services").

- **Psychological mechanisms:**
  These mechanisms "take advantage of the ways that people think about situations they encounter", for example **loss aversion** ("refers to people's psychological tendency to evaluate potential losses as larger and more significant than equivalent gains" to "encourage participation among infrequent or inactive users" or to "get users to act by suggesting that something is available for only a limited time", e.g. "only three airline

---

[254] Fitz-Walter, Zachary; Tjondronegoro, Dian (2011): Exploring the Opportunities and Challenges of Using Mobile Sensing for Gamification. In: UbiComp 11: Proceedings of the 2011 ACM Conference on Ubiquitous Computing, ACM Press, Beijing, pp. 1-5. Online: http://eprints.qut.edu.au/48632
[255] Oracle Gamification Guidelines. Online: http://www.oracle.com/webfolder/ux/Applications/uxd/assets/sites/gamification/index.html [20.07.2016]
[256] Phase 3: Select Gamification Elements. In: Oracle Gamification Guidelines. Online: http://www.oracle.com/webfolder/ux/Applications/uxd/assets/sites/gamification/phase_3.html [20.07.2016]

seats left") and **appointment dynamics** ("require users to access the system or flow or take some action at a particular time or place for either a positive effect or to avoid a negative effect", e.g. "when a game rewards players for returning to the game regularly and punishes users who don't return at specific intervals").

Game mechanics from marketing to health & work

Using such elements of game design in other contexts than games is not new. **Well-known examples** include classroom grades, Boy Scout badges, happy hour drink specials or loyalty points (see Whitson 2013). Oracle's list is by far not complete, but many of these mechanisms can be found in today's **mobile and web apps**, especially in fitness apps. Oracle's gamification guidelines seem to be written for all kinds of use cases in business, from online sales and customer retention to employee performance management. Indeed, such game mechanisms are increasingly used in many fields from **marketing and sales to education, health and work**. Major social media platforms also use elements of gamification: The number of friends and Likes on *Facebook*, the number of followers and tweets on *Twitter*, badges on *Foursquare* and many more.

Gamification…

According to *Oracle*, it is "essential to analyze how the gamification system is doing", to "use analytics and to track performance" and to "measure, track, aggregate, and report the gamification system data" in order to "determine whether the specific game mechanics are altering user behavior to the degree that you hope".[257]

…and surveillance

Jennifer Whitson (2013) argues that today's technology-based practices of gamification are "rooted in surveillance" because they provide "real-time feedback about users' actions by amassing large quantities of data". According to her, gamification is "reliant on quantification", on "monitoring users' everyday lives to measure and quantify their activities". **Gamification practices based on data collection and quantification** are "leveraging surveillance to evoke behavior change", having as objectives for example "weight loss, workplace productivity, educational advancement, consumer loyalty". While **self-quantification** promises to "make daily practices more fulfilling and fun" by adopting "incentivization and pleasure rather than risk and fear to shape desired behaviours", it also became "a new driving logic in the technological expansion and **public acceptance of surveillance**".

4.3.2 Example: Fitbit's devices and apps

Fitbit was founded in 2007 and is now, according to IDC, the **global market leader in wearables** with a market share of 26.9% and 21 million units shipped in 2015.[258] The U.S.-based company offers a variety of fitness and activity trackers from wristbands and watches to a "smart" scale. Revenue was about $ 1.8 billion in 2015.[259]

Recording activity

Fitbit's activity trackers use sensors such as an accelerometer to record the "frequency, duration, intensity, and patterns of movement" of users to determine "**steps taken, distance traveled, calories burned, and sleep quality**".[260] When data from the devices

[257] Phase 5: Tracking and Analyzing the Progress of a Gamification System In: Oracle Gamification Guidelines. Online: http://www.oracle.com/webfolder/ux/Applications/uxd/assets/sites/gamification/phase_5.html [20.07.2016]
[258] IDC (2016): The Worldwide Wearables Market Leaps 126.9% in the Fourth Quarter and 171.6% in 2015, According to IDC. Press Release, 23.02.2016. Online: http://www.idc.com/getdoc.jsp?containerId=prUS41037416 [20.07.2016]
[259] See Fitbit (2016): Annual report 2015
[260] https://help.fitbit.com/articles/en_US/Help_article/1143 [21.07.2016]

is synchronized with online dashboards and mobile apps, it is transferred to *Fitbit*'s servers that are located in the United States.[261]

Track yourself and reach goals

Most of *Fitbit*'s devices track "daily steps, calories burned, distance traveled, and active minutes" as well as "floors climbed, sleep duration and quality", some of them also gather "**heart rate and GPS-based information** such as speed, distance, and exercise routes".[262] According to their *Product Manual*[263], users can also enter **body size and weight, birthdate and gender** as well as manually track **mood, allergies, blood pressure, glucose and food.** Every single meal can be entered. Users can also add "custom trackers" to track "anything" they "want", for example "cigarettes, push-ups, beers". [264] Based on all recorded and manually entered data, different reports, graphics and diagrams are generated. Users can set **goals** (e.g. weight loss), set up **fitness plans** or earn **badges**, e.g. for 10,000 daily steps or a "lifetime distance" of 250 miles. Progress bars are used to visualize how much activity is still needed in order to reach the defined goals.[265]

Competing with friends and others

In addition, many functions for **social networking** are integrated into the software. Users get a **profile page**, including a picture and information about their activities such as badges, steps, distances, calories burned or sleeping duration. Depending on the user's privacy settings, this profile page may also be **publicly available**. Several features are motivating users to compete with "friends", who can be invited via *Facebook* and email, and with other *Fitbit* users in forums and groups.[266] Activities can be **shared** via *Facebook*, *Twitter* and with "thousands"[267] of other apps, for example with *Microsoft*'s health data platform *HealthVault*[268], with an account at *Weight Watchers* or with the popular fitness app *MyFitnessPal*.[269]

Fitbit sells "de-identified" data

According to *Fitbit*'s **privacy policy**[270], they "may share or sell aggregated, de-identified data", personally identifiable information may be disclosed or transferred in "connection with the sale, merger, bankruptcy, sale of assets or reorganization of our company". It is not clear, how data is de-identified, and whether unique identifiers such as "hashed" email addresses are seen as de-identified (see Chapter 5.6).

>10 third-party companies receive data

In its additional **cookie policy**[271] *Fitbit* mentions a list of third-party companies, whose services are integrated with *Fitbit* and who certainly somehow receive data based on the interactions of users: *AppNexus*, *DataXu*, *DoubleClick* (*Google*), *DoubleClick Floodlight* (*Google*), *Google Adwords Conversion*, *AdRoll*, *Twitter Advertising*, *LiveRamp* (*Acxiom*), *Advertising.com* (*AOL*), *Bidswitch*, *Facebook Custom Audiences*, *Genome* (*Yahoo*), *SearchForce*, *MixPanel , Google Analytics, New Relic, KissInsights* and *Optimizely. Fitbit*

---

[261] Fitbit (2014): Privacy Policy. Last updated December 9, 2014. Online: http://www.fitbit.com/legal/privacy-policy [21.07.2016]
[262] Fitbit (2016): Annual report 2015
[263] Fitbit Tracker Product Manual. Online: https://www.fitbit.com/manual [21.07.2016]
[264] Despite prominently featured in Fitbit's "Product Manual" it is, according to a post in Fitbit's community forums, no longer possible to log "Blood Pressure, Custom Trackers, Body Measurements, Heart Rate, Journal, Glucose" since August 2015: https://community.fitbit.com/t5/Fitbit-com-Dashboard/Old-manual-logging-pages-will-be-retired-on-8-31/m-p/894230#U894230 [21.07.2016]
[265] Fitbit Tracker Product Manual
[266] Fitbit Tracker Product Manual
[267] Fitbit (2016): Annual report 2015
[268] https://www.fitbit.com/user/profile/share/healthvault [21.07.2016]
[269] http://www.fitbit.com/apps [21.07.2016]
[270] Fitbit (2014): Privacy Policy. Last updated December 9, 2014. Online: http://www.fitbit.com/legal/privacy-policy [21.07.2016]
[271] Fitbit (2014): Cookie Policy. Last updated December 9, 2014. Online: http://www.fitbit.com/legal/cookie-policy [21.07.2016]

provides links to several other privacy policies, makes recommendations such as "we encourage you to read the Google Privacy Policy" and mentions that interactions with "social media tools, like widgets and plug-ins" are "governed by the privacy policy of the company providing them, not by Fitbit's Privacy Policy".

4.3.3 Transmitting data to third parties

Apart from manifold issues with data security on many levels[272] one of the main concerns since the introduction of fitness trackers is that at some point, the recorded health data could be accessed and analyzed by data brokers or even by insurance companies and employers. As summarized in the precedent chapter, *Fitbit* may transmit data to more than 10 third-party companies including *LiveRamp*, a subsidiary of the data broker *Acxiom*.

Ad networks and marketing data brokers

A study from 2013 analyzed 43 popular *Android* and *iOS* health and fitness apps. They found that **39% of free apps and 30% of paid apps** sent data to third parties not mentioned in the app or in any privacy policy. 43% of free apps and 5% of paid apps shared personally identifiable information with advertisers (see Ackerman 2013). In 2014, the IT security firm *Symantec* found that popular self-tracking devices and fitness apps **contacted five unique domains on average**, a "significant number" of them "contacted 10 or more different domains" – from service providers to ad networks and marketing data companies such as *Tapjoy, Doubleclick, Apsalar, Localytics, Apptentive, Flurry* and *Admob*.

Data on workouts, diets and medical searches

Latanya Sweeney (2014), chief technologist of the U.S. Federal Trade Commission, quotes a study from *Evidon*, which found in 2013 that 20 popular health and fitness apps disclosed information to 70 third-party companies. When Sweeney herself conducted a similar analysis on 12 apps and two wearable devices, she found that information was transmitted to **76 different third-party companies**. One of the tested apps disclosed information from consumer specific identifiers to diet and workout information, to 18 third parties. Reversely, one of the third-party companies received **common identifiers**, gender and workout information from four of the analyzed apps. 18 third parties received device-specific identifiers and 14 received names, usernames or email addresses. 22 of them received additional information on **exercises, meal and diet information, geolocation and medical/symptom searches**.

Health and fitness apps share user data

Taken together, it is largely unclear which kinds of user data both activity trackers and fitness and health apps are providing or selling to third parties. U.S. companies can analyze and share data collected by fitness trackers quite freely, because this type of data is not classified as "health" data in the U.S. (see Hilts et al 2016). A report by the Norwegian Consumer Council (2016) found that "health and fitness apps share user data with partners and advertisers" and revealed that both *Runkeeper* and *Endomondo* retrieve the user's location even when apps are not in use[273]. After the study was published, several apps changed their terms and practices (see also chapter 4.2.1).

---

[272] From on-device and transmission to cloud storage risks, see e.g. Barcena, Mario Ballano; Candid Wueest, and Hon Lau (2014): How safe is your quantified self? Symantec, August 11, 2014. Online: https://www.symantec.com/content/dam/symantec/docs/white-papers/how-safe-is-your-quantified-self-en.pdf

[273] Norwegian Consumer Council (2016): Health and fitness apps violate users privacy. Press Release, 25.02.2016. Online: http://www.forbrukerradet.no/side/health-and-fitness-apps-violate-users-privacy [21.07.2016]

4.3.4 Health data for insurances and corporate wellness

Apart from the practice of transmitting user data to third-party companies, which users are often not aware of, fitness and health platforms are increasingly cooperating with employers and insurance companies.

„Increase employee productivity“

Market leader ***Fitbit*** offers its devices and services to employers and helps them to "plan, track, manage and execute" their **corporate wellness programs**.[274] Activity trackers are sold to companies at quantity discounts.[275] These can either give the devices to their employees for free or offer them for a very low price. According to *Fitbit*'s *Group Health* website,[276] the company takes care of "orders, payment collection & shipping" to employees. During device setup, employees are invited to sign into a company-specific version of *Fitbit*'s software platform to "create immediate employee engagement", where employees can "track their progress" and compete in "corporate challenges". *Fitbit* advertises its corporate wellness products to companies with slogans like "increase employee productivity", "get employees more active, and potentially reduce healthcare costs" and "Fitbit Group Health lets you monitor individual, team, and company-wide progress".

Less insurance costs by tracking

Fitbit corporate wellness customers include the ***Bank of America***, ***IBM*** and ***Time Warner***. In April 2016, ***Target*** announced that it would offer 335,000 devices to its employees, while *Barclays* offered 75,000 to its employees.[277] *Fitbit* claims that more than 50 of the Fortune 500 companies belong to their customers.[278] Some companies already successfully adopted corporate wellness programs not just to increase employee's health, but also to reduce insurance costs. In 2014, the CEO of the U.S.-based company *Appiro* told Bloomberg that he negotiated **"$300,000** off his company's roughly $5 million in annual insurance costs", when about 400 of his employees participated in a "voluntary fitness program that includes uploading their activity with Fitbit" and "sharing the data with the company's health care provider".[279]

$ 1,200 discount on insurance premium

Equally, 14,000 employees of the oil corporation ***BP*** decided to let free *Fitbit* tracker record their steps in 2013. All those who achieved a million steps "gained points that could go towards a lower insurance premium".[280] Bloomberg reported that one BP employee **saved $1,200** on his annual health insurance bill due to participating in this program and reaching **1 million steps**.[281] BP is self-insured, pays directly for health-related expenses of its employees, and thus has a strong interest to keep them as low as possible. On the other hand, $1,200 is a considerable amount of money, which could practically force certain employees to wear a fitness tracker and let it monitor their lives.

---

[274] Fitbit (2015): Fitbit for Corporate Wellness. Infosheet. Online: http://content.fitbit.com/rs/493-CEF-482/images/FitbitWellness_InfoSheet.pdf [21.07.2016]
[275] See Fitbit (2016): Annual report 2015
[276] https://www.fitbit.com/group-health [22.07.2016]
[277] Farr, Christina (2016): How Fitbit Became The Next Big Thing In Corporate Wellness. Fast Company, 18.04.2016. Online: http://www.fastcompany.com/3058462/how-fitbit-became-the-next-big-thing-in-corporate-wellness [22.07.2016]
[278] https://investor.fitbit.com/press/press-releases/press-release-details/2015/Fitbit-Wellness-Adds-Over-20-New-Enterprise-Customers-Including-Barclays-PLC/default.aspx [22.07.2016]
[279] Satariano, Adam (2014): Wear This Device So the Boss Knows You're Losing Weight. Bloomberg. Online: http://www.bloomberg.com/news/2014-08-21/wear-this-device-so-the-boss-knows-you-re-losing-weight.html [22.07.2016]
[280] Olson, Parmy (2014b): Wearable Tech Is Plugging Into Health Insurance. Forbes, 19.06.2014. Online: http://www.forbes.com/sites/parmyolson/2014/06/19/wearable-tech-health-insurance [22.07.2016]
[281] Satariano, Adam (2014)

Corporate wellness at BP

In 2016, *BP* still offers its corporate wellness program including free *Fitbit* devices. When employees complete one million steps they earn 500 "wellness points", and 250 points for every additional million steps. *BP*'s wellness program also rewards other health-related activities from participating in "telephonic lifestyle management" or a "comprehensive health questionnaire" (both 250 points) to a "biometric screening" (125 points).[282] Employees who want to participate in specific health options which include the chance to receive a **$1,000 contribution** to their "Health Savings Account" have to reach 1,000 points to "remain eligible". [283] This would correspond to **three million steps**.

Data managed by neutral parties

BP's corporate wellness program is managed by *Fitbit*'s partner ***StayWell***, according to Forbes, a "population-management firm" who manages the collected health data as a "neutral third party".[284] *StayWell* describes itself as a "health engagement company"[285], whose "population-specific programs" are "backed by decades of experience and **deep expertise in the science of behavior change**".[286] It is at least questionable whether a company, which specializes in the "science on behavior change", can really be considered as a "neutral party" regarding health data of employees. *Fitbit* claims to partner with corporate wellness vendors with health plans "who cover more than **50% of the US population**"[287]. In 2015, *Fitbit* announced that it now supports **HIPAA compliance** "to more effectively integrate with HIPAA-covered entities, including corporate wellness partners, health plans and self-insured employers".[288] The U.S. *Health Insurance Portability and Accountability Act* (HIPAA) protects the privacy of certain health-related information, when this information is managed by organizations and companies that fall under the remit of HIPAA (see FTC 2014, p. 14).

Insurance programs incorporating wearables

Large U.S. insurance companies like ***United Health***, ***Humana***, ***Cigna*** and ***Highmark*** started voluntary programs that involve wearables years ago. Consumers wear tracking devices and their activity data is submitted to online systems, in which they gain points.[289] Initially, such points could be traded for small rewards like coupons or cinema tickets.

15% discount on life insurance

In 2015, ***John Hancock***, one of the largest life insurers in the U.S., went one step further.[290] [291] They teamed up with *Vitality,* a corporate wellness provider, to offer policy holders a discount when they let a **free *Fitbit*** device track their activities. Consumers receive "personalized health goals and can easily log their activities using online and automated

---

[282] http://hr.bpglobal.com/LifeBenefits/Sites/Core/BP-Life-benefits/BP-Wellness-Programs/2016-BP-wellness-program.aspx [22.07.2016]
[283] http://hr.bpglobal.com/LifeBenefits/Sites/Core/BP-Life-benefits/BP-Wellness-Programs/How-the-BP-wellness-program-works.aspx [22.07.2016]
[284] Olson, Parmy (2014): The Quantified Other: Nest And Fitbit Chase A Lucrative Side Business. Forbes, 05.05.2014. http://www.forbes.com/sites/parmyolson/2014/04/17/the-quantified-other-nest-and-fitbit-chase-a-lucrative-side-business [22.07.2016]
[285] http://staywell.com/about-staywell [22.07.2016]
[286] http://staywell.com/employer-solutions [22.07.2016]
[287] https://www.fitbit.com/group-health/partners [22.07.2016]
[288] Fitbit (2015): Fitbit Extends Corporate Wellness Offering with HIPAA Compliant Capabilities. Press Release, 09/16/2015. Online: https://investor.fitbit.com/press/press-releases/press-release-details/2015/Fitbit-Extends-Corporate-Wellness-Offering-with-HIPAA-Compliant-Capabilities/default.aspx [22.07.2016]
[289] Satariano, Adam (2014)
[290] John Hancock (2015): John Hancock Introduces a Whole New Approach to Life Insurance in the U.S. That Rewards Customers for Healthy Living. April 8, 2015. Online: http://www.johnhancock.com/about/news_details.php?fn=apr0815-text&yr=2015 [22.07.2016]
[291] http://www.thevitalitygroup.com/john-hancock-enters-exclusive-partnership-with-vitality [22.07.2016]

tools". By gaining "Vitality Points" they can get a **discount of up to 15% on their life insurance policy**. Other rewards like gift cards, discounted hotel stays and flights are available as well.[292] According to *John Hancock*, a "45 year old couple (of average health)" buying a life insurance policy of $500,000 each "could potentially save more than $25,000 on their premiums by the time they reach 85", as long as they earn enough points in all years.[293] Steps and activity recorded by the *Apple Watch* and *iPhones* can also be used to gain points, because the program also integrates with *Apple*'s *HealthKit* platform.[294]

Discovery's Vitality program

In 2016, *John Hancock*'s Canadian parent company ***Manulife*** announced a similar program for Canadian consumers.[295] Their partner *Vitality*, which is part of the **South Africa**-based insurance company ***Discovery***, additionally lists supported health devices and apps such as *Polar*, *Garmin*, *Withings*, *Jawbone* and *Samsung*'s *S-Health*.[296] *Vitality* is also offered in the **UK**, branded as "VitalityHealth" and "VitalityLife".[297] In addition, *Vitality* has built partnerships with insurance companies all over the world to introduce similar programs, for example with *AIA* in **Asia** and *Ping An Health* in **China**, and, lately, with *Generali* in Europe.[298]

Vitality in Europe

According to *Discovery*, the *Generali Group* has now "exclusive rights" to the *Vitality* program in Europe. In **Germany** it is available to policyholders since July 2016.[299] According to their German website, it works similar to *John Hancock*'s program. Members have to pay a monthly fee of €5, but it is only available in connection with a life or occupational disability insurance policy. Points are collected by participating in health questionnaires and by recording their activity with a fitness tracker. Besides rewards such as discounts on sport shoes and refunds for travels, they promise a **discount up to 16% on insurance premiums**.[300] According to an interview with *Generali*'s Giovanni Liverani, members could also "allow" fitness centers and supermarket chains to inform *Generali*, how often they are attending and which products they are buying, respectively.[301] After Germany, launches in France and Austria are planned.[302]

---

[292] Mearian, Lucas (2015): Insurance company now offers discounts -- if you let it track your Fitbit. Computerworld, Apr 17, 2015. Online: http://www.computerworld.com/article/2911594/insurance-company-now-offers-discounts-if-you-let-it-track-your-fitbit.html [22.07.2016]

[293] John Hancock (2015)

[294] John Hancock (2015): John Hancock Vitality Life Insurance Solutions Launches HealthKit-enabled App for iPhone and iPod touch; allows policyholders to get rewarded for recording healthy activities on iPhone and the Apple Watch. Apr 28, 2015. Online: http://www.prnewswire.com/news-releases/john-hancock-vitality-life-insurance-solutions-launches-healthkit-enabled-app-for-iphone-and-ipod-touch-allows-policyholders-to-get-rewarded-for-recording-healthy-activities-on-iphone-and-the-apple-watch-300073300.html [22.07.2016]

[295] Evans, Pete (2016): Manulife to offer Canadians discounts for healthy activities. CBC News. Feb 09, 2016. Online: http://www.cbc.ca/news/business/manulife-fitness-insurance-1.3439904 [22.07.2016]

[296]https://www.discovery.co.za/discovery_coza/web/linked_content/pdfs/vitality/vitality_news/vitality_fitness_points.pdf [22.07.2016]

[297] https://www.vitality.co.uk/our-journey [01.08.2016]

[298] Discovery (2016): Discovery Vitality launches in Europe through Generali. 23 June 2016. Online: https://discovery.co.za/portal/individual/vitality-launches-in-europe-through-generali [22.07.2016]

[299] Ibid.

[300] https://www.generali-vitalityerleben.de [22.07.2016]

[301] Honsel, Gregor (2016): Tracking durch die Versicherung: "Wir werden Sie nicht bestrafen". Technology Review, 27.08.2015. Online: http://www.heise.de/tr/artikel/Tracking-durch-die-Versicherung-Wir-werden-Sie-nicht-bestrafen-2791079.html [01.08.2016]

[302] Ralph, Oliver (2016): Insurer to sell data-driven product in privacy-conscious Germany. Financial Times, 23.06.2016. Online:

Also in 2016, ***United Health*** announced a corporate wellness program which is not affiliated with *Vitality.* The program provides free fitness trackers to employees of customer companies and offers them the opportunity to "earn up to $1,460 per year by meeting certain goals for the number of daily steps".[303]

Punish, not reward

Other companies are experimenting with punishment schemes instead of rewards. The U.S. startup ***StickK***[304] offers an app that incorporates data from wearables, but instead of collecting "wellness points", points are deducted if users do not achieve their activity goals. *StickK's* offer to consumers is based on a kind of "contract", in which users commit to donate a certain amount of money to specific charities when they are not achieving their goals. *StickK* argues that their approach is "far more effective than offering rewards" and already had 13 corporate customers in 2014, including three Fortune 500 companies.[305]

Mandatory discrimination?

Corporate wellness programs, health plans or insurances accessing data from fitness and activity trackers, have often been criticized by media, privacy advocates and scholars. Possible **risks for both individuals and society** include:

- Data security issues[306]
- Activity data or inferred health scores could be disclosed or sold to third parties.[307]
- Companies might use the detailed information that fitness trackers are recording about their employee's work and private lives, for purposes other than corporate wellness.[308]
- People who fail to achieve goals might have to pay higher premiums (see Lupton 2014) or could even be placed on blacklists[309]. This could lead to discrimination against people who are not young and healthy.[310]
- While people who voluntarily participate in these programs are rewarded, people who don't want to participate could be penalized.[311]

---

http://www.ft.com/cms/s/0/b539ec08-3897-11e6-9a05-82a9b15a8ee7.html [23.07.2016]

[303] UnitedHealth Group (2016): UnitedHealthcare and Qualcomm Collaborate to Launch New Wellness Program That Links Financial Incentives with the Use of Wearable Devices. Mar. 01, 2016. Online:
http://www.unitedhealthgroup.com/newsroom/articles/feed/unitedhealthcare/2016/0301qualcommunitedhealthcaremotion.aspx [22.07.2016]

[304] http://www.stickk.com

[305] Olson, Parmy (2014)

[306] e.g. 80 million people were affected by a data breach at Anthem, a large health insurer, see: http://www.npr.org/sections/alltechconsidered/2015/02/05/384099135/anthem-hack-renews-calls-for-laws-to-better-prevent-breaches [22.07.2016]

[307] See also chapter 5.3

[308] See e.g. Haggin, Patience (2016): How Should Companies Handle Data From Employees' Wearable Devices? The Wall Street Journal, May 22, 2016. Online: http://www.wsj.com/articles/how-should-companies-handle-data-from-employees-wearable-devices-1463968803 [01.08.2016]

[309] E.g. "In other words, privacy experts fear that there's very little to stop a life insurance company from putting those who fail to stay active on a blacklist of sorts, which is akin to the experience of having a low credit score", see: http://www.npr.org/sections/alltechconsidered/2015/04/09/398416513/weighing-privacy-vs-rewards-of-letting-insurers-track-your-fitness [22.07.2016]

[310] E.g. "cutting insurance costs for lower risk customers, raising them for higher risk ones", see: http://www.zdnet.com/article/yes-insurers-want-your-health-data-but-not-for-the-reason-you-think [22.07.2016]

[311] E.g. "Companies have increasingly used a combination of carrots (free vacation days!) and sticks (higher premiums) to coerce employees into participating in health screenings and wellness programs—a practice that the Equal Employment Opportunity Commission has fought with varying success", see: http://www.fastcompany.com/3058462/how-fitbit-became-the-next-big-thing-in-corporate-wellness [22.07.2016]

- When the incentives offered are considerably valuable, it could become less voluntary to participate (see Lupton 2014), it becomes a necessity and "normal", in particular for those that are financially less well off.

From self tracking to the virtual prison

A student app goes one step further than any other available product. ***Reclamate*** is a smartphone app that sees itself as a "safer, cheaper alternative to traditional prisons" and wants to give "offenders access to a variety of services while monitoring their actions and encouraging pro-social behaviors".[312] It "nudges nonviolent offenders to keep up on their post-release job training, drug testing, parole visits". Good-behavior leads to rewards such as an "extended curfew on weekends".[313] A report by *Deloitte University Press* has already taken up a similar idea and suggests "pairing smartphone technology with existing electronic monitoring practices", to create a "new model of virtual incarceration".[314]

4.4 Ubiquitous surveillance in an Internet of Things?

"The Internet will disappear ... There will be so many IP addresses...so many devices, sensors, things that you are wearing, things that you are interacting with that you won't even sense it. It will be part of your presence all the time."
Eric Schmidt, 2015 [315]

It is generally believed that the term **Internet of Things (IoT)** was coined by Kevin Ashton in 1999.[316] Six years later, the *International Telecommunications Union (ITU)* predicted that the "creation of the Internet of Things will entail the connection of everyday objects and devices to all kinds of networks".[317]

Anytime connectivity for anything

There is still no common **definition** of this term (see Minerva et al 2015). According to some prominent organizations, the Internet of Things "links the objects of the real world with the virtual world, thus enabling anytime, anyplace connectivity for anything and not only for anyone" (Sundmaeker et al 2010). The term describes an "interconnected environment where all manner of objects have a digital presence and the ability to communicate with other objects and people" (FTC 2015). It refers to a "world-wide network of interconnected objects uniquely addressable, based on standard communication protocols" (Botterman 2009).

More and more physical objects and spaces are connected to the Internet, from printers, fridges, cars and doors to objects located in offices, industrial plants or in public space. All these objects are equipped with sensors, processing power, network connections and actuators, which may enable them to act in certain ways. They are able to process information and to communicate with other objects and networks, with their environment and with people. Gartner estimates that 4.9 billion "connected things" have been in use in

---

[312] http://www.appitupchallenge.com/stats.html [22.08.2016]
[313] http://articles.philly.com/2016-04-23/business/72540075_1_mobile-phone-app-ben-franklin-technology-partners-uber [22.08.2016]
[314] http://dupress.com/wp-content/uploads/2013/03/DU220_Beyond-the-Bars_vFINAL-3.5.pdf [22.08.2016]
[315] http://www.hollywoodreporter.com/news/google-chairman-eric-schmidt-internet-765989 [01.08.2016]
[316] Ashton, Kevin (2009): That 'Internet of Things' Thing. In the real world, things matter more than ideas. In: RFID Journal, 22.06.2009. Cited am 21.12.2016 von http://www.rfidjournal.com/articles/view?4986
[317] ITU (2005): ITU Internet Reports. The Internet of Things. International Telecommunications Union, November 2005. Online: https://www.itu.int/net/wsis/tunis/newsroom/stats/The-Internet-of-Things-2005.pdf

2015, thereof 3 billion in the consumer space, and predicts an increase to 20.8 billion connected objects by 2020.[318]

RFID tags leading the way

For many years, the debate about the Internet of Things was focused on **radio-frequency identification (RFID)**, a technology that uses "radio waves to automatically identify and track individual items" (ITU 2005). RFID **transponders or tags** located on objects can carry information which RFID **readers** can access remotely. RFID technology partly replaced barcodes in many fields. Tags are attached to different objects from shopping items and passports to containers. RFID transponders do not solely store codes that can uniquely identify an object, but can also store additional information such as fingerprints or photos. The use of tags within goods from **clothes to medicine and ID cards** has already caused many debates regarding privacy. For example, concerns were raised that data on RFID tags attached to goods or ID cards could be **accessed without the consumers' knowledge**, through materials and from a distance (see Sterbik-Lamina et al 2009). In 2015, an estimated number of 9.1 billion RFID tags were sold.[319]

Invisible like the wires in the walls

However, RFID is just one of a variety of technologies used today to connect the physical world with digital networks. Apart from a wide range of **wireless technologies** such as GSM, UMTS, LTE, Wi-Fi, Bluetooth, NFC and **localization technologies** such as GPS tracking, the development of sensors plays an important role. **Sensors** help to record data from the physical world and to make it usable for digital processing, often in real-time. Due to integration of small computers with network connections and a variety of sensors in everyday objects, these computers become ubiquitous. Consequently, this gives rise to the use of terms such as **Ubiquitous Computing** and **Pervasive Computing** (see Spiekermann and Pallas 2005). As Mark Weiser (1991) stated, all these computers will become more and more invisible – "like the wires in the walls".

Fields of application

Smartphones and laptop computers are sometimes excluded from debates concerning the Internet of Things. Since today's smartphones, wearables and many other devices often share a very similar set of sensors, network connections and software, the lines between them become more and more blurred. But within the Internet of Things, many other devices and domains are discussed. The following list describes several fields and future areas of application, based on a survey on the Internet of Things by Atzori et al (2010) and a report by the *Pew Research Center* (2014):

- **Personal life, body and healthcare:** Many people will not only wear devices that "give them feedback on their activities, health and fitness", but they will also "monitor others" – for example, their "children or employees", who are "also wearing sensors, or moving in and out of places that have sensors" (Pew 2014). In hospitals, applications could include the tracking of staff, patients and medical inventory, identifying and authenticating people, the monitoring of health indicators and medication, and the remote monitoring of patients at home. In personal life, sensor-equipped places, goods or other objects could be integrated with apps and social networking. Also the location of personal belongings could be tracked in order to prevent loss and theft (Atzori et al 2010).

[318] Gartner (2015): Gartner Says 6.4 Billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015. Press Release, 10.11.2015. Online: http://www.gartner.com/newsroom/id/3165317 [24.07.2016]

[319] IDTechEx (2015): IDTechEx Research finds RFID market exceeds $10bn milestone in 2015. 19.10.2015. Online: http://www.idtechex.com/research/articles/idtechex-research-finds-rfid-market-exceeds-10bn-milestone-in-2015-00008567.asp [25.07.2016]

- **Homes, offices and other buildings**: Many devices and facilities in buildings will soon be equipped with sensors connected to the Internet such as fridges, ovens and coffee machines, door locks, heating, air condition, lighting, water pipes and fire alarms (Pew 2014). This is not only true for homes, but also for offices, industrial plants or leisure environments. In addition, electrical devices could automatically switch on and off based on energy prices, which are dynamically changed by the energy providers (Atzori et al 2010) – as a result of "self regulating power grids" (Pew 2014).
- **City, infrastructure and transportation:** Not only "cars, trains, buses as well as bicycles" become equipped with sensors, tags and network connections (Atzori et al 2010), but also streets, buildings and bridges. These objects could capture data on their condition or on pollution levels. The sensor data may be used for public safety or traffic control, and might even be synchronized with data about people's "eating and commuting habits and their day-to-day calendars". Other examples mentioned include municipal trashcans that "signal when they need to be emptied" and paper towel dispensers in restrooms which "signal when they need to be refilled" (Pew 2014).
- **Manufacturing and commerce**: In factories, both machines and production parts are equipped with RFID tags, sensors and other elements of information technology. Similarly, the whole supply chain is designed to track every single event from the manufacturing of goods to logistics and retail. Managers can not only oversee the entire production process, but can also get a "global view on all the elements and the possible side effects of a production line delay due to shop-floor device malfunctions" (Atzori et al 2010). In Germany, this is discussed under the term "Industrie 4.0".[320]

Societal implications

A report about "The Internet of Things: Opportunities for Insurers" by a consulting firm explains that insurers could "use IoT-enriched relationships to connect more holistically to customers and influence their behaviors".[321] Similarly, many of the 1,606 experts interviewed by Pew Research (2014) expect that "incentives to try to get people to change their behavior" will become a "major driver" of the Internet of Things, for example to motivate people to purchase a product, to act in a more healthy or safe manner or to improve their performance at work. They state that the "realities of this data-drenched world raise substantial concerns about privacy and people's abilities to control their own lives". If "everyday activities are monitored and people are generating informational outputs, the level of profiling and targeting will grow and amplify social, economic, and political struggles".

Algorithms and control

Some of the experts interviewed expressed their concerns that the automated feedback and stimulation loops used, as well as the algorithms making decisions on humans could have negative social consequences. Sensors and wearables are usually introduced because of "some company's business strategy" and not "necessarily to advance a collective good". The Internet of Things could be "an incredible, powerful tool for controlling populations". Therefore, it is important to discuss who is going to be in control of it. In addition, the "kind of complexity caused by such a large network" could be "too difficult to maintain and evolve well" and may lead to "complicated, unintended consequences".

Security risks

In 2013, the U.S. **Federal Trade Commission** hosted a workshop titled "The Internet of Things: Privacy and Security in a Connected World". According to the resulting report

[320] https://www.bmbf.de/de/zukunftsprojekt-industrie-4-0-848.html [26.07.2016]
[321] ATKearney (2014): The Internet of Things: Opportunity for Insurers. December 2014. Online: https://www.atkearney.com/digital-business/ideas-insights/featured-article/-/asset_publisher/Su8nWSQlHtbB/content/internet-of-things-opportunity-for-insurers/10192 [01.08.2016]

which was released in 2015, participants warned that the Internet Of Things presents "a variety of potential security risks that could be exploited to harm consumers" – ranging from "unauthorized access and misuse of personal information" to companies which might use the recorded data to "make credit, insurance, and employment decisions". Consequently, the main discussion focused on long-standing principles such as "security, data minimization, notice, and choice" (see FTC 2015).

Everywhere, but nowhere

Natasha Lomas from the tech industry blog *TechCrunch* wrote in 2015: "Imagine what kind of surveillance opportunities are opened up by an 'invisible' Internet — which is both everywhere but also perceptually nowhere, encouraging users to submit to its data-mining embrace without objection". Subsequently, she concludes: "**In the offline world we have cars and roads. We also have speed limits — for a reason.** The key imperative for regulators now, as we are propelled towards a more densely-packed universe of connected devices, is coming up with the sensornet's speed limits. And fast".[322]

4.4.1 Examples – from body and home to work and public space

In the context of the Internet of Things already billions of physical objects include sensors and network connections. Devices that monitor activities, bodies and health of human beings are the most relevant but also may put individuals at risk. Smartphones and apps, health and fitness trackers and the "connected car" have been examined in previous chapters. With a focus on privacy aspects, the following section lists additional examples for devices, platforms and applications in several fields of life:

Google at home

Connected thermostats and smoke alarms: In 2014, *Nest Labs*[323] was acquired by *Google* for 3.2 billion dollars.[324] They offer a "learning" thermostat including Wi-Fi and Bluetooth connections and sensors for temperature, humidity, ambient light, near-field and far-field activity.[325] It records data about the everyday behavior of residents and offers to manage the room temperature accordingly. *Nest* also offers indoor cameras and "smoke + CO" alarms, which are also equipped with network connections and various sensors, including a microphone and an occupancy sensor.[326] According to the Wallstreet Journal, *Nest Labs* started to share user data with *Google* a few months after the acquisition, including information about "when Nest users are at home or not".[327]

Behavioral patterns from energy usage

Smart meters are "networked metering devices" for the measurement of electrical energy consumption, sometimes also water and gas, which report the recorded data back to the utility provider, sometimes "with an interval of as low as 2 seconds". There has been a debate on the privacy implications of these devices for several years. Data from smart meters "could be used to very accurately identify the behavioral patterns for individual

---

[322] Lomas, Natasha (2015): What Happens To Privacy When The Internet Is In Everything? TechCrunch, Jan 25, 2015. Online: https://techcrunch.com/2015/01/25/what-happens-to-privacy-when-the-internet-is-in-everything [02.08.2016]
[323] https://nest.com [24.07.2016]
[324] Wohlsen, Marcus (2014): What Google Really Gets Out of Buying Nest for $3.2 Billion. Wired, 14.01.2014. Online: http://www.wired.com/2014/01/googles-3-billion-nest-buy-finally-make-internet-things-real-us [24.07.2016]
[325] https://store.nest.com/product/thermostat/
[326] https://store.nest.com/product/smoke-co-alarm/
[327] Winkler, Rolfe; Alistair Barr (2014): Nest to Share User Information With Google for the First Time. WSJ Digits, 24.06.2014. Online: http://blogs.wsj.com/digits/2014/06/24/nest-to-share-user-information-with-google-for-first-time [24.07.2016]

household members".[328] The European Union plans to "replace at least 80% of electricity meters with smart meters by 2020 wherever it is cost-effective to do so".[329]

Telescreens

Smart TVs: According to an investigation by a UK-based organization in 2014 Smart TVs from *LG*, *Samsung*, *Sony*, *Panasonic* and *Toshiba* were tracking the viewing habits of consumers "to some degree". Some of them also transferred information about websites visited, files watched from USB sticks, location and postcode data to manufacturers.[330] In 2015, the voice recognition features of *Samsung*'s devices caused a large public debate. *Samsung* warned users in its privacy policy to "be aware" that if their "spoken words include personal or other sensitive information, that information will be among the data captured and transmitted to a third party through your use of Voice Recognition". Presumably, this feature was perceived as especially invasive, because TVs which are listening to what is spoken bring the "telescreens" in George Orwell's famous novel "1984" into mind.[331] Currently, *Samsung* allows users to deactivate voice recognition.[332] Still, Smart TVs can be a rich source for detailed information about the lives of consumers.

Analyzing readers

E-readers: Many devices transfer detailed information about users' reading behavior to companies, for example, which books they read, how far readers get, how long they spend reading, which passages were read and what was highlighted or annotated.[333] That not only allows deep insights about readers, but could also influence how books are written in the future, due to the detailed methods of analysis of what is popular and what isn't.

Small wearables

Biometric headphones: *BioSport*, a headphone model developed by *SMS Audio* in cooperation with *Intel*, measures the heart rate of users with a built-in optical sensor.[334] The German company *Bragi* offers a small wireless headphone model[335], which is equipped with "27 sensors" to measure steps, distances, breaths and heart rate.[336]

Connected insurance plan

Toothbrushes that include a tracking-based dental insurance plan: Today, several toothbrushes that can transfer data to apps are available on the market, for example by major brands such as *Oral-B*.[337] The U.S.-based company *Beam* goes one step further. They offer dental insurance plans including a toothbrush device, which "measures progress and rewards users for improving their dental health". According to their website members can get up to 25% discount depending on their "brushing score", which is calculated on the

---

[328] Jobst, Martin Erich (2013): Security and Privacy in the Smart Energy Grid. Seminars FI / IITM / ACN SS2013, Network Architectures and Services, August 2013. DOI: 10.2313/NET-2013-08-1_21. Online: https://www.net.in.tum.de/fileadmin/TUM/NET/NET-2013-08-1/NET-2013-08-1_21.pdf
[329] https://ec.europa.eu/energy/en/topics/markets-and-consumers/smart-grids-and-meters [25.07.2016]
[330] Laughlin, Andrew (2014): Smart TV spying – are you watching TV, or is it watching you? Which? Magazine, 20.08.2014. Online: http://blogs.which.co.uk/technology/tvs/smart-tv-spying-weve-investigated [25.07.2016]
[331] Orwell, George (1949): 1984. Secker and Warburg, London.
[332] http://www.samsung.com/uk/info/privacy-SmartTV.html [25.07.2016]
[333] Alter, Alexandra (2012): Your E-Book Is Reading You. The Wall Street Journal, 19.07.2012. Online: http://online.wsj.com/news/articles/SB10001424052702304870304577490950051438304 [25.07.2016]
[334] Intel (2014): Biometric Headphones Will Optimize Workouts for Ultra-marathoners, Aspirational Exercisers and Everyone in Between. Pressemeldung, 14.08.2014. Cited 21.02.2016 http://newsroom.intel.com/community/intel_newsroom/blog/2014/08/14/intel-and-sms-audio-to-supercharge-fitness-wearables
[335] http://www.bragi.com
[336] http://cdn.bragi.com/www/2016/07/12141138/2.0PR-1207-LATEST.pdf [25.07.2016]
[337] http://connectedtoothbrush.com

basis of the recorded data.[338] The company's CEO explained in 2013 that they were "actually not interested in toothbrushes at all", but in "health data".[339]

Breathing and state of mind

Respiratory Monitoring: *Spire*[340] is a small wearable device, which monitors breathing patterns and steps. They describe respiration as an "information-dense data stream", which has "many components to it such as rate, depth, inhalation-to-exhalation ratio (IER), durations of inhalation, retention, exhalation, and hold, consistency, smoothness, transition, and so on". In combination with the corresponding app *Spire* promises not only give insights into how often users "sit, stand, and lay down" and their "daily activity and state of mind", but also insights into how to maintain "balance and focus, preventing burnout".[341]

Google Glass at Work

Smart Glasses: The introduction of *Google Glass* had triggered a lot of attention and debate. It contained a built-in computer, network connections via Wi-Fi and Bluetooth and several sensors, including microphone and camera.[342] *Google* stopped offering its headset to consumers in 2015.[343] But still, similar head-worn displays are relevant in several contexts from entertainment to manufacturing and medicine. Their main purpose is to "provide users with information and services relevant for their contexts and useful for the users to perform their tasks".[344] Therefore, these devices have to be capable of recognizing other persons, objects and their behavior and also possess the ability to draw conclusions. Currently, *Google* is offering *Glass at Work* for business customers.[345]

Electronic ankle bracelets for babies

Sensor-equipped clothes for babies: *Owlet Baby Care* offers a "smart sock", which measures the heart rate and oxygen levels and transmits the recorded information to an app.[346] The *Baby Monitor* of *Rest Devices* is built into a bodysuit, tracks a baby's "breathing, sleeping temperature, body position, activity level, and whether they are awake and asleep" and transmits this information to the parents smartphone, along with "live audio". Parents can "share the information with as many caregivers as they like" and connect it to the thermostats and indoor cams from *Google*'s *Nest*.[347]

Sensors control drinking and eating behavior

Controlling body functions: In cooperation with the design and innovation consultancy *Smart Design*, the *PepsiCo* subsidiary *Gatorade* developed a range of products that measure and track individual data. The company also works on biosensors that are used to measure athletes' nutrition levels.[348] "[T]he brand has developed a suite of products and technologies that work together to measure and track individuals' data. […] A patch, like a near-field communication chip-enabled Band-Aid, will analyze a player's sweat and

---

338 https://www.beam.dental [25.07.2016]
339 Kaye, Kate (2015): Your Toothbrush Data Will Get You a Deal at the Dentist (Depending). AdvertisingAge, 18.05.2015. Online: http://adage.com/article/datadriven-marketing/toothbrush-data-a-deal-dentist/298655 [25.07.2016]
340 https://spire.io
341 https://www.spire.io/faq.html [26.07.2016]
342 https://support.google.com/glass/answer/3064128?hl=en [26.07.2016]
343 Eadicicco, Lisa (2015): See the New Version of Google's Wildest Product. Time, 29.12.2015. Online: http://time.com/4163067/google-glass-2-photos-2015 [26.07.2016]
344 Bertarini, Marica (2014): Smart glasses: Interaction, privacy and social implications. Ubiquitous Computing Seminar, FS2014, Student report, ETH Zurich. Online: https://www.vs.inf.ethz.ch/edu/FS2014/UCS/reports/MaricaBertarini_SmartGlasses_report.pdf
345 https://developers.google.com/glass/distribute/glass-at-work [26.07.2016]
346 http://www.owletcare.com [25.07.2016]
347 http://mimobaby.com/product/ [25.07.2016]
348 http://www.wsj.com/articles/gatorade-sets-its-sights-on-digital-fitness-1457640150 [11.08.2016]

communicate with the digital platform to identify his sweat type—which will determine sodium, electrolyte, and additional fluid-intake needs."[349]

Monitoring customers and workers

Locating and monitoring retail and sales staff: *Theatro* markets itself as a system for "in-store communication and hourly worker productivity" based on a wearable for employees in retail, hospitality, and manufacturing.[350] In addition to voice communication and indoor location tracking it promises to measure "social interaction data to understand what's impacting productivity and who the top performers are" and to give managers "unprecedented insights into what [their] employees do".[351]

In-store tracking in retail: *Brickstream* offers devices that combine video, Wi-Fi and iBeacon technology for "traffic counting, labor optimization, and in-store analytics". It provides "detailed data on wifi enabled devices in or near the store site" that can be used to analyze customer behavior, for example, to feed "loyalty program and incentive data".[352]

Smart city surveillance

Street lighting, including video monitoring: *PennSMART*[353] provides street lighting devices, which "discreetly monitor, detect and analyze activity that takes place in the vulnerable areas under the trees" including "360-degree motion sensor video cameras", facial recognition, license plate readers and gunshot and glass breaking sensors.[354]

Sensor equipped insurance risk assessment

IBM's IoT for Insurance: IBM has announced a service, which offers insurers a "full 360-degree context of their policyholders" including information retrieved from the Internet of Things. Insurers can "utilize the data derived from all types of devices as well as external sources, such as weather data" to perform "real time risk assessments". They mention "intelligent wellness/workers", "intelligent home & buildings", "intelligent cars/fleet" and "intelligent assets & equipment" as examples of sensor-equipped environments, which insurers could base their programs on.

---

[349] http://www.fastcompany.com/3054919/tech-forecast/gatorade-gets-in-the-game [11.08.2016]
[350] http://theatro.com/leading-heads-hands-free-mobile-workforce-revolution [25.07.2016]
[351] http://theatro.com/what-we-provide [25.07.2016]
[352] http://brickstream.com/home-3DPlus.html [02.08.2016]
[353] http://www.pennsmartlighting.com
[354] http://www.sensorsmag.com/news/tech-product/news/smart-iot-lighting-brings-surveillance-and-safety-21662 [26.07.2016]

5. Data Brokers and the Business of Personal Data

"It's your data. You have the right to control it, share it and use it how you see fit."
How the data broker *Lotame* is addressing its corporate clients on its website, 2016 [355]

"the power of personal information lies at the heart of surveillance"
Neil M. Richards (2013), Harvard Law Review

As we have seen in the previous chapters, more and more devices, apps, platforms and services are collecting enormous amounts of personal data about our everyday life. Big Data analytics makes it possible to infer personal details and even predict future behavior based on transactional or behavioral data about individuals, which seem to be rather insignificant and meaningless. During the last few years, the quantity of data collected by companies rapidly increased. Consequently this data became a valuable economic asset. A whole new economy emerged around the monetization and exploitation of personal data. This chapter focuses on today's personal data ecosystem, on companies selling access to personal data or to information derived from it to other companies, and on the implications and risks these practices bring along for consumers.

5.1 The marketing data economy and the value of personal data

This section focuses on the structure of this personal data ecosystem, the main corporate actors, and the data they collect, buy and sell. It is "not easy to draw an accurate and reliable picture of the scope, structure and connections" of this industry, "not least because of its secrecy" (Bria et al 2015, p. 36).

Loyalty programs and database marketing

The history of the commercial use of digital personal data ranges back to the 1970s, when direct marketing grew rapidly, and to the 1980s, when companies started to apply **database marketing** concepts. During the 1980s businesses did not just learn "their customer's names and addresses", but also "began to collect detailed personal and purchasing information" (see Petrison et al 1993). Beginning with ***American Airlines***' *AAdvantage* program in 1981,[356] companies started to introduce loyalty programs. These were "intended to help retailers build a more loyal customer base, but also provided them with detailed data on their customers and their purchasing preferences" (CMA 2015, p. 22).

Billions from targeted ads

Loyalty programs "continue to be an important source of customer data for businesses". However, the rise of digital communication technology "has led to a substantial shift in the ability of firms to gather data on actual and potential customers" (ibid.). Today's Internet giants such as ***Google*** or ***Facebook*** generate large parts of their turnovers with targeted advertising based on personal data. Google's advertising revenue was $59.6 billion in 2014[357]. *Facebook*'s advertising revenue was $11.5 billion in 2014.[358] These prominent companies are very visible to consumers and to the general public. But there are

[355] https://www.lotame.com/resource/its-your-data-you-should-be-able-to-do-what-you-want-with-it [01.08.2016]
[356] https://www.aa.com/i18n/aboutUs/corporateInformation/facts/history.jsp [25.01.2016]
[357] https://investor.google.com/financial/tables.html [25.01.2016]
[358] http://investor.fb.com/releasedetail.cfm?ReleaseID=893395 [25.01.2016]

thousands of less known businesses which collect, analyze and sell personal profiles containing hundreds of attributes about consumers.

Consumer data worth $156 billion?

A study by John Deighton and Peter Johnson[359] found in 2013 that data-driven marketing relying on "individual-level consumer data" accounts for at least $156 billion in value-added revenue in the U.S. alone, 71% of it through "services directly or indirectly dependent on data exchanged or rented among firms". Only 29% of value-added revenues are based on data services within single firms (Deighton et al 2013, p. 7). The study, commissioned by the **Direct Marketing Association**'s "Data-Driven Marketing Institute", is "summing what firms spent on data and data services", excluding any "benefits that firms received in exchange for spending on data and data services, which commonly exceed data costs by 20% to 60%".

The following table summarizes the segment-specific revenues based on "individual-level consumer data" as estimated in the Deighton study, and offers a categorization of businesses in the marketing data economy – from a marketing perspective:

[359] DDMI (2013): The Value of Data: Consequences for Insight, Innovation, and Efficiency in the U.S. Economy. Summary of a study commissioned by DMA's Data-Driven Marketing Institute (DDMI), October 14, 2013. Online: https://thedma.org/wp-content/uploads/DDMI-Summary-Analysis-Value-of-Data-Study.pdf [25.01.2016]

	Business Segment	Description	Value-added revenues (in billion $)		
			Total contri-bution	Directly dependent on data exchanged or rented	Indirectly dependent on data exchanged or rented
Strategic Marketing Services	Agency Holding Companies	Large firms spanning a broad range of services including creative/media agencies, direct marketing agencies, market research suppliers, database management, analytics	7	1	4
	Agencies	Other independent general agencies	6	1	4
	Digital Agencies	Digital agencies, born in the middle 1990s.	2	0	1
	Direct/CRM Agencies	Direct Agencies "advise their clients on how to aggregate markets of specific individuals".	2	1	1
	Measurement, Analytics	Firms, which analyze data for marketing and "collect data from public sources, private 'panels', purchased third party sources, and from marketers' 'owned' data"	3	1	0
Audience Assembly and Targeting	Digital Audience Assembly	Targeted advertising by online publishers, which "relies substantially on individual-level data traded among firms", e.g. display, mobile & social advertising publishers like Google, Facebook, Yahoo, MSN, AOL and Twitter.	14	7	4
	Search Audience Assembly	Targeted advertising based on web search results, still depending on "traded or exchanged data", but less.	19	2	2
	Audience Targeting	"Demand Side Platforms (DSPs), Supply Side Platforms (SSPs), Data Management Systems (DMSs), Behavioral Data Providers, and Ad Exchanges"	4	4	0
Prospect / Customer Relationship Marketing	Direct/CRM Customer Targeting	List brokers and database marketing service providers ("data brokers") including analytics, segmenting, scoring, matching, appending and database management services.	7	3	4
	Postal Media and Direct Mail	The "individually addressed, direct response advertising mail and catalog production and delivery subsystem"	32	1	24
	Email Marketing	Email marketing service providers (ESPs)	1	1	0
	Telephone Sales	"Outbound telemarketing" and "inbound call center activity, which involves upselling or cross-selling efforts".	10	2	6
	Mobile Targeting	"Mobile SMS and app-based CRM services"	2	0	0
Data-driven Commerce & Fulfillment	eCommerce	Estimated data-driven revenues of online retailing, e.g. Amazon, Staples, Apple iTunes, Wal-Mart and many "small niche retailers" (excluding online advertising payments).	34	4	22
	Loyalty	Estimated data-driven revenues of "brick and mortar retailing", particularly based on data collection and exchange through loyalty programs	5	2	1
	Fulfillment	"Delivery of offline goods into the hands of purchasers", only indirectly dependent on data exchanged by other players.	9	0	4
	Total		156	32	78

Table 18: Revenues of companies in the marketing data economy that rely on "individual-level consumer data", adapted from Deighton et al (2013, p. 8)

The *European Federation of Direct Marketing Associations (FEDMA)* estimates that the direct marketing sector "represents an annual expenditure of over 60 billion euros" within the EU.[360]

[360] http://www.fedma.org/index.php?id=34 [25.01.2016]

What is the value of personal data?

In an interview, Deighton estimated that an average household pays $200 a year more for products when "declining to join supermarket frequent shopper programs". Participating on "airline frequent flyer and hotel frequent guest programs" would "amount to discounts of 1 to 5 percent over the prices paid by non-subscribers".[361] Another study suggests that Deighton "may understate the individual value of customer's data to companies and to the economy" (McConville et al 2014, p. 57). It could be "much higher", when the customer lifetime value (CLV) is "taken into account". McConville et al set out that there are **"good" customers** who "generate substantial lifetime value, are loyal, and 'promote' the company" – and **"bad" customers**, who are "disloyal, consume a disproportionate amount of company resources, and 'detract' from the company" (ibid., p. 58). In the financial services and banking industry, loyal "promoter" customers generated "about 6.7x the lifetime value" of a "standard, neutral customer". Based on Deighton and other research, McConville et al (2014, p. 64) suggest that a "top-of-the-line loyal, affluent customer's data" could be worth $126K to companies in the US" – assuming that the company is "in a sector where customer loyalty matters most". By simply dividing up this value based on the "individual-level consumer data" of $156 billion and the number of consumer units, they suggest that even a "low-value, disloyal customer, has **data worth approximately $880 per year**" in the U.S.

Consumer lifetime risk?

What they did not discuss is a **"lifetime risk"** which consumers may be exposed to when sharing personal data. Negative implications could range from not receiving certain offers or discounts, getting worse conditions or higher prices than others, to rejected loan, apartment and job applications. What could be the **long-term costs for someone**, who experiences negative implications based on being categorized and rated in a certain way by firms?

As opposed to the concept of a customer lifetime value, we introduce the term "Customer Lifetime Risk" in an attempt to further concretize and categorize the different risks a customer might bear when exchanging his (personal) data with companies.

There is limited academic research about the value of personal data, not just from a company's perspective but also from a consumer's perspective[362]. Possible hidden costs and long-term risks when sharing data have rarely been investigated in detail.

[361] Salls, M. and Silverthorne, S. (2003): Should You Sell Your Digital Privacy? Harvard Business School. Online: http://hbswk.hbs.edu/item/should-you-sell-your-digital-privacy [25.01.2016]
[362] For example see Roosendaal (2014)

5.2 Thoughts on a 'Customers' Lifetime Risk' – an excursus

Seen modern data analyses and the growing number application fields, it becomes reasonable to ask in which ways the described data collection, processing and dissemination might impose risks for individual customers as well as society at large. The draft for a concept we call "Customers' Lifetime Risk" is by no means exhaustive but is rather a first attempt to sketch a concept, which could be further researched.

What is Customers' Lifetime

Customers' Lifetime Risk (CLR) compiles and categorizes the potential risks a customer faces during and after the exchange of (personal) data with a company as well societal risks that arise from the aggregation of such individual risks.

- Risks may be based on licit or illicit use of the customer data
- Risks may result from the customer's own or from the company's handling of personal data
- Risks may relate to various dimensions of negative consequences

Similar to the Customer Lifetime Value (CLV), a well-established marketing instrument (see Venkatesan and Kumar 2004), we developed the CLR to describe the integrity of risks that may occur over the entire time of a business-to-consumer relationship. However, the nature of CLR significantly differs from that of CLV.

CLR as opposed to CLV

The calculation of CLV aims to generate a single, monetary value for one person. A similar attempt would be neither insightful nor practical for the CLR. The risks of data sharing personal data go beyond financial losses and go beyond one person's wellbeing. Impacts on human dignity or the inhibition of free speech cannot be measured in monetary terms or at an individual level. Moreover, the perceived impact of a risk category is ***relatively subjective*** and can change ***depending on the context as well as over time***. For example, the disclosure of one's gender can be a humiliating exposure to some individuals while it is completely unimportant to others.

Although CLR cannot be measured in monetary values, like the CLV, it shares some similarity in that it is additive in nature and builds up over the course of a customer-company relationship. It is amplified by a long lasting, intensive relationship including many transactions.

Which factors influence

Several ***antecedent conditions*** increase the risk probability, including:

- Duration of the data exchange
- Number of companies the data is shared with (resp. devices used)
- Intimacy of the information shared
- Security measures taken by user
- Privacy rights and regulation followed by the data collector
- Technical options provided to access and analyze personal data

Daniel Solove (2006, p.490) provides a classification of privacy harms from a legal perspective, which serves as a basis for the development of a CLR. Transferring the insights from Solove's *Taxonomy of Privacy* into our given context, harms that focus on the relationship between individuals and their governments are deliberately waived here. The remaining harms for individuals and society at large have been regrouped. They have then been complemented by harms, which seem to be reasonably expectable against the background of this report.

Non exhaustive list of drivers relevant in Customers'	
Discrimination	▪ Price discrimination ▪ Offer discrimination ▪ Limited access to insurance ▪ Limited access to credit ▪ Limited access to employment ▪ Disconnection of data from its context, leading to ・ Misinterpretation and false conclusions ・ Unfair judgements based on false or outdated data
Manipulation	▪ Being targeted when vulnerable ▪ Behavioral control (with rewards / punishments) ▪ Personalized information campaigns with political or commercial interests
Security Threats	▪ Identity theft ▪ Illicit use of personal data (i.e. stolen credit card data) ▪ Cyber attacks ▪ Information loss and Information leaks ▪ Breaches of confidentiality
Societal changes	▪ Altering of the behavior ・ Self-censorship ・ Inhibition ▪ Conformism ▪ Inhibiting impacts on ・ Creativity ・ Free speech ・ Individual autonomy : Self-development and self-determination ▪ Digital wildfires ▪ Valuation of self & others based on data ▪ Ability to speak and act anonymously is inhibited by Identification methods ▪ "Architectural problems"[363]: systemic and structural harms, causing the enhancement of risk probabilities or a shift in the balance of social or institutional power
Individual privacy	▪ Public disclosure of private facts and secrets ▪ Revelation of information far beyond the expectation of the user ▪ Misuse and dissemination of data without user's (informed) consent ▪ Stalking, harassment, Cyber mobbing ▪ Spread of false or misleading information and rumours ▪ Exposure, condemnation ▪ Damage of reputation ▪ Blackmailing ▪ Invasive advertising (spam, pop ups or telemarketing)
Market imbalances	▪ Industrial espionage ▪ Creation of monopolies ▪ Less market diversity ▪ Intransparency ▪ Advantages for national economies where servers are harboured ▪ Unfair information practices for customers: ・ Limited knowledge about existence of data records ・ Limited ability to reveal what information is in a record and how it is used ・ Limited ability to correct or amend a record of identifiable information

[363] Solove, D. J. 2004, p.97

5.3 From marketing data to credit scoring and fraud detection

A broad range of companies from very different business segments are active in the data marketing ecosystem, generating billions of revenue for services that rely on "individual-level consumer data". Yet, the report of Deighton et al (2013), which was published on behalf of a marketing industry group, **misses some important parts** of the personal data ecosystem.

Risk management and fraud detection

The typology developed by the FTC (2014) in its report on data broker covers marketing data, too. On the one hand it has a narrower scope than Deighton's summary about "data-driven marketing", because it focuses primarily on the data broker industry. For this reason, business sectors such as customer relationship management (CRM) and e-commerce are missing in the FTC's typology. However, the FTC's overview additionally includes "people search" companies and the sector of risk management services such as identity verification and fraud detection:

Type	Subtype	Offers of Data Brokers
Marketing	**Direct marketing**	Marketing lists: Data brokers provide lists of consumers with specific attributes ("list broking").
		Data append: Data brokers offer clients the ability to add attributes and profile information to their existing customer data.
	Online marketing	Registration targeting: So-called "registration websites" that "allow consumers to register or log in to obtain services, such as retail, news, and travel sites" send a list of registered users/customers to data brokers – either to receive additional information on them or to offer targeted advertising space.
		Collaborative targeting: The data broker serves two clients. On the one hand the registration website sends a list of users to a potential advertiser, and on the other hand an advertiser looking for targeted advertising on the registration website sends its customer and prospect list.
		Onboarding: A client sends data about consumer's offline activities to a data broker who adds offline data to profiles tied to cookies and enables clients to either "find" and target existing customers "anywhere on the Internet" or target consumers with similar characteristics. Onboarding typically includes three steps – segmentation, matching and targeting.
	Marketing analytics	Data brokers offer clients the ability to analyze their customer data in order to gain insights about attitudes and preferences – sometimes based on "hundreds or thousands of data elements". Several kinds of marketing "scores" rank the client's customers and predict future behavior.
Risk mitigation	**Identity verification**	Some data brokers "assist clients in confirming the identity of an individual", often in the form of "scores" indicating the risk associated with a transaction. Some also offer employment verification products, e.g. "that X consumer works for Y employer".
	Fraud detection	Some data brokers help their clients to "identify or reduce fraud" and to verify contact information and transaction histories by "detecting patterns associated with attempted fraud" and in general by "verifying the reliability or truthfulness of information" submitted by consumers – for example, "if a public benefit is contingent on a consumer's level of income".
People search		Some data brokers offer "information about consumers obtained from government and other publicly available sources, such as social media sites" mainly intended "for use by individuals, although they can be used by organizations as well".

Table 19: Typology of data brokers, adapted from FTC (2014)

Credit bureaus and credit scoring?

The FTC's typology still misses a very relevant field. Credit bureaus, credit reporting agencies and credit scoring companies are not covered at all. In their comprehensive report on the "Identity Ecosystem" Bria et al (2015, p. 38 et seq.) have created the following typology of data brokers:

Category	Description
Identity and fraud services, including credit scoring	These services help companies to manage risk, including fraud prevention, identity theft products, credit reports, credit scores and credit models. Sometimes also pre-employment drug screening solutions, credential verification services, and background checks are provided. Examples: Experian, ID Analytics and Equifax.
Customer relations and customer care, including loyalty programs	Besides loyalty programs which are "one of the main systems to gather consumer information" and "part of the core business" of many data brokers, these services help others to "get and retain customers". They provide list marketing data, strategy, marketing technology, creative services, media reach, and personalization of online, offline and mobile marketing campaigns. Examples: Epsilon, Bluekai.
Marketing and advertising	Linked to customer care, these companies offer marketing, lead generation, digital advertising, and targeting: Example: Criteo.
Predictive analytics	These services provide, for example, "consumer, financial and property information, analytics and services" and develop "predictive decision analytics by combining public, contributory and proprietary data". Examples: Corelogic, eBureau.
Other	Many companies specialize in very different services, ranging from online/offline matching, e-mail intelligence and people search to threat intelligence based on the indexing of web content. Examples: Intelius, PeekYou, Rapleaf, Recorded Future.

Table 20: Typology of data brokers, adapted from Bria et al (2015)

According to Bruce Schneier (2015), four basic surveillance streams existed before the Internet: companies keeping records of customers, direct marketing, credit bureaus and public records from government. Today's data brokers combined these four streams. But there are many types of companies offering several types of services in the personal data economy – from large generalists to small specialists.

Nontransparent networks of data brokers

That is why the categories offered by the previous typologies overlap with each other in parts. There are companies that have emerged from payment, credit scoring and fraud detection but then entered the **marketing data sphere**. Others originated from market research but started to aggregate more and more data on an individual level, and ended up developing **predictive analytics and scoring products** – including financial scores on individuals. Major database and software vendors like *Oracle* became data brokers (see chapter 5.7). The social networking giant ***Facebook*** doesn't just market its user data, but has also, for example, registered a patent about credit scoring.[364] Although they are not mentioned in the above typologies, mobile carriers also "manage, package and sell" their customer data – the "global market for **telco data** as a service is potentially worth $24.1 billion this year".[365] But it is not only large corporations that are active in this market. There are millions of small companies uploading their customer data to nontransparent networks of data brokers and merging it with data gathered online in real-time, often without the knowledge of consumers.

[364] Meyer, R. (2015): Could a Bank Deny Your Loan Based on Your Facebook Friends? The Atlantic. Online: http://www.theatlantic.com/technology/archive/2015/09/facebooks-new-patent-and-digital-redlining/407287/ [25.01.2016]
[365] http://adage.com/article/datadriven-marketing/24-billion-data-business-telcos-discuss/301058/ [25.01.2016]

5.4 Observing, inferring, modeling and scoring people

Types of personal data collected

Personal information processed by companies can be grouped in different ways, for example based on how it is obtained:

- **Volunteered data** (also: **declared data**, **opt-in data**) is "created and explicitly shared by individuals, e.g., social network profiles" (WEF 2011, p. 37). Users provide it "when they sign up for service". It is the data type "which users are most aware" of. Users provide it, for example "when transacting, or registering for a service" (CMA 2015, p. 21). Volunteered data could be, for example, simply an e-mail-address, but also an "array of demographic information" (Busby et al 2012).

- **Observed data** is "captured by recording the actions of individuals, e.g., location data when using cell phones" (WEF 2011, p. 37). Consumers "generate and supply it passively" (CMA 2015, p. 21). In the online context, **first-party observed data** are "gathered as users surf the Web" while **third-party observed data** are "purchased from other websites that have done the collecting" (Busby et al 2012).

- **Inferred data** are "data about individuals based on analysis of volunteered or observed information, e.g., credit scores" (WEF 2011, p. 38). They are "assumptions that third-party ad networks and agencies make" based on observed data combined with volunteered data, and they are "notoriously unreliable" (Busby et al 2012).

The content of the data

The CMA (2015, pp. 24-25) compiled a non-exhaustive list of types of personal information that are directly or indirectly collected by companies:

- **financial** – such as information on income and credit ratings;
- **contact** – such as an individual's home or work address, their email address, and phone number;
- **socio-demographic** – such as age, ethnicity, gender, occupation and social class;
- **transactional** – such as purchases made with loyalty cards or transactions completed online and the prices paid;
- **contractual** – such as service details and history maintained by utility suppliers;
- **locational** – such as location data shared by mobile devices, vehicle telematics, GPS data, planned journeys entered into satnavs, and sensor data collected from radio-frequency identification (RFID) tags;
- **behavioral** – such as websites visited and adverts clicked on, data on consumers' use of games apps, and telematics data captured by motor insurance companies;
- **technical** – such as Internet Protocol (IP) addresses and device data such as the IMEI (International Mobile Equipment Identity);
- **communications** – such as entries in social media and in email exchanges;
- **social relationships** – such as the links between family members and friends

Online user data

In the *Boston Consulting Group*'s paper "The evolution of online-user data" they listed **demographic** data including attributes such as age, gender and income, **behavioral or contextual** data including user's interests and attitudes, **purchase intention** data measuring "a person's plans to make a specific purchase", **user location** data, and **social** data describing "the relationship a person has with other people" – in marketing it is often assumed that "people who are connected" have similar attributes (see Busby et al 2012).

Personal data can also be categorized on the basis of the type of relationship that the collecting company has with the consumer (CMA 2015, p. 34):

- **First-party data** is collected by businesses, which have a direct relationship with consumers. It is collected "directly and exclusively from consumers through interactions", for instance "during a transaction for a product or service in a shop".

Examples include: electronic Point of Sale (ePOS) data, collected by retailers in combination loyalty card data

- **Third-party data** is acquired either "from a first party or another third party through purchase, licensing or exchange", or it is collected "by gathering publicly available data from public records or by analyzing social media". Third parties, however, sometimes use "their own cookies which are installed on a user's device when they visit a first-party's website". Companies that process and analyze data on behalf of other companies are also considered as "third parties". Companies that "acquire data from first parties are sometimes called 'second parties'".

Where do firms collect data from?

Accenture asked about 600 businesses in a survey which sources they are "routinely collecting data" from. The result was that 79% are collecting data "directly from individuals themselves" and 42% are collecting it "directly from other organizations (e.g. through commercial or data-sharing agreement". 33% are collecting data "from connected devices" or "purchase" it "from third-party data suppliers" (Cooper et al 2016, p. 8).

Analyzing consumers

The data broker report of the *U.S. Senate Committee on Commerce, Science, and Transportation* (2013, p. 20) additionally differentiates between **actual data**, including "factual information about individuals, such as their date of birth, contact information, and presence of children in a household", and **modeled data,** which "results from drawing inferences about consumer characteristics or predicted behavior based on actual data". Many companies offer **segments**, which are "groupings of consumers defined by shared characteristics and likely behaviors". The idea of segmenting consumers dates back to the 1970s, when market research and geodemographic segmenting products like PRIZM came up. While in its early ages, consumer segmentation was mainly based on large-scale information such as census data, today's segmenting systems can use detailed individual-level data about billions of consumers and apply advanced analysis technologies.

Predicting future behavior

Another type of data product offered by data brokers and analytics companies is **scoring**. According to the Senate Committee's report, scoring products utilize data "to make predictions about likely consumer behavior" (ibid., p. 23). They are "designed to provide marketers insight about existing and prospective customers by assigning a number or range that signifies each consumer's likelihood to exhibit certain characteristics or perform certain actions". The idea of assigning a number to individuals to predict future behavior is well known from **credit scoring**, which has been around for decades.

Problems of credit scoring

The **FICO score**, one of the earliest credit scoring products and still one of the most important ones in the U.S., is today based on the consumer's payment history, the amounts owed, the length of credit history, the "mix of credit cards, retail accounts, installment loans, finance company accounts and mortgage loans" in use, and the type and frequency of opening "new accounts" or applying for "new credit".[366] Citron and Pasquale (2014, p.10) summarized the problems of credit scores for consumers: "their opacity, arbitrary results, and disparate impact" on different population groups, often systematizing discriminatory practices.

Most credit scores are based on information collected by credit report agencies. In 2012, the FTC reported that 26% of survey participants in its study of credit report accuracy "identified at least one potentially material error on at least one of their three credit reports" (FTC 2012, p. *i*). According to a German study published by both the Federal Ministry of Justice and Consumer Protection and the Federal Ministry of the Interior,

[366] http://www.myfico.com/CreditEducation/WhatsInYourScore.aspx [25.01.2016]

credit scores are often based on estimations and their validity on an individual level is questionable[367] (see ULD, 2014).

Rate, rank and segment consumers

The *World Privacy Forum*'s report "The Scoring of America" (Dixon et al 2014, p. 8) summarizes the history of scoring and offers a detailed analysis on how **consumer scores** are widely used beyond credit scoring today – in many fields from marketing to healthcare. According to their definition, a consumer score "describes an individual or sometimes a group of individuals (like a household), and predicts a consumer's behavior, habit, or predilection." Consumer scores "rate, rank, or segment consumers" based on information about "consumer characteristics, past behaviors, and other attributes in statistical models". These models are used by businesses and governments "to make decisions about individual consumers and groups of consumers", and the "consequences can range from innocuous to important". Consumer scores are used for many different purposes "from predicting fraud to predicting the health care costs of an individual to eligibility decisions to almost anything".

Hundreds of secret consumer scores

While in 2007, the *World Privacy Forum* identified less than 25 types of consumer scores, their research in 2014 "uncovered hundreds of scores, with the strong likelihood that thousands of custom scores exist". Some examples of consumer scores that they examined (Dixon et al 2014, p. 19):

- **Consumer profitability scores** "predict, identify, and target marketing prospects in households likely to be profitable and pay debt".
- The **Job Security Score** "claims to predict future income and capacity to pay".
- **Charitable Donor Scores** "seek to classify and rank those who donate to charities".
- **Churn scores** "seek to predict when a customer will move his or her business or account to another merchant".
- The **Medication Adherence Score** predicts if people are "likely to take" their "medication according to [their] doctor's orders".
- Some **Frailty Scores** can "predict mortality within one year". While these scores "can predict care needs, the scores can also be used to simply project costs, and this raises questions about possible misuse in non-health scores or marketing activities".
- **Fraud Scores** are "used everywhere from the Post Office at point of sale to retailers at point of sale to behind-the-scenes credit card transactions".

The U.S.-based company *Social Intelligence* offers even a "Smoker Assessment Score" and a "Substance Abuse Score", which "provides a real-time assessment of an applicant's substance abuse risk".[368]

„Alternative" scoring

In 2014, the FTC hosted a seminar on "alternative scoring products" where "the speakers described a variety of predictive analytics products offered by many data brokers to predict trends and consumer behavior in a variety of contexts, ranging from identity verification and fraud prevention to marketing and advertising". The FTC explains that "consumers are largely unaware of these scores". Thus, these predictive scores "raise a variety of potential privacy concerns and questions".[369]

---

[367] In German: „Es wird festgestellt, dass die Datenschutznovelle für Banken und Auskunfteien Auswirkungen hinsichtlich ihres Auskunftsverhaltens hatte, dass Beeinträchtigungen der Verbraucherrechte aber weiter bestehen. [...] Scores basieren auf Schätzungen, deren individueller Aussagegehalt oft fragwürdig ist."

[368] http://socialintel.com/life/ [22.08.2016]

[369] https://www.ftc.gov/news-events/events-calendar/2014/03/spring-privacy-series-alternative-scoring-products [25.01.2016]

5.5 Data brokers and online data management platforms

According to the *United States Government Accountability Office,* **data brokers** or **information resellers** are companies "that collect information, including personal information about consumers, from a wide variety of sources for the purpose of reselling such information to their customers, which include both private-sector businesses and government agencies" (GAO 2006, p. 1).

Data brokers, the unknown force

Until recently, these companies, which often have extensive dossiers on large parts of the population, were little known to the public. Despite the low affinity for data protection and the non-existent "right to informational self-determination" as it is defined in Europe, an increasing media coverage and public debate about the practices of these companies came up, also in the United States.

Massive amounts of data

After a report on data brokers by the *Senate Committee on Commerce, Science, and Transportation* (2013) the U.S. *Federal Trade Commission* published a report on data brokers in 2014, which examined nine companies, *Acxiom, Core Logic, Datalogix, eBureau, ID Analytics, Intelius, PeekYou, Rapleaf* and *Recorded Future* (see FTC, 2014). As a result, the FTC stated that **data brokers** collect "massive amounts of data" about consumers from "online and offline sources" and "combine them into profiles about each of us", "largely without consumers' knowledge" (FTC 2014, p. C-3). Information collected includes data such as purchase behavior, browsing history, social media activities, health status, religious and political affiliation. The following table shows some examples of how data about consumers is collected:

Consumers…	Their data is collected by…
… post information publicly online	data brokers
…shop online	online stores
…register on websites	websites
…shop at stores	stores
…fill out warranty cards	companies
…buy houses	local governments

Table 21: Data Brokers in the U.S.: examples for the ways of personal data. Source: FTC 2014, p. 2

The nine companies examined in the FTC report obtain their data from public authorities as well as from public and commercial sources. Some examples (FTC 2014, p. 11 et seq.):

Government sources	Public sources	Commercial sources
Professional licenses e.g. pilots, doctors, lawyers, architects	Telephone and other directories	Information from telephone companies
Recreational licenses e.g. hunting, fishing	Press reports	Information from automobile dealers
Real property and assessor records e.g. taxes, assed values, deeds, mortgages etc.	Publicly available information from the internet e.g. social media sites and blogs (via web crawler)	Purchase history from merchants
Voter registration information e.g. name, address, date of birth, party affiliation		Online or offline marketing surveys, warranty registrations and contests
Motor vehicle and driving records		
Court records e.g. criminal records, birth, marriage, divorce, death records, civil actions and judgments		

Table 22: Examples for sources, which data brokers in the U.S. collect data from. Source: FTC, 2014

Information from public authorities has played an important role for data brokers in the U.S. for decades. Sometimes, this data is not directly obtained from the authorities but purchased from companies or even collected manually through visits to local authorities. Since there is no central population register in the U.S., the data retrieved from voter registrations or driver's license data is an important source of verified, basic information about individuals. In addition, data brokers obtain information from other data brokers that, for example, aggregate the "purchase history of 190 million individual consumers from more than 2600 merchants" (ibid., p. 14).

Sensitive inferences

Data brokers combine and analyze the collected data to make inferences about them. They use this data to create products that **"define consumers in categories"** (Senate Committee on Commerce, Science, and Transportation 2013, p. *ii*). According to the FTC, potentially sensitive information is inferred about individuals, for example about their ethnicity, income, health and parenting. Some data brokers store the collected data for an **unlimited period**, and combine **online and offline data** to market to consumers online. The data collected is often passed on by several companies. Seven of the nine data brokers investigated by the FTC **provide data to each other**. In consequence, it would be "virtually impossible for a consumer to determine how a data broker obtained his or her data"(FTC 2014, p. 14).

Online tracking

In addition to **large data brokers** that often already exist for many decades, such as *Acxiom*, there are many **new players** in the fields of online tracking and targeted advertising which collect vast amounts of personal information. Many of the companies involved are largely unknown to the public, but often record every click on the Internet and every interaction on mobile devices. Some websites and mobile apps transmit information to more than 200 different companies at once. These third parties, to whom information about website visits is transferred to, are often largely unknown ad networks and web analytics services, but also prominent companies like Google, *Facebook* or *Twitter* (see chapter 0).

How to send user data to 100 services

The U.S. company ***Segment*** claims to "help thousands of companies collect data"[370], and promotes its service as follows: "Send your data to over 100 apps with the flip of a switch". Developers can easily embed *Segment*'s service into their website and mobile apps. After installation, the embedded services automatically send data about the users' behavior to more than 100 third-party companies, without this being visible to users in any way. As each integration offered by *Segment* comes with a certain amount of effort, their available third-party tracking services probably cover some of the most popular ones in the market, ranging from advertising and analytics to CRM services.[371]

3,000 tracking companies

Actually, there are thousands of companies and services, to whom personal data from both website visits and from the use of smartphone apps is transferred. At this point in time, the sector is rather nontransparent and little is known about most of these companies, which might be due to a lack of systematic research as well. On its website, the privacy service *Ghostery* lists nearly 3,000 companies, to whom data from websites or apps is transferred to on a regular basis. The list contains activities and short descriptions of the companies and, in some cases, information on the use of the data, links to privacy policies and possibilities to Opt-Out.[372]

Data Management Platforms (DSPs)

Many different types of companies exist in the **online advertising and tracking business**, such as ad servers, ad networks, ad exchanges, supply-side platforms (SSP),

[370] https://segment.com/ [25.01.2016]
[371] https://segment.com/integrations [25.01.2016]
[372] http://www.ghosteryenterprise.com/company-database/ [25.01.2016]

demand-side platforms (DSP), and data management platforms (DMP).[373] It is beyond the scope of this report to cover all of these concepts, however the latter are especially relevant here. **Data management platform (DMPs)** are essentially real-time online data brokers, they are the "central hub" used to "seamlessly (and rapidly) collect, integrate, manage and activate those large volumes of data" and to "aggregate, integrate, manage and deploy disparate sources of data".[374] According to a *Gartner* blog article, they offer companies the ability to:[375]

- **import data**, for example "information about customers, such as their customer ID or email address (to identify them), what they have bought or looked at, their loyalty status" and "demographic and other characteristics", for example from marketing systems, e-mail, e-commerce and customer loyalty platforms
- **match customer IDs**, for example "two data sources with a common field like a customer ID or email address (or anonymized ID)" can be "stored by the DMP as belonging to the same person"
- **collect new data**, for example by putting "tags" on the company's website, emails, advertisements, and mobile apps
- **provide access to data vendors,** for example "pre-defined or custom segments of (anonymous) people to target"
- **find segments** with specific characteristics, sometimes also called "audiences" (by analyzing and categorizing users)
- **suggest new groups of people to target,** for example "by finding people who look like your current customers", so-called "lookalikes"
- **send instructions** about "who to target", "with what message", and "in what channel" or on "what device", for example to target ads or to personalize websites

Example DMPs

Datanyze, a website offering market share data based on a "web crawler to detect the presence of a technology"[376], lists the following data management platforms:[377]

> *LiveRamp (Acxiom), DataLogix (Oracle), eXelate (Nielsen Display Ads), Lotame, [x+1] (Rocket Fuel), Bluekai (Oracle), AudienceScience, Krux, Acxiom, Digilant, Flxone, Navegg, TailTarget, Platform 161, I-Behavior, Eyeota, Sojern, Brilig, NuggAd, Enreach, Adobe Audience Manager, Blueconic, Crowd Science, Epsilon*

List brokers and online lead generation

According to a report by Led Astray (2015, p. 2), a rather controversial business sector relying on the online advertising ecosystem and data brokers is the field of **online lead generation**, which "is the business of selling leads — pieces of evidence that a consumer is interested in a product or service". Lead generators "encourage consumers to provide information about themselves" and often "sell consumers' data to businesses that offer risky financial products and other controversial services". They collect "sensitive financial information from vulnerable and often desperate consumers" to offer them, for example, payday loans. Data brokers have collected and sold extensive **lists of names and addresses** of consumers grouped by specific characteristics for decades, including lists of people "suffering from conditions including cancer, diabetes, and depression, and the

---

[373] For basic explanations see: https://www.clickz.com/clickz/column/1931527/dsps-ssps-rtbs-dmps-online-medias-alphabet-soup [25.01.2016]
[374] Winterberry Group (2012): The Data Management Platform: Foundation for Right-Time Customer Engagement. A Winterberry Group Whitepaper. Online: http://www.iab.net/media/file/Winterberry_Group_White_Paper-Data_Management_Platforms-November_2012.pdf [25.01.2016]
[375] http://blogs.gartner.com/martin-kihn/data-management-platform [25.01.2016]
[376] http://www.datanyze.com/faq/ [26.01.2016]
[377] http://www.datanyze.com/market-share/dmp/ [26.01.2016]

medications used for those conditions; another is offering lists naming consumers, their credit scores, and specific health conditions" (Senate Committee on Commerce, Science, and Transportation 2013, p. 5)[378]. Traditionally, these lists have been compiled, for example, from mail order customers, magazine subscribers and sweepstake entries, but they are nowadays also created or enriched by analyzing, segmenting and scoring the extensive databases from online data brokers.

5.6 Cross-device tracking and linking user profiles with hidden identifiers

As described in the previous chapter, online data management platforms allow companies to import their customer data, combine it with millions of detailed third-party user profiles from online and offline sources, to identify their own customers or to target other individuals through online or offline channels. These platforms often offer to analyze, segment and score consumers, and they are connected to other data brokers and advertising companies.

„Anonymous" identification?

To recognize website visitors across several tracking companies the platforms cooperate with each other and use for example **cookie synching**, which refers to "the process of mapping user Ids from one system to another".[379] This way, they can **match user identifiers across different systems** such as ad networks, ad exchanges and data providers. But today's data management platforms offer more than just the identification of people surfing the web. They promise to identify consumers in many life situations by matching profile data from different sources and mostly claim that this matching is "anonymous".

Matching is a crucial point for consumer privacy

Unique identifiers

Unique identifiers for consumers are often derived from their e-mail address or their phone number. Based on these identifiers, data records from corporate customer databases can be linked with third-party profiles from data brokers, social network profiles, online and mobile behavior gathered via tracking services or cookies, and any other device, platform or service consumers are using in everyday life. To create these unique identifiers, most vendors use **hashing**. They convert, for example, an email address into an alphanumeric string by using a cryptographic function such as MD5.[380]

Anonymous matching via hashing?

In theory, hashing is a one-way operation and cannot be reversed. But, aside from many other possible ways of de-anonymization[381], when we imagine real-time data sharing between all kinds of companies collecting hundreds of millions of e-mail addresses and **all of them use the same "one way" operation** to "anonymize" these e-mail addresses, these "anonymized" email addresses can be matched across different datasets. Consumers can therefore be recognized again, as soon as they use a service linked with the same email address. Although some of the companies and organizations involved in these data sharing processes may not know the name or address of consumers, they can always identify them as the same person in many situations and link their profile to comprehensive information based on volunteered, observed, inferred and modeled data.

---

[378] The data broker report from the Senate Committee on Commerce, Science, and Transportation (2013) provides a comprehensive overview on list brokers. In the author's German 2014 report "Kommerzielle Digitale Überwachung im Alltag" German list brokers have been investigated.
[379] https://www.admonsters.com/blog/cookie-synching [25.01.2016]
[380] https://www.clickz.com/clickz/column/2288689/whats-an-email-hash-anyway [25.01.2016]
[381] See chapter 2.3

Hashed identifiers can be personal data

According to the major U.S. privacy compliance company *TRUSTe*, a **hash is in fact personally-identifiable information (PII)**, when the "entire reason for keeping the hashed data is to be able to identify a discrete user the next time they return to the site". It may be a "good security rationale" when a service "cannot recover the user's email address and name" and "associated data will only be recovered when the user next enters their email address and name", but it "fails to understand the privacy implications by ignoring the definition of PII".[382]

The Marketing scholar Joseph Turow concluded:

> *Industry claims of anonymity surrounding all these data may soften the impact of the sorting and labeling processes. But in doing so, it seriously undermines the traditional meaning of the word. If a company can follow and interact with you in the digital environment – and that potentially includes the mobile phone and your television set – its claim that you are anonymous is meaningless, particularly when firms intermittently add offline information to the online data and then simply strip the name and address to make it "anonymous."*[383]

There is no anonymous identification

Data management platforms and their clients often describe the data used in their tracking and sharing processes as "anonymized" or "de-identified". Apparently, hashing is in fact **pseudonymization** rather than anonymization. In *Adobe*'s digital marketing magazine CMO, Ruth Boardman, a "leading privacy lawyer", suggests that "marketers should stop trying to convince themselves they are working with anonymised data, rather than personal information"[384]. Iain Bourne from the *Information Commissioner's Office*, the UK's privacy regulator, adds: "It's not really worth having a long debate about whether this is not personal information when it's aimed at identifying people" (ibid.)

Singling out people without knowing their names

In his paper "Singling out people without knowing their names" the privacy scholar Frederik Borgesius (2016, p. 2) concluded that European data protection authorities "take the view that a company processes personal data if it uses data to single out a person, even if it cannot tie a name to these data". A name would accordingly be "merely one of the identifiers that can be tied to data about a person, and it is not even the most practical identifier" for today's online tracking economy.

Deterministic and probabilistic cross-device tracking

Since consumers are increasingly using multiple devices, **matching data from different devices** is considered as a major challenge for tracking and data companies. When Internet users are surfing the web, websites can identify them again as the same user, for instance, via cookies (see chapter 4). But browser cookies can easily be deleted or blocked, computers can be used by multiple persons, people can use multiple computers, and cookies don't help to track the use of mobile apps and other upcoming devices, from **game consoles** and **smart TVs** to **fitness trackers**. To match these different devices, the tracking industry needs more than cookies. In 2015, Omar Tawakol, the general manager of *Oracle*'s *Data Cloud* stated[385] that marketers had "these first-party data assets: data tied

---

[382] http://www.truste.com/blog/2013/04/16/data-anonymization/ [25.01.2016]
[383] Joseph Turow (2011): The Daily You. Cited from: Senate Committee on Commerce, Science, and Transportation (2013, p. 32)
[384] CMO. by Adobe (2015): Adobe Summit EMEA: Brands Advised To Always Assume It's Personal. Online: http://www.cmo.com/features/articles/2015/4/29/adobe-summit-emea-brands-advised-to-always-assume-its-personal.html [16.08.2016]
[385] eMarketer (2015): Cross-Device Targeting and Measurement Will Impact Digital Display Advertisers in 2015. Online: http://www.emarketer.com/Article/Cross-Device-Targeting-Measurement-Will-Impact-Digital-Display-Advertisers-2015/1012081 [25.01.2016]

to email, data tied to a physical address, data tied to cookies—and they're all massively disconnected." That is why "cross-linking everything across screens and devices" would be "the biggest and most important trend this year". In 2015, the FTC has initiated an investigation on cross-device tracking and consumer privacy.[386]

Linking user accounts and device identifiers

The most important approach is **deterministic cross-device matching**, for example by identifying users when they have logged into one of the major platforms with millions of users, such as *Google, Facebook* or *Twitter*.[387] This way, it is possible to link information gathered on users across platforms and devices, based on identifiers such as hashed email addresses, mobile identifiers and cookies from different tracking companies. To link mobile device identifiers like the **Apple Identifier for Advertisers** (IFA/IDFA) or the **Google Advertising ID** can be used.[388] **Microsoft**'s "Advertising ID" helps tracking *Windows* phone and *Windows* users.[389] During the past years, many companies providing platforms and services introduced unique identifiers for users. ***Verizon*** introduced its "Precision ID", a "unique hashed ID" which has, according to an article published by the trade magazine *adexchanger*, partnerships with data brokers like *Experian* and *Oracle* "to enable anonymous matches between the Precision ID identifier and third-party data".[390] Some data brokers, data management platforms and other online advertising vendors also introduced their own unique identifiers, such as *Acxiom*'s **AbiliTec Link**[391] and the **Oracle ID Graph**[392].

Probabilistic cross-device matching

According to adexchanger's "Marketer's Guide To Cross-Device Identity",[393] another approach is to use **probabilistic cross-device matching**, offered by companies such as *Tapad, Drawbridge* and *Experian*. Probabilistic matching is

> *"achieved by algorithmically analyzing thousands of different anonymous data points – device type, operating system, location data [...], time of day [...] – to create statistical, aka likely, matches between devices. For example, if a phone, a tablet and a laptop connect to the same networks or Wi-Fi hotspots in the same places every weekday, it's safe to surmise that all three devices belong to a specific commuter."*

The guide lists the following providers for cross-device tracking:

> *Acxiom, Adelphic, Adobe, AOL, BlueKai (acquired by Oracle), Conversant (acquired by Epsilon/Alliance Data), Criteo, Crosswise, Iris Mobile, Krux, Lotame, MediaMath, Neustar Aggregate Knowledge, Turn, 4INFO, [x+1] (acquired by Rocket Fuel).*

Ultrasonic audio signals to link devices

Another way to link two devices belonging to the same consumer with each other is to use "apps that can hear TV sounds, QR codes, NFC" and other "data links".[394] Companies even

---

[386] https://www.ftc.gov/news-events/events-calendar/2015/11/cross-device-tracking [25.01.2016]
[387] https://iapp.org/news/a/cookies-are-so-yesterday-cross-device-tracking-is-insome-tips/ [25.01.2016]
[388] http://adexchanger.com/data-exchanges/a-marketers-guide-to-cross-device-identity/ [25.01.2016]
[389] https://msdn.microsoft.com/en-us/library/windows/apps/windows.system.userprofile.advertisingmanager.advertisingid [25.01.2016]
[390] http://adexchanger.com/data-exchanges/can-you-identify-me-now-a-deep-dive-on-verizons-data-practices/ [25.01.2016]
[391] https://developer.myacxiom.com/code/api/endpoints/abilitec-link [25.01.2016]
[392] Oracle Inc. (2015): Oracle buys Datalogix. Online: http://www.oracle.com/us/corporate/acquisitions/datalogix/general-presentation-2395307.pdf [25.01.2016]
[393] http://adexchanger.com/data-exchanges/a-marketers-guide-to-cross-device-identity/ [25.01.2016]

use "ultrasonic audio through the use of the speakers on the computer or device", which is then "recognized and received on the other smart device". They use it not just to match users "cross-device", but also "cross-channel", and embed "**audio beacon signals into TV commercials**", which are received by tracking apps and allow users who were exposed to a specific TV program to be identified.[395]

Cross-device identifiers by major platforms?

Google, Twitter and Facebook's Atlas

However, the most relevant data sharing occurs between companies connecting their **customer and CRM data** with the online tracking universe. Both *Google*[396] and *Twitter*[397] started to allow CRM matching in 2015. *Facebook* started to offer companies the ability to match their customer data already back in 2012.[398] *Facebook*'s **Custom Audiences** product allows companies to upload "hashed" email addresses or phone numbers to target these customers online.[399] In early 2013, *Facebook* started to partner with data brokers such as *Acxiom, Epsilon, Datalogix* and *BlueKai* (the latter now both owned by *Oracle*). *Facebook* also acquired the "giant ad-serving and measurement business"[400] ***Atlas***, which will, according to a company statement, "solve the cross-device problem" and help companies "reach real people across devices, platforms and publishers"[401] An *Atlas* representative explained that the "data that Facebook has on its 1.3 billion users is data that we can use in Atlas"[402]. An *Atlas* whitepaper states that "Facebook syncs the Atlas and Facebook cookies" by writing "a version of the user's Facebook ID into the Atlas cookie"[403].

Facebook's cross device ID

According to the Wall Street Journal, *Facebook*'s *Atlas* could "help tie online ads to offline sales". For example, a "consumer who purchases a pair of shoes in a store might volunteer her email address at the checkout. If the email address is linked to a *Facebook* account, *Facebook* could inform the retailer, if, when and where the consumer saw its ads across the web".[404] A marketing magazine summarized that ***Facebook*** had in fact introduced a **"cross-device ID based on logged in users"**, which would not only work on *facebook.com*, *Facebook*'s mobile app and *Instagram*, but also on "thousands of other websites and apps". They would use a "combination" of their "Facebook ID" and "mobile device identifiers such as Apple's Identifier for Advertising (IDFA) and Android's

---

394 http://adexchanger.com/data-driven-thinking/cross-device-tracking-dont-believe-the-hype/ [25.01.2016]
395 Calabrese, C., McInnis, K. L., Hans, G. S., Norcie, G. (2015): Comments for November 2015 Workshop on Cross-Device Tracking. Center For Democracy & Technology. Online: https://cdt.org/files/2015/10/10.16.15-CDT-Cross-Device-Comments.pdf [25.01.2016]
396 http://adexchanger.com/mobile/google-allows-targeted-ads-based-on-first-party-data/ [25.01.2016]
397 Weber, H. (2015): Twitter's new 'partner audiences' will help more advertisers track you outside Twitter. VentureBeat. Online: http://venturebeat.com/2015/03/05/twitters-new-partner-audiences-will-help-more-advertisers-track-you-outside-twitter/ [25.01.2016]
398 Constine, J. (2012): First Results Are In: Facebook's New Custom Audience CRM Ads Increase Conversions And Lower Costs. TechChrunch. Online: http://techcrunch.com/2012/10/11/facebook-custom-audience-ads/ [25.01.2016]
399 https://www.facebook.com/ads/manage/customaudiences/tos.php [25.01.2016]
400 Edwards, J. (2013): This Is What Facebook Thinks The Future Of Cookies Look Like. Business Insider. Online: http://www.businessinsider.com/facebook-cookie-ads-from-atlas-2013-12 [25.01.2016]
401 http://atlassolutions.com/2014/09/29/meet-the-new-atlas/ [25.01.2016]
402 eMarketer (2014): Bye-Bye, Cookies-Atlas Tracks Consumers Online and Offline via Facebook IDs. Online: http://www.emarketer.com/Article/Bye-Bye-CookiesAtlas-Tracks-Consumers-Online-Offline-via-Facebook-IDs/1011661 [25.01.2016]
403 Atlas Solutions, LLC (2015): The Case for People-based Measurement & Delivery. Online: https://atlassolutionstwo.files.wordpress.com/2014/12/the_case_for_people_based_measurement_final_1-15.pdf [25.01.2016]
404 Marshall, Jack (2014): What Marketers Need to Know About Facebook's Atlas. Wallstreet Journal, Sep 29, 2014. Online: http://blogs.wsj.com/cmo/2014/09/29/what-marketers-need-to-know-about-facebooks-atlas/ [25.01.2016]

Advertising ID". However, according to the article, *Atlas* still does "not allow its cross-device data to leave Facebook's walls".[405]

5.7 Case studies and example companies

There is little solid research about the practices of companies doing business with personal data of consumers. In addition, this ecosystem is evolving and changing very fast. Within a few months, a company that is examined by researchers or media, may have been acquired, merged with other companies or rapidly changed the services offered.

Companies in different fields

Based on the different typologies of the personal data ecosystem as described in the previous chapters and on a review of literature and media articles, several companies were selected for a more detailed examination. Starting with Busby et al (2012), Chester (2014), CMA (2015), Deighton (2013), Dixon et al (2014), FTC (2015), McConville (2014) and Senate Committee on Commerce, Science, and Transportation (2013) also several lists of clients and partners of companies which were mentioned in these publications, have been examined. We selected large, mid-sized, and small companies, ranging from generalists offering diverse portfolios to smaller specialists. Most of them are based in the U.S., but some are offering their services globally. We also included companies based in Germany, Netherlands and India. *Experian*'s headquarters are in Dublin, Ireland.

The goal of this review of products and services in the case studies and exemplary companies is to answer questions such as: What types of personal data about consumers do they collect, analyze and sell? How many consumers are affected? Which products do these companies offer? How "anonymous" is the collection, matching and exploitation of personal data really when companies claim to use anonymization? And what are the implications and risks for consumers?

5.7.1 Acxiom – the world's largest commercial database on consumers

Up to 3,000 attributes on 700 million consumers

The U.S. company *Acxiom* claims to have "data and insight" into "700 million consumers worldwide", including into over "3,000 propensities[406] for nearly every U.S. consumer".[407] According to the New York Times, "few consumers have ever heard of *Acxiom*, but analysts say it has amassed the world's largest commercial database on consumers".[408] The company was founded in 1969, under the name *Demographics Inc*., and initially performed direct mail campaigns based on the data of public phone books, including advertisement for election campaigns. Today *Acxiom*, according to their annual report 2014, manages **2.5 billion "customer relationships"** and maintains 3.7 "billion prospect records" for clients. They operate 15,000 customer databases for 7,000 companies from sectors such as finance, insurance, retail, healthcare, travel and automotive, and work for "47 of the Fortune 100" companies, but also for U.S. government agencies. Shortly after the FBI had

---

[405] Rodgers, Z. (2014): With Atlas Relaunch, Facebook Advances New Cross-Device ID Based On Logged In Users. AdExchanger. Online: http://adexchanger.com/platforms/with-atlas-relaunch-facebook-advances-new-cross-device-id-based-on-logged-in-users/ [25.01.2016]

[406] According to Merriam-Webster, a propensity is a "strong natural tendency to do something" or an "often intense natural inclination or preference", see: http://www.merriam-webster.com/dictionary/propensity [01.08.2016]

[407] Acxiom Corporation (2014): Annual Report 2014. Online: http://files.shareholder.com/downloads/ACXM/0x0x763250/A1DBFBD8-E136-4701-B0F2-3DC695E5ED08/acxiom2014_Annual_Report_FINAL_RRD_PDF_.pdf [22.01.2016]

[408] Singer, Natasha (2012): You for Sale. Mapping, and Sharing, the Consumer Genome. New York Times, June 16, 2012. Online: http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html [22.01.2016]

released the names of the 19 hijackers on 11 September 2001, Acxiom had identified eleven of them in their own databases.[409] *Acxiom* has been active in **Germany since 2004** and has already collected data on 44 million Germans.[410] According to a talk by Acxiom's Managing Director for Germany and Poland, the company possesses "offline profile data" on "nearly every household in Germany".[411] ***Acxiom UK*** claims to have data on "over 40 million consumers" and "1,000 lifestyle demographics and behavioral variables".[412]

Data catalog

Acxiom's "Consumer Data Products Catalog" from 2011[413] lists hundreds of "data elements" which corporate clients can obtain about individuals or households to complete their customer databases. In addition to basic information such as name, age, gender, phone numbers, email addresses, education, occupation, children, income and credit card use, detailed records on housing and vehicle ownership are available. In the "geography and address" category, 25 different attributes are available, in the "ethnicity" category ten attributes – for example several **"race codes"**. In addition, data on voting party and "interests" such as "dieting/weight loss", "casino", "gambling", "lotteries" or "smoking/tobacco" are available. Data on **health "needs"** such as "allergy related", "arthritis / mobility", "disabled individual in the household" and "diabetic focus" is "derived from purchases and self-reported sources".

Acxiom's Audience Operating System

According to their data catalog, *Acxiom* offers several **"scores"** to categorize people and predict their future behavior, such as "NetWorth Gold", the "Charitable Giving Score", the "Life Insurance Purchase Propensity", the "Consumer Prominence Indicator" and the "Inferred Household Rank". As part of their **analysis and segmentation system** "Personicx", households are assigned to one or more of 1,270 groups describing their lifestyle, based on "specific consumer behavior and demographic characteristics". Sub modules provide data on specific target groups, for example "Personicx Hispanic", "Personicx Insurance Groups" or "Personicx Life Changes", which predicts "consumer life stage changes". "Personicx" is apparently no longer offered separately, but was part of the "Acxiom Audience Operation System (AOS)" until 2015. According to an analysis[414] from *Gartner in 2013*, this system offered a "comprehensive view of an audience" across "channels, devices and media sources" with "unduplicated data, enriched by detailed demographic, contextual, behavioral and social profiles" from "both online and offline activities".

Connecting individual profiles?

Gartner emphasized that *Acxiom*'s technology could "connect individual profiles across devices and channels" and "relate them to a common record containing personally

---

[409] Behar, Richard (2004): Never Heard Of Acxiom? Chances Are It's Heard Of You. Fortune Magazine, 23.02.2004. Online: http://archive.fortune.com/magazines/fortune/fortune_archive/2004/02/23/362182/index.htm [22.01.2016]

[410] Mclaughlin, Catriona (2013): Acxiom. Die Besserwisser. Die Zeit, 05.07.2013. Online: http://www.zeit.de/2013/28/acxiom/komplettansicht [22.01.2016]

[411] YouTube-Video from the panel discussion "Die Strategien der großen Daten-Anbieter" on d3con 2014 conference on 11.02.2014 in Hamburg. In German he said: "Heute haben wir in Deutschland Offline-Profildaten verfügbar für nahezu jeden Haushalt", from Minute 14:50: https://www.youtube.com/watch?v=W41HcRo-3P8 [22.01.2016]

[412] http://dq2qu0j6xxb34.cloudfront.net/wp-content/uploads/2014/01/Facebook-Case-Study-Final-no-bleed.pdf [22.01.2016]

[413] Acxiom: The Power of Insight: Consumer Data Products Catalog. Online: https://www.hashdoc.com/documents/8135/data-products-catalog [22.01.2016]

[414] Frank, Andrew; Kihn, Martin (2013): Acxiom's Audience Operating System Could Reinvent Data-Driven Marketing. Gartner, 26.09.2013. Online: https://www.gartner.com/doc/2597521?ref=ddisp [22.01.2016]

identifiable information" from a "customer database of a company". *Acxiom*'s *AOS* would **eliminate the need for third-party cookies**, the common "method of connecting behavior across websites", which hundreds of companies in online ad targeting still rely on. As third-party cookies are controversial, this would have "some privacy appeal". However, it would be "unclear how the privacy community" would "respond" to *AOS*' use of personally identifiable information and first-party data". *Gartner* stated, that "risk remains high that some marketers and consumers may misuse, mistrust or misunderstand" this technology. In conclusion, it is recommended that businesses "consider strategic options for the increasingly likely scenario" that third-party cookies will be "replaced by large data exchanges operated by companies such as Acxiom and Google".

Matching customer data with online identifiers

In 2015, *Acxiom* announced[415] its new service **LiveRamp Connect**, which "combines the very best elements of *Acxiom*'s Audience Operating System" with the services of *LiveRamp*, a "data onboarding" company acquired by *Acxiom* in May 2014.[416] Back in 2012, *LiveRamp* explained on their corporate blog[417] that client companies send them customer records, which are "keyed by some sort of identifier, such as an email address, postal address, or geographical code". *LiveRamp* matches these customer records to "online identifiers" that are "associated with a browser or device". In 2016, *LiveRamp* offers companies the ability to "use CRM data, purchase histories, and third-party data to address consumers at every stage of their customer journey"[418] and to use "CRM, sales, and third-party data across more than 200 marketing platforms". They claim to "onboard **20 billion records** each month" and offer companies to "match" their "anonymized data to online devices and digital IDs" using "advanced recognition technologies, including **exact one-to-one matching** and Acxiom AbiliTec". In parallel, *LiveRamp* emphasizes that they "anonymize" company's "customer data files through a de-identification process that removes all personally identifiable information".[419]

Exact one-to-one matching with hashed identifiers

Acxiom claims to "anonymize" data records, but offers "exact one-to-one matching" to its clients. While this could be merely a questionable, inconsistent rhetoric, it could also signify that single persons can easily be identified, which is strongly opposed to the principle of "anonymized" data records. According to *Acxiom*'s "Data Services API"[420] their **AbiliTec Link** product does not just "bring together only similar customer records", but is able to "link all customer records", while "assigning one link for each record to allow the most complete view of each customer". It allows "instant recognition of customers" and even "identifies the household of which an individual is a member". *Acxiom's* **"hashed entity representation"** seems to be used as a unique identifier for an individual, based on email addresses, phone numbers or nearly any combination of the former and name, address, city and zip code.[421] The marketing blog *ClickZ* summarized[422] that "marketers are starting to realize that CRM data – specifically the hash of the email address – is an amazing cross-device, cross-platform, cross-browser key". This would overcome many of

---

[415] http://www.acxiom.com/introducing-liveramp-connect/ [22.01.2016]
[416] Kaye, K. (2015): Why Acxiom Killed AOS and Used LiveRamp Name for New Platform. Advertising Age, Feb 24, 2015. Online: http://adage.com/article/datadriven-marketing/acxiom-kills-aos-brand-launches-combined-targeting-platform/297276/ [22.01.2016]
[417] http://liveramp.com/engineering/data-onboarding-system-overview/ [22.01.2016]
[418] http://liveramp.com/why-data-connectivity/ [22.01.2016]
[419] http://liveramp.com/why-liveramp/liveramp-onboarding/ [22.01.2016]
[420] https://developer.myacxiom.com/code/api/endpoints/abilitec-link [22.01.2016]
[421] https://developer.myacxiom.com/code/api/endpoints/hashed-entity [22.01.2016]
[422] Hendricks, D. (2014): What Acxiom „Hash" Figured Out. Clickz, May 21, 2014. Online: https://www.clickz.com/clickz/column/2345821/what-acxiom-hash-figured-out [22.01.2016]

the "challenges of tracking cookies" and that's why "working with *LiveRamp* to verify matches" would enable marketers to reach consumers "across devices, browsers, and services".

Health interests and persuadability

Acxiom's **Data Services API** offers insights into how to "request data for people, places, households, or entities" [423] based on "unique identifiers for person, place, and household documents"[424], for example about "the consumer's insurance behaviors, propensities, and preferences"[425], health "interests"[426], and about the "likelihood" of someone "to be influenced by social media"[427].

Bank data, offline shopping, Google and Facebook

During the past few years, *Acxiom* has started to cooperate with the dominant online companies. They have partnered with both ***Facebook*** and ***Twitter***, for example to "target ads to users on the social networks based on their purchases in stores". Together with ***Google*** they have been working on ways to "match how clicks on *Google*'s ad network tie to in-store sales".[428] In 2016, *Acxiom*'s *LiveRamp* announced a "new integration with Google Customer Match".[429] Regarding **offline behavior of consumers,** *Acxiom* reported to be able to recognize a consumer's identity, when a store clerk captures the shopper's name from a check or credit card at the point of sale and then asks for the shopper's ZIP code or phone number.[430] According to Chester et al (2014, p. 39), the company explained that it could take "bank data" and combine it with information *Acxiom* and "data broker partners provide about a consumer's 'behaviors,' 'email opens,' social media, 'search,' and 'offline' activity. Detailed information regarding an individual" could be "scored and segmented", for example, "knowing that an individual is a 'female with small children, searched on site for travel rewards' and also was 'served … a card ad'".

5.7.2 Oracle and their consumer data brokers Bluekai and Datalogix

According to the trade magazine *adexchanger*, *Oracle*, the world's second-largest software vendor with $29.6 billion in software revenue during 2013[431], has recently become a "strong data-management contender" and a "leader" in "cloud marketing".[432] In 2012 and 2013, *Oracle* acquired *Eloqua*, a marketing automation company, for $871 million and

[423] https://developer.myacxiom.com/code/api/data-bundles/main [22.01.2016]
[424] https://developer.myacxiom.com/code/api/data-bundles/level1bundle/identification-and-linkage [22.01.2016]
[425] https://developer.myacxiom.com/code/api/data-bundles/bundle/insurance [22.01.2016]
[426] https://developer.myacxiom.com/code/api/data-bundles/bundle/healthAndMedical [22.01.2016]
[427] https://developer.myacxiom.com/code/api/data-bundles/bundle/socialMedia [22.01.2016]
[428] Dwoskin, Elizabeth (2014): Data Broker Acxiom Moves to Tie Physical World to Online Data. Wall Street Journal, May 05, 2014. Online: http://blogs.wsj.com/digits/2014/05/14/data-broker-acxiom-moves-to-tie-physical-world-to-online-data [22.01.2016]
[429] http://www.businesswire.com/news/home/20160526005180/en/LiveRamp-Extends-Data-Connectivity-Partnership-Google [01.08.2016]
[430] Singer, Natasha (2012): You for Sale. Mapping, and Sharing, the Consumer Genome. New York Times, June 16, 2012. Online: http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html [22.01.2016]
[431] Kanaracus, C. (2014); Oracle overtakes IBM as second-largest software vendor, Gartner says. Computerworld. Mar 31, 2014. Online: http://www.computerworld.com/article/2489278/it-management/oracle-overtakes-ibm-as-second-largest-software-vendor--gartner-says.html [22.01.2016]
[432] Rodgers, Z. (2015): Answering Your Questions About Google's Forthcoming DMP. Adexchanger, April 24, 2015. Online: http://adexchanger.com/data-exchanges/answering-your-questions-about-googles-forthcoming-dmp/ [22.01.2016]

Responsys, a cloud-marketing platform, for $1.5 billion.[433] In 2014, they acquired the data management platform ***BlueKai*** for a reported $400 million[434] and also ***Datalogix***, one of the companies which was part of the FTC's investigation on data brokers (see FTC 2014), for reported $1.2 billion[435]. In 2016, *Oracle* bought ***AddThis***, a "data company" known for "harvesting updated behavioral data through its share buttons" on more than **15 million websites** worldwide,[436] and ***Crosswise***, which provides "machine-learning based cross-device data" and claims to process "user and device activity data from billions of unique devices every month".[437]

BlueKai, an online data management platform (DMP)

According to a corporate presentation[438] *BlueKai*'s **data management platform (DMP)** allows companies to combine first-party data with third-party data for personalized marketing and online targeting. It also allows "partners to securely share" their customer data "in a mutually beneficial way". Together with *Oracle*, *BlueKai* would enable companies to "build more complete customer profiles, enriched with detailed 1st party data, easily accessible 3rd party data, and new 2nd party partner data". *BlueKai* claims to offer the "world's largest data marketplace for digital marketers" with "access to the largest aggregation of licensed 3rd party data providers available anywhere". Their "Audience Data Marketplace" provides clients "more than 30,000 data attributes including intent, B2B, past purchases, geo/demo, interest/lifestyle, branded and qualified demographics" and "over **700 million global profiles**" from "more than 200 data providers".

Datalogix, purchasing data worth $2 trillion

Datalogix, the second company that **Oracle** acquired, has data "on $2 trillion in consumer spending from 1,500 data partners across 110 million US households" and it "connects offline purchasing data to digital media", according to a presentation[439]. Together with *Oracle,* they could provide marketers "with the richest understanding of consumers" based on "what they do, what they say, and what they buy". According to *EPIC*, *Datalogix* mainly collects data "by forming partnerships with stores who offer membership or loyalty cards".[440] *Oracle* indicates that data originates from sources such as "10+ billion SKU-level[441] transactions across **1500 leading retailers**", "UPC level[442] purchases from 50+ retailers across grocery, club, mass and drugstore" and "3+ billion donation records across US households" (Oracle 2015, p. 14). *Oracle Datalogix* (DLX) offers "data types" such as **"DLX Finance"**. It is used to "reach audiences based on financial behavior including credit cards, home value, net worth, income and more" and is based on

---

[433] Heine, Christopher (2013): Oracle Buys Responsys for $1.5 Billion. Adweek, December 20, 2013. Online: http://www.adweek.com/news/technology/oracle-buys-responsys-15-billion-154630 [22.01.2016]
[434] Dignan, L. (2014): Oracle acquires BlueKai, rounds out its marketing cloud. ZDNet, Feb 24, 2014. Online: http://www.zdnet.com/article/oracle-acquires-bluekai-rounds-out-its-marketing-cloud/ [22.01.2016]
[435] Chernova, Y. (2015): Oracle Paid More Than $1.2 Billion for Datalogix. Wall Street Journal. Feb. 4, 2015. Online: http://www.wsj.com/articles/oracle-paid-more-than-1-2-billion-for-datalogix-1423083774 [22.01.2016]
[436] Kaye, K. (2016): With AddThis Buy, Oracle Gets Pipeline to Continually Update Audience Data. Advertising Age, Jan. 05, 2016. Online: http://adage.com/article/datadriven-marketing/addthis-buy-oracle-pipeline-audience-data/301998/ [22.01.2016]
[437] https://www.oracle.com/corporate/acquisitions/crosswise/index.html [01.08.2016]
[438] http://www.oracle.com/us/corporate/acquisitions/bluekai/general-presentation-2150582.pdf [22.01.2016]
[439] http://www.oracle.com/us/corporate/acquisitions/datalogix/general-presentation-2395307.pdf [22.01.2016]
[440] https://epic.org/privacy/facebook/facebook_and_datalogix.html [22.01.2016]
[441] SKU = stock-keeping unit
[442] UPC = universal product code

information originating from "offline data on 110+ million households from U.S. Census, public record housing & deeds" and "permissible credit header sources" (ibid., p. 15). *Datalogix* categorizes people into "over 1,800 segments" based on "purchase-based data, rich demographics and deep financial insights".[443]

Partnership with Facebook

Datalogix was among the first data brokers partnering[444] with ***Facebook*** in 2012 to allow advertisers to target *Facebook* users based not only on their online behavior, but also on offline data[445], for example based on data about "known buyers" of specific brands, "known online & offline purchase[s]" and "known demographics" such as age, gender and income.[446] *Datalogix's* "Facebook Audience Guide"[447] provides an overview on the Facebook-related product portfolio of the company. One product offered allows companies to "reach consumers likely to respond to mortgage offers" on *Facebook*, because these consumers have "similar profile characteristics" as those "who **applied for mortgages online**". Another one allows targeting consumers on *Facebook* who "have similar profile characteristics of consumers who applied for and purchased auto insurance via an online channel". The **"DLX TV"** product allows *Facebook* users to be targeted based on data from "set-top-box TV exposure data" on 4.2 million U.S. households and on "tracked data from both live & recorded (DVR) viewing". With "DLX OnRamp" companies can pass on their customer data ("any CRM file") to *Datalogix*, which then "matches and converts the file to Facebook users". They explain that this product allows companies to "identify users with multiple email addresses", but in general they "match on 20+ variables including postal address and multiple email addresses".[448]

Oracle Data Cloud: 3 billion profiles

All these products and services now seem to be part of the **Oracle Data Cloud**. A press release[449] explains this product consists of "Oracle Data as a Service for Marketing" (with "access to more than 1 billion profiles globally" and "more than 300 data partners"), and "Oracle Data as a Service for Social" (which "derives insights from more than 700 million social messages daily, across more than 40 million social media and news data sites"). In April 2016, *Oracle* stated that it is "aggregating more than 3 billion profiles from over 15 million websites in its data marketplace".[450] They also use the brand name "Oracle Data Management Platform (DMP)", which is part of the "Oracle Marketing Cloud", and "powered by" the "Oracle Data Cloud".[451] *Oracle*'s **"Data Directory"** offers insights on the services and data types provided by *Oracle*'s affiliated entities and partner companies. The directory includes several of Oracle's own services, which are still branded as *BlueKai*, *Datalogix* and *AddThis*, but also a detailed overview on consumer data offered by more than **40 "data partners"** such as *Acxiom, Alliant Data, AnalyticsIQ, comScore, Experian, Forbes, GfK, Lotame, MasterCard, Neustar, TransUnion* and *TruSignal* (see Oracle 2015).

---

[443] http://www.datalogix.com/audiences/online/syndicated-segments/ [22.01.2016]
[444] Constine, J. (2012): Facebook Will Use Datalogix Offline Purchase Records To Show Ads The Perfect Number Of Times. Techchrunch, Oct 01, 2012. Online: http://techcrunch.com/2012/10/01/facebook-ads-frequency/ [22.01.2016]
[445] Koetsier, J. (2013): Facebook now lets ads target you based on what you do outside Facebook. Venturebeat, April 10, 2013. Online: http://venturebeat.com/2013/04/10/facebook-launches-partner-categories-to-help-advertisers-target-demand-not-just-demographics/ [22.01.2016]
[446] http://www.datalogix.com/wp-content/uploads/2013/10/DLX_Facebook_Audience_Guide.pdf [22.01.2016]
[447] Ibid.
[448] Ibid.
[449] http://www.oracle.com/us/corporate/pressrelease/data-cloud-and-daas-072214 [22.01.2016]
[450] https://www.oracle.com/corporate/acquisitions/crosswise/index.html [01.08.2016]
[451] https://www.oracle.com/marketingcloud/products/data-management-platform/index.html [22.01.2016]

A unique ID for consumers

According to a corporate presentation[452], the **Oracle Identity Graph** (also "Oracle ID Graph") "unites all [consumer] interactions across various channels to create one addressable consumer profile". It allows companies to "unify addressable identities across all devices, screens and channels" and to "identify customers and prospects everywhere". *Oracle* mentions several kinds of IDs such as a "postal ID", "cookie ID", "email ID", "mobile ID", "registration ID" and a "set-top ID". In another presentation[453] they state that "the Oracle ID Graph connects an individual customer to all channels & devices", and claim to have access to **229 million "device ID's"**. *Oracle*'s developer website explains how all kinds of personal information collected by clients are linked with the *Oracle ID Graph*:[454] On the one hand, "data ingest", which is the "process of collecting and classifying user data" into *Oracle*'s platform, "entails extracting user attributes from your online, offline, and mobile source". In addition, **"offline match integration"** allows to "onboard data from a data warehouse, a Customer Relationship Management (CRM) database, or an email-based offline source", which then can be used to "target, optimize, analyze, and model your users based on their offline attributes".[455]

Details on the matching process

The developer website provides further details on the user information matching process.[456] To **"identify" users "in both the online and offline space"** clients should send their "match keys", which could be "any unique user id", to *Oracle*. The "most common match key" is an "encrypted/hashed email address", because it could be "collected offline during the point of sale (POS) and online when the user signs on to your site". Clients can either use "Oracle Hashed IDs", which are "generated from raw personally identifiable information (PII)" such as e-mail addresses or phone numbers "using Oracle BlueKai code", or "encrypted/hashed UUIDs" based on "e-mail addresses, phone numbers, physical addresses, and client account numbers", and even IP addresses. After receiving these match keys, *Oracle* will "synchronize them to the network of user and statistical IDs that are linked together in the Oracle ID Graph (OIDG), which is used to manage IDs and user attributes for all Oracle BlueKai customers". Clients can also "onboard the mobile data stored" in their "data warehouse, CRM database, or any other offline source" to "monetize those audiences" and contribute **mobile identifiers** such as *Apple*'s IDFA, the *Android ID* and the *Google Advertising ID*. Even a "unique identification header (UIDH)" can be used to "offer marketers and advertisers the ability to target users on, for example, the ***Verizon* mobile network** based on their online behavior".[457] This is apparently referring to *Verizon*'s "perma cookie", which was controversially discussed already.[458]

Combining master data with social data

In a whitepaper published in 2013[459] *Oracle* recommends that client companies should integrate their "enterprise data", for example customer, sales and CRM data, with "social data". They suggest the following "data-integration process": After identifying the

---

[452] http://www.oracle.com/us/corporate/acquisitions/datalogix/general-presentation-2395307.pdf [22.01.2016]
[453] https://www.oracle.com/us/assets/lad-2015-ses16178-toledo-2604862.pdf [22.01.2016]
[454] https://docs.oracle.com/cloud/latest/daasmarketing_gs/DSMKT/GUID-160C5787-1226-4CE2-A418-24454DA3EC36.htm [22.01.2016]
[455] Ibid.
[456] Ibid.
[457] Ibid.
[458] Hoffman-Andrews, J. (2014): Verizon Injecting Perma-Cookies to Track Mobile Customers, Bypassing Privacy Controls. Eff, Nov 03, 2014. Online: https://www.eff.org/de/deeplinks/2014/11/verizon-x-uidh [22.01.2016]
[459] Oracle (2013): The value of social data. Integrated Social and Enterprise Data = Enhanced Analytics. Oracle white paper, December 2013. Online: http://www.sponsored.com.au/app/webroot/uploaded_files/media/SRM_US_EN_WP_SocialData_1.pdf [13.01.2016]

availability and formats of different data from a "mix" of "traditional sources" (e.g. "customer profile data and transactional data, including orders, service requests") and "social data" (e.g. "unified social profiles, Tweets, posts, pictures, videos"), companies should "plug that data into a data exchange" and "enrich the combination of traditional data and social data to gain insights based on a more complete view of the customer". *Oracle* is, along with *IBM* and *SAP*, a major player in **master data management (MDM)**, which has, according to *Gartner*, "become a critical discipline required for dealing with the challenges of social data, 'big data' and data in the cloud".[460]

5.7.3 Experian – expanding from credit scoring to consumer data

Data on 235m people in the U.S., 45m in the UK and 68m in Germany

Experian is one of the three major credit reporting agencies in the U.S.[461], and a global player in credit services, analytics, fraud detection, and marketing data. With around 17.000[462] employees in 39 countries, the total revenue for 2015/2016 was $4.5 billion.[463] *Experian* maintains credit information on about 220 million U.S. consumers, "demographic information" on about 235 million people in 117 million "living units across the U.S.", and information on "more than 650 million vehicles" in the U.S. and Canada.[464] In the **UK,** their "consumer database" holds 45 million records, and they process 1.5 million credit reports per week.[465] In **Germany,** *Experian* is able to categorize 68 million adults along lifestyle groups.[466] On a global level, *Experian* claims to have "insights on **2.3 billion consumers**".[467] The company runs 18 consumer credit bureaus around the world, which contribute 49% to its global revenue.[468] Marketing services, which contribute 18% to global revenue[469], include products like the ***Identity Manager*** to "identify who your customers are regardless of channel or device", the ***Intelligence Manager*** to "understand your customer behaviors and preferences" and the ***Interactions Manager*** to "engage with your customers with the right message".[470]

Predict, identify and target

According to Dixon et al (2014, p. 43 et seq.), *Experian* offers several types of consumer scores. Their ***ChoiceScore*** which "helps marketers identify and effectively target under-banked and emerging consumers"[471] is created "from consumer demographic, behavioral, and geo-demographic information". *Experian*'s ***Median Equivalency Score*** allows corporate customers to "identify areas that may be more or less likely to have future derogatory credit activity". The ***ConsumerView Profitability Score***, which is "designed to predict, identify, and target marketing prospects in households likely to be profitable and pay debt", is based on *Experian*'s "ConsumerView" database. A special version of this score[472] is marketed to healthcare companies, who can "leverage information about consumer's lifestyles, interests and activities" and "bolsters health risk assessments"[473].

---

[460] http://www.gartner.com/newsroom/id/1886314 [13.01.2016]
[461] https://www.usa.gov/credit-reports [13.01.2016]
[462] https://www.experianplc.com/media/2733/experian-ar2016.pdf (p.50) [16.08.2016]
[463] https://www.experianplc.com/media/2733/experian-ar2016.pdf (p.27) [16.08.2016]
[464] http://www.experian.com/corporate/experian-corporate-factsheet.html [13.01.2016]
[465] http://www.experian.co.uk/about-experian/capabilities.html [13.01.2016]
[466] http://www.experian.de/assets/presse/brochures/p-2010-03-01-so-lebt-der-kunde-konsumentensegmentierung.pdf [13.01.2016]
[467] http://www.experian.com/assets/marketing-services/brochures/marketing-suite-brochure-04-2015.pdf [13.01.2016]
[468] https://www.experianplc.com/about-us/our-business-lines/credit-services/ [13.01.2016]
[469] https://www.experianplc.com/about-us/our-business-lines/marketing-services/ [13.01.2016]
[470] http://www.experian.com/marketing-services/marketing-suite.html [13.01.2016]
[471] Dixon et al (2014) cited a reseller of the data
[472] https://www.experian.com/small-business/listdetails/consumerview-healthcare.html [05.08.2016]
[473] Dixon et al (2014) cited Experian's website

Reversely, *Experian*'s **Never Pay score** based on "credit reporting data" can be used to "ensure that consumers who have a high never-pay risk are not included in" the client company's "marketing efforts".[474] *Experian*'s ***Veriscore*** predicts "response and lifetime value of new customers generated from alternate media sources such as call centers and registration forms"[475] .

Data from social media

The company's ***Social Intelligence Platform***[476] allows "social profiling" and "profile analysis combining customer, Experian consumer and social attributes" by harnessing data from social media platforms. It consists of the ***Social Data Linkage*** service, which obtains "individual-level public Facebook behavioral data" through "list-based email address matching", and the ***Social Analytics Engine*** which gathers "individual-level private opt-in Facebook behavioral data" from *Facebook* "as consumers provide permission via the Facebook Open Graph Protocol". Social data includes "name, address, gender, fan pages, including possible competitors, birthday, relationship status, posts, posting date" and allows for example the creation of "social engagement scores".

Credit risk and marketing data

Besides credit scoring products like ***Delphi for Customer Management***, which returns "over 200 variables" and provides "multiple scores to target each specific area of customer management"[477], *Experian* UK offers products in the fields of identity verification, fraud prevention, age verification, online document verification, and employee screening.[478] ***Delphi for Marketing*** combines the "wealth of consumer credit and marketing data" to "generate scores based on an individual's credit risk" and to "avoid targeting those who are already under financial stress".[479]

Experian UK's "Data Directory"

According to *Experian* UK's "Data Directory" brochure,[480] the "ConsumerView" marketing database contains **49 million names and addresses** "for enrichment", 42 million names and addresses "for prospecting", 33 million email addresses, 20 million mobile numbers and 25 million landline numbers. In addition, "consumer characteristics, lifestyles and behaviours" with "**over 500 variables**" from demographics to "financial attitudes and behaviours" are available. Several "propensity models" can for example "indicate the likelihood of an individual or household to own a particular product, or use a particular service". Furthermore, "daily, weekly or monthly **life event triggers**" which can be "based on important life events like moving home or having a baby", are offered. Their "Club Canvasse" offers information about the "**buying habits** of over 23 million individuals that have purchased" from home shopping companies.[481]

Linking offline data to online behavior

Experian's **Hitwise** product is able to "report on millions of unique internet users, hundreds of millions of monthly site visits and tens of millions of monthly searches". **ChannelView** allows "combining" offline postal addresses with 33 million online email contacts, and to enrich "existing customer records" of companies with "any of our 500+ lifestyle variables or social-demographic models" to "bring email and mobile data to life". By "linking" the "ConsumerView database of 49m individuals and 27m households to

---

[474] http://www.experian.com/newsletters/fraud_advisor/0409/cc_qa.html [22.08.2016]
[475] Dixon et al (2014) cited Experian's website
[476] http://www.experian.com/marketing-services/social-intelligence.html [13.01.2016]
[477] http://www.experian.co.uk/consumer-information/delphi-for-customer-management.html [13.01.2016]
[478] http://www.experian.co.uk/identity-and-fraud.html [13.01.2016]
[479] http://www.experian.co.uk/assets/business-strategies/brochures/delphi-for-marketing-product-sheet-final.pdf [13.01.2016]
[480] https://www.experian.co.uk/assets/marketing-services/brochures/brochure-data-directory.pdf [13.01.2016]
[481] Ibid.

client or publisher 1st party data, […] **99% of the UK's targetable population**" can be reached.[482]

Political views, ethnicity and medication preferences

According to the "List Services Catalog"[483] from *Experian* U.S. (2011) the ***ConsumerView*** database contains attributes from occupation and political affiliation to "children by age, month, day or year of birth". Via ***Ethnic Insight*** corporate customers can select from 181 ethnicities, religions and countries "of origin". The ***BehaviorBank*** database includes "responsive consumers who have purchased items or have completed surveys on their leisure activities, brand preferences, computer ownership, occupations, ailments, diet and fitness, financial products, reading preferences and more." It is updated monthly and contains data including "known transactional data, printed surveys via direct mail and online surveys".

Attributes include whether someone has a "dry" or "oily" skin type, prefers champagne or scotch, is a smoker or not, or is an "active military member" or a "veteran". Nearly **100 medication preferences** from Insulin to Prozac are available, "ailments" listed include Alzheimer's disease, cancer, clinical depression, heart disease, multiple sclerosis and "wheelchair". Another product called ***Transactional Data on ConsumerView*** is based on "actual retail (catalog, Internet, and brick and mortar) purchase history" and provides many categories from "low price home décor" and "extreme snow sports" to "high price jewelry and accessories". *Experian* also offers a "New Homeowners Database", a "New Movers Database", and a **"New Parents Database"**.[484]

An "underclass of the working poor"

According to the U.S. Senate Committee on Commerce, Science, and Transportation (2013, p. 24), *Experian* has also offered "targeting products **identifying financially vulnerable populations**", for example a consumer cluster named "Hard Times", which was described by *Experian* as: "This is the bottom of the socioeconomic ladder, the poorest lifestyle segment in the nation. Hard Times are older singles in poor city neighborhoods. Nearly three-quarters of the adults are between the ages of 50 and 75; this is an underclass of the working poor and destitute seniors without family support".

Fraud detection technology enables linking of marketing data

In 2014, *Experian's* subsidiary *AdTruth* introduced its ***AdTruth ID***, an identifier that enables companies to link consumers across devices.[485] A company representative explained in an interview[486] that they aim "to build a platform to manage all datasets in one place" and to "connect users across all data sets". The *AdTruth* technology came from *41st Parameter*, a **fraud detection company** *Experian* acquired in 2013. According to that interview, *AdTruth* provides "a number of variables to identify a single user and how many applications are coming from a device" to help predicting whether "this is a single person or not". According to *Experian*, this technology "empowers the world's most progressive brands to identify, link and engage audiences across all digital touch points" today.[487] In 2015, *Experian* announced[488] ***AdTruth Resolve***, which is able to "reconcile and

---

[482] Ibid.
[483] Experian (2011): List Services Catalog. Online: http://www.experian.com/assets/data-university/brochures/ems-list-services-catalog.pdf [13.01.2016]
[484] Ibid.
[485] Joe, R. (2014): How AdTruth Adds Truth To Cross-Device Connections. Adexchanger, Jul 02, 2014. Online: http://adexchanger.com/omnichannel-2/how-adtruth-adds-truth-to-cross-device-connections/ [13.01.2016]
[486] Ibid.
[487] http://www.experian.com/decision-analytics/identity-and-fraud/adtruth.html [30.08.2016]
[488] PRNewswire (2015): Experian solves industry-wide challenge of engaging audiences across all devices and environments with the launch of AdTruth Resolve. Mar. 03,2015, Online: http://www.prnewswire.com/news-releases/experian-marketing-services-solves-industry-wide-

associate" a company's "existing digital identifiers — including cookies, device IDs, IP addresses and more". As a part of *Experian*'s "Marketing Suite", this would represent "another milestone in Experian Marketing Services' long-term strategy to provide marketers with a ubiquitous, consistent and persistent link across all channels".[489]

5.7.4 arvato Bertelsmann – credit scoring and consumer data in Germany

Avoiding high-risk customers

Owned by the German corporate group *Bertelsmann*, ***arvato*** is a large service provider in digital marketing, financial services, customer relationship management, supply chain management and IT services. Their 70.000 employees in 40 countries are generating a business volume of nearly 5 billion[490] and their CRM division is serving **600 million consumers**.[491] *arvato*'s "Financial Solutions" division manages "around 10,000 customers, specializing primarily in the retail/e-commerce, telecommunications, insurance, banking and healthcare sectors".[492] Besides finance and accounting, factoring, collection and payment processing they are also offering several risk management and credit scoring products, stating that "[h]igh-risk customers should not be developed at all".[493]

Scoring and predicting future behavior

Having "40 million characteristics with negative information about 7.8 million persons" in Germany, they claim to perform "100 million credit checks" per year.[494] Their ***application scoring***, for example, offers companies a "reliable prediction of the expected customer behavior (e.g. payment of the purchase or repayment of a loan)", because "[p]otentially profitable customers should be acquired while customers with a high risk should be avoided from the very beginning".[495] The ***Informa-Storno-Score*** allows companies to "predict a customer's natural loyalty and, therefore, the probability of a cancellation".[496] ***Behavior scoring*** provides "a consistent measure of risk for the entire portfolio". It is based on the "historic behaviour of each customer and allows reliable predictions for the future".[497]

Integrate all existing internal and external data

With the company's ***"infoRate+"*** system, "all existing internal and external data can be densified and integrated". Data sources include "information from credit agencies, telephone and bank data registers as well as data from the AZ Direct address database, a company of arvato Financial Solutions"[498]. The "infoRate+" system can be used for "[c]ontrolling payment methods and credit limits"[499], and it allows "[f]lexible online evaluation of customers".[500] Available modules include address verification, checking "negative lists", validation of phone numbers and bank details, detecting fraud, scoring and "[m]icro-geographic analysis".[501] Lists like *arvato*'s ***Telecommunications Pool*** contain "information on consumers with negative payment behavior". Participating

---

challenge-of-engaging-audiences-across-all-devices-and-environments-with-the-launch-of-adtruth-resolve-300044838.html [13.01.2016]

[489] https://www.experianplc.com/media/news/2015/adtruth-resolve/ [17.08.2016]

[490] https://www.arvato.com/en/about/facts-and-figures.html [15.01.2016]

[491] https://crm.arvato.com/en.html [15.01.2016]

[492] https://www.arvato.com/finance/en.html [15.01.2016]

[493] http://www.arvato-infoscore.de/en/services/risk-management/ [15.01.2016]

[494] http://www.arvato-infoscore.de/en/company/facts-figures/ [15.01.2016]

[495] http://www.arvato-infoscore.de/en/services/risk-management/application-scoring/ [15.01.2016]

[496] http://www.arvato-infoscore.de/en/services/risk-management/informa-storno-score-for-cancellations/ [15.01.2016]

[497] http://www.arvato-infoscore.de/en/services/risk-management/behaviour-scoring/effective-instrument/ [15.01.2016]

[498] http://www.arvato-infoscore.de/en/services/risk-management/inforate/ [15.01.2016]

[499] http://www.arvato-infoscore.de/en/services/risk-management/inforate/benefits/ [15.01.2016]

[500] http://www.arvato-infoscore.de/en/services/risk-management/inforate/ [15.01.2016]

[501] http://www.arvato-infoscore.de/en/services/risk-management/inforate/online-evaluation-of-customers/ [15.01.2016]

companies receive information from the pool when they deliver their own data to the pool.[502] *arvato* also offers **tenant screening**.[503]

Cross-device tracking and hash IDs

Their ***Profile Tracking*** module is able to "identify particular internet access devices on the basis of this hash-ID clearly and in real-time", because "[n]o matter" if the device is a "PC, tablet, smartphone or game console: each of these devices leaves a unique and identifiable trace, the so-called hash-ID".[504] The company's **risk management product for e-commerce** is based on "experience with payments, information from shopping baskets and external data on credit ratings". It can make use of "current and historical customer information" and provides "analysis and intelligent linking of customer data". To avoid making their "decision-making system" sound like a fully automated system, *arvato* emphasizes that online shops receive the "results of these checks in the form of a recommendation for action (e.g. offer for a method of payment)". Follow-up processes can then be "triggered by the eShop/customer system".[505]

Marketing data: 600 attributes on 70 million consumers

At the same time, *arvato* runs ***AZ Direct***, a leading direct marketing and data broker company in German-speaking countries. According to a corporate presentation[506], they offer 600 attributes on 70 million consumers and 40 million households in Germany, amongst other data sources based on 300 million shopping transactions. In their consumer database ***AZ DIAS***, an "ID" is assigned to every single person, household and building. 32 million people can be reached via direct mail, 33 million people via targeted email, and 27 million people via "data-driven advertising" online. According to their "Merkmalskatalog" (see AZ Direct, 2015) they offer profile attributes like age, sex, lifestyle, social status, children, income and even the ethnical origin of names[507]. In addition, people can be categorized in terms of online usage, financial behavior, and for example, whether they focus on security/stability or tend to risky behavior regarding insurance.[508] All these attributes are also available for the **enrichment of existing customer databases** on different aggregate levels (for example 5, 20 or 70 households). The "Informa-Geoscore", which predicts good or bad future payment behavior, is available on an aggregate of 20 households on average (see AZ Direct 2015).

Online data management platform (DMP)

Furthermore, *arvato* runs the targeting and data management platform ***adailty***[509] which offers[510] **"data partners"** to capitalize their "offline data", for example "master data" or "transaction data"[511]. To their so-called **"matching partners"**, *adailty* offers to support the

---

[502] http://www.arvato-infoscore.de/en/services/risk-management/data-pools/telecommunications-pool/closed-data-pool/ [15.01.2016]
[503] http://www.arvato-infoscore-mieterauskunft.de/ [15.01.2016]
[504] http://www.arvato-infoscore.de/en/services/risk-management/profile-tracking/ [15.01.2016]
[505] http://www.arvato-infoscore.de/en/services/risk-management/risk-solution-services/ [15.01.2016]
[506] Hüffner, W. (2015): Datenschutzkonformes Smart Data und Data Pooling. arvato Digital Marketing, Mar. 05, 2015. Online: https://www-950.ibm.com/events/wwe/grp/grp006.nsf/vLookupPDFs/H%C3%BCffer_IBM_SPSS_2015/$file/H%C3%BCffer_IBM_SPSS_2015.pdf [15.01.2016]
[507] In German: "Namensherkunft: […] Hier können über den Vornamen Rückschlüsse auf die Herkunft des Vornamens, d.h. die Nationalität der Person, gemacht werden". Available options include "Deutsch klingend", "Ausländisch klingend", "Assimiliert".
[508] "Versicherungstypologie", Available options include "Sicherheitsorientierter Typ" and "Risikobereiter Typ".
[509] One to One New Marketing (2013): Arvato bündelt CRM und Dialog-Dienstleistungen. Online: (http://www.onetoone.de/Arvato-buendelt-CRM-und-Dialog-Dienstleistungen-23590.html [15.01.2016]
[510] http://adality.de/partner/ [15.01.2016]
[511] Ibid., in German: „Als Datenpartner kapitalisieren wir Ihre Offline-Daten (z. B. Stammdaten, Transaktionsdaten)"

matching of offline data to cookies, for example by incorporating "tags" or "pixels" into their clients' websites and email newsletters[512]. Data products offered include socio-demographic data, interests, spending capacity and income.[513] According to a talk[514] by CEO Christian Vennemann, *adality* is also able to access the *AZ Direct* database containing 250 attributes about 70 million "persons".[515]

5.7.5 LexisNexis and ID Analytics – scoring, identity, fraud and credit risks

LexisNexis

The controversially discussed[516] data broker *Choicepoint,* which had extensive data records about 220 million people, was acquired by *LexisNexis* more than ten years ago, and is now part of the risk management division of *RELX Group* (formerly known as *Reed Elsevier)*. ***LexisNexis Risk Solutions***[517] claims to have data on 500 million consumers[518], and they work for all 50 of the 50 largest U.S. banks, for 70% of U.S. local government authorities and for 80% of U.S. federal agencies.[519] They provide risk management solutions for insurance, finance, retail, travel, government, gaming and for the healthcare sector. In 2015, the director of their government division told the New York Times: "Because of our identity information, we know more than the government entities", and he added: "We know where virtually every individual over 18 is".[520]

Biometric data loyalty cards and social media

LexisNexis provides data about consumer creditworthiness[521], insurance scores[522], background checks for employers on both applicants and employees, as well as "resident screening" services to "protect […] property from problem renters".[523] Their identity and authentication system ***TrueID***[524] offers to "link" **biometric data** from photos to fingerprints "to other user data to track transactional behavior throughout the customer lifecycle". The identity of persons can be verified using a database of "34 billion records from over **10,000 sources**" and "of nearly 4,100 ID types from nearly 200 countries". Identity can also be linked with **"payment cards, checks, loyalty cards and other customer data"**. Moreover, even biometric services for voice recognition using "the sound, pattern and rhythm of an individual's voice" are offered.[525] Their ***Social Media Monitor*** which is part of their product "LexisNexis Accurint® for Law Enforcement" offers to "identify posts and/or tweets within specific geographical locations" and to "discover risks and threats" in order to "unlock the value of big data from social media".[526]

---

[512] Ibid., in German: „Als Matchingpartner können Sie uns dabei unterstützen, anonymisierte Offline-Daten mit Cookies anzureichern. Sie integrieren hierfür ein Tag (bzw. Pixel) in Ihre reichweitenstarke Website oder E-Mail-Aussendungen"
[513] http://adality.de/produkte/ [15.01.2016]
[514] YouTube video, from minute 1:50: https://www.youtube.com/watch?v=W41HcRo-3P8 [15.01.2016]
[515] Ibid., in German: Adality hätte „aktuell exklusiv Zugriff auf die Daten" von AZ Direct, die „70 Millionen Personen mit über 250 kombinierbaren Merkmalen" in ihrer „Datenbank" hätten.
[516] O'Harrow, Robert (2005): They're Watching You. Bloomberg Businessweek, 23.01.2005. Online: http://www.businessweek.com/stories/2005-01-23/theyre-watching-you [15.01.2016]
[517] http://www.lexisnexis.com/risk [22.01.2016]
[518] http://www.lexisnexis.com/risk/about/data.aspx [22.01.2016]
[519] http://www.lexisnexis.com/risk/about/default.aspx [22.01.2016]
[520] Singer, N. (2015): Bringing Big Data to the Fight against Benefits Fraud. New York Times, Feb. 20, 2015. Online: http://www.nytimes.com/2015/02/22/technology/bringing-big-data-to-the-fight-against-benefits-fraud.html [22.01.2016]
[521] http://www.lexisnexis.com/risk/products/riskview-credit-risk-management.aspx [22.01.2016]
[522] http://www.lexisnexis.com/risk/products/insurance/attract.aspx [22.01.2016]
[523] https://www.lexisnexis.com/government/solutions/literature/screening.pdf [22.01.2016]
[524] http://www.lexisnexis.com/risk/downloads/literature/trueid.pdf [22.01.2016]
[525] http://www.lexisnexis.com/risk/products/voice-biometrics.aspx [22.01.2016]
[526] http://www.lexisnexis.com/risk/newsevents/press-release.aspx?id=1381851197735305 [22.01.2016]

Telematics, insurance scores and marketing

In 2014 *LexisNexis* acquired *Wunelli*, a **telematics** provider, in order to "empower insurers to leverage telematics" [527]. They promise that their combined "datasets" will result "in one of the largest provider-held insurance telematics databases in the world" to "support insurers as they assess risk". *LexisNexis* also provides insurance scores based on credit report data, which can be "applied at the time of quote, at underwriting, at renewal and for prescreening".[528] Their "PowerView Score"[529] is based on data from various sources, including "telecom/utility payment data", property data and asset ownership. It allows auto lenders to predict creditworthiness and to perform "incremental segmentation to upgrade or downgrade terms". In addition, *LexisNexis* also provides solutions for **marketing**. For example, their ***Lead Optimizer*** product[530] "scores insurance leads in real-time" and offers insurers to "save time and money by eliminating unproductive leads early in the process". Their ***DirectLink(SM)*** product[531] for insurers "seamlessly integrates all components of prospecting and customer contact campaigns into a complete system" to "optimize responses and conversion" and acquire and retain "profitable customers". It allows the integration of mail, email and telemarketing, and the use of "individual customer and prospect data attribute selections" as well as "predictive models" for segmentation and targeting.

ID Analytics

The U.S.-based scoring and data company ***ID Analytics*** offers products for identity verification, credit scoring, fraud risk and payments[532] and was one of the nine companies examined in the FTC's data broker study (FTC, 2014). It is a subsidiary of *LifeLock Inc.*, which had 669 employees in 2014.[533] In 2012, their ***ID Network*** contained "more than 700 billion instances of PII, like names, addresses, SSNs, DOBs, phone numbers and emails", providing insights about "more than 315 million unique people in the U.S.". [534] It has "aggregated more than **1.7 billion consumer transactions** that contain this PII, including 2.9 million reported fraud events". According to another document[535], the "ID Network" is a "consortium of consumer behavioral data built through the contributions of more than 250 enterprise clients". In 2014, "six of the top ten U.S. financial service institutions, three of the top four U.S. wireless carriers, and seven of the top ten U.S. credit card issuers" have contributed data.

Online, call center, mail and in-store

ID Analytics offers an ***ID Score***, which "assesses the likelihood that an application will result in fraud".[536] In addition, *ID Analytics* provides access to an "identity repository" in which "54 million identity elements" are "updated daily". The company's ***ID Network Attributes*** are a "set of derived data points", that are implementable "across all points of customer contact including online, call centers, mail, and in-store". It "examines **a consumer's identity elements**, individually and in combination, across eight categories

[527] http://www.lexisnexis.com/risk/newsevents/press-release.aspx?id=1400513019730653 [11.01.2016]
[528] http://www.lexisnexis.com/risk/products/insurance/attract.aspx [22.08.2016]
[529] http://www.lexisnexis.com/risk/products/credit-risk-management/powerview-score.aspx [22.08.2016]
[530] http://www.lexisnexis.com/risk/products/insurance/lead-optimizer.aspx [22.08.2016]
[531] http://www.lexisnexis.com/risk/products/insurance/directlink.aspx [22.08.2016]
[532] http://www.idanalytics.com/ [11.01.2016]
[533] https://www.lifelock.com/about/ [11.01.2016]
[534] http://www.idanalytics.com/media/ID-Analytics-I-See-Fraud-Rings-White-Paper1.pdf [11.01.2016]
[535] http://www.idanalytics.com/media/Exploring-the-Impact-of-SSN-Randomization.pdf [11.01.2016]
[536] http://www.idanalytics.com/solutions/fraud-risk-management/id-score/ [11.01.2016]

of behavior" including "confirmed negative behavior" and "demographics/mode of living" and the "historical use of internet-enabled devices".[537]

Credit scoring

In addition, *ID Analytics* offers "credit risk solutions" to help companies improve their "lending" and "approval and pricing decisions" by "pairing traditional credit data with powerful alternative insights from the wireless, banking and sub-prime markets".[538] According to *ID Analytics* own statement, their ***Credit Optics*** product uses the "unique blend of traditional and alternative consumer credit data" in the *ID Network* to "inject new, predictive information into existing credit bureau and custom models".[539] It can also be used for profiling existing customers[540], to "prescreen" and to "[i]dentify the right prospects", or to send "direct-mail offers to risk-appropriate consumers" only.[541] In 2010, ***TransUnion*** announced to offer a scoring product, which includes *TransUnions*'s credit data as well as "alternative" data from *ID Analytics*.[542]

5.7.6 Palantir – data analytics for national security, banks and insurers

Palantir Technologies is not a typical data broker in a sense that the company trades personal data. However, *Palantir* is an important data intelligence company, providing its sophisticated analytical services to both public and private customers.

Some of the world's most sensitive sets of data

Palantir was founded in 2004 by Alexander Karp and Peter Thiel. The latter is also the founder of the online payment company *PayPal* and the first investor in *Facebook*. The company was originally designed to "uncover terror networks using the approach PayPal had devised to fight […] cybercriminals".[543] By linking and simultaneously querying large numbers of databases, *Palantir* provided a valuable service for the intelligence and national security agencies. In 2009, the company supplied its software and services to the ***Central Intelligence Agency (CIA)***, the ***Pentagon*** and the ***Federal Bureau of Investigation (FBI)*** within more than 50 projects.[544] In 2013, the software solutions were used by police departments and by at least 12 groups within the US Government, including CIA, FBI, NSA, the Marine Corps and the Air Force and dealt "with some of the world's most sensitive sets of data".[545]

Public debates

Palantir raised significant public awareness in 2011, when the company was exposed by a hacker group "to be in negotiation for a proposal to track labor union activists and other critics of the U.S. Chamber of Commerce, the largest business lobbying group in Washington"[546]. The proposal lead to public debates, and *Palantir* was **accused of**

---

[537] http://www.idanalytics.com/media/Fraud-ID-Network-Attributes-Datasheet.pdf [11.01.2016]
[538] http://www.idanalytics.com/solutions/credit-risk-solutions-and-risk-analytics [11.01.2016]
[539] http://www.idanalytics.com/solutions/credit-risk-solutions-and-risk-analytics/alternative-credit-data/ [11.01.2016]
[540] http://www.idanalytics.com/solutions/credit-risk-solutions-and-risk-analytics/credit-optics-portfolio-management/ [11.01.2016]
[541] http://www.idanalytics.com/solutions/credit-risk-solutions-and-risk-analytics/credit-optics-prescreen/ [11.01.2016]
[542] http://newsroom.transunion.com/transunion-unveils-new-credit-opticstm-plus-score [11.01.2016]
[543] Gorman, S. (2009): How Team of Geeks Cracked Spy Trade. Wall Street Journal, Sept.4, 2009. Online: http://www.wsj.com/articles/SB125200842406984303 [30.08.2016]
[544] Ibid.
[545] Burns, M. (2015): Leaked Palantir Doc Reveals Uses, Specific Functions And Key Clients. TechCrunch, Jan. 11, 2016. Online: https://techcrunch.com/2015/01/11/leaked-palantir-doc-reveals-uses-specific-functions-and-key-clients/ [30.08.2016]
[546] Fang, L. (2016): The CIA is Investing in Firms that Mine Your Tweets and Instagram Photos. The Intercept, Apr.14, 2016. Online: https://theintercept.com/2016/04/14/in-undisclosed-cia-investments-social-media-mining-looms-large/ [30.08.2016]

abusing its power. According to The Nation[547], it targeted activists, reporters, labor unions and political organizations and - according to a leaked report[548] – suggested to investigate activists' families and even used "sophisticated hacking tools to break into computers". Later in 2011, a proposal called "The WikiLeaks Threat" and related email conversations were leaked by the Hacker Group *Anonymous*.[549] The document, which is still available online,[550] was prepared by the three data intelligence firms *Palantir Technologies*, *HBGary Federal*, and *Berico Technologies*. The presentation was publicly criticized for being unethical as it mentioned "potential proactive tactics against WikiLeaks includ[ing] feeding the fuel between the feuding groups, disinformation, creating messages around actions to sabotage or discredit the opposing organization, and submitting fake documents to WikiLeaks and then calling out the error."[551]

In a comprehensive report[552], the *Infosec Institute* named *Palantir* as one of the principal technological partners for the *PRISM* program and indicated that the company may play a role in financing *Facebook*. The German *manager magazin* describes *Palantir* founder Peter Thiel as one of the most successful investors in Silicon Valley and first ***Facebook*** financier.[553] Whether Peter Thiel's investment in and relationship with *Facebook* plays a role in *Palantir*'s intelligence services is not publicly known.

Insurance, financial services and healthcare

Today, *Palantir* is valued at about $20 billion and earns 75% of its revenue from corporate clients, to whom the company delivers fraud detection services, studies of consumer behavior and analyses of the competition.[554] Its "two main products, ***Gotham*** and ***Metropolis***, serve the same basic purpose—bringing together massive, disparate data sources and scouring them for connections and patterns that aren't obvious to the human eye".[555] *Palantir*'s clients come from a variety of industries and sectors, such as financial services, retail, legal intelligence, pharmaceutical companies, insurance analytics, healthcare delivery, disease response, biomedical research, federal and local law enforcement agencies, defense, intelligence and accountability.[556]

5.7.7 Alliant Data and Analytics IQ – payment data and consumer scores

Connecting purchases with online tracking

The marketing data company *Alliant Data* claims to be the "industry's largest source of detailed micropayment data" offering information on "payments, returns, billings, and write-offs" as well as aggregating "consumer response behavior" from "more than 400

---

[547] Fang, L. (2013): How Spy Agency Contractors Have Already Abused Their Power. The Nation, Jun. 11, 2013. Online: https://www.thenation.com/article/how-spy-agency-contractors-have-already-abused-their-power/ [30.08.2016]
[548] Ibid.
[549] Ragan, S. (2011): Data Intelligence firms proposed a systematic attack against WikiLeaks. The Tech Herald, Feb. 10, 2011. http://www.thetechherald.com/articles/Data-intelligence-firms-proposed-a-systematic-attack-against-WikiLeaks/12751/ [30.08.2016]
[550] https://wikileaks.org/IMG/pdf/WikiLeaks_Response_v6.pdf [30.08.2016]
[551] Ragan, S. (2011) Firm targeting WikiLeaks cuts ties with HBGary – apologizes to reporter. http://www.thetechherald.com/articles/Firm-targeting-WikiLeaks-cuts-ties-with-HBGary-apologizes-to-reporter/12767/ [30.08.2016]
[552] Infosec Institute (2013): The Palantir Technologies model, lights and shadows on a case of success. Posted in General Security, Jul. 9, 2013. Online: http://resources.infosecinstitute.com/the-palantir-technologies-model-lights-and-shadows-on-a-case-of-success/ [30.08.2016]
[553] Rungg, A. (2015): Palantir und die dunkle Seite der Macht. Manager magazin, Jan. 14, 2015. Online: http://www.manager-magazin.de/unternehmen/it/palantir-und-die-dunkle-seite-der-macht-a-1013000-2.html [30.08.2016]
[554] Lev-Ram, M. (2016): Palantir Connects the Dots With Big Data. Fortune, Mar. 1, 2016. Online: http://fortune.com/palantir-big-data-analysis/ [30.08.2016]
[555] Ibid.
[556] https://www.palantir.com/solutions/ [30.08.2016]

subscription, continuity, and one-shot brands". They offer "600+ Audience Selects" on 270 million U.S. consumers.[557] Data resources include "transaction-level behavioral, demographic and lifestyle data on more than 270 million consumers" (Oracle 2015, p. 29). *Alliant*'s "Online Audiences" are available "through most major platforms via partnerships with over 70 DSPs, DMPs, and ad exchanges" – and contain data on 115 million U.S. households, 180 million "30-day Unique IDs (desktop devices)", and 47 million "30-day Unique IDs (mobile devices)".[558] These "30-day Unique IDs" could be cookie and mobile identifiers of web browsers and mobile devices, which are used at least one time in 30 days. All in all, *Alliant* claims to have "over two billion match keys for consumers, including email addresses, mobile numbers and device IDs".[559]

Database enrichment and credit scoring

Alliant also offers "database enrichment" to enhance other companies' data with "newly updated emails, mobile device identifiers, postal addresses and predictive/descriptive variable".[560] Furthermore, their ***TransactionBase*** product is a "source of alternative data for credit decisions, thin file scoring, and billing management" and contains "detailed payment information" on "over 90 million consumers". It offers "financial services and insurance marketers a full range of credit-scoring solutions" and "provides full prescreening services for qualification of lead lists".[561] According to Chester et al (2014), *Alliant* has been selling information on "Financially Challenged", "Credit Card Rejects", "Credit Challenged", and "Risky Consumers".

AnalyticsIQ

AnalyticsIQ is a consumer data analytics company based in Atlanta whose products are, for example, offered by *Oracle* (2015, p. 30). The company claims to provide data about 210 million individuals and 110 million U.S. households from 120 "unique data sources"[562], including "aggregated credit, demographics, purchase, lifestyle and real-estate information, econometrics, financial and proprietary data".[563] Their "consumer financial intelligence" portfolio includes several scoring products to predict "consumer financial behavior". Beside of several ***GeoCredit scores***, they offer "affluence scores" like "Spendex", "InvestorIQ", "WealthIQ" and "IncomeIQ", and "home and mortgage scores" like "Home ValueIQ " and "Home EquityIQ".[564] The company's "demographic data products" include ***EthnicIQ*** and "**Political & Religious Affiliation**". Furthermore *AnalyticsIQ* offers "consumer lifestyle and behavioral data" like the social media influence score ***SocialIQ***, which predicts "consumer social media activity and influence", and the loyalty score "ChurnIQ", which "predicts a consumer's likelihood to be loyal" to a brand.[565]

5.7.8 Lotame – an online data management platform (DMP)

Buying and selling data

Lotame is a data management platform (DMP), which allows corporate customers to buy and sell data.[566] They provide "access to a pool of more than **three billion cookies** and **two billion mobile device IDs**", which they categorize into "thousands" of "segments".[567]

---

[557] http://alliantinsight.com/solution-sets/alliant-consumer-audiences/ [11.01.2016]
[558] http://alliantinsight.com/solution-sets/alliant-online-audiences/ [11.01.2016]
[559] http://alliantinsight.com/solution-sets/alliant-engage/ [31.01.2016]
[560] http://alliantinsight.com/solution-sets/alliant-data-marts/ [11.01.2016]
[561] http://alliantinsight.com/solution-sets/alliant-transactionbase/ [11.01.2016]
[562] http://analytics-iq.com/ [11.01.2016]
[563] http://analytics-iq.com/data-solutions/data-foundation/ [11.01.2016]
[564] http://analytics-iq.com/data-solutions/consumer-financial-intelligence/ [11.01.2016]
[565] http://analytics-iq.com/data-solutions/demographics-lifestyle/ [19.01.2016]
[566] "Buying Data" and "Selling Data": http://www.lotame.com/data-exchange/ [19.01.2016]
[567] Ibid.

The following table shows examples of the number of "unique" web browsers and mobile devices per country they provide access to:[568]

France	Germany	Italy	Netherlands	Pakistan	Poland	Spain	Turkey	UK	Russia
56.8m	59.7m	42.9m	13.8m	11m	25.3m	21.3m	77.2m	146.3m	121.2m

U.S.	Canada	Mexico	Argentina	Brazil	China	India	Indonesia	Japan	Australia
891m	83.1m	29.2m	14.3m	98.7m	3.5m	70.1m	32.4m	34.2m	35.6m

Table 23: How many "million monthly uniques" Lotame provides access to, per country. Source: Lotame

Third-party and CRM data

As they, for example, claim to provide access to 891 "million monthly uniques" in the U.S., it seems that different web browsers and mobile devices of users are separately counted. In addition, *Lotame* offers "direct integrations with over 20 of the world's largest third-party data providers".[569] On its website, the company gives insights about the data strategies it offers to clients. [570] Clients can "collect first-party data" from across their "sites, apps and ad campaigns", and combine it with "other first-party sources, such as email data or data housed" within their "CRM system". *Lotame* could then "create audience segments", selecting "specific demographics, interests and actions", and enrich these "by using third-party data". Integrations with many other companies and services in online marketing (with "every major DSP[571], ad server, exchange and SSP[572]") allow corporate customers to use the audience segments for targeting. Finally, clients can use ***Lotame Syndicate***, a "private marketplace" for the "secure exchange of first-party data" to "access rich second-party data not available in the open marketplace".[573] *Lotame Syndicate*[574] is especially designed for companies who are "targeting the same affluent audiences" but are "not directly competing with each other". As an example *Lotame* mentions a "luxury auto brand" that could "share select audience data with an app that profiles 5-star travel resorts".

"100%" declared data, matched with offline sources

According to *Oracle* (2015, p. 82), "Lotame Smart Data" categorizes "100% declared and demonstrated data (NOT panel-based, modeled, or inferred) into over 2200 audience segments". Their partners "place proprietary Behavioral Collection Pixels", allow them to "collect demographic, interest, action, search, purchase intent, and other data points". **Demographic data** would be "100% Self-declared by a user on a profile or registration, and matched with offline sources". **Behavioral data** is based on "articles read, on-site searches, clicked on, searched for and any other action a user could complete on a page", but also information about "in-store purchases" is collected in partnership with companies "who anonymously match in-store purchases to online cookies for targeting". Finally, also **social data** is available – from "users that frequently complete social actions that others online can see, such as sharing, rating, posting, or commenting".

5.7.9 Drawbridge – tracking and recognizing people across devices

3.6 billion devices from 1.2 billion consumers

Several companies are specialized in cross-device tracking to ensure that consumers are recognized as the same person when using different devices like their PC or their

---

[568] "Million Monthly Uniques", Ibid.
[569] Ibid.
[570] http://www.lotame.com/platform/ [19.01.2016]
[571] DSP = demand-side platform
[572] SSP = supply-side platform
[573] Ibid.
[574] http://www.lotame.com/resource/qa-what-is-lotame-syndicate-and-how-can-it-add-value-to-my-data-strategy/ [19.01.2016]

smartphone. One of them is ***Drawbridge***, which claims to have about 1.2 billion "consumers connected across more than 3.6 billion devices".[575] According to their privacy policy,[576] they receive user data from "various advertising exchanges, platform and ad networks" and combine it with "additional demographic, geolocation and interest-based segment data" from third-party providers. Subsequently, *Drawbridge* uses "probabilistic modeling" to "determine the probability that a desktop web cookie and a mobile device belong to the same User" and "share this device matching information" with their corporate clients "to enable them to provide advertising, analytics or other services". The information they receive includes visited websites (including date and time of visits), IP addresses, mobile device identifiers such as Apple IDFA or Google Advertising ID, geolocation (including GPS data), browser type, carrier, referring/exit pages, device model, operating system, gender, age, clickstream data and cookie information.

„Connected Consumer Graph"

In a corporate presentation[577], *Drawbridge* describes its device and behavior fingerprinting technology the ***Connected Consumer Graph***, which is "made up of interconnected Device Graphs". Each of these graphs consists of "collected and inferred demographic and behavioral information". This technology enables the company to "paint a granular portrait of each individual consumer" and to "make educated predictions about users and their devices". *Drawbridge* indicates their **cross-device reach** to be 400 million people in North America, 150 million in Latin America, 350 million in EMEA (Europe, the Middle East and Africa), 200 million in APAC (Asia-Pacific) and 20 million in AUS/NZ. According to *Drawbridge*, *Nielsen* had analyzed their technology, compared it to other cross-device identity approaches, and found it to be "97.3% accurate in indicating a relationship between two or more devices". Information *Drawbridge* receives comes from more than **50 partners**, including "mobile and desktop exchanges, advertisers, publishers, data management platforms, and other data providers"[578] – for example: *xAd* and *Factual* ("location data"), *Oracle*'s *Bluekai* and *Exelate* ("3rd party DMPs"), *Adobe* and *Acxiom* ("1st party DMPs") and *Acxiom*'s *LiveRamp* (CRM Data).[579]

From advertising to finance?

In an interview[580], *Drawbridge*'s CEO explains that they are no longer purely focusing on advertising. There would be other companies beyond advertising, including those in the finance and travel industries that "want to understand the consumer journey across devices". Recently *Drawbridge* announced that they have partnered with *TVTY* to "enable marketers [to] sync their digital reach across smartphones, tablets, and personal computers with **TV programming in real-time**". Their technology would be "more accurate and faster than audio recognition".[581]

5.7.10 Flurry, InMobi and Sense Networks – mobile and location data

Tracking across 1.4 billion devices

The mobile analytics and ad platform ***Flurry***, acquired by *Yahoo* in 2014, maintains a system that collects information about smartphone users' behavior, and offers it to app vendors in order to analyze their users and earn money with targeted advertising and

---

[575] http://drawbridge.com/ [15.01.2016]
[576] http://drawbridge.com/privacy [15.01.2016]
[577]https://gallery.mailchimp.com/dd5380a49beb13eb00838c7e2/files/DB_White_Paper_011216.pdf [15.01.2016]
[578] Ibid.
[579] https://gallery.mailchimp.com/dd5380a49beb13eb00838c7e2/files/DB_MediaKit_011216.pdf [15.01.2016]
[580] Ha, A. (2015): Drawbridge Adds Offline Purchases To Its Cross-Device Marketing Data. TechCrunch, May 4, 2015. Online: http://techcrunch.com/2015/05/04/drawbridge-cross-device/ [15.01.2016]
[581] http://drawbridge.com/news/p/drawbridge-enables-marketers-to-sync-cross-device-ads-with-tv-programming-in-real-time-with-tvty-integration [15.01.2016]

other methods. *Flurry* is, according to themselves, embedded in **540,000 different apps on iOS, Android and other platforms**, and installed on more than 1.4 billion smartphones and tablets[582]. Thus, the company "has built unique profiles on more than 1.4 billion devices worldwide"[583] and, according to *Forbes*[584] a "trove of mobile-app-user data that is bigger in reach than Google and Facebook".

Hardcore gamers, new mothers and LGBT

Flurry claims to measure one-third of the global app activity and "sees an average of 7 apps per device on over 90% of the world's devices". Because user behaviour can be analyzed across apps, *Flurry* would be able to "paint a rich picture about a person's interests".[585] The platform offers to **categorize users** into segments and to target users based on attributes such as interests, gender, age, language, device, operating system[586] and in accordance to so-called ***Personas*** like "hardcore gamers", "financial geeks", "new mothers", "slots player" and even "LGBT" (lesbian, gay, bisexual and transgender).[587] These "personas" and other data are calculated from the app-usage patterns.

350 profile attributes

Since 2014, *Flurry* cooperates[588] with the market research and consumer data company ***Research Now***, which conducts surveys and sees itself as the "world's leading digital data collection company".[589] *Flurry* combined their data with its own knowledge about the app users and since then offers additional "350 profile attributes including demographic, interest, lifestyle" information[590] including "hundreds of offline data points" such as "household income, number of children and travel preferences" for targeting purposes.[591] *Flurry* also offers app developers to "[l]ocate and [t]arget] [s]pecific [d]evice IDs" to retarget users, and to identify the "[m]ost valuable customers" or "users who have made a purchase on your mobile website but not in your app".[592]

InMobi, tracking 1 billion app users

InMobi is a mobile ad network with 17 offices across the globe covering 200 countries. They claim to generate 138 billion "monthly ad impressions" across 1 billion "monthly active users". They offer to categorize these users into "20,000+ refined audience segments", which can be "validated through a consumer panel of seven million users".[593] According to their privacy policy from January 2016,[594] they may collect extensive data about the user's devices[595] and the ads viewed, as well as information about "post-click activity in relation to the ad", and information "mobile publishers or app developers" have

---

582 http://www.flurry.com/solutions/advertisers/brands [22.01.2016] Archived version: https://web.archive.org/web/20160125204729/http://www.flurry.com/solutions/advertisers/brands [22.08.2016]
583 Ibid.
584 Olson, Parmy (2013): Meet The Company That Tracks More Phones Than Google Or Facebook. Forbes, 30.10.2013. Online: http://www.forbes.com/sites/parmyolson/2013/10/30/meet-the-company-that-tracks-more-phones-than-google-or-facebook/ [22.01.2016]
585 http://www.flurry.com/solutions/advertisers/brands [22.01.2016]
586 Ibid.
587 http://www.flurry.com/sites/default/files/resources/Personas%20vF.pdf [22.01.2016]
588 Bergen, Mark (2014): Flurry Launches Service to Track Mobile App Users, Offline The Analytics Firm Partners With Research Now, As the Race to Target Inside Apps Picks Up. Advertising Age, 24.03.2014. Online: http://adage.com/article/digital/flurry-research-build-mobile-app-advertising-database/292287/ [22.01.2016]
589 http://www.researchnow.com/about-us [22.01.2016]
590 https://www.flurry.com/sites/default/files/resources/FlurryEnhancedPersonas.pdf [22.01.2016]
591 http://www.flurry.com/solutions/advertisers/brands [22.01.2016]
592 Ibid.
593 http://www.inmobi.com/company/ [16.01.2016]
594 http://www.inmobi.com/privacy-policy/ [16.01.2016]
595 E.g. device type, operating system, network provider, IP address, browser version, carrier user ID, iOS identifiers, mac address, IMEI, phone model, session start/stop time, locale, time zone, WiFi network status, geo-location, unique device identifiers

"separately collected". They seem to consider all this data not to be personally identifiable data[596], and claim to "anonymise this information using one-way hashing" before sharing with third-parties. The categorization of users is "based on **purchase history**, engagement levels, app launches" and includes segments, that help app developers to identify "[h]igh [v]alue [u]sers" who "don't spend enough money in your app", "[d]ormant [u]sers" who "don't spend enough time in your app", and "[s]ocial [i]nfluencers".[597] In June 2006, *InMobi* was penalized by the U.S. *Federal Trade Commission* to pay $ 950,000, because the company "deceptively tracked the locations of hundreds of millions of consumers – including children – without their knowledge or consent".[598]

Sense Networks, predicting where people will go

Sense Networks is a mobile and location data analytics company owned by the marketing giant *YP*, which generated $1 billion in revenue in 2013.[599] According to their website, they use "mobile location data to understand consumer behavior"[600] and to predict where people will go.[601] By analyzing "location patterns" – for example, where "consumers shop, eat and hang out" – they build "anonymous, individual user profiles" containing "over 1,000 behavioral attributes including shopping, dining and lifestyle habits".[602] They claim to "have profiles built on over 150 million mobile users"[603], and to process "170 billion location points per month into profiles", more than "any company other than Google or Facebook".[604] In an interview with *Wired* magazine[605], the CEO of Sense Networks stated that "location data, created all day long just by having a phone in your pocket, is probably the richest source of information in the world today".

Their ***Retail Targeting*** product "analyzes mobile travel patterns" to identify and "target prospects" who are frequently near particular stores, or "when they are at other locations near the retailer, such as home or work".[606] In addition to targeting "people who shop at specific retailers, frequent quick-serve restaurants, visit banks and go to car dealers" also **demographic data** (e.g. age, income, education or ethnicity) and **lifestyle information** are available.[607] According to their privacy policy,[608] *Sense Networks* also "build[s] anonymous profiles for 3rd party mobile publishers". These publishers "provide" them with "location data and possible other data such as application usage and demographic information", which "may be tied to an anonymous identifier". In an additional "privacy principles" section,[609] *Sense Networks* states that "all data collection should be 'opt-in'".

---

[596] "This information does not enable us to work out your identity in real life"
[597] http://www.inmobi.com/products/analytics-segments/ [16.01.2016]
[598] https://www.ftc.gov/news-events/press-releases/2016/06/mobile-advertising-network-inmobi-settles-ftc-charges-it-tracked [01.08.2016]
[599] Gelles, David (2014): YP, a Mobile Search Firm, Buys Sense Networks. The New York Times, Jan. 06, 2014. Online: http://dealbook.nytimes.com/2014/01/06/yp-a-mobile-ad-firm-buys-a-rival-sense-networks/ [16.01.2016]
[600] https://www.sensenetworks.com/life-happens-outside-of-the-geo-fence/ [16.01.2016]
[601] Fitzgerald, M. (2008): Predicting Where You'll Go and What You'll Like. The New York Times, Jun. 22, 2008. Online: http://www.nytimes.com/2008/06/22/technology/22proto.html [16.01.2016]
[602] https://www.sensenetworks.com/life-happens-outside-of-the-geo-fence/ [16.01.2016]
[603] https://www.sensenetworks.com/customers/advertisers/ [16.01.2016]
[604] https://www.sensenetworks.com/products/macrosense%20technology%20platform/ [07.01.2016]
[605] http://www.wired.co.uk/article/the-hidden-persuaders-mining-your-mobile-phone-log [07.01.2016]
[606] https://www.sensenetworks.com/retail-retargeting/ [07.01.2016]
[607] https://www.sensenetworks.com/audience-segments-and-results/ [07.01.2016]
[608] https://www.sensenetworks.com/principles/privacy-policy/ [07.01.2016]
[609] https://www.sensenetworks.com/principles/privacy-principles/ [07.01.2016]

However, many users consent to mobile apps asking for permission to access user data "without understanding the agreement or appreciating the consequences".[610]

5.7.11 Adyen, PAY.ON and others – payment and fraud detection

Many new players in the field of online payment are also developing risk management and fraud detection technologies, and thus analyzing vast amounts of data about consumer behavior and about their devices. In addition, these companies often also offer credit scoring and algorithms to make automated decisions on consumers, for example on payment methods offered – or even to exclude consumers from shopping.

Adyen, creating a „holistic view" of shoppers

The Amsterdam-based payment company ***Adyen***, for instance, describes its "risk mitigation" platform "RevenueProtect" as a tool for corporate customers to "maintain the perfect balance between fraud defense and optimized conversions".[611] It utilizes "lists of known good and bad shopper attributes (e.g. card numbers)", external risk checks, and "device fingerprinting" to "identify the same machine across multiple sessions, despite the user changing login identities, clearing cache and cookies, and attempting other obfuscation techniques".[612] A feature called ***ShopperDNA*** claims to build "a holistic view of the shopper behind each transaction by using advanced linking algorithms, proprietary device fingerprinting and network intelligence to track devices, networks and online persona".[613] It allows the "creation of automated rules that monitor the behavior" of shoppers across different transactions.[614]

PAY.ON

Germany-based payment service provider and *Bertelsmann* affiliate ***PAY.ON***'s fraud prevention tools include "**more than 120 risk checks**". Besides "device fingerprinting" and "black and white listing"[615] they offer access to "third-party databases, such as address verifications and credit scores" regarding "which payments shall be accepted, denied or manually reviewed"[616]. Examples for third-party providers mentioned are *ThreatMetrix, ReD Shield, Datacash Gatekeeper, Schufa, Telego! creditPass, Deltavista, Deutsche Post Address Services, Intercard, Creditreform Boniversum, Arvato infoscore* and more.[617] *PAY.ON* also provides a system to offer shoppers "the right set of payment methods" according to the "shopper risk group" based on "risk and fraud checks, historic customer information, [...] external data (e.g. credit agency records), identity checks and the differentiation of new and existing customers as well as shopping basket information and dynamic limit management".[618]

Customer profiles based on bank transactions

Chester et al (2014, p. 11) summarized a white paper produced by *TSYS*, another leading payment processor, stating that companies in the **financial services industry** now enjoy "unprecedented levels of insight to use in their consumer decision-making". The original TSYS whitepaper[619] provides details on how transaction histories could "provide banks

---

[610] Dwoskin, E. (2014): In Digital Ads, It's Location, Location, Location. The Wall Street Journal, Jan. 06, 2014. Online: http://blogs.wsj.com/digits/2014/01/06/in-digital-ads-its-location-location-location/ [07.01.2016]
[611] https://www.adyen.com/our-solution/risk-management [17.08.2016]
[612] https://docs.adyen.com/developers/revenueprotect [17.08.2016]
[613] https://www.adyen.com/our-solution/risk-management [17.08.2016]
[614] https://docs.adyen.com/developers/revenueprotect [17.08.2016]
[615] https://www.payon.com/fraud-prevention [07.01.2016]
[616] https://www.payon.com/sites/www.payon.com/files/downloads/Product sheet External Risk Providers.pdf [17.08.2016]
[617] Ibid.
[618] https://test.payon.com/sites/www.payon.com/files/downloads/PAYON%20Productsheet-Active-payment-method-selection.pdf [17.08.2016]
[619] Hudson, R. (2013): How Card Issuers Can Leverage Big Data to Improve Cardholder Retention Efforts. TSys People-Centered Payments, Jun 2013, Online:

with a robust customer profile, including an indication of the customer's approximate annual income, spending habits, online usage patterns and transaction types, along with how he or she typically makes payments". **Transaction data** is, according to *TSYS*, "extremely valuable for predicting future customer behaviours and transactions" and it "will provide a more complete picture of cardholder behavior and, in turn, identify which cardholders are most profitable".[620]

5.7.12 MasterCard – fraud scoring and marketing data

Fraud scoring and risk scores

MasterCard provides "fraud scoring" technologies to financial institutions.[621] They have developed predictive fraud models to reveal the risk based on spending patterns using a "vast repository of globally integrated authorization and fraud data". In the context of credit scoring *MasterCard* has even utilized data from mobile phones to calculate risk. According to a report by the company[622], they have "developed models showing that prepaid-mobile history and phone usage are predictive of ability — and willingness — to repay loans".

MasterCard's data for marketing

MasterCard also offers their data for marketing purposes. They provide "access to relevant and actionable intelligence based on 95 billion anonymized, real transactions from **2 billion cardholders** in 210 countries worldwide" to "[f]orecast consumer behavior" and to "[h]elp clients make better decisions".[623] Their product ***Propensity Models for Marketing***[624] enables clients to use **scores** that "reflect a cardholder's likelihood to engage in a behavior or respond to an offer". These models are "available on consumer debit, consumer credit and commercial portfolios". They explain that a "propensity model rank" would order the "best prospects" within the client's "cardholder population".[625]

Connecting purchases to online data

Another product called ***MasterCard Audiences***[626] allows companies to reach "online audiences based on insights drawn from aggregate purchase behavior" for "more precise online marketing". According to *MasterCard*, transaction data is "[a]nonymous – no name or contact information of any kind", but "[i]ncludes transaction amount, merchant, online/offline, location, date and time". This data is analyzed in order to "create millions of segments" and to build "hundreds of unique audiences" by "aggregating segment propensities and applying them to third-party consumer populations" available through "ad networks and data aggregators".[627]

Partnerships with DMPs and Facebook

According to *Nielsen*'s data platform ***eXelate***, which is offering *MasterCard*'s data to marketers, it is "collected from online & offline anonymized transactions and associated with online populations through the use of proprietary analytics".[628] Another partner is ***Oracle***, who explains in its "Data Directory" (Oracle 2015, p. 84) that *MasterCard*'s "behavioral based segments" like "used vehicle sales", "luxury" or "professional services" (e.g. "electricians, accounting, tax, legal") are available in several categories like "top tier spenders" or "frequent transactors". Data is "**associated with cookie populations** through a proprietary 'privacy by design' double blind matching process", but "[n]o PII is

---

http://tsys.com/Assets/TSYS/downloads/wp_How Card-Issuers-Can-Leverage-Big-Data-pc.pdf (p.4) [17.08.2016]

[620] Ibid. (p. 5)[07.01.2016]

[621] https://www.mastercard.com/us/company/en/docs/ems_hosted_sell_sheet.pdf [07.01.2016]

[622] compendium.mastercard.com/app/SKU_pdfs/alternativeData.pdf [01.08.2016]

[623] www.mastercardadvisors.com/information-services.html [07.01.2016]

[624] www.mastercardadvisors.com/solutions/product_list/propensity_models_for_marketing/ [07.01.2016]

[625] Ibid.

[626] www.mastercardadvisors.com/solutions/media/customer_insights/mastercard_audiences.html [07.01.2016]

[627] Ibid.

[628] http://partners.exelate.com/media/1/_orig/1441043383-5557.pdf [07.01.2016]

collected or leveraged in MasterCard's processes". In 2014, the media reported that ***MasterCard*** and ***Facebook*** "signed a two-year deal to share data".[629] According to the *Daily Mail*, a *Facebook* spokesman said: "'We are working with them to create targeting clusters using Custom Audiences — a tool that matches anonymised data from Facebook with their own anonymised data for optimising ad delivery on Facebook to their users."[630]

Will MasterCard, Visa and AmEx become data companies?

According to the trade magazine *payment week*[631], *MasterCard* reported $2.177 billion in revenue from payment processing in Q1 2014 and $341 million for "information products, including sales of data" already. However, the "rate of growth for the latter was 22 percent versus 14 percent for payments". The article suggests that "selling products and services created from data analytics could become" *MasterCards*'s "core business". Besides ***American Express***, which has also started to offer "audience segments for use in online ad targeting"[632], ***Visa*** recently launched its "Visa Integrated Marketing Solutions", which allows "[c]ard issuers and partners" to "combine information about their own customers with powerful insights from more than 100 billion transactions" per year, as well as "other third-party demographic, lifestyle and economic data to inform their programs".[633]

[629] Michael, S. (2014): MasterCard is mining Facebook users' conversations data to get consumer behavior information it can sell to banks. DailyMail Online, Oct. 06, 2014. Online: www.dailymail.co.uk/news/article-2782937/MasterCard-mining-Facebook-users-conversations-data-consumer-behaviour-information-sell-banks.html [07.01.2016]
[630] Ibid.
[631] Genova, J. (2014): For MasterCard, Processing and Analytics go Hand in Hand. Paymentweek, Jun. 16, 2014. Online: http://paymentweek.com/2014-6-16-for-mastercard-processing-and-analytics-go-hand-in-hand-4908 [07.01.2016]
[632] Kaye, K. (2013): Mastercard, AmEx Quietly Feed Data to Advertisers. AdvertisindAge, Apr. 16, 2013. Online: http://adage.com/article/dataworks/mastercard-amex-feed-data-marketers/240800/ [07.01.2016]
[633] http://investor.visa.com/news/news-details/2015/Visa-Launches-New-Platform-to-Help-Card-Issuers-Market-and-Grow-Their-Business/default.aspx [07.01.2016]

6. Summary of Findings and Discussion of its Societal Implications

"If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place"
Eric Schmidt, Google, 2009[634]

"Surveillance is not about knowing your secrets, but about managing populations, managing people"
Katarzyna Szymielewicz, Vice-President EDRi, 2015[635]

Around the same time as *Apple* introduced its first smartphone[636] and *Facebook* reached 30 million users[637] in 2007, online advertisers started to use individual-level data to profile and target users individually (Deighton and Johnson 2013, p. 45). Less than ten years later, ubiquitous and real-time corporate surveillance has become a "convenient by-product of ordinary daily transactions and interactions" (De Zwart et al 2014, p. 746). We have entered a **surveillance society** as David Lyon foresaw it already in the early 1990s; a society in which the practices of "social sorting", the permanent monitoring and classification of the whole population through information technology and software algorithms, have silently become an everyday reality (see Lyon 1994, Lyon 2003).

Ubiquitous, invisible and pervasive

This surveillance society is enabled by a number of phenomena, which we summarize and reflect on in this chapter. At the core of our current surveillance society is the technical idea that computing should be "ubiquitous", "invisible" and "pervasive". The acknowledged founding father of this thinking is Mark Weiser, an American research scientist who used to work at *Xerox*, who once wrote: "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it [...] we are trying to conceive a new way of thinking about computers in the world, one that takes into account the natural human environment and allows the computers themselves to vanish into the background" (Weiser 1991, p.1).

Who is in control?

Ever since this vision was formulated, computer scientists and engineers around the world have been working towards realizing it; interpreting the aspect of disappearance as a ubiquitous, sensor-based and networked digital infrastructure. Few engineers have probably been expecting that the "mind-children" of this vision and their subsequent work on it would be abused by economic and governmental forces in the way that it is today. Bathing themselves in the shallow reassurance that "technology is neutral" they have been laying powerful tools in the hands of many players. As this report shows it is not obvious that all of the players are able to live up to the responsibility required for them, because responsible use of data would include an ethical questioning and partial refraining from practices we observe today. Few tech people have taken the warnings that Marc Weiser voiced seriously. In 1999, he wrote that "the problem [associated with ubiquitous computing] while often couched in terms of privacy is really one of control" (Weiser et al. 1999, p.694). Very slowly, computer scientists and engineers around the world are

634 Esguerra, Richard (2009): Google CEO Eric Schmidt Dismisses the Importance of Privacy. Electronic Frontier Foundation. Online: https://www.eff.org/de/deeplinks/2009/12/google-ceo-eric-schmidt-dismisses-privacy [01.08.2016]

635 Grossman, Wendy M (2016): Democracy, film review: How the EU's data protection law was made. ZDNet UK Book Reviews, June 9, 2016. Online: http://www.zdnet.com/article/democracy-film-review-how-the-eus-data-protection-law-was-made [01.08.2016]

636 http://www.apple.com/pr/library/2007/01/09Apple-Reinvents-the-Phone-with-iPhone.html [01.08.2016]

637 Phillips, Sarah (2007): A brief history of Facebook. The Guardian, 25 July 2007. Online: https://www.theguardian.com/technology/2007/jul/25/media.newmedia [01.08.2016]

realizing that they might have summoned technological 'spirits' that now ignore their command. The following sections summarize the state-of-the-art as we see it, based on the facts accumulated in the chapters above.

Status quo

In a nutshell, the **ubiquity** of data collection and sharing through a vast globally networked digital infrastructure has led to a loss of control over data flows and a **sacrifice of contextual integrity** of personal data. As Helen Nissenbaum (2004) has argued, contextual integrity of data is a cornerstone for the protection of peoples' privacy. With the vast and uncontrolled sharing practices outlined in previous chapters, privacy is undermined at scale without people noticing it.

We assume that companies do not strive to consciously harm their customers. But they are confronted with the fact that data has become such an important and strategic part of many business models that they can hardly see a way out of this lucrative personal data market and the dynamics it has created. To deal with the rising criticism voiced by the public, the media and political institutions, companies now look for strategies to deal with the data business and the associated ethical and legal challenges. An important compromise in this effort could be to create more **transparency** around their data-intensive processes. Not surprisingly, transparency has been a core concern for the new European General Data Protection Regulation (GDPR).[638] That said, as of today the status quo is: Transparency is not provided, but avoided. Ambiguous business practices are still the norm and even misleading rhetoric is used to trick people into one-sided and disadvantageous data contracts.

The lack of transparency is one enabler of **power imbalances** between those parties that possess data and those who don't. Democratic as well as economic thinkers have always been suspicious of information and power asymmetries. And the current abuses of personal data that are highlighted in our report support their suspicion: Data richness is systematically used to **discriminate** against people. Companies „turn individuals into ranked and rated objects" (Citron and Pasquale 2014). Everyone is constantly sorted and addressed on the basis of their economic potential; a practice that is undermining the core values of democracy: people's equality and dignity.

Against this background, consumers using digital services are advised to consider what we call their **"customer lifetime risk"** when starting to interact with a digital service provider. The question is whether they will ever do so? - When in fact people embrace the data-rich services that betray their trust. As we discuss at the end of this chapter, perhaps, we are entrained to love embracing the soft digital controls rising around us and are thus on the verge of becoming perfect self-censors.

6.1 Ubiquitous data collection

Networks of corporate surveillance

Our report shows that the collection of data concerning people and their daily lives has become **ubiquitous**. As more and more devices and objects include sensors and network connections, data collection is happening **invisibly**. Information recorded by websites, smartphone apps, fitness trackers and many other platforms is often transferred to a wide range of **third-party companies.** A network of major online platforms, publishers, app providers, data brokers and advertising networks is now able to recognize, profile and judge people at nearly every moment of their lives. By using pseudonymous identifiers based on phone numbers, email addresses and other attributes, profiles are matched

[638] European Commission (2016)

cross-device and cross-platform with digital records in customer databases of a myriad of other businesses (see chapter 5). More and more physical objects and spaces are connected to the Internet, ranging from printers, fridges, cars and doors to objects located in offices, industrial plants or in public space. The Internet of Things envisions billions of networked sensors that are recording our lives, in cities, vehicles, offices, factories, at home and even in our bodies.

Example data ecosystem

Many parties are interested in the recorded data. The extent to which invisible sharing across devices and platforms is possible can be demonstrated by revisiting the example of the "connected car", as compiled by FIPA (2015). The car data created is attractive not only for automakers and their partners but also for car dealers, insurance companies, lenders, telematics service providers, call center operators, third-party app developers, vehicle infotainment content providers, mobile network operators or mobile device system providers like *Google* and *Apple*. Also, third parties outside the telematics industry itself are standing in line to acquire telematics data, including local retailers and merchants, online advertising agencies, data brokers, law enforcement agencies, debt collectors, fraud investigators, litigants and many more can be added to the list of potential bidders for the data.

As many other examples in our report show, personal data already is, and will increasingly be, used in completely different contexts or for different purposes than it was initially collected and this is done at ubiquitous scale.

6.2 A loss of contextual integrity

As data is used for other purposes than the ones stated at the time of its collection, it may lose its contextual integrity.

Marketing, risk assessment and employment

Data that has been collected in the context of online **fraud prevention, credit scoring or payment processing** is used for customer relationship management, online targeting and other marketing purposes. For example, advertising gets pre-filtered according to risk judgments and "high risk" customers are treated differently, or even excluded from the beginning. Conversely, data collected in marketing contexts, by smartphone apps or social networks, is used for risk assessment. Generally, companies and practices in marketing and risk assessment are increasingly merging (see chapters 3.5 and 5.7). In the **realm of work**, information that is collected to improve business processes – for example, customer satisfaction reports, location tracking in logistics, in-store-tracking for customer analytics, data from project management tools – is also used to monitor, judge and control employees (see chapter 3.3).

Since the 1990s, scholars have used the term **function creep** to describe when systems, which are recording digital data about people, are later being used for tasks other than those originally intended (see Lyon 2010, p. 330). For example, the UK database of school children, which was set up in 1997 to collect "general aggregate data to plan for services", later "became a means of amassing detailed information on children—how they arrive at school, who eats meals at school, who has special needs" (ibid). This phenomenon can be observed in many areas today. More and more businesses are collecting vast amounts of information about people without even knowing yet in which context or for what purpose

they want to use it for later. As the founder of a credit scoring start-up stated: "We feel like all data is credit data, we just don't know how to use it yet".[639]

Loss of context

A player that demonstrates a powerful way for the decontextualization of data is *Facebook*. Facebook encourages its users to provide information as accurate, real, valid and complete as possible (Dumortier 2009, p. 1). *Facebook* is not known for directly selling the personal profiles it collects; its core asset. However, the company allows a large group of marketers and app developers to leverage user data for targeted advertising and other purposes. Consequently, a *Facebook* user's data may reappear or may be re-used in very different contexts than expected by the users. For example, the data broker Experian on its website offers its corporate clients the ability to harness individual-level social data from Facebook, including names, fan pages, relationship status and posts (see chapter 5.7.3). Oracle recommends that companies integrate their enterprise data with social data (see chapter 5.7.2). In practice, this means that the social networks of Facebook users are for instance used very successfully by car insurers for their fraud prevention algorithms. Telecom operators use Facebook status data to double-check whether contract-relevant data provided to them is correct (i.e. whether it corresponds to what a person has stated about themselves on Facebook).[640] These examples demonstrate how *Facebook* has created an "asymmetry between a user's imagined and actual audience" that "threatens the possibility of the individual to act as a contextual and relational self" (Dumortier 2009, p. 11). The same is true for many other platforms, services and apps. The loss of one's contextual and relational self again deeply undermines personal dignity.

Corporate data accessed by governments

Another example of extensive de-contextualization is **governmental surveillance** that is enabled by corporate databases (see e.g. De Zwart et al 2014). Many documents revealed by Edward Snowden showed that governmental authorities are excessively accessing information about citizens, which was originally collected by corporate players. In addition, companies and institutions that investigate insurance claims or social benefits often use the same analytics tools that are used to investigate terrorism or delinquency (see chapter 3.5).

6.3 The transparency issue

Both, Edward Snowden's revelations and critical investigations of data markets (such as this report) are slowly opening peoples' eyes to the massive data-sharing and decontextualizing that is going on. As a result, established policy bodies have started to call for more transparency around data flows, including the World Economic Forum (WEF 2014), the U.S. Federal Trade Commission (FTC) and the European Parliament. Together with the Parliament, the European Commission passed the GDPR, a new regulation on data protection which will come into effect in 2018 and tries to sanction companies when they collect too much data about individuals and use it out of context.

Lack of transparency

That said, at this point in time, transparency is missing. Companies are collecting data about billions of consumers from various sources (see chapter 5), "largely without consumers' knowledge" (FTC 2014, p. IV). As data brokers often share data with others, it is "virtually **impossible for a consumer** to determine how a data broker obtained" their

---

639 Hardy, Quentin (2012): Just the Facts. Yes, All of Them. New York Times, 24.03.2012. Online: http://www.nytimes.com/2012/03/25/business/factuals-gil-elbaz-wants-to-gather-the-data-universe.html [27.07.2016]

640 According to a presentation of the author Yvonne Hofstetter held at the Club of Vienna in early 2016

data (ibid). Most consumers have "no way of knowing that data brokers may be collecting their data" at all (Senate Committee on Commerce, Science, and Transportation 2013, p. 32). Consumers are often neither aware of what personal information about them and their behavior is collected, nor how this data is processed, with whom it is shared or sold, which conclusions can be drawn from it, and which decisions are then based on such conclusions (see chapters 2-5, IWGDPT 2014, Tene and Polonetsky 2013). Both dominant platforms and smaller providers of websites, services, apps and platforms – generally speaking – act in a largely non-transparent way when it comes to the storage, processing and the utilization of personal data.

Data brokers sharing data with each other

To determine how a company obtained someone's data, consumers need to be able to "retrace the path of data through a series of data brokers" (FTC 2014, p. IV), which is very challenging. In fact, a number of technical tools that would support this have been proposed and even standardized in the past. These include, amongst others, the **P3P Protocol** (Cranor 2003; Cranor et al. 2006) and academic work around **sticky policies** (Casassa Mont et al. 2003). A good overview of current works in this direction can be gained form a special issue of the journal Electronic Markets on "Personal Data Markets and Privacy" co-edited by one of the authors of this report (Spiekermann et al. 2015). However, most industry players have so far refused to co-operate in the development and avoided the use of existing technical standards, such as the W3C P3P standard; a refusal that can be well recapitulated when looking into the failed debates around a potential **Do No Track** standard at the W3C or CMU Professor Lorrie Cranor's account of past debates with industry (Cranor 2012).

Instead of enabling transparent data collection and data flows, businesses often use ambiguous and misleading rhetoric, both in user interfaces and documents; i.e. in their terms and conditions. As the Norwegian Consumer Council (2016, p. 4) observed, many terms of apps use "difficult and obscure language" and have "generally unclear and complicated terms dominated by hypothetical language", such as "may" and "can". Many apps "reserve the right to share personal data with unspecified third parties for poorly specified purposes". Some "treat personal and identifiable data as non-personal data" or use "unclear or ambiguous definitions of personal data" (ibid, p. 16). Danah Boyd et al (2010, p.5) observed that companies are even "tricking [people] into clicking through in a way that permission is granted unintentionally".

Hidden options and limited choices

An ethically ambiguous way to receive consent for various secondary uses of data is also to hide or omit choices, which would actually be very important to protect one's privacy. *Facebook,* for example, offers a set of prominently placed "privacy checkup" options. Users can easily access options such as "who can see my stuff", "who can contact me" and "how to stop someone from bothering me".[641] Although *Facebook* hides some additional settings relevant to privacy under the title "ads", which allow users to control some aspects of how third-party companies can make use of their data on and off the *Facebook* platform, the company doesn't give users a simple option to disallow third-party companies from making use of their data.[642] Instead, by focusing on options such as "who can see my stuff" *Facebook*'s "privacy checkup" promotes a very limited concept of privacy, which does not include *Facebook*'s own utilization of the collected data.

Data security vs. data privacy

Misleading or limited rhetoric can also be observed when it comes to the description of how data is treated. Companies often indicate that data will be **"anonymized" or "de-identified"** when they are in fact using pseudonymous identifiers to track, match, profile,

[641] Facebook user interface, accessed from a personal account on August 13, 2016
[642] Ibid.

and target individuals (see chapter 5.6). In addition, businesses sometimes seem to **intentionally confuse data privacy and data security**. For example, when it is emphasized that employers or insurers don't have access to raw fitness data recorded by activity trackers, because it would be managed by a "neutral" third party (see chapter 4.3.4). This seems to be clearly beneficial from a data security point of view. However**,** from a data privacy point of view the crucial question is not so much, who has access to the raw data, but who has access to the enriched information, which is derived from it, such as activity indices or health scores.

Users are informed inaccurately

Taken together, users are often informed **incompletely, inaccurately or not at all** about which data is being collected and shared with third parties (see also chapters 4.2.1 and 4.3.3). Many companies do not even allow users to access their own data, and they consider their algorithms as trade secrets (see Citron and Pasquale 2014, Weichert 2013). Most importantly, companies normally don't provide consumers with the ability to access **information inferred from collected data**, although it is used to sort and categorize them on an individual level (see Senate Committee on Commerce, Science, and Transportation 2013). Consumer scores, which are "derived from many data sources" and "used widely to predict behaviours like spending, health, fraud, profitability" are opaque and "typically secret in some way". The "existence of the score itself, its uses, the underlying factors, data sources, or even the score range may be hidden" (Dixon and Gellman 2014, p. 6).

6.4 Power imbalances

A Big Data divide

While users become more and more transparent, corporate data mining practices remain largely obscure. This is resulting in an imbalance of power between users and companies (IWGDPT 2014, p. 9). Mark Andrejevic (2014, p. 1673) stated that this "asymmetric relationship between those who collect, store, and mine large quantities of data, and those whom data collection targets" leads to a **Big Data divide**. He points out that the "systemic, structural opacity" (ibid, p. 1677) of today's practices in data mining creates a divide between those "with access to data, expertise, and processing power" (ibid, p. 1676), who are able to analyze, categorize and sort people, and those "who find their lives affected by the resulting decisions" (ibid, p. 1683). His reflections are based on Boyd and Crawford (2012, p. 674), who recognized a "new kind of digital divide" between the **Big Data rich** and the **Big Data poor**. Tene and Polonetsky (2013, p. 255) compared the relation between users and large data owners to a "game of poker where one of the players has his hand open and the other keeps his cards close". The player whose hand is open will always lose. Consumers have very limited power to influence how companies behave and they cannot **democratically participate** in decisions about how the systems and platforms work.

Opting out is difficult

This is especially true because opting out from data collection becomes increasingly difficult, or nearly impossible. Today, consumers can "hardly avoid privacy contracts: almost all banks, software and hardware vendors, social networking sites, digital content services, retail loyalty programmes and telecommunications providers employ them" (Rhoen 2016, p. 2). For example, to use a standard mobile phone people have to link it to a user account at one of the major platforms such as *Google* and *Apple*, at least as long as they don't have special technical knowledge (see chapter 4.1). The terms, which allow companies to collect and use personal data on many levels, are almost non-negotiable for consumers. The programmer and web entrepreneur Maciej Cegłowski (2016) pointed out that "opting out of surveillance capitalism is like opting out of electricity, or cooked foods — you are free to do it in theory". In practice, it would mean "opting out of much of modern life". Whether the new European GDPR is really effectively able to improve this

with its Article 7 on the conditions of consent remains to be seen. Either way it would only be an improvement for European citizens and not a worldwide solution.

Sometimes, corporate leaders and others argue that "privacy is dead" and that really people do not care about privacy any more. Their most important argument for this conclusion is the wide use of *Facebook* and other popular services. We do not agree with this argument. In contrast, research shows that Internet users do perceive the power asymmetries online and react to them. Mark Andrejevic's (2014, p. 1685) qualitative research shows that users feel "frustration over a sense of powerlessness in the face of increasingly sophisticated and comprehensive forms of data collection and mining". He argues that users "operate within structured power relations that they dislike but feel powerless to contest" (ibid, p. 1678). A recent study of over 1300 Facebook users by one of the authors of this report showed that 90 to 95% think twice before they post anything (Spiekermann et al. 2016). Shoshana Zuboff (2015, p. 82) points to the chilling effects of "**anticipatory conformity**", which "assumes a point of origin in consciousness from which a choice is made to conform for the purposes of evasion of sanctions and social camouflage".

Given that the World Wide Web started out as a powerful technology for communication and knowledge, where people could open up freely, it is a pity where it has arrived now. Today, it "has become a system that is often subject to control by governments and corporations", as the New York Times frankly stated in an article about Tim Berners-Lee, its creator.[643]

6.5 Power imbalances abused: systematic discrimination and sorting

If companies and corporate leaders would take ethics and social responsibility as serious as they sometimes claim, then power imbalances could possibly be more acceptable for users. Unfortunately, the abuse of information asymmetries in personal data markets shows only the contrary. The available information tends to be leveraged only for economic corporate advantage. Ethical reflections play no role. Even legal boundaries have been widely ignored where a lack of sanctions allowed it. In countries where laws and directives exist that protect consumer's privacy, those regulations have been bent, undermined and misinterpreted frequently.

Limiting chances and choices

When companies use predictive analytics to judge, address or treat people differently or even to deny them opportunities, the chances and choices of the individuals become limited (see Lyon 2003). A classic example is the practice known as **redlining**, when financial institutions use information about an individual's neighbourhood to predict risk and creditworthiness. While companies may know that redlining is biased against poor neighbourhoods and fails "to capture significant variation within each subpopulation", they might use it anyway, because profits are higher than costs from inaccuracy (Barocas and Selbst 2016, p. 689). In the field of employment, systems that automatically **score, sort and rank resumes** may lead to the unfair discrimination and exclusion of applicants, depending on which individual attributes are included in which kinds of predictive models (see chapter 3.3).

[643] Hardy, Q. (2016): The Web's Creator Looks to Reinvent It. New York Times, June 7, 2016. Online: http://www.nytimes.com/2016/06/08/technology/the-webs-creator-looks-to-reinvent-it.html [30.08.2016]

Amplifying existing inequalities?

Scores about consumers sometimes even *create* the "financial distress they claim merely to indicate", and thus become **self-fulfilling prophecies** (Citron and Pasquale 2014, p. 18). The "act of designating someone as a likely credit risk" may raise the cost of future loans, insurance rates and/or decrease employability for this individual. Similarly, automatic judgments in hiring decrease future employability (ibid). But scoring is not only used in such crucial areas as banking, insurance and employment today. Data brokers and other businesses offer scores that segment, rate and rank consumers in many areas of life. **Consumer scores** predict, for example, the profitability of individuals and their future income, the likeliness that someone will take medication or not, possible care needs and even mortality (see chapter 5.4).

Many small decisions on consumers

Compared to crucial fields of application such as banking, insurance, health, employment or law enforcement the use of personal data for marketing purposes has often been considered as less relevant for rights and justice. However, as this report shows, the spheres of marketing and risk management are increasingly merging. Advertising can now be personalized on an individual level, and people are recognized, profiled and matched in real-time – across devices, platforms and customer databases from myriads of companies. Businesses are **constantly sorting and categorizing** both customers and prospects, when they are surfing the web or using mobile devices, according to how **valuable or risky** they are. Consequently, businesses can, for example, calculate the exact minimum action necessary to keep customers loyal. Today, data about consumer's lives and behavior is used to make many different small **decisions** about them every day – ranging from how long someone has to wait when calling a phone hotline (Graham 2005, p. 569) to which contents, ads, offers, discounts, prices and payment methods someone gets (see chapters 3.6 and 5.7).

Cumulative disadvantage

When, as suggested by a major data broker, the **top 30%** of a company's customers are categorized as individuals who could add 500% of value, and the **bottom 20%** of customers are categorized as individuals who could actually cost 400% of value, the company may "shower their top customers with attention, while ignoring the latter 20%, who may spend 'too much' time on customer service calls, cost companies in returns or coupons, or otherwise cost more than they provide" (Marwick 2013, p. 5). These "low-value targets" have been categorized as "**waste**" by data brokers (ibid). In contrast, Internet users, whose **customer lifetime value** has been recognized as high based on a wide range of data, increasingly receive personalized offers, calls and discounts via web and mobile ads, email and other channels (see chapter 3.6). Privacy expert Michael Fertik (2013) stated that the "rich" would already see "a different Internet" than the "poor". Subsequently, when consumers are categorized as non-valuable or risky they experience many small disadvantages in their everyday lives, each of them not very significant on their own, but accumulated resulting in a significant disadvantage in life. Perhaps this phenomenon could be labelled as a **cumulative disadvantage**. Originating from inequality theory in sociology[644], Oscar Gandy (2009, p. 1) used this term to describe the "application of probability and statistics to an ever-widening number of life-decisions", which "shape the opportunities people face" and "reproduce, reinforce, and widen disparities in the quality of life that different groups of people can enjoy".

How to prove unfair discrimination?

In addition, the different treatment of individuals based on opaque, automated decisions and a wide range of data entails another problem. As long as data and algorithms are

[644] See e.g. Ferraro, Kenneth F., and Tetyana Pylypiv Shippee (2009): Aging and Cumulative Inequality: How Does Inequality Get Under the Skin? The Gerontologist 49.3, 333–343, PMC. Online: https://www.researchgate.net/publication/281549834_Cumulative_inequality_theory_for_research_on_aging_and_the_life_course

secret, it is not possible to even notice or prove discrimination. For example, existing studies on personalized pricing show that it is challenging – if not impossible – to accurately investigate, whether online shops offer different products or prices to different consumers based on individual attributes or user behavior (see chapter 3.6). Under these circumstances **consumers have no chance** to understand, what their individual offers and prices are based on, or whether they get individual offers and prices at all.

Internet of Things as a key to behavioral change

With the Internet of Things and ever more data collected through our incredibly "smart" environments, such sorting practices are likely to increase further. Tim O'Reilly stated in 2014 that "advertising turned out to be the native business model for the internet", but he expects that "insurance is going to be the native business model for the Internet of Things"[645]. In a recent report about "The Internet of Things: Opportunities for Insurers" a consulting firm explains that insurers could "use IoT-enriched relationships to connect more holistically to customers and influence their behaviors".[646] Many experts interviewed by Pew Research (2014, p. 8) expect that "**incentives to try to get people to change their behavior**" will become a "major driver" of the Internet of Things – for example, to motivate people to purchase a good, to act in a more healthy or safe manner or to perform in a certain way at work. They conclude that the "realities of this data-drenched world raise substantial concerns about privacy and **people's abilities to control their own lives**" (ibid, p. 9).

6.6 Companies hurt consumers *and* themselves

Inaccurate data

As in fairy tales, bad practices fire back. Companies which are most successful in their personal data business often have a miserable public image and trust ratings are low. Moreover, while they may be successful businesses at first sight, they often trade a big chunk of out-dated data In the U.S., 26% of participants in a survey identified at least one error on at least one of their three credit reports (FTC 2012, p. i). In Germany the validity of credit scores has been assessed as questionable and often based on estimations (see chapter 5.4). (Credit scoring is amongst the best-regulated realms for algorithmic judgment on individuals). "Data quality problems plague every department, in every industry, at every level, and for every type of information [...] Studies show that knowledge workers waste up to 50% of time hunting for data, identifying and correcting errors, and seeking for confirmatory sources for data they do not trust", writes David Redman in Harvard Business Review in 2013 (Redman 2013, p. 2). In fact, even commercial computer programs rarely come without bugs. Typically, there are at least a few mistakes in every 10,000 lines of code, even in professionally commercialized software products.[647]

---

645 Myslewski, Rik (2014): The Internet of Things helps insurance firms reward, punish. The Register, 24.05.2014. Online http://www.theregister.co.uk/2014/05/23/the_internet_of_things_helps_insurance_firms_reward_punish [01.08.2016]

646 ATKearney (2014): The Internet of Things: Opportunity for Insurers. December 2014. Online: https://www.atkearney.com/digital-business/ideas-insights/featured-article/-/asset_publisher/Su8nWSQlHtbB/content/internet-of-things-opportunity-for-insurers/10192 [01.08.2016]

647 Martin C. Libicki, Lillian Ablon, Tim Webb (2015): The Defender's Dilemma: Charting a Course Toward Cybersecurity. RAND Corporation, p. 42. Online: http://www.rand.org/content/dam/rand/pubs/research_reports/RR1000/RR1024/RAND_RR1024.pdf

Inaccurate conclusions

While errors in collected or transmitted data, inaccurate classifications and assessments based on flawed prediction models and data analytics can massively impact the lives of individuals (see IWGDPT 2014, National Consumer Law Center 2014, FTC 2016), Big Data analytics is far from objectivity and accuracy in general (see Boyd et al 2012). Predictions are blurry by design, because they are based on **correlations and probabilities**. Someone, for example, knowing the wrong people, living in the wrong district or swiping in a wrong way when using a mobile app may get categorized and judged in a certain negative way. Companies carry the risk that their digital profiles de-contextualize and misinterpret the interactions they recorded about their customers. The underlying "motivations for particular actions are never explained or understood" (De Zwart 2014, p. 718). Big Data analytics can also lead to "more individuals mistakenly being denied opportunities based on the actions of others" just because they share some characteristics with other consumers (FTC 2016, p. 9).

Accurate, but biased

It is not only inaccuracy that may harm consumers, but also data and predictions that are "too accurate" (see Barocas and Selbst 2016). When companies incorporate sensitive personal attributes such as gender, age, ethnic or religious affiliation, poverty or health into their automated decisions, this can lead to discrimination or even to the exclusion of entire parts of the population. For example, an insurance company could "classify a consumer as higher risk", when this individual was categorized to have an "interest" in diabetes before he is actually suffering from it (FTC 2014, p. vi). This can also happen when sensitive attributes are not obtained directly from individuals, but calculated by algorithms based on statistical analysis or machine learning. **Refusal to participate** in digital tracking may have consequences too. If not enough data about a person is available, the risk of a customer relationship may be considered as too high, also in the case where this would have been a good customer.

Security threats

Finally, another problem for companies is that large data volumes also come with a certain liability. Data security and effective data protection becomes a costly and risky issue for them. Wherever large amounts of data are stored there is a risk of data abuse and loss (see also chapter 4). According to *DatalossDB*, 3,930 security incidents were reported in the year 2015, exposing more than 736 million records about individuals such as email addresses, passwords and usernames.[648] The operational costs of the IT needed are huge.

6.7 Long term effects: the end of dignity?

Online marketing aims to target individuals, who could become valuable customers or loyal users, and in many cases to avoid individuals, who won't. However, the goal is not only to reach people, who could be interested, but also to persuade them to act in certain ways, for example to click on an ad, participate in a survey, register for a service, or purchase a product. Online marketing aims to increase "conversion rates", which describe the percentage of people, who acted exactly in the way marketers or app developers wanted them to act. Marketers also want to prevent the loss of valuable customers ("customer churn") or they want them to purchase complementary products ("cross-selling") or more expensive products ("up-selling").[649]

---

[648] https://blog.datalossdb.org/2016/02/11/2015-reported-data-breaches-surpasses-all-previous-years/ [01.08.2016]

[649] See e.g. SCN Education (2001): Customer Relationship Management: The Ultimate Guide to the Efficient Use of CRM. Vieweg+Teubner Verlag, Wiesbaden.

Influencing behavior

These techniques of marketing and customer relationship management (CRM) are not new. But as opposed to former times, marketers are now able to track and measure **every single interaction** on an individual level. After measuring and analyzing behavior they try to "optimize" conversion rates, to make more "targets" act in the way they want them to act. They test different versions of functionalities, user interface designs, messages, wordings or even different discounts and prices, and then measure and analyze again how they can successfully influence behavior. Elements of "gamification" are used to reward and incentivize desired behavior (see chapter 4.3.1). The more a company knows about individuals, for example about their "personal biases and weaknesses" (Helberger 2016, p. 15), the better they can "change people's actual behavior at scale" (Zuboff 2016). Based on digital tracking companies sometimes even "overcharge [people] when the data collected indicates that the buyer is indifferent, uninformed or in a hurry".[650]

Pavlovian dogs

Kaptein et al. (2011, p. 66) pointed to the concept of **persuasion profiles**, which are "sets of estimates on the effectiveness of particular influence-strategies on individuals, based on their past responses to these strategies". Many businesses in marketing openly admit that they aim to achieve behavioral change.[651] The question is whether in business we might have gotten so used to these strategies that we overlook their questionable ethic. Can it be right to use consumers as Pavlovian dogs? Do persuasive strategies – when they become as personalized, pervasive and permanent as they do now – undermine the human **autonomy** that is central to human **dignity**? Or have we indeed created a digital environment in which people and masses are "managed" based on surveillance in the way Katarzyna Szymielewicz is quoted at the beginning of this chapter?

Personalized manipulation?

Manipulation of opinions and desires is unfortunately not limited to the realm of product and service marketing. Similar methods are used for **voter targeting**. For example, a campaign of a U.S. politician communicated political issues in different ways to different people, based on extensive data about 220 million U.S. citizens, which were categorized along their views on issues from environment, gun rights, national security or immigration (see chapter 3.1). Political manipulation through mass media and advertising has been discussed since ages. Scholars in communication studies have long challenged the idea of plain top-down manipulation as inappropriate and too simplistic[652], insisting that humans are able to use different individual appropriation of communication strategies. The shift to completely personalized interactions based on extensive individual profiles possibly creates new and unknown degrees of manipulation.

Reward and punishment

In addition to opaque forms of manipulation, we also observe open strategies to make people change their behavior. **Insurers** reward consumers when they voluntarily agree to wear devices, which permanently track their steps and everyday life activities, or when they consent to the digital recording of their driving behavior. When they reach arbitrary, but transparent goals – such as a million steps or not driving too long during the night – they receive financial incentives or discounts. Such programmes can be beneficial, as long as consumers **can choose to not participate** in digital tracking and have attractive alternative options. However, when the incentives offered are considerably more valuable in comparison to insurance programs without tracking, then it could effectively become mandatory for many people to participate; in particular for poorer people who cannot afford to miss out on the discounts. When people feel forced to participate in self-

---

650 Zarsky T (2004): Desperately Seeking Solutions: Using Implementation-Based Solutions for the Troubles of Information Privacy in the Age of Data Mining and the Internet Society. 56(1) Maine Law Review 13, p. 52. See also p. 30-31. Quoted from: Borgesius (2015)

651 See e.g. the results of a simple Google search after "marketing" and "behavioral change"

652 See e.g. Fiske, John. Introduction to Communication Studies. London: Routledge, 1990.

surveillance programmes due to social inequalities, the human right to be treated as equals is undermined. Dignity is undermined. A more detailed discussion of this aspect can be found in the next chapter.

Kafkaesque experience

Finally, within the current networks of digital tracking and corporate surveillance "we can never be sure, how we are being read or what the consequences are" (Lyon 2010, p. 332). People can at times confront a Kafkaesque experience. We don't know why we see a specific ad, why we receive a specific offer, or why we had to wait hours on the phone hotline. Was it because we acted in a specific way before? Did the newsletter of that political candidate contain issues, which were personalized in a specific way? Was it because we visited a specific website, used a specific mobile app, bought a specific product in the supermarket or watched a specific TV program? Could it happen that we get a loan denied someday, when we visit the online gambling website once too often today?

Humans as numbers

Under the conditions of today's opaque and non-transparent networks of digital tracking **individuals do not know**, which data about their lives is recorded, analyzed and transferred – and which decisions are being made based on this information. They can't see how advertising networks and data brokers are continuously calculating their "customer value" or "risk score", how these ratings are updated with each of their interactions, and how this influences the contents and options they see.

Markets for behavioral control?

As outlined above, Shoshana Zuboff (2015) pointed towards the effects of "anticipatory conformity". But she also went one step further than many other researchers and scholars, by pointing to the fact that we do not only see the rise of "markets for personal data", but "**markets for behavioral control**". She coined the concept of the **Big Other** to describe a new "distributed and largely uncontested new expression of power" (ibid, p. 75), which would be different from Big Brother's "centralized command and control" (ibid, p. 82). The Big Other refers to a "ubiquitous networked institutional regime that records, modifies, and commodifies everyday experience from toasters to bodies, communication to thought" (ibid, p. 81). It started within **surveillance capitalism**, a new "emergent logic of accumulation in the networked sphere", which led to "new markets of behavioral prediction and modification" (ibid, p. 1). According to Zuboff, these "**markets of behavioral control**" are "composed of those who sell opportunities to influence behavior for profit and those who purchase such opportunities" (ibid, p. 85).

Freedom, autonomy, democracy

In a newspaper essay she concludes that surveillance capitalism could "demean human dignity" (Zuboff 2016). She warns that society would enter "virgin territory" and would face a drastic challenge that "threatens the **existential and political canon of the modern liberal order** defined by principles of self-determination", for example "the sanctity of the individual and the ideals of social equality; the development of identity, autonomy, and moral reasoning; the integrity of contract, the freedom that accrues to the making and fulfilling of promises; norms and rules of collective agreement; the functions of market democracy; the political integrity of societies; and the future of democratic sovereignty" (ibid).

6.8. Final reflection: From voluntary to mandatory surveillance?

As described in the context of power imbalances above, opting out from digital tracking becomes increasingly difficult. Individuals can hardly avoid consenting to data collection without opting out of much of modern life. In addition, persons who don't participate in data collection, who don't have social networking accounts or too thin credit reports, could be judged as "suspicious" and "too risky" in advance.

Mandatory consent?

Today's personal data ecosystem raises many concerns about data being collected, analyzed, transmitted and used without informed consent of individuals. But in many

cases, people are required to consent, either because offers and services that are not based on invasive digital tracking are not available, or because non-participation would lead to disadvantages in life. Scott Peppet (2010, p. 1) expects that "those with valuable credentials, clean medical records, and impressive credit scores will want to disclose those traits to receive preferential economic treatment. Others may then find that they must also disclose private information to avoid the negative inferences attached to staying silent". He points to an **unraveling effect** that could make disclosure of personal data "from a consensual to a more coerced decision" (ibid).

From taking care of the self...

It has been pointed out many times that there is a wide variety of reasons for individuals to participate in self tracking, which are embedded in developments of individualization and societal core values such as self-understanding, self-knowledge, self-optimization, self-improvement, self-responsibility, self-control and self-management. Based on the work of the French philosopher Michel Foucault, many scholars in sociology have shown that the way power works in society has shifted from "hard" authority and punishment to "soft" control. Individuals voluntarily become "entrepreneurs of the self", who are taking "care of the self" and "governing the self" (see Ajana 2005, Bröckling 2007, Whitson 2013, Lupton 2014).

...to insurers and employers taking control

But while individuals are taking "care of the self", insurers or employers take control of the collected data and create rules to incentivize desired behavior or penalize non-desired behavior. Subsequently, "private self-tracking" becomes "pushed or imposed self-tracking". It becomes "harnessed" to broader "commercial, economic or social imperatives" (Lupton 2014). Thus the circle is complete. Voluntary self-tracking cultures and classical, bureaucratic population management based on invisible surveillance are complementing each other. A publication by *Ernst & Young* asked whether tracking-based insurance could already be the "new normal" and suggests insurers to introduce "Pay-As-You-Live (PAYL)" programs.[653]

[653] Walter Poetscher (2015). Usage Based Insurance. The New Normal? EY, July 2015. Online: http://www.ey.com/Publication/vwLUAssets/EY-usage-based-insurance-the-new-normal/$File/EY-usage-based-insurance-the-new-normal.pdf [24.07.2016]

7. Ethical Reflections on Personal Data Markets (by Sarah Spiekermann)

As we have seen in the above chapters, personal data markets have reached a massive scale. Can we accept these in their current form as we move on with a digital economy based on information technology?

Ethical theories applicable to personal data markets

Ethical reflections can help to find new ways for the design of personal data markets and identify rules for the actors operating within them. We therefore include them in this second to last section of our book. I point of course to the limitation that the normative theories applied hereafter are those dominating in the Western world. I take three of them here and apply them to personal data markets: The Utilitarian calculus, which is the original philosophy underlying modern economics (Mill 1863/1987). The Kantian duty perspective, which has been a cornerstone for what we historically call "The Enlightenment" (Kant 1784/2009), and finally Virtue Ethics, an approach to life that originates in Aristotle's thinking about human flourishing and has seen considerable revival over the past 30 years (MacIntyre 1984).

7.1 A short Utilitarian reflection on personal data markets

The Utilitarian calculus, which originates in works by John Stuart Mill and Jeremy Bentham, tries to weigh the beneficial and harmful consequences of actions. As a result of such a weighing process one can come to a conclusion on what is best to do (Mill 1863/1987). What is important to know is that originally Utilitarianism focused on the maximization of people's happiness. While economic theory tended for the past 160 years to only emphasize the monetary consequences of actions (utility in money terms). John Stuart Mill actually wrote: "…the creed which accepts as the foundation of morals, Utility, or the Greatest Happiness Principle, holds that actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness ... and that standard is not the agent's own greatest happiness, but the greatest amount of happiness altogether" (1863/1987, pp. 278, 282). Hence, when reflecting hereafter on personal data markets from a utilitarian perspective, we must consider many more factors than just the financial benefits or harms of these markets. I do so by considering for instance the market's effects on human knowledge, power and relationships presuming that these values are important for happiness (based on (Maslow 1970)). For complexity reasons I only use General Utilitarianism as my discussion framework (as opposed to Act-, or Rule Utilitarianism).

Financial Benefits

From a Utilitarian perspective, monetary value is considered a benefit. Investors and organizations collecting personal data can monetize it and certainly have a 'plus' on their Utilitarian balance sheet. Profits are especially justified, when companies redistribute some of their profits to pay for the common good through their taxes and create employment. Yet, profits need to be made on legitimate and fair grounds. As game theory has shown for decades, purely one-sided financial benefits of players are perceived as unfair by those market participants that do not share in the profits (Tisserand 2014). And so, if companies and data brokers continue leveraging financial benefits of personal data trading without data subjects' active share in the profits, they might see a destabilization of their business in the mid- to long term. Behavioral economics clearly suggests that people or "data subjects" would need to be financially compensated for their data somehow.

If we now assumed that at some point personal data markets found forms of profit sharing, then on economic grounds personal data markets would appear advantageous or catering to some form of "happiness". Yet, as outlined above, Utilitarianism embeds a

more holistic decision spectrum than just financial benefits. So I take another important value created in personal data markets, which is likely to be impacted by personal data markets: knowledge extractable from Big Data.

Knowledge and Power

So far, the knowledge created about people's behavior is asymmetrical. In our current personal data market design, knowledge has become a powerful value from which only a few service monopolies benefit; for instance data brokers like Acxiom, BlueKai, and large data collectors, such as Google, Apple or Facebook. We have talked about "power imbalances" in the chapters above. Hence, the social utility potentially created through Big Data knowledge among a few players is counterbalanced by the drawback of power asymmetries.

Power asymmetries breed distrust and undermine cooperation. In personal data markets, distrust can be observed in at least two forms: between economies, and between corporates and people. Firstly, most European economies don't have powerful data brokers. Their legal framework (i.e. the EU Directive on Data Protection 95/46/EC) have not allowed for the rise of this phenomenon at the same scale as was the case in the US. As a result, European companies and institutions don't benefit from the potential knowledge aggregation inherent in personal data markets to the extent that American players do. The result is a rising political tension between the US and the EU on this matter. Secondly, as we have shown above, people don't know how much corporates know about them. The people who provide their data don't learn anything about themselves from the knowledge that others hold about them. In contrast, they are exposed to potential manipulation and economic disadvantages (Christl 2014). So, taken together, the utility created through personal data markets' knowledge potential is neutralized (or even negatively outweighed) by power asymmetries and their consequences.

Political and technical design could change this calculus though! If it were possible to build personal data markets as symmetrical knowledge structures, in which people get full insight into what companies know about them, societies might become more knowledgeable and thoughtful. What would happen if Facebook were willing to give me feedback on the entire data pool they hold about me, telling me not only about the raw data they have, but also what my data tells them and others about me? Who I am psychologically, emotionally as well as socially according to their current analytical models? The learning and sensitivity I might gain from this insight could be beneficial for me as an individual. I might grow and become more humble upon such feedback. However, I might also be so shocked about the conclusions Facebook's algorithms make about me that I would like to ensure nobody knows about all of this except me. I might demand choice and control over my data as a result. As a European I might also prefer to have my data stored and processed in Europe. Moreover, I could also feel that self-tracking is not good for my development as a person and I might therefore prefer to not participate in it at all. If all of this were granted, including exit- and control rights, then knowledge asymmetries between users and corporates could largely disappear.

Taken together: The political and technical design of personal data markets has the potential to assure two-sided knowledge and symmetry of power. If market players and policy makers went down this balanced 'knowledge-creation' path, then a positive argument would be created on the Utilitarian balance sheet.

Belongingness and Quality of Human Relationships

Knowledge, power and money are not all we care about. Other crucial values for consideration in a Utilitarian calculus are the importance of honest and free communication between humans and our need for belongingness. Some parts of this belongingness can be nourished through our exchanges with others online. How do current data markets play into this dimension?

The digital realm has a huge potential for honest communication. Scientists talk about a 'disinhibition effect' online (Suler 2004). People tend to be more willing to say what they

think online and overcome their possible shyness. Except for excesses of disinhibition (i.e. trolling behavior), peoples' opening-up behavior can be considered as a positive side of the Web. It can add to people's inner peace, freedom and chances to make friends. In virtual worlds for instance it has been recognized that sometimes friendships develop, which are more honest and straightforward from the start (Yee 2014). However, data markets are currently designed such that they systematically monetize our personal exchanges and sell and analyze our relational data. Being in a virtual world I can never be sure that my behavior and my discussions there with others will not be analyzed, monitored, sold or added to a psychological profile. As a result, the darker or idiosyncratic part of my identity cannot be expressed or strive online. I hold myself back. The Facebook studies we conducted at WU Vienna have shown that over 90% of the users on the social network "think twice" before they post something about themselves (Futurezone 2012). We have discussed in the above chapter to what extent people self-censor and might engage in "anticipatory conformity" (Zuboff 2015).

Holding oneself back in the way it is done today may just be the start. If personal data markets advance and people become more aware of being watched or their communication being potentially used against them, it might be that strategic communication could become the norm online. Even more so, if personal data markets allowed people to make money on their data and their online conversations, communication could become strongly calculus-driven. Already today, people engage in 'impression management' online. Trying to trick machines into paying higher prices for keywords used in artificial communication online seems far-fetched these days, but cannot be excluded as a potential scenario. If this happened, then the human relationships involved in this online communication could seriously suffer as a result.

Such negative effects could be mitigated through good technical design. If we ensured truly private rooms in the digital realm where our data was neither monitored nor sold, but instead encrypted, anonymized and deleted (when no longer useful to the individual), then we could have more honest and deliberate communication online; potentially building very truthful relationships on digital platforms. The digital realm could contribute to our freedom and autonomy where needed.

Taken together, a few short Utilitarian reflections on personal markets show that their ethicality depends crucially on their technical and organizational design. Unfortunately, their currently observable design as reported in this study with their one-sided financial gains, knowledge asymmetries and lack of privacy undermine their ethicality from a Utilitarian perspective.

Utilitarian philosophy is only one angle to think about the ethicality of an act or a phenomenon. As outlined above other philosophical perspectives can complement Utilitarian reasoning. Therefore, the next section is going to look at personal data markets from a deontological perspective.

7.2 A short deontological reflection on personal data markets

The word "deontology" roots in "deon", a Greek word that stands for duty. Deontology is a philosophy of obligation, which flourished in 18th century Europe. One of the main fore thinkers of deontology was Immanuel Kant. Kant (1724–1804), a German philosopher, is regarded as one of the most influential thinkers of "The Enlightenment in Europe". He wanted to create a universal justification for moral actions. In order for moral justifications to be rational, he argued that the consequences of an act might be too much subject to the volatile ideas of human happiness and could therefore not serve as a reliable moral guideline. So he effectively questioned the Utilitarian kind of reasoning I have used above.

Kant's Categorical Imperative

A moral obligation, which he called a "categorical imperative," can be justified only by something that is a universal principle in itself. So Kant formulated a *Categorical Imperative* that more specific actions should conform to. The first part of this imperative reads as follows: "Act only in accordance with that maxim through which you can at the same time will that it become a universal law" (Kant 1785/1999, p. 73, 4:421). Note the use of the word "maxim" here. For Kant, maxims are not just values that one can try to live up to. Instead, maxims are a kind of subjective law or 'principle of action' that can be universalized and upon which one has the *duty* to act. Take the example having the maxim to never lie to anyone. *Wanting* to tell the truth would not be enough for Kant. In Kant's sense, I have the duty to never lie or to always tell the truth ("Act *only* according to that maxim"). Why is Kant so strict? Because of the ethical confidence we can then have in our surroundings. If the above maxim would be a universal law then we could fully trust that everyone tells the truth.

Kant also argued that there should be a universal principle that guides my considerations on what are worthwhile maxims: this is that in our maxims human beings should always be treated as ends in themselves and never only used as a means to something. For this reason, he completed his Categorical Imperative with a second part that stressed human dignity: "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means" (Kant 1785/1999, p. 80, 4:429).

In Kant's philosophy, maxims are always *subjective* principles that are supposed to be held by a person; notably by the person who is supposed to take an ethical decision. When a person needs to make an ethical decision, Kant reasons that they should behave as if they would not only decide for themselves, but as if they were a "universal lawmaker". Note the difference between Utilitarianism and Deontology in this regard: Utilitarianism allows for reasoning at an abstract level, weighing pros and cons of something without taking any actual *subjective* stance. For instance, I can argue that the pros of personal data markets are more knowledge, but the cons are more power asymmetries. The argument is valid, but the individual decision maker or lawmaker or analyst of personal data markets that is formulating this utilitarian argument is not personally touched or involved in this observation. He or she does therefore not perceive any duty to act towards anyone upon the analysis. Utilitarianism is hence analysing personal data markets 'neutrally'. Thomas Nagel would have critically referred to this neutrality as practicing "a view from nowhere" (Nagel 1992).

Scott Howe, CEO of Acxiom

Kant in contrast requires ethical decisions to be taken from a personal position that involves subjective principles. If we want to use the Categorical Imperative to ethically judge personal data markets, then we therefore need to put ourselves in the shoes of a concrete person who is personally involved with the object of analysis. When it comes to personal data markets this could be any fictitious or real person who is involved in them. Ideally we choose a person who is coming close to being a true "universal lawmaker" in these markets. This could be a person who is running a personal data market, such as Scott Howe, current president and CEO of Acxiom, Lawrence J. Ellison, head of Oracle and Bluekai or someone else in a comparatively powerful position. To make the following reasoning didactically entertaining I allow myself to engage in the deontological analysis of personal data markets by taking Scott Howe of Acxiom as an exemplary 'universal lawmaker' who could be asked to decide on the ethicality of practices that his proprietary personal data company engages in.

The maxim of Scott Howe could be to have Acxiom engage in or abstain from certain means of data collection, aggregation, analysis, sharing and monetization. From a deontological perspective the question is what universal law Scott wants for himself and others. He needs to personally want that the data-activities in which Acxiom engages in should always take place in the way *he* designs them for his company (or signs them off).

For this purpose it is crucial to discuss personal data markets' practices one by one. Potential duties arising with data collection are very different from duties relevant in data aggregation or analysis. Deontological analysis forces us to look at all market activities separately. For reasons of length of this report I will do so in the following only for the activity of data collection.

Scott Howe could theoretically hold the maxim "Always collect the data of the people such that the people do *not* know what has happened (because they are not asked for their explicit and informed consent)." But deontological ethics questions whether such a position is realistic. Deontological analysis seeks to understand whether Scott Howe can really want such invisible collection to become a universal law. Instead of taking a "view from nowhere" the philosophical theory demands to reflect on what the CEO would want for himself as well (the philosopher puts himself into the shoes of the universal lawmaker). So, would Scott Howe want to have his own personal data collected without his knowing? How much time would he want to invest into reading and understanding terms and conditions? What intimate data would he want to have collected at all about his personal life and the lives of his friends? I could presume that rationally and in his own self-interest Scott Howe can actually not want that the data Acxiom processes about himself and others to be collected without his conscious and explicit knowing and free consent; for instance information about his body weight, health status and diet, sleep quality and mental stress; his private phone number and place of residence, favourite leisure spots, personal relationships, secret passions, collections and weaknesses? In contrast, it seems rational that Scott Howe would want his personal data to be collected by parties he engages with only with his fully conscious, informed and explicit consent. Ethics from a deontological perspective demands that Scott's duty resides in now applying this subjective principle to his corporate decisions. If we think this to the end, Scott and his team at Acxiom would now need to think about how to make informed and conscious ethical data collection possible from a technical and organizational perspective. Many potential actions could be taken. For instance, Acxiom could require its data suppliers to prove that the data subjects' informed consent was given for all of Acxiom's data processing purposes. To ensure that informed consent was given, it might support standards for controlled and policy-based data flows, such as sticky policies (Casassa Mont et al. 2003). Acxiom might offer end-users *full* access to their personal records, the analysis that is done on the basis of these records and allow for dynamic consent procedures (Kaye et al. 2014). Acxiom might start working only with partners that are certified for fair data collection practices, etc. I do not want to expand here on the full list of potential actions Acxiom could engage in to enable Scott Howe's maxim. But what we see from the line of arguments is that the ethical reflection leads to organizational and technical design measures available to fulfil the maxim.

Can we presume that Acxiom would be ethically on the safe side if it thus followed Scott Howe's new maxims? Unfortunately, not yet from a deontological view.

People serving as a means to an end

According to deontological thinking, ethical judgments need to consider the 2nd formula of the Categorical Imperative as well. This formula condemns practices where people serve *only* as a means to an end. Are they only a means in the data collection process to reach a certain end? The answer to this question is straight forward: If we use people just as a means to get their signature under a data-sharing contract, then the 2nd formula of the Categorical Imperative is *not* fulfilled. This is what mostly happens today. People's notice and choice (if it is granted) does not aim to easily inform people as we have shown above. In contrast: Often people are just used to give their signature so that companies then have a free ticket to use their data and make money. Current data collection may then be permissible from a legal perspective, but from a duty-ethical perspective it is not appropriate.

So in order to fulfil the 2nd part of the Categorical Imperative what would be Scot Howe's duty? One strategy could be to position the data collection process in a different way than it is being done today. It is possible to view data collection as a means of deliberate participation in the crafting of a better world that thrives on more knowledge (as discussed in the Utilitarian analysis above) and that sees less fraud. Personal data markets can give people the opportunity to participate and share in these goals. Companies like Acxiom could collect data from fully conscious individuals who are willing to share their data for explicit causes. The key to this path is one thing: People would need to *autonomously* buy into these goals. Autonomy is a key to Kant's understanding of ethical conduct.

Autonomous and free consent to data collection would mean that, first, data subjects would need to learn about all of the purposes pursued with their data and they would then need to consent to these one by one. Most importantly, this fine-grained consent would need to be given freely. Data subjects today are often forced into data sharing, because if they deny sharing, they are denied service delivery. Such enforcement for data sharing contradicts Kant's Categorical Imperative. Enforcement can also be very subtle; i.e. psychological pressure can be put on people by repeating mantras to them, such as "sharing is caring", etc. Data collectors need to abstain from any of such manipulative tactics. They need to be very frank, and let the people decide as they want to. They need to be ready to forgo many opportunities for collecting data from people who simply don't want to share. And they need to be willing to provide non-sharers with the same service as everyone else (even if this implied less profit for them). Only if data collecting companies act this way will they enter the ethical white-zone; at least from Kant's deontological perspective.

7.3 A short virtue ethical reflection on personal data markets

The virtue ethical approach to decision-making and behavior is a 20th century rediscovery of Aristotelian philosophy (Aristotle 1915; Aristotle 2000; Hursthouse 1999; MacIntyre 1984). Virtue-ethical thinking focuses on the long-term flourishing or wellbeing of people; a wellbeing that might become affected by certain behaviors, technology, or the existence of personal data markets. In Aristotle's view a virtuous life is a necessary condition for flourishing; or achieving what he calls "eudemonia" (often translated as "wellbeing"). Two concepts are particularly constitutive of virtuousness, and enable eudemonia: these are arête and phronesis (Hursthouse 2012).

The effects of actions on people

Arête stands for an excellent character expressed in well-balanced behaviors towards oneself and others (golden-mean behaviors). Aristotle pulls out the nature of arête in his "Nicomachean Ethics" where he describes a number of concrete virtues such as courage, temperance, high-mindedness, veracity, generosity, etc. (Aristotle 2000). These virtues are examples, which illustrate the meaning of arête. A noteworthy aspect of arête is that virtuous behavior is generally *not instrumental* to anything. Instead it is driven out of an inner compass for what is right and good. The world of literature and film is full of examples of arête: for instance the fairy tale character Cinderella, Jane Bennett in Jane Austin's novel Pride and Prejudice or the protagonist Jake Sully in James Cameron's recent film 'Avatar'.

A core virtue leading to people's flourishing (also called "eudemonia") is phronesis. Phronesis stands for practical wisdom. It is the knowledge and ability of a person to take right and just decisions. Phronesis is *not* about rules that can directly be applied (such as legal regulations). Instead phronesis implies the ability to recognize in a situation what it is that does justice to the virtues, people, and goods involved. Phronetic leaders are good in prioritizing the right actions and recognizing a relatively complete spectrum of consequences; including the "soft" or long-term consequences of decisions for virtues,

persons and goods. Phronesis seems vital for instance to make good judgments on the utilitarian weights of the costs and benefits of personal data markets that I outlined above in the Utilitarian analysis.

Unlike Utilitarianism that focuses on consequences and deontology that focuses on universal law makers, virtue ethical analysis focuses on the effects of actions on people. In the personal data market context virtue ethical analysis asks whether the technical, social and economic manifestations of data markets will influence people's lives such that they impede or hinder them to become the kind of person that possesses arête and phronesis. Could personal data markets lead to subtle unconducive conditions of oppression that bar people from cultivating their virtues and develop phronesis and arête? In doing so, could they impede people's flourishing in any way? In the words of Lisa Tessman: Could personal data markets create a condition of "systemic constitutive bad luck"? (Tessman 2005) A bad luck that then undermines the long-term goodness of those affected.

A concrete person

To answer these questions, it is helpful to envision a concrete person (actor) who might live in a future world in which personal data markets strive. Let's take a fictitious person called Bill who is seriously overweight and has therefore started to use a health-tracking device. The device measures his weight, transpiration, heart rate, cholesterol, fat, steps and movements, calories, body posture, etc. Bill has acquired the device as part of his plan to do a lot of sports in order to bring his body back into a healthy condition. Yet, this plan turns out to be extremely hard and the device's data suggests that Bill's plan has failed.

Projecting today's technical architectures into the future, all or most of Bill's data would probably flow uncontrolled to the health app provider who might sell it on and share the data with 3rd parties, including insurance companies, data brokers, employers, etc. The health app may be free, but it is very likely that Bill's data then turns against him. His health insurance rate might go up more steeply than expected. The number of invitations he might get as a result of his applications for sales jobs might be lower than he expected. Bill may not know that his health app data is behind this 'systemic constitutive bad luck.' The virtuousness of his character might not be directly impacted by the fact that invisible forces make life more difficult for him. However, what could happen is that he becomes depressed or angry. The chances to live a good life and to benefit from the flourishing his good character actually deserves are reduced. The data-driven circumstances might lead to a character change in Bill who might turn from a positive character into a frustrated one. That said, it could also be that Bill's positive character is extremely resistant and that his person and behavior does not change much when faced with a data-world driving his life into a negative spiral. Virtue ethical analysis projected into a likely future does not give definite answers. It just helps to envision scenarios with a potential likelihood.

Society at large

Virtue ethics also allows for analysis at both the organizational and societal level. Let's therefore take a step back from Bill as a person and look at society at large: We must ask the question how an economy and a society evolves in which people start feeling discriminated because of their data profiles. Their feelings and perceptions towards anyone they meet (e.g. employers, the state), or any service they use (e.g. health apps) could become increasingly marked by distrust and ambiguity. People might start presuming that the vis-à-vis knows more about them than they do about him; that no matter where they turn, they confront a knowledge asymmetry that puts them into a weaker position than they could be in if there was no data sharing. If this evolution is permitted to happen, we will see a society reigned by distrust and lack of loyalty; or as Hume anticipated it: A society in which everyone is everyone else's wolf. This is indeed a very negative virtue ethical outlook.

A second scenario could be considered that combines personal analysis with societal implications. Let's presume people would receive property rights in personal data and could financially benefit from data markets. This is what the Utilitarian analysis above recommends. In such a scenario, Bill would be very well aware that his health data is

shared and with whom and under what conditions. Let's say that Bill is not too rich. Therefore he has made a deal with the health app provider and licensed out the usage of his health data for the next five years to come. He also struck a deal with his health insurance company that receives the data, tracks his progress and allows him in return a 10% discount on this rate over the next 5 years. At first sight, this looks much better than the kind of intransparent data-world we are in right now. Bill actually might have taken a prudent decision by selling his data, because this deal motivates him to a certain degree to really change his fitness behavior. Through experience he might be conscious of the fact that he will not endure a fitness plan if he does not put himself under some financial pressure to succeed. I also assume that Bill knows everything that his insurance company knows about him. Loyalty and trust is created due to such knowledge symmetry. From a virtue ethical perspective (which looks at his *personality* development) at first sight there seems to be no risk to Bill.

However, there is a serious virtue ethical risk in this scenario: What happens if Bill loses weight and becomes quite sportive within a year. He has reached his health goals. He has formed a good health-habitus. But still he is forced to continue sharing his data. He is not allowed to stop using the health device, as he would naturally want to. Instead, he is forced to continue using it and bravely upload his data, because otherwise he would experience considerable financial loss. Naturally, he will become aware of the fact that he sells himself, or parts of himself. He realizes that he has made a deal of himself. He might become aware that there is a long-term monitoring of his person happening at real-time. If such awareness arises, it might not be healthy for Bill's perception of the self. He might start seeing himself with different eyes. He might see himself as directed by others, as being forced to serve interests, which are not his own. And he might start disliking the service provider, the insurance provider and all those parties who deprive him of his liberty to use or not use technical services as he likes to. If data deals are so designed , that people cannot easily opt out of them anymore, liberty is most likely to suffer. And people who have sold their digital identity might lose faith in their own virtuousness. Their behavior could start to become instrumental instead of natural. Virtue would suffer. At least this is a possibility.

The kind of envisioning of the future that is necessary for virtue ethical analysis is of course speculative. The true effects might come out completely differently. Still, they are likely and in technology design and policy making it is finally only this exercise of informed anticipation and envisioning that allows us to make virtuous decisions (Nonaka et al. 2008).

7.4 Conclusion on ethical reflections

Taken together, I would argue that the short ethical analysis of personal data markets in this chapter suggests that these markets have negative ethical implications in the way they currently operate. They must therefore be regarded with great caution. Property rights might alleviate the effects of personal data markets to some extent. But they are dangerous from a virtue ethical perspective, because they might lead us into an extensive commercialization of the self. The technical and legal design of personal data markets must also be carefully crafted to ensure the long-term liberty of people. In particular, it must be ensured that people can exit data deals at any time and that such exits will not lead to negative consequences for them. Personal data should never become an asset that people are forced to sell. They must freely and autonomously consent to data collection and have the ability to recall this consent at any time. They should have full transparency of their data and most importantly the knowledge that is created from it. Only if these requirements are met by the relevant industry players and all the small market participants (such as app providers) can personal data markets receive some ethical legitimacy.

8. Recommended Action

"You have to fight for your privacy or you will lose it"
Eric Schmidt, Google, 2013[654]

Serious concerns

A society based on ubiquitous digital tracking that is happening in an intransparent manner and systematically discriminating people for economic advantage raises serious concerns about the future of freedom, democracy, autonomy and human dignity. We have argued above that there is a massive power imbalance between individuals and networks of companies processing vast amounts of information about the lives of billions of people. To date, individuals have limited ways to protect themselves from corporate surveillance; even if they take the unrealistic step to remain largely offline.

So where do we go from here?

Networks of control

Very often reports such as this one or academic work come to the conclusion that regulators need to do something. As the film "Democracy"[655] shows very well, the regulator is not in an easy position. It is being said that over 2,000 lobbyists were hired exclusively to turn the new European framework for personal data regulation, the so-called "GDPR", into a weak piece of legislation. The corporate power in favor of corporate surveillance is a "network of control" in itself. Over 4,000 amendments were made to the GDPR's version as proposed by Jan Albrecht, its rapporteur in the European Parliament. Many of the good suggestions that got ratified by the European Parliament were then twisted and weakened in the Council. Member states are known to have been sending representatives to Council negotiations who were known to have highly personal relationships with lobbyists.

The lobby network operating in the IT industry is so vast and so powerful and so unscrupulous that we believe the only way to move into a better future is to **thoroughly publish any personal encounter between policy makers and industry representative** as well as the lawyers that work for them. Tools such as *LobbyPlag*[656] and organizations such as *Transparency International* should be heavily supported by donors, crowdfunding initiatives and governments themselves. **Governments should support co-operations between NGOs and universities** so that NGOs can leverage the existing educational infrastructure and universities' educational programmes can benefit from the positive energy and knowledge that activists often hold. In fact, governments have an interest in supporting such co-operations, because the power of states is equally diminished by the data/information imbalances created by personal data markets (i.e. in the field of identity management).

A selection of suggestions

Below, we provide a selection of recommendations based on our findings, ranging from regulation to crucial challenges regarding transparency, education and knowledge across society. In addition, we present a technical and legal model for a privacy-friendly digital economy, which has recently been introduced by one of the authors of this report.

---

654 Colvole, R. (2013): Eric Schmidt interview: 'You have to fight for your privacy or you will lose it'. The Telegraph, May 25, 2013. Online: http://www.telegraph.co.uk/technology/eric-schmidt/10076175/Eric-Schmidt-interview-You-have-to-fight-for-your-privacy-or-you-will-lose-it.html [30.08.2016]
655 http://www.imdb.com/title/tt5053042/
656 http://lobbyplag.eu [30.08.2016]

8.1 Short- and medium term aspects of regulation

In the **United States** strict regulation exists in some select sectors such as health, finance and education, but the corporate processing of personal data is widely permitted in many other contexts (Solove and Schwartz 2015). Most consumer data is not covered by any specific privacy law (CDT 2010, p. 5). The U.S. continues to follow a "harms based" approach to privacy. And as virtual harms are not immediately felt by individuals and are hardly ever provable, U.S. law has been widely ineffective to protect its 'virtual citizens' who are processed in its databases. In the **European Union**, in contrast, data protection is regarded as a fundamental right and legally, companies can process personal data only under strict conditions.[657] The most important pieces of legislation to reflect on are the recently published **EU General Data Protection Regulation (GDPR)** and the **e-Privacy Directive.** Can we place trust into these pieces of regulation?

European outlooks

After years of negotiation and discussion the **GDPR** will finally come into effect in 2018 and replace the *Directive* from 1995. The GDPR aims to harmonize privacy legislation in all EU member states. Also companies from outside the EU will have to comply with the GDPR, when they offer services, process data or "monitor behaviour" of European citizens. Infringements can be subject to fines up to 20 million Euro or 4% of the "total worldwide annual turnover".[658] The GDPR is certainly not perfect. Both privacy advocates and industry have concerns, but accept it as a compromise.[659] The digital rights organization *EDRi* called the GDPR "lacking ambition but saving the basics".[660] A representative of the *Interactive Advertising Bureau (IAB)* called it an "imperfect piece of legislation", but there would be "no use in crying over spilled milk".[661] Still, both sides see "foggy" notions[662], unpredictability[663], "undefined" terms, "ambiguities", "mechanisms that are not clear", "sweeping provisions that could be interpreted in widely different ways".[664]

GDPR criticism

From a privacy perspective, the GDPR has some major drawbacks, including, but not limited to:

- the "right not to be subject" to automated decisions and to profiling is limited to cases, which "significantly" affect individuals.[665] But who defines this significance?
- "explicit" consent to data collection is only required for the processing of *sensitive* personal data, while "consent" is enough for all other kinds of data.[666] Since the list of sensitive data is limited this could weaken the consent requirements for normal ubiquitous data collection we report on above.

---

[657] http://ec.europa.eu/justice/data-protection [30.08.2016]
[658] GDPR final text
[659] See e.g. Hughes, T. (2016): General Data Protection Regulation: A Milestone Of The Digital Age. TechCrunch, Jan. 10, 2016. Online: https://techcrunch.com/2016/01/10/the-biggest-privacy-law-in-the-world-has-arrived/ [30.08.2016]
[660] https://edri.org/eu-data-protection-package-lacking-ambition-but-saving-the-basics [30.08.2016]
[661] http://www.iabeurope.eu/all-news/press-releases/statement-on-the-adoption-of-the-general-data-protection-regulation/ [30.08.2016]
[662] https://edri.org/eu-data-protection-package-lacking-ambition-but-saving-the-basics/ [30.08.2016]
[663] http://policyreview.info/articles/analysis/beyond-consent-improving-data-protection-through-consumer-protection-law [30.08.2016]
[664] Dockery, S. (2016): Uncertainty Abounds in Europe's Data privacy overhaul. The Wall Street Journal, April 25, 2016. Online: http://blogs.wsj.com/riskandcompliance/2016/04/25/uncertainty-abounds-in-europes-data-privacy-overhaul/ [30.08.2016]
[665] https://www.eaid-berlin.de/?p=930 [30.08.2016]
[666] Ibid.

- The GDPR allows companies to process personal data of individuals without consent, when it is "necessary for the purposes of the legitimate interests pursued by the controller or by a third party". The "legitimate interests" listed include fraud prevention, network security and direct marketing.[667] This could mean that much data collection and processing done today may actually continue as is.

Complexity

Finally, the GDPR has lost the simple power the 95/46/EC Directive had (its predecessor). It is very hard to handle by normal legal offices. Figure 5 shows the complexity of articles within the GDPR that is created only through the practice of constant cross-referencing between articles within it. We fear that only specialized legal practices will be successfully able to handle this piece of legislation. These legal offices are again in the dilemma that they live on paying customers, which typically pay them for legal trials or defences of their dubious practices. We fear that this corporate reality will lead to a systematic interpretation of the GDPR that is not in line with the original ideas of privacy idea that were embedded in it.

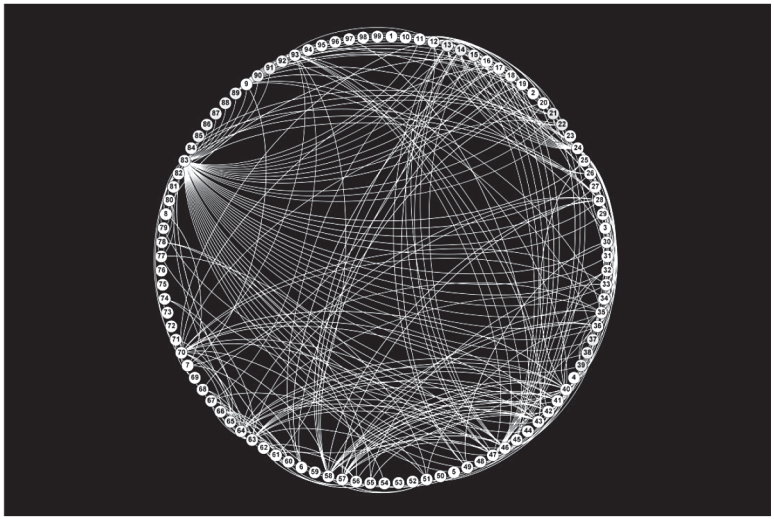

Figure 5: Crossreferencing of articles within GDPR (created by Sushant Agarwal, Institute for Management Information Systems, 2016)

We therefore recommend the development of one or several **digital tools that support the easy analysis and comprehension of the GDPR** (as well as other privacy laws). Such tool(s) may visually enhance the legal text and turn it into a click-and-learn project. The tool might link for instance to simple documents, which provide a privacy-friendly version of interpretations that can be derived from the original text. Such tools should be freely available to the public and help people to learn about their rights. Governments and NGOs could set up public help-desks that support people in using these tools and learn about their rights. Universities could develop, maintain and refresh such tools that are also usable also in education.

At least equally important for privacy is the upcoming European **ePrivacy Directive**, which "particularises and complements" the GDPR by "setting-up specific rules concerning

[667] GDPR final text (European Commission 2016)

the processing of personal data in the electronic communication sector".[668] Not unexpectedly, privacy advocates have stated drastic expressions of concern. For example, EDRi, which is representing more than 30 civil and human rights organizations all over Europe[669], stated that the "huge lobby against the GDPR" would now be "hard at work, to completely destroy the ePrivacy Directive".[670]

Crucial issues for privacy regulation

No matter the concrete arrangements of today's European legal landscape, we generally recommend to focus on the following legal issues and suggestions, which NGOs, academics and privacy-friendly experts have identified as important for a long time:

- We need better **enforcement** of existing (and upcoming) law. This includes not only high sanctions for data breaches, but also an infrastructure that allows citizens access to legal trials in this field.
- Profiling and targeting individuals based on **pseudonymous identifiers** (e.g. device IDs, cookie IDs, "hashed" identifiers derived from email addresses or phone numbers, and other "internal" pseudonymous identifiers) is not "anonymous" as often expressed, and should therefore be considered as processing of personal data (see chapter 5.6).
- When an identifiable individual is **attributed, labeled, classified, scored, rated or judged** in any way, this should be considered as processing of personal data, also when methods of analytics are not merely or not at all based on data about the same identifiable individual. As soon as such an attribute is attached to an individual and "used to single out a person, regardless of whether a name can be tied to the data" (Borgesius 2016), it should be considered as personal data. For example, when a sensitive attribute such as health status or sexual orientation is not collected from a person, but predicted based on data from others, and then attached to this person, it is personal data.
- The **re-identification** of individuals from anonymized data sets should be forbidden and integrated into criminal law.

Legitimate interests of companies

- We believe that individuals should have the right to know which data about them is being collected and which purposes it is being used for. Subsequently, letting companies process data about individuals for purposes such as direct marketing **without informed consent** and just based on their "legitimate interests" is problematic. In the case of "legitimate interests" such as fraud prevention and network security it is crucial to guarantee 1) some form of accountability for inaccurate decisions 2) ways to object inaccurate decisions 3) data collected for fraud prevention and network security under the premise of a "legitimate interest" must not be used in other contexts, in particular also not in contexts such as marketing, customer relationship management and online targeting.
- We believe that the **traditional European approach** to regulate the collection and processing of personal data based on informed and explicit consent for specific purposes as a general rule is important. We recognize that it is highly important to have more digital tools developed (personal privacy agents) which support the process of constant choice and consenting. These are outlined in more detail and with a focus on what should be done in the last chapter below.

Consumer protection, anti-discrimination and anti-trust

- Regulatory instruments such as **anti-discrimination law** should be harnessed to challenge unfair discrimination and exclusion based on the processing of personal data.

---

668 https://ec.europa.eu/digital-single-market/en/news/public-consultation-evaluation-and-review-eprivacy-directive [30.08.2016]
669 https://edri.org/about/ [30.08.2016]
670 https://edri.org/epd-revision-analysis [30.08.2016]

- To challenge information asymmetries and power imbalances between individuals and networks of corporate surveillance, **consumer protection law** should "help shift power back to consumers by improving participation and accountability" (Rhoen 2016). This could be particularly relevant in situations, where consumers hardly have the choice to avoid common services of today's digital world such as dominant social networks or smartphone operating systems. For example:
  - Every digital service providers should be obliged by law to offer a privacy-friendly service version. .
  - Tables with standardized fines for data breaches should be developed (as proposed i.e. by Traung (2012)).
- Consumers should have the right to **withdrawal from data contracts** without experiencing any disadvantages.
- Given the dominance of very few corporations in specific fields of application, it is important to also consider regulatory realms such as **competition law**. Shoshana Zuboff, who points to "markets of behavioral prediction and modification" suggests to regulate "excessive and exclusive concentrations" of "behavioral surplus" collected by companies "as free raw material" (Zuboff 2016).

Algorithmic transparency

- It is not enough to give individuals the right to access to personal data, which companies process about them. More and more companies are calculating **opaque and arbitrary scores** about many aspects of personal life. As Citron and Pasquale (2014) stated, regulators "should be able to test scoring systems to ensure their fairness and accuracy" and individuals "should be granted meaningful opportunities to challenge adverse decisions based on scores miscategorizing them". We share this view and believe that the evaluation and implementation of algorithmic transparency and accountability should have a high priority, not only in the field of credit scoring. The GDPR will create a "right to explanation" of algorithmic decisions, which "significantly" affect individuals.

A public registry

- A mandatory **public registry** of applications processing personal data should be set up (where it does not already exist) to enable both consumers and authorities in data and consumer protection to inspect, which data is collected and for which purposes it is being used. In Austria, a data processing registry has existed for decades.[671] Similar approaches have also been proposed in other countries. For example, the US-based World Privacy Forum suggested a "mandatory public registry of consumer scores" (Dixon and Gellmann 2014). One could argue that such a registry sounds like a bureaucratic barrier for innovation. We believe that such a registry could be set up in an easily accessible form and allow the registering of applications via online interfaces – much more simple than setting up any data collection process.

Standardized icons

- One of the most difficult challenges with data transparency is how to enable innovative data-driven applications while *accurately* informing consumers 1) without making them continuously untick privacy notification boxes and 2) without overburdening them with hundreds of pages of terms at the same time. The GDPR mentions "standardized icons that companies "may" provide to give consumers a "meaningful overview" of data processing in an "easily visible, intelligible and clearly legible manner".[672] This promising approach should be pursued further. It is reasonable to **enforce the use of such standardized icons** and respective standardization efforts are probably under way.

---

[671] https://web.archive.org/web/20160829035743/http://www.dsb.gv.at/site/6262/default.aspx [30.08.2016]

[672] GDPR final text (European Commission 2016)

Standardization efforts

- That said, we believe that all standards that come out of standardization efforts foreseen in the GDPR (in the course of delegated acts initiated by the EU Commission) should be signed off by the ***Article 29 Working Party*** or the privacy board that succeeds it. The *Article 29 Working Party* (Art. 29 WP) is made up of a representative from the data protection authority of each EU Member State, the *European Data Protection Supervisor* and the *European Commission*. It should have the right to block proposals coming from standardization bodies that do not meet privacy standards. The EU Commission should have a limited and transparent role in such negotiations.
- It should be ensured that **standardization efforts** expected from the delegated acts in the GDPR are actively accompanied by NGO representatives, chaired by neutral bodies and that public representatives balance the views of companies in these bodies. The people chairing standardization efforts and controlling the drafts' wording should not be paid by industry, but should be recognized as a neutral party; truly representing various stakeholder perspectives. Participation in standardization efforts should be free of charge, open for participation, transparent and run – at least in part – virtually, so that a wider public can participate.

Some of these recommendations are not new; certainly not to insiders, academics, policy-makers and lobbyists in the field of privacy. However, as of today they were not put into practice. We therefore believe that it is essential to create much more transparency around personal data markets. This report is an effort in this transparency endeavor.

8.2 Enforcing transparency from outside the "black boxes"

We feel that we, as a society, cannot wait until today's networks of ubiquitous digital tracking will, whether voluntarily or forced by law, disclose comprehensive information about their mostly nontransparent practices. As this study and other reports show, there is information available. But most of it is incomplete, fragmented or out of date.

Further research needed

To challenge the existing information asymmetries and the lack of transparency we strongly recommend conducting and supporting further research from "outside the black boxes" of corporate surveillance, including, but not limited to the following topics and approaches:

- How do ad tech companies, data brokers, data management platforms and many other businesses actually collect, analyze, link and use personal data? How do the actual personal data flows across companies look like?
- To what extent do data-driven practices and technologies in marketing, scoring, fraud detection and risk management merge? Where is data on individuals used in completely other contexts or for other purposes than it was collected for?
- Which kinds of algorithmic decisions are being made on individuals, based on which data and which kinds of analyses – in marketing as well as in finance, insurance, employment, education, law enforcement? How do companies try to change behavior based on digital tracking?
- How could this affect people's lives? Which specific practices could have problematic implications for individuals - such as discrimination, social exclusion or amplification of existing inequalities? Could many small disadvantages in everyday life cumulate to a major disadvantage? How could this affect equality, freedom, autonomy, human dignity - on both an individual and societal levels? What impacts exist for the individual psyche and development as a personality?

- Technical approaches for the investigation of algorithmic systems from the outside, so-called "black box testing", can help to map data flows and discover, how which data and behavior is influencing which kinds of decisions such as personalized ads and prices.[673] [674] [675]
- Online databases about smartphone apps, including automated analyzing of used permissions and embedded third-party services could help to make mobile tracking more transparent (see chapter 4.1).
- Documentation and monitoring projects could continuously map data flows as well as investigate and summarize developments in the field of digital tracking and corporate surveillance – in different sectors, countries and regions.

Initiatives by privacy advocates and NGOs have also effectively enhanced transparency and helped to make providers accountable. For example, the *Norwegian Consumer Council* carried out a campaign in 2016, which was based on a research report about smartphone apps, its data sharing practices and privacy policies. They found that a major fitness app transmitted data to third parties even when the app or phone was not in use. As a result of research, global media outreach and legal action several apps changed its practices and privacy policies.[676]

What should be public knowledge?

However, we acknowledge that within the current corporate environment it is almost impossible to find comprehensive answers to the issues listed above and to do research on it. Companies regard this knowledge as their corporate secrets or internal knowledge and if they co-operate with selected academics at all, then only under strict non-disclosure agreements. At a higher level we therefore recommend that governments **face the debate on a new balance** on what is 'corporate knowledge' and what should be 'public knowledge'**.** Should governmental agencies such as ***independent*** (!) **data protection authorities** not have access to data processing facilities at companies like *Facebook*, *Google* and *Apple*? Should companies not be obliged to document their data flows according to a common standard and publish those parts of them, which are in a common interest? Our recommendation is that such questions must be asked and debated and not avoided.

8.3 Knowledge, awareness and education on a broad scale

A democratic information society

We recommend making comprehensive knowledge about data-driven practices and its societal, ethical and personal implications accessible and understandable much better; for the general public, but also for experts such as policymakers, civil society, journalists and corporate stakeholders. A lack of understanding of these technologies and its implications limits not only the ability of individuals to make informed decisions in today's digital world, but also makes a democratic debate about our future information society impossible.

Teaching of **digital literacy in schools** should not be limited to practical skills, but also focus on critical thinking about opportunities and risks of digital technology and encourage pupils to reflect their own usage. For all levels of education this includes **enhancing knowledge about tools** to protect one's privacy such as browser extensions to prevent tracking, knowledge about how app permissions on smartphones allow control

673 https://www.cs.cmu.edu/~mtschant/ife [30.08.2015]
674 https://roxanageambasu.github.io/01-research/ [30.08.2016]
675 http://www.nsf.gov/awardsearch/showAward?AWD_ID=1514509 [30.08.2016]
676 http://www.forbrukerradet.no/appfail-en [30.08.2016]

and knowledge about alternative apps and services such as privacy-preserving search engines and messenger apps.

University education

Particularly, university education has a serious problem in catching up with the ethical gap that we observe in data markets today and in business in general. Business students in the area of marketing, innovation management and entrepreneurship are often instructed to collect as much data as possible, seen that current business models are often built on data. Economics students are literally brainwashed for years to work with models that have profit maximization as the primary or even sole goal function. Unfortunately though the digital economy's profits are largely driven by data assets today (Varian et al. 1999). Corporate social responsibility is often treated as an elective subject in the universities or it is not offered at all. The result is that "Bad Management Theories Are Destroying Good Management Practice" (Ghoshal 2005). This comes on top of a **lack of general education** in subjects as philosophy, ethics or the actual foundations of economic theory (which go far beyond profit maximization).

In companies, graduates and professionals mingle with people who have been trained in software engineering, business informatics or computer science in general. Here students typically learn that 'more data is more knowledge'. Standard works in software engineering such as Ian Sommerville's standard volume on "Software Engineering" teach them in the 10th edition that feasibility analysis should be "cheap and quick" (Sommerville 2011, p. 37). As a privacy academic recently pointed out, we now see a "**Technology Mindset**" among engineering students that is taught to them and that differs very much from the "**Privacy Mindset**" that is slowly embraced by the legal word, the public and policy makers.

Universities, we believe, are not really on top of this educational gap. If they are, they tend to argue that data protection matters can be taught in the legal departments where data protection now starts to be integrated into curricula. Ethical decision-making however is a matter that should be at the core of any university curriculum. Matters of privacy and security must be known to every software engineer and management student today as **data governance** is at the heart of the digital economy, which again drives most of our traditional business by now as well.

The very idea to delegate ethics and data protection into the legal department is – not surprisingly – mirrored by (or carried on) into companies. Here, marketing and IT managers regularly delegate ethical decisions into the legal department; thereby freeing themselves of the responsibility that they should actually co-shoulder. The legal departments are then in the difficult role to legitimize what isn't legitimate; an unfortunate position to be in.

Taken together, **the university system provides hardly any basis for a proper education of young people in such important matters as data protection, privacy, ethical- or value-based design of IT systems.** The university system completely fails to prepare young people for these matters in corporations and undermines a constructive and ethical corporate positioning in the digital economy. Our recommendation is therefore to **develop a global initiative to change this**. Every ministry of education worldwide should work top down and at all levels of the educational system to develop the ethical sensibility that is required to shape this digital economy driving our lives. They should ***honestly*** support and strongly fund initiatives in this direction.

Digital civil society

Besides the academic and corporate worlds we would like to acknowledge that all over the planet, versatile communities of thinkers, writers, activists, hackers, privacy advocates and non-profit organizations have emerged which could be subsumed under the label of a "digital civil society". These communities have existed since the early days of the Internet and continuously provide substantial contributions to the ongoing debate about how to shape our future information society beyond corporate and governmental interests, from

claiming digital rights and developing open source tools to empower users to working on alternative concepts and technologies for public good. Individuals and organizations within digital civil society constantly lack resources, but can offer a wealth of expertise that society should absolutely make use of.[677] Additional support and resources could leverage existing and new efforts, in particular, when collaboration between universities and NGOs in digital rights, consumer rights and civil rights is stimulated.

8.4 A technical and legal model for a privacy-friendly digital economy

In a recent edition of the 'Computer Law and Security Review' one of the authors of this report, Sarah Spiekermann, has laid out a proposal for how more privacy and trust could be created in personal data markets while embracing an economic rational around data, the current legal landscape as well as timely privacy enhancing technologies. The proposal is meant as a bridge between players in the current personal data ecosystem and data protection proponents. It is entitled "A vision for global privacy bridges: Technical and legal measures for international data markets" and it accumulated a long list of technical and organizational enablers of a privacy-friendly digital economy. A vision, supported, refined and critically reviewed by 13 leading experts working for major corporations (including *IBM, HP, Metro Group*), standardization bodies (including the W3C DNT working group), data protection authorities (including *ULD Schleswig Holstein*), data brokers (including a major credit-scoring agency), industry associations (including *IAB*), legal counselling groups and one NGO (EDRi). This report embraces and re-emphasizes the technical and economic recommendations made in that piece of research.

Four spheres

In a nutshell, we distinguish four "spheres" in which we recommend to group the activities happening in personal data markets and in the start-up scene and outlaw some of them (Spiekermann et al. 2015): The first market space is the "**customer relationship space**", which should include customers and companies that are directly and ***visibly*** involved in a service exchange; i.e. through a contract. For example, *Amazon* or a fitness tracker. The second market space is a part of the market, which alongside the EU GDPR could be labelled as the **„controller-processor space**". It includes the distributed computing and service infrastructures that enables today's electronic business relationships. This space includes all companies providing services to those companies that directly enable and enrich the customer relationship. For example, a company like *Deutsche Telekom* can be a cloud service provider, which handles the purchase data for a fitness tracker. The third market space is a new part of future data markets, which we called "**customer-controlled data space**" and which is now slowly emerging. This space embraces services that enable customers to exercise ownership of their personal information. Companies in this space manage and control data on users' behalf in a privacy-friendly way. Such companies or organizations (to be founded) could be the long-envisaged trusted party included in so many academic security- and privacy papers. Finally, the fourth market space, which we have called "**safe harbor for big data**", grants equal access to anonymized ‚people data' to all market entities that need such data. For example, a market research agency could download aggregated people data from this safe big data space to analyze and forecast future consumer trends. Participants in this "safe harbor for big data"

[677] See e.g. Dobusch Leonhard (2014): Digitale Zivilgesellschaft in Deutschland. Stand und Perspektiven 2014. Freie Universität Berlin, Fachbereich Wirtschaftswissenschaft, Diskussionsbeiträge. Online: http://edocs.fu-berlin.de/docs/servlets/MCRFileNodeServlet/FUDOCS_derivate_000000003411/discpaper2014_7.pdf

can provide and process as much data as they want, but the data they handle must be anonymized with the best available techniques. The safe harbor for big data is filled with data originating from users. Yet, each time personal information from a user is transferred to this safe harbor, it must pass an "anonymity frontier". In fact, when personal information leaves the context of an identified customer relationship and is transferred to an entity that is not involved in the customer relationship context, the personal data must cross the anonymity frontier.

The reason for organizing personal data markets in this way is linked to one of the many criticisms voiced in this report: the lack of transparency currently dominating in data markets.

Of course transparency through more order alone is not sufficient. Within each of the four market spheres we need to leverage a number of technical, legal and organizational controls that can enable a proper market functioning. All of these controls have been proposed and researched by various research groups around the world. Hereafter we want to recapitulate the essential recommendations for each of these spheres:

A privacy-friendly customer relationship space

The core of the customer relationship space should be are re-establishment of the one-to-one business relationships we know from the past. Companies that invest in a customer relationship need and want identified customer relationships (Spiekermann et al., 2003). And many customers are willing to provide their personal information in a service context if it is necessary for service delivery or if they receive appropriate returns for their data. Therefore, our vision departs from the traditional data protection call for anonymity vis-a-vis directly used services (Gritzalis, 2004; Bella et al., 2011). Currently, however, an individual online customer often deals with multiple parties collecting personal information simultaneously during an electronic transaction. For instance, an average of 56 tracking companies monitor people's transactions on commercial news portals (Angwin, 2012). Thus, companies that are the legitimate owners of a customer relationship often serve as gateways to the shadow market of companies we have analyzed in this report. We believe that only the one company that is visible to a customer is legitimately allowed to collect personal information in the context of an exchange. However, we also think that users have a right to know that there is a 'data-deal' in addition to the service deal; at least in today's business models. This deal must be made transparent to users.

The companies visible to the consumer

The **'one-visible-partner rule'** and **transparent data-deals** only work if they are automatically monitored by **technical accountability-management platforms** that are regularly audited and safeguarded by legal sanctions. A company engaged in a primary customer relationship should become liable for the proper handling of the personal information it collects; a request quite well addressed now by the EU GDPR. All personal information companies receive from their direct customers should be recognized as being **owned by the respective customer** and should be used by them only for purposes set down in digital **information usage policies**, which should accompany each piece of data exchanged. Personal data and policy exchanges can be automated with the help of **privacy exchange protocols**, such as P3P (Cranor, 2012) or HTTP-A (Seneviratne, 2012). User-friendly and operational protocols are still in the making. A core issue for them is that they require minimal user configuration. Once policies are exchanged, these are then the basis of a technically enabled and legally enforceable accountability scheme governing later data exchanges between controllers and processors.

The controller-processor space

The company initially collecting personal data is often not the only party involved in the delivery of services and products; we have shown this extensively above. Subcontracting,

outsourcing, and strategic alliances across multiple organizations are today's default. This complex service web of subcontractors receiving people's personal information reduces the transparency and security of data markets. For this reason, we think that the service web behind a data collector should be "controlled" by that data collector. The data collector is the main entity, which is finally seen as accountable for data breaches by data subjects. The new EU GDPR largely supports this direction of thinking.

Subcontractors controlled by data collectors

The controller-processor space in our definition should comprise only those sub-contractors under control of the controller, which **directly contribute to a serve delivery.** To ensure **contextual integrity**, subcontractors' contributions must be such that their receipt of information can in fact be **anticipated by or justifiable to customers**. Companies for which such contextual integrity cannot be justified should not qualify to receive personal information! Controllers should then be made liable and **accountable for their subcontractors.** For example, all application service providers that reach out to *Facebook* customers would be part of *Facebook*'s controlled space. Facebook in turn would become accountable and liable for any data breaches that occur within their partner network. Context-based trust between customers, controllers and service providers would need to be supported technically again by some kind of **accountability management platform**. Such a platform would manage the collection and sharing of personal data based on the usage policies negotiated and exchanged with customers. Through such platforms, accountability would be effectively created technically and authorization, non-repudiation, separation, and auditability of sharing practices could be ensured. Legal safeguards should back up the appropriate use and setup of accountability management platform. Proposals for such platforms have been made by prominent companies, such as *Microsoft* (Maguire et al. 2015; Nguyen et al. 2013).

The customer-controlled data space

Privacy scholars and some start-up companies have suggested relocating personal data into a sphere solely controllable by customers. They are working on **identity management platforms** (e.g., PRIME and PRIMELife Project) that could help customers to manage their personal data and the various identities linked to them. In an entrepreneurial effort in Silicon Valley, Kaliya Hamlin, who calls herself "Identity Woman", has established the Personal Data Ecosystem (Pdec, 2014), which supports small companies with venture capital access and knowledge to pursue a user-centric data-handling strategy. In an extremely visionary way, Doc Searls proposed the establishment of an "Intention Economy" (Mitchell et al., 2008) where customers use personal agents or third parties to pull services from companies rather than companies pro-actively approaching customers and offering their services. **Trusted third parties and personal data vault technologies** are key for such ideas to thrive.

Trusted third parties

The proposed platforms could act as intermediaries for customers. They could take notes of what customers have revealed to whom. In a more sophisticated scenario, they could mine personal activities and begin to understand customers' preferences based on **user-sided data mining** (Lodder and Voulon, 2002). User preferences could then be used to support customers in their online activities such as their searches for product offerings with providers that are ethical enough to deserve the exchange. The customer-controlled data space would need to be enabled by trusted third parties and personal data vaults. Compared to the above described controller-processor space, trusted third parties and data vault providers would offer customers increased **control over their personal information storage location, intelligence applied to their data, and data deletion**. Of course, they could also be operated by non-profit organizations.

A prerequisite for such technologies to be effective are that companies are willing to engage in this kind of customer-driven exchange. They would need to make their product- and service offerings public and accessible for user-sided agents.

Additional safeguards

By establishing identity management platforms and personal data vaults, consumers benefit from increased privacy because they can control where data is stored, how it is analyzed, when it is deleted and how much is revealed. However, as we discussed in the ethical reflections above, data ownership (or legal property rights) embedded in data exchange policies bear the risk of an extensive commodification of the self. Our discussion of the societal implications showed that power imbalances might turn "voluntary" into "mandatory" and force consumers into data contracts. Additional legal safeguards could address these risks, for example, the "right to a privacy-friendly service" by outlawing a coupling of corporate services with data contracts. Clear provisions of which kinds of information employers, banks or insurers are allowed to demand from applicants prevent that individuals are forced to provide all available data to get a job, loan or insurance policy. A right to data portability prevents that consumers are caught within data contracts and cannot afford to leave a service without losing the results of past efforts.

A Safe Harbor for Big Data

A core recommendation for markets is that safe harbous for Big Data are established. Here, anonymized information could be collected for all players and not just data monopolies as we observe them today. This is essential to re-establish innovation and competition among digital companies and economies. We use the term 'people data' to denote **anonymized personal information**. We do acknowledge of course that anonymized data can be re-identified. We would therefore recommend considering any **re-identification practices as criminal acts**. We also believe that anonymization can only work if a timely and common standard is established, which outlines **‚Best Available Techniques' (BATs) for anonymization**. Such BATs need to be supervised by independent technical bodies, such as the new EU Privacy Board. A body overseeing BATs for anonymization would also undermine the frequent abuse of the word „anonymization" which we have criticized above.

Beyond the ‚anonymity frontier'

Customers could voluntarily '**donate' their data** to the safe harbor for big data; a schema actually propagated by Kaliya Hamlin and called „data raindrops" by her. For example, individuals may share their navigation patterns with the safe harbor for big data so anyone can benefit from traffic congestion information (and not just Google, Apple and Facebook). Data controllers and trusted third parties may transfer data to the safe harbor for big data on behalf of their customers. Each time any data is transferred to the safe harbor for big data, it must cross the 'anonymity frontier' though. Based on a **principle of reciprocity**, everyone might get access to the data stored. The space might even be designed to grant access to data for those who also contribute proportionally to it (both in quantity and/or quality).

List of tables

Table 1: The five dimensions of the "Big Five" personality model. Source: McCrae and Joh 1992.....16

Table 2: Recorded mobile phone data to predict personality traits. Source: Chittaranja et al 2011.....16

Table 3: Pairwise correlations between features and traits having p<0.01, ranked by absolute value of r Source: Chittaranjan et al 2011.....17

Table 4: Accuracy of predicting personality traits from phone data. Source: Chittaranjan et al 2011.....17

Table 5: Evaluated mobile phone data. Source: Montjoye et al 2013.....18

Table 6: Accuracy of predicting personality traits from phone data. Source: Montjoye et al 2013.....18

Table 7: "Big Five" profiles of average visitors of three websites. Source: Kosinski et al, 2012.....19

Table 8: Predicting gender, age, level of education and occupation from website visits. Source: De Bock and Van den Poel 2010.....19

Table 9: Keyboard input evaluated. Source: Epp et al, 2011.....20

Table 10: Accuracy of predicting emotional states from keystroke dynamics. Source: Epp et al, 2011.....20

Table 11: Data models to predict political ideology of voters. Source: Cambridge Analytica.....27

Table 12: Data models to predict political opinions. Source: Cambridge Analytica.....27

Table 13: Risky mobile app behaviors (Source: Appthority, 2014).....49

Table 14: Sensitive data free mobile apps send to third parties (Source: Zang et al 2015).....50

Table 15: Trackers users of smartphone apps are exposed to (Source: Seneviratne et al 2015).....51

Table 16: How Progressive is scoring safe or risky car driving behavior in Ohio. Source: Progressive (2016).....54

Table 17: A set of sensors used for activity recognition (Source: Su et al, 2014).....58

Table 18: Revenues of companies in the marketing data economy that rely on "individual-level consumer data", adapted from Deighton et al (2013, p. 8).....78

Table 19: Typology of data brokers, adapted from FTC (2014).....82

Table 20: Typology of data brokers, adapted from Bria et al (2015).....83

Table 21: Data Brokers in the U.S.: examples for the ways of personal data. Source: FTC 2014, p. 2.....87

Table 22: Examples for sources, which data brokers in the U.S. collect data from. Source: FTC, 2014.....87

Table 23: How many "million monthly uniques" Lotame provides access to, per country. Source: Lotame.....111

List of figures

Figure 1: Types of data offered by VisualDNA. Source: Screenshot VisualDNA website *26*

Figure 2: Data sets and models to predict health. Source: GNS Healthcare website, screenshot 31.07.2016 ..... *37*

Figure 3: "Full-spectrum transaction analysis" for online fraud prevention. Source: Trustev website, screenshot from 29.07.2016. ..... *39*

Figure 4: Earning points for desired behavior at Discovery's usage-based insurance offer. Source: Discovery. *55*

Figure 5: Crossreferencing of articles within GDPR (created by Sushant Agarwal, Institute for Management Information Systems, 2016) *141*

References

Achara, Jagdish Prasad; Gergely Acs, and Claude Castelluccia (2015): On the Unicity of Smartphone Applications. ACM CCS Workshop on Privacy in Electronic Society (WPES), Oct 2015, Denver, Colorado, USA, France. Online: https://hal.inria.fr/hal-01181040/document

Ajana, Btihaj (2005): Surveillance and Biopolitics. Electronic Journal of Sociology, vol 7. Online: http://www.sociology.org/content/2005/tier1/ajana_biopolitics.pdf

Almalki, Manal, Kathleen Gray, and Fernando Martin Sanchez (2015): The Use of Self-Quantification Systems for Personal Health Information: Big Data Management Activities and Prospects. Health Information Science and Systems 3. Suppl 1 (2015): S1. PMC. Web. 20 July 2016. Online: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4437547/

Andrejevic, Mark (2014): The Big Data Divide. International Journal of Communication 8 (2014), 1673–1689. Online: http://ijoc.org/index.php/ijoc/article/download/2161/1163

Angwin J. (2012): Online tracking ramps up e popularity of user-tailored advertising fuels data gathering on browsing habits. Wall Street Journal. June 18 2012:B1.

Appthority (2014): App Reputation Report, 4.8.2014. Online: http://www.nomasis.ch/fileadmin/user_upload/flyer/produkte/Appthority/App_reputation_report.pdf

Aristotle (1915): Magna Moralia Oxford, Carendon Press.

Aristotle (2000): Nichomachean Ethics Cambridge, Cambridge University Press.

Astray, Led (2015): Online Lead Generation and Payday Loans. Upturn, October 2015. Online: https://www.teamupturn.com/static/reports/2015/led-astray/files/Upturn_-_Led_Astray_v.1.01.pdf

Atzori, Luigi; Antonio Iera, and Giacomo Morabito (2010): The Internet of Things: A survey. Comput. Netw. 54, 15 (October 2010), 2787-2805. DOI: 10.1016/j.comnet.2010.05.010. Online: https://www.elsevier.com/__data/assets/pdf_file/0006/97026/The-Internet-of-Things.pdf

AZ Direct (2015): AZ DIAS PROFILDATEN. Merkmalskatalog. Personal copy of file with author Wolfie Christl.

Backstrom, Lars; Kleinberg, Jon. (2014): Romantic partnerships and the dispersion of social ties: a network analysis of relationship status on facebook. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14). ACM, New York, NY, USA, pp. 831-841. Online: http://arxiv.org/pdf/1310.6753v1.pdf

Barcena, Mario Ballano; Candid Wueest, and Hon Lau (2014): How safe is your quantified self? Symantec, August 11, 2014. Online: https://www.symantec.com/content/dam/symantec/docs/white-papers/how-safe-is-your-quantified-self-en.pdf

Barocas, Solon and Selbst, Andrew D. (2016): Big Data's Disparate Impact (2016). 104 California Law Review 671 (2016). Online: http://ssrn.com/abstract=2477899

Bella G, Giustolisi R, Riccobene S. (2011): Enforcing privacy in ecommerce by balancing anonymity and trust. Comput Secur 2011;30(8):705e18.

Bergen, Mark (2014): Flurry Launches Service to Track Mobile App Users, Offline The Analytics Firm Partners With Research Now, As the Race to Target Inside Apps Picks Up. Advertising Age, 24.03.2014. Online: http://adage.com/article/digital/flurry-research-build-mobile-app-advertising-database/292287/

Borgesius; Frederik J. Zuiderveen (2015): Online Price Discrimination and Data Protection Law (August 28, 2015). Forthcoming as a conference paper for the Amsterdam Privacy Conference 23-26 October 2015; Amsterdam Law School Research Paper No. 2015-32; Institute for Information Law Research Paper No. 2015-02. Online: http://ssrn.com/abstract=2652665

Borgesius, Frederik J. Zuiderveen (2016): Singling Out People Without Knowing Their Names – Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation (February 16, 2016). Online: http://ssrn.com/abstract=2733115

Botterman, Maarten (2009): Internet of Things: an early reality of the Future Internet. Workshop Report, European Commission Information Society and Media, May 2009. Online: http://cordis.europa.eu/pub/fp7/ict/docs/enet/iot-prague-workshop-report-vfinal-20090706_en.pdf

Bouk, Dan (2015): How Our Days Became Numbered. Risk and the Rise of the Statistical Individual. University of Chicago Press.

boyd, danah; Urs Gasser; John Palfrey (2010): How the COPPA, as Implemented, Is Misinterpreted by the Public: A Research Perspective. The Berkman Center for Internet & Society, Research Publication No. 2010-12, April 29, 2010. Online: http://ssrn.com/abstract=1794223

boyd danah; Crawford, Kate (2012): Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon, Information, Communication & Society 15:5, pp. 662-679. Online: http://www.tandfonline.com/doi/abs/10.1080/.VB8Tz_l_uCk

Brat, Eric; Stephan Heydorn, Matthew Stover, and Martin Ziegler (2013): Big Data: The Next Big Thing for Insurers? Boston Consulting Group, March 25, 2013. Online: https://www.bcgperspectives.com/content/articles/insurance_it_performance_big_data_next_big_thing_for_insurers

Bria, Francesca; Javier Ruiz, Gemma Galdon Clavell, José Maria Zavala, Laura Fitchner, Harry Halpin (2015): D3.3 Research on Identity Ecosystem. Report by D-CENT, Decentralised Citizens Engagement Technologies, 31 June 2015, Version Number: 2. Online: http://dcentproject.eu/wp-content/uploads/2015/10/research_on_digital_identity_ecosystems.pdf

Bröckling, Ulrich (2007): Das unternehmerische Selbst: Soziologie einer Subjektivierungsform, Frankfurt, M.: Suhrkamp Verlag, 2007

Bujlow, Tomasz; Valentín Carela-Español, Josep Solé-Pareta and Pere Barlet-Ros (2015): A survey on Web Tracking: Mechanisms, Implications, and Defenses. Under review, July 2015. Online: http://arxiv.org/abs/1507.07872

Busby, Ed; Tawfik Hammoud, John Rose, Ravi Prashad (2012): The evolution of online-user data. The Boston Consulting Group, 2012. Online: https://www.bcgperspectives.com/content/articles/marketing_technology_evolution_of_online_user_data/

Casassa Mont, M., Pearson, S., and Bramhall, P(2003): "Towards Accountable Management of Identity and Privacy: Sticky Policies and Enforceable Tracing Services," HP Laboratories Bristol.

CDT, Center for Democracy and Technology (2010): Comments of the Center for Democracy & Technology. In the Matter of Information Privacy and Innovation in the Internet Economy. June 14, 2010. Online: https://www.cdt.org/files/pdfs/20100613_doc_privacy_noi.pdf

Cegłowski, Maciej (2016): The Moral Economy of Tech. Text version of remarks, SASE conference, Berkeley, June 26, 2016. Online: http://idlewords.com/talks/sase_panel.htm

Chester, Jeff; Edmund Mierzwinski (2014): Big Data Means Big Opportunities and Big Challenges. Promoting Financial Inclusion and Consumer Protection in the “Big Data” Financial Era. U.S. PIRG Education Fund and Center for Digital Democracy, March 2014. Online: http://www.uspirg.org/sites/pirg/files/reports/USPIRGEFandCDDBigDataReportMar14_1.3web.pdf

Chittaranjan, G.; Blom, J. & Gatica-Perez, D. (2011): Who's Who with Big-Five: Analyzing and Classifying Personality Traits with Smartphones. In: ISWC, IEEE, pp. 29-36. Online: http://infoscience.epfl.ch/record/192371/files/Chittaranjan_ISWC11_2011.pdf

Christl, Wolfie (2014): "Kommerzielle Digitale Überwachung im Alltag," Cracked Labs - Institut für Kritische Digitale Kultur, Wien.

Citron, Danielle Keats and Pasquale, Frank A. (2014): The Scored Society: Due Process for Automated Predictions. Washington Law Review, Vol. 89, 2014; U of Maryland Legal Studies Research Paper No. 2014-8. Online: http://ssrn.com/abstract=2376209

CMA (2015): The commercial use of consumer data. Report on the CMA's call for information. Competition and Markets Authority UK, CMA38, June 2015. Online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/435817/The_commercial_use_of_consumer_data.pdf

Cooper, Tim and Ryan LaSalle (2016): Guarding and growing personal data value. Accenture, 2016. Online: https://www.accenture.com/_acnmedia/PDF-4/Accenture-Guarding-and-Growing-Personal-Data-Value-POV-Low-Res.pdf

Cranor, L.F. (2003): "P3P: Making Privacy Policies More Useful," in: IEEE Security & Privacy, pp. 50-55.

Cranor, L.F., Dobbs, B., Egelman, S., Hogben, G., Humphrey, J., and Schunter, M. (2006): "The Platform for Privacy Preferences 1.1 (P3P1.1) Specification - W3C Working Group Note 13 November 2006," R. Wenning and M. Schunter (eds.), World Wide Web Consortium (W3C) - P3P Working Group.

Cranor LF (2012): Necessary but not sufficient: standardized mechanisms for privacy notice and choice. J Telecomm High Tech L 2012 (10:2), pp. 273-308.

Crawford, Kate; Jessa Lingel and Tero Karppi (2015): Our Metrics, Ourselves: A Hundred Years of Self-Tracking From The Weight Scale to The Wrist Wearable Device, European Journal of Cultural Studies. Online: http://ecs.sagepub.com/content/18/4-5/479.full.pdf+html

De Bock, K., Van den Poel, D. (2010): Predicting website audience demographics for web advertising targeting using multi-website clickstream data. FUNDAMENTA INFORMATICAE, 98(1), pp. 49–70. Online: http://hdl.handle.net/1854/LU-967442

De Domenico, M.; Lima, A.; Musolesi, M. (2012): Interdependence and Predictability of Human Mobility and Social Interactions.Proceedings of the Nokia Mobile Data Challenge Workshop Newcastle, United Kingdom. June 2012. Online: http://www.cs.bham.ac.uk/research/projects/nsl/mobility-prediction/mdc12.pdf

Deighton, John; Peter A. Johnson (2013): The Value of Data: Consequences for Insight, Innovation & Efficiency in the U.S. Economy. October 8, 2013. Online: https://www.ipc.be/~/media/documents/public/markets/the-value-of-data-consequences-for-insight-innovation-and-efficiency-in-the-us-economy.pdf

Deloitte (2010): Predictive Modeling for Life Insurance. April 2010. Online: https://www.soa.org/files/pdf/research-pred-mod-life-batty.pdf

Deterding, Sebastian; Khaled, Rilla; Nacke, Lennart; Dixon, Dan (2011): Gamification: Toward a Definition, Proc. Workshop on Gamification at the ACM Intl. Conf. on Human Factors in Computing Systems (CHI). Online: http://gamification-research.org/wp-content/uploads/2011/04/02-Deterding-Khaled-Nacke-Dixon.pdf

De Zwart, Melissa; Humphreys, Sal; Van Dissel, Beatrix (2014). Surveillance, big data and democracy: lessons for Australia from the US and UK, UNSW Law Journal. Online: http://www.unswlawjournal.unsw.edu.au/sites/default/files/final_t3_de_zwart_humphreys_and_van_dissel.pdf

Dixon, Pam; Robert Gellman (2014): The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future. World Privacy Forum, April 2, 2014. Online: http://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf

Dobusch Leonhard (2014): Digitale Zivilgesellschaft in Deutschland. Stand und Perspektiven 2014. Freie Universität Berlin, Fachbereich Wirtschaftswissenschaft, Diskussionsbeiträge. Online: http://edocs.fu-berlin.de/docs/servlets/MCRFileNodeServlet/FUDOCS_derivate_000000003411/discpaper2014_7.pdf

Dwoskin, Elizabeth (2014): Data Broker Acxiom Moves to Tie Physical World to Online Data. Wall Street Journal, 14.05.2014. Online: http://blogs.wsj.com/digits/2014/05/14/data-broker-acxiom-moves-to-tie-physical-world-to-online-data

Duhigg, Charles (2012): Die Macht der Gewohnheit. Berlin Verlag.

Dumortier, Franck (2009). Facebook and Risks of "De-contextualization" of Information. In: Monograph "5th Internet, Law and Politics Congress. The Pros and Cons of Social Networks", Universitat Oberta de Catalunya. Online: http://journals.uoc.edu/index.php/idp/article/viewFile/n9_dumortier/n9_dumortier_eng

Enck, W.; Gilbert, P.; Chun, B.; Cox, L.; Jung, J.; Mc-Daniel, P.; Sheth, A. (2010): TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones. In Proc. of USENIX Symposium on Operating Systems Design and Implementation (OSDI). Online: http://static.usenix.org/event/osdi10/tech/full_papers/Enck.pdf

Epp, C.; Lippold, M.; Mandryk, R.L. (2011): Identifying Emotional States Using Keystroke Dynamics. In Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (CHI 2011), Vancouver, BC, Canada. Pp. 715-724. Online: http://hci.usask.ca/uploads/203-p715-epp.pdf

European Commission (2016): REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). E. Commission (ed.), European Commission, Brussels. Online: http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf

Experian (2011): List Services Catalog. Online: http://www.experian.com/assets/data-university/brochures/ems-list-services-catalog.pdf [13.01.2016]

Fertik, Michael (2013): The Rich See a Different Internet Than the Poor. Ninety-nine percent of us live on the wrong side of a one-way mirror. Scientific American, 14.01.2013. Online: http://www.scientificamerican.com/article/rich-see-different-internet-than-the-poor/

Fitbit (2016): Annual report pursuant to section 13 or 15(d) of the Securities Exchange Act of 1934. Form 10-K. For the Fiscal Year Ended December 31, 2015. Online: http://d1lge852tjjqow.cloudfront.net/CIK-0001447599/f7a564df-cf62-4681-90a4-a3ef813b0d83.pdf?noexit=true [21.07.2016]

FIPA / B.C. Freedom of Information and Privacy Association (2015): The Connected Car: Who is in the driver's seat? A study on privacy and onboard vehicle telematics technology. March 2015. Online: https://fipa.bc.ca/wordpress/wp-content/uploads/2015/03/CC_report_lite.pdf

FTC (2012): Report to Congress Under Section 319 of the Fair and Accurate Credit Transactions Act of 2003, December 2012. Online: https://www.ftc.gov/sites/default/files/documents/reports/section-319-fair-and-accurate-credit-transactions-act-2003-fifth-interim-federal-trade-commission/130211factareport.pdf

FTC, US Federal Trade Commission (2014): Data Brokers. A Call for Transparency and Accountability. Online: http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

FTC, U.S. Federal Trade Commission (2015): Internet of Things. Privacy and Security in a Connected World. FTC Staff Report, January 2015. Online: https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

FTC, U.S. Federal Trade Commission (2016): Big Data – A Tool for Inclusion or Exclusion? Understanding the Issues. FTC Report January 2016. Online:: https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf

Futurezone (2012): This is how Austrian Facebook users tick! In: Futurezone - Technology News, Kurier.at, Vienna.

Gandy, Oscar (2006): Data Mining, Surveillance, and Discrimination in the Post-9/11 Environment. In: Haggerty, K., Ericson, R. (2006): The New Politics of Surveillance and Visibility, Toronto, University of Toronto Press

Gandy Jr., O.H. (2009): Coming to Terms With Chance: Engaging Rational Discrimination and Cumulative Disadvantage, London: Ashgate.

GAO, United States Government Accountability Office (2006): Personal Information. Agency and Reseller Adherence to Key Privacy Principles. GAO-06-421, April 2006. Online: http://www.gao.gov/new.items/d06421.pdf

Ghoshal, S. (2005): "Bad Management Theories Are Destroying Good Management Practices," Academy of Management Learning and Education (4:1), pp 75-91.

Graham, Stephen D.N. (2005): Software-sorted geographies. Progress in Human Geography 29, 5 (2005) pp. 562–580. Online: http://www.dourish.com/classes/readings/Graham-SoftwareSortedGeographies-PHG.pdf

Gritzalis S. (2004): Enhancing web privacy and anonymity in the digital era. Inf Manag Comput Secur 2004;12(3):255e87.

Han, Jiawei; Kamber, Micheline; Pei, Jian (2011): Data Mining: Concepts and Techniques, 3rd ed. The Morgan Kaufmann Series in Data Management Systems.

Hannak, Aniko; Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson (2014): Measuring Price Discrimination and Steering on E-commerce Web Sites. In Proceedings of the 2014 Conference on Internet Measurement Conference (IMC '14). ACM, New York, NY, USA, 305-318. Online: http://personalization.ccs.neu.edu/papers/price_discrimination.pdf

Hannak, Aniko; Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson (2014): Paper Overview. Online: http://personalization.ccs.neu.edu/PriceDiscrimination/Press

Hansen, Marit (2012): Überwachungstechnologien. In: Schmidt, Jan-Hinrik; Weichert, Thilo (Ed.) (2012): Datenschutz. Grundlagen, Entwicklungen und Kontroversen, 1. Ed., Bundeszentrale für politische Bildung, Bonn, pp. 23-32. Online: http://www.bpb.de/system/files/dokument_pdf/1190-Datenschutz-X3.pdf

Hardy, Quentin (2012): Just the Facts. Yes, All of Them. New York Times, 25.03.2012. Online: http://www.nytimes.com/2012/03/25/business/factuals-gil-elbaz-wants-to-gather-the-data-universe.html

Harford, Tim (2014): Big data: are we making a big mistake? Financial Times, 28.03.2014. Online: http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz3DK9lcAdI

Helberger, Natali (2016): Profiling and Targeting Consumers in the Internet of Things – A New Challenge for Consumer Law (February 6, 2016). Online: http://ssrn.com/abstract=2728717

Hildebrandt, M. (2015): Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology Cheltenham, UK, Edward Elgar Pub.

Hill, Kashmir (2014): 10 Other Facebook Experiments On Users, Rated On A Highly-Scientific WTF Scale. Forbes, 10.07.2014. Online: http://www.forbes.com/sites/kashmirhill/2014/07/10/facebook-experiments-on-users/

Hilts, Andrew; Christopher Parsons, and Jeffrey Knockel (2016): Every Step You Fake: A Comparative Analysis of Fitness Tracker Privacy and Security. Open Effect Report, 2016. Online: https://openeffect.ca/reports/Every_Step_You_Fake.pdf

Hursthouse, R. (1999): On Virtue Ethics Oxford, UK, Oxford University Press.

Hursthouse, R. (2012): "Virtue Ethics," in: The Stanford Encyclopedia of Philosophy, E.N. Zalta (ed.), The Metaphysics Research Lab Stanford.

Information Resources Management Association (2012): Data Mining: Concepts, Methodologies, Tools, and Applications. IGI Global

Ito, Aki (2013): Hiring in the Age of Big Data. Bloomberg Businessweek, 24.10.2013. Online: http://www.businessweek.com/articles/2013-10-24/new-way-to-assess-job-applicants-online-games-and-quizzes

IWGDPT, International Working Group on Data Protection in Telecommunications (2014): Working Paper on Big Data and Privacy. Privacy principles under pressure in the age of Big Data analytics. International Working Group on Data Protection in Telecommunications, 55th Meeting, 5 – 6 May 2014, Skopje. Online: http://www.datenschutz-berlin.de/attachments/1052/WP_Big_Data_final_clean_675.48.12.pdf

Kanngieser, Anja (2013): Tracking and tracing: geographies of logistical governance and labouring bodies. In: Environment and Planning D: Society and Space 2013, volume 31, pages 594 – 610. DOI: 10.1068/d24611. Online: http://anjakanngieser.com/wp-content/uploads/2012/07/Kanngieser_Tracking-and-tracing.pdf

Kant, I. (1784/2009): An Answer to the Question: "What Is Enlightenment?" London, Penguin Books.

Kant, I. (1785/1999): "Groundwork for the Metaphysics of Morals," in: Practicle Philosophy, M.J. Gregor and A.W. Wood (eds.), New York, Cambridge University Press.

Kaptein, Maurits; Dean Eckles, and Janet Davis (2011): Envisioning Persuasion Profiles: Challenges for Public Policy and Ethical Practice. 9/10 Interactions 66-69, 66; Online: https://www.semanticscholar.org/paper/Envisioning-persuasion-profiles-challenges-for-Kaptein-Eckles/fe5f2029df491bdea2cf46697b2e4145c1e226f2/pdf

Karapiperis, Dimitris; Birny Birnbaum, Aaron Brandenburg, Sandra Castagna, Allen Greenberg, Robin Harbage, Anne Obersteadt (2015): Usage-Based Insurance and Vehicle Telematics: Insurance Market and Regulatory Implications. Study by National Association of Insurance Commissioners (NAIC) and Center for Insurance Policy and Research (CIPR). March 2015. Online: http://www.naic.org/documents/cipr_study_150324_usage_based_insurance_and_vehicle_telematics_study_series.pdf

Kaye, J., Whitley, E.A., Lund, D., Morrison, M., Teare, H., and Melham, K. (2014): "Dynamic consent: a patient interface for twenty-first century research networks," European Journal of Human Genetics), pp 1-6.

Kaye, Kate (2014): Acxiom Acquires LiveRamp to Boost Offline-to-Online Data Capability. Advertising Age, 14.05.2014. Online: http://adage.com/article/datadriven-marketing/acxiom-buys-liveramp-offline-online-data-capability/293212/

Koh Hian, C., Chan Kin Leong, G. (2002). Data Mining and Customer Relationship Marketing in the Banking Industry. Singapore Management Review, Vol. 24, No. 2, p. 4

Kollaten Venne, Patrick; Eikenberg, Ronald; Schmidt; Jürgen (2012), Selbstbedienungsladen Smartphone, c't, Heft 7/2012, S. 114.

Kosinski, Michal: Stillwell, David; Kohli, Pushmeet; Bachrach, Yoram; Graepel, Thore (2012): Personality and Website Choice, in ACM Web Sciences 2012, ACM Conference on Web Sciences, 2012. Online: http://research.microsoft.com/pubs/163547/person_WebSci_final.pdf

Kosinski, Michal; Stillwell, David; Graepelb, Thore (2013): Private Traits and Attributes Are Predictable from Digital Records of Human Behavior. PNAS, March 2013. Online: http://www.pnas.org/content/110/15/5802

Kramer, Adam D. I.; Guillory, Jamie E. & Hancock, Jeffrey T. (2014): Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences 111 (24), pp. 8788–8790. Online: http://www.pnas.org/content/111/24/8788.full

Krishnan, Krish (2013): Data Warehousing in the Age of Big Data. Morgan Kaufmann.

Kulyk, Oksana; Paul Gerber, Michael El Hanafi, Benjamin Reinheimer, Karen Renaud, Melanie Volkamer (2016): Encouraging Privacy-Aware Smartphone App Installation: Finding out what the Technically-Adept Do. In: USEC Workshop, San Diego, California, 21 Feb 2016. Online: http://eprints.gla.ac.uk/116161/

Laughlin, Andrew (2014): Smart TV spying – are you watching TV, or is it watching you? Which? Magazine, 20.08.2014. Online: http://blogs.which.co.uk/technology/tvs/smart-tv-spying-weve-investigated

Lin, J., J.I. Hong, N. Sadeh (2014): Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings. In Symposium on Usable Privacy and Security (SOUPS 2014). Online: http://cmuchimps.org/publications/modeling_users_mobile_app_privacy_preferences_restoring_usability_in_a_sea_of_permission_settings_2014

Lodder AR, Voulon MB (2002): Intelligent agents and the information requirements of the directives on distance selling and ecommerce. Int Rev Law Comput Technol 2002;16(3):277e87.

Lopes, Hezal and Rahul Lopes (2013): Comparative Analysis of Mobile Security Threats And Solution. In: Int. Journal of Engineering Research and Application, Vol. 3, Issue 5, Sep-Oct 2013, pp.499-502. Online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.399.2948&rep=rep1&type=pdf

Lupton, Deborah (2016): The diverse domains of quantified selves: self-tracking modes and dataveillance, Economy and Society, 45:1, 101-122, DOI: 10.1080/03085147.2016.1143726. Online: http://www.tandfonline.com/doi/pdf/10.1080/03085147.2016.1143726

Lyon, David (1994): The Electronic Eye: The Rise of Surveillance Society, Minneapolis, University of Minnesota Press.

Lyon, David (2003) Surveillance as social sorting: Computer codes and mobile bodies. In: Lyon, D. (Ed.): Surveillance as social sorting: Privacy, risk, and digital discrimination. Routledge, London, New York.

Lyon, David (2007): Surveillance Studies: An Overview. Cambridge: Polity Press

MacIntyre, A. (1984): After Virtue: A Study in Moral Theory, (2nd Edition ed.) Notre Dame, Indiana, University of Notre Dame Press.

Maguire, S., Friedberg, J., Nguyen, C., and Haynes, P. (2015): "A metadata-based architecture for user-centered data accountability," Electronic Markets (25:2), pp 155–160.

Marwick, Alice E. (2013): Big Data, Data-Mining, and the Social Web. Talk for the New York Review of Books Event: Privacy, Power & the Internet, October 30, 2013. Online: http://www.tiara.org/blog/wp-content/uploads/2013/10/marwick_2013_datamining_talk.pdf [14.08.2016]

Maslow, A. 1970 Motivation and Personality, (2nd edition ed.) New York, Harper & Row Publishers.

Mattioli, Dana (2012): On Orbitz, Mac Users Steered to Pricier Hotels. Wall Street Journal, 23.08.2012. Online: http://online.wsj.com/news/articles/SB10001424052702304458604577488822667325882

Mayer-Schönberger, Viktor; Cukier, Kenneth (2013): Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt

McConville, Ryan; Steven Mong (2014): The (Pro) Consumer Genome: The Rise of Customer Agents in the Personal Data Market. Mack Institute for Innovation Management, Wharton University of Pennsylvania,

February 3, 2014. Online: https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2013/01/McConville-Mong_Mack_Institute_Paper_Final_vFinal.pdf

Mclaughlin, Catriona (2013): Acxiom. Die Besserwisser. Die Zeit, 05.07.2013. Online: http://www.zeit.de/2013/28/acxiom/komplettansicht

Mikians, Jakub; Gyarmati, László; Erramilli, Vijay; Laoutaris, Nikolaos (2012): Detecting price and search discrimination on the internet. In Proceedings of the 11th ACM Workshop on Hot Topics in Networks (HotNets-XI). ACM, New York, NY, USA, 79-84. Online: http://conferences.sigcomm.org/hotnets/2012/papers/hotnets12-final94.pdf

Mill, J.S. (1863/1987): "Utilitarianism," in: Utilitarianism and Other Essays, A. Ryan (ed.), London, Penguin Books.

Minerva, Roberto; Abyi Biru, and Domenico Rotondi (2015): Towards a Definition of the Internet of Things (IoT). IEEE Internet Initiative, May 27, 2015. Online: http://iot.ieee.org/images/files/pdf/IEEE_IoT_Towards_Definition_Internet_of_Things_Revision1_27MAY15.pdf

Mitchell A, Henderson I, Searls D. (2008): Reinventing direct marketing - with VRM inside. J Direct Data Digit Mark Pract 2008;10(1):3e15.

Montjoye, Yves-Alexandre de; Quoidbach, Jordi; Robic, Florent; Pentland, Alex (2013): Predicting personality using novel mobile phone-based metrics. In: Ariel M. Greenberg, William G. Kennedy, and Nathan D. Bos (Ed.): Proceedings of the 6th international conference on Social Computing, Behavioral-Cultural Modeling and Prediction (SBP'13), Springer-Verlag, Berlin, Heidelberg, pp. 48-55. Online: http://web.media.mit.edu/~yva/papers/deMontjoye2013predicting.pdf

Montjoye, Yves-Alexandre de; Hidalgo, César A.; Verleysen, Michel; Blondel, Vincent D. (2013b): Unique in the Crowd: The Privacy Bounds of Human Mobility. Scientific Reports, March 2013, No. 1376. Online: http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html

Mudholkar, Smita S.; Shende, Pradnya M.; Sarode, Milind V. (2012): Biometrics Authentication Technique for Intrusion Detection Systems Using Fingerprint Recognition. In: International Journal of Computer Science, Engineering and Information Technology, Vol.2, No.1, February 2012. Online: http://airccse.org/journal/ijcseit/papers/2112ijcseit06.pdf

Munson, Sean A.; Sunny Consolvo (2012): Exploring Goal-setting, Rewards, Self-monitoring, and Sharing to Motivate Physical Activity. In: 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops. Online: https://www.semanticscholar.org/paper/Exploring-goal-setting-rewards-self-monitoring-and-Munson-Consolvo/9e3ea24b4492f58568d5a8e73b99ceb3fd5ee610/pdf

Myslewski, Rik: The Internet of Things helps insurance firms reward, punish. The Register, 24.05.2014. Online: http://www.theregister.co.uk/2014/05/23/the_internet_of_things_helps_insurance_firms_reward_punish/

Nagel, T (1992): Der Blick von Nirgendwo Frankfurt/Main Suhrkamp.

Narayanan, Arvind; Shmatikov, Vitaly (2008): Robust De-anonymization of Large Sparse Datasets. In: Proceedings of the 2008 IEEE Symposium on Security and Privacy (SP '08). IEEE Computer Society, Washington, DC, USA, pp. 111-125. Online: https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf

National Consumer Law Center (2014): Comments to the Federal Trade Commission Big Data: A Tool for Inclusion or Exclusion? Workshop, Project No. P145406. Submitted August 15, 2014. Online: https://www.ftc.gov/system/files/documents/public_comments/2014/08/00018-92374.pdf

Nguyen, C., Haynes, P., Maguire, S., and Friedberg, J. (2013): "A User-Centred Approach to the Data Dilemma: Context, Architecture, and Policy," in: The Digital Englightment Yearbook 2013, M. Hilebrandt (ed.), Brussels, IOS Press.

Nissenbaum, H. 2004. "Privacy as Contextual Integrity," Washington Law Review (79:1), pp 101-139.

Nonaka, I., Toyama, R., and Toru, H. (2008): Managing Flow - A Process Theory of the Knowledge-Based Firm New York, Palgrave MacMillan.

Nonaka, I., and Takeuchi, H. (2011): "The Wise Leader," Harvard Business Review. May 2011, pp 58-67

Norwegian Consumer Council (2016): APPFAIL. Threats to Consumers in Mobile Apps. March, 2016. Online: http://fbrno.climg.no/wp-content/uploads/2016/03/Appfail-Report-2016.pdf

Norwegian Consumer Council (2016b): Attachment to the report 'Appfail – Threats to Consumers in Mobile Apps'. Updated 1. July 2016. Online: http://fbrno.climg.no/wp-content/uploads/2016/03/appwin-update-of-changes-1.-july.pdf

Office of the Privacy Commissioner of Canada (2014): Global Privacy Enforcement Network (GPEN) Privacy Sweep. Online: https://www.priv.gc.ca/media/nr-c/2014/bg_140910_e.asp

Ohm, Paul (2009): Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization (August 13, 2009). UCLA Law Review, Vol. 57, 2010; U of Colorado Law Legal Studies Research Paper No. 9-12. Online: http://www.uclalawreview.org/pdf/57-6-3.pdf

Olson, Parmy (2013): Meet The Company That Tracks More Phones Than Google Or Facebook. Forbes, 30.10.2013. Online: http://www.forbes.com/sites/parmyolson/2013/10/30/meet-the-company-that-tracks-more-phones-than-google-or-facebook/

Olson, Parmy (2014): The Quantified Other: Nest And Fitbit Chase A Lucrative Side Business. Forbes, 05.05.2014. Online: http://www.forbes.com/sites/parmyolson/2014/04/17/the-quantified-other-nest-and-fitbit-chase-a-lucrative-side-business

Olson, Parmy (2014b): Wearable Tech Is Plugging Into Health Insurance. Forbes, 19.06.2014. Online: http://www.forbes.com/sites/parmyolson/2014/06/19/wearable-tech-health-insurance/

O'Neill, Cathy (2016): Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown, 2016.

Oracle (2013): The value of social data. Integrated Social and Enterprise Data = Enhanced Analytics. Oracle white paper, December 2013. Online: http://www.sponsored.com.au/app/webroot/uploaded_files/media/SRM_US_EN_WP_SocialData_1.pdf [13.01.2016]

Oracle (2015): Oracle Data Cloud. Data Directory. Online: http://www.oracle.com/us/solutions/cloud/data-directory-2810741.pdf [22.01.2016]

Pasquale, Frank (2015): The Other Big Brother. The Atlantic, Sep 21, 2015. Online: http://www.theatlantic.com/business/archive/2015/09/corporate-surveillance-activists/406201

Pdec. Personal data ecosystem consortium (2014): Empowering people with their personal data. Available from: http://pde.cc/ [20.01.2014]

Petrison, L. A., Blattberg, R. C. and Wang, P. (1993): Database marketing. Past, present, and future. J. Direct Mark., 7: 27–43. doi: 10.1002/dir.4000070306

Peppet, Scott R. (2011): Unraveling Privacy: The Personal Prospectus & the Threat of a Full Disclosure Future. Northwestern University Law Review, 2011. Online: http://ssrn.com/abstract=1678634

Pew Research Center (2014): The Internet of Things Will Thrive by 2025. May 2014. Online: http://www.pewinternet.org/2014/05/14/internet-of-things

Pfitzmann, Andreas; Hansen, Marit (2010): A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. Version 0.34, August 2010. Online: http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf

Ptolemus (2016): Usage-based Insurance Global Study 2016. January 2016. Download of free abstract and purchase of full report: http://www.ptolemus.com/ubi-study/#downloadpoint

Redman, T.C. (2013): "Data's Credibility Problem," Harvard Business Review), December 2013, pp 2-6.

Rhoen, Michiel (2016): Beyond consent: improving data protection through consumer protection law. Internet Policy Review, 5(1). DOI: 10.14763/2016.1.404. Online: http://policyreview.info/articles/analysis/beyond-consent-improving-data-protection-through-consumer-protection-law

Richards, Neil (2013): The Dangers of Surveillance. Harvard Law Review 126, 1934, 1953.Online: http://cdn.harvardlawreview.org/wp-content/uploads/pdfs/vol126_richards.pdf

Robbins, Rebecca (2015): Insurers want to nudge you to better health. So they're data mining your shopping lists. Stat, 15.12.2015. Online: https://www.statnews.com/2015/12/15/insurance-big-data

Roesner, Franziska; Tadayoshi Kohno, and David Wetherall (2012): Detecting and Defending Against the Web. In: NSDI'12 Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, Pages 12-12, USENIX Association Berkeley, CA, USA. Online: https://www.usenix.org/system/files/conference/nsdi12/nsdi12-final17.pdf

Roosendaal, Arnold; Marc van Lieshout, Anne Fleur van Veenstra (2014): Personal Data Markets. Earth, Life & Social Sciences, TNO 2014 R11390. Online: http://publications.tno.nl/publication/34612412/riZsP9/TNO-2014-R11390.pdf

Rosenblat, Alex, Kneese, Tamara and Boyd, Danah (2014): Networked Employment Discrimination (October 08, 2014). Open Society Foundations' Future of Work Commissioned Research Papers 2014. Online: http://www.datasociety.net/pubs/fow/EmploymentDiscrimination.pdf

Rothmann, Robert; Sterbik-Lamina, Jaro; Peissl, Walter; Čas, Johann (2012) Aktuelle Fragen der Geodaten-Nutzung auf mobilen Geräten – Endbericht. Bericht-Nr. ITA-PB A63; Institut für Technikfolgen-Abschätzung (ITA): Wien; im Auftrag von: Österreichische Bundesarbeitskammer. Online: http://epub.oeaw.ac.at/ita/ita-projektberichte/d2-2a63.pdf

Rouvroy, Antoinette (2016): Of Data and Men. Fundamental Rights and Freedoms in a World of Big Data. Bureau of the consultative commitee of the convention for the protection of individuals with regard to automatic processing of personal data. Council of Europe, Strasbourg, 11 January 2016. Online: https://www.coe.int/t/dghl/standardsetting/dataprotection/TPD_documents/T-PD-BUR(2015)09REV_Big%20Data%20report_A%20%20Rouvroy_Final_EN.pdf

Satariano, Adam (2014): Wear This Device So the Boss Knows You're Losing Weight. Bloomberg. Online: http://www.bloomberg.com/news/2014-08-21/wear-this-device-so-the-boss-knows-you-re-losing-weight.html

Schneier, Bruce (2015): Data and Goliath. The Hidden Battles to Collect Your Data and Control Your World. W. W. Norton & Company.

Scism, Leslie; Maremont, Mark (2010): Insurers Test Data Profiles to Identify Risky Clients. The Wall Street Journal, 19.11.2010.Online: http://online.wsj.com/articles/SB10001424052748704648604575620750998072986

Senate Committee on Commerce, Science, and Transportation (2013): A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes. Staff Report for Chairman Rockefeller, December 18, 2013. Online: https://www.commerce.senate.gov/public/_cache/files/bd5dad8b-a9e8-4fe9-a2a7-

b17f4798ee5a/D5E458CDB663175E9D73231DF42EC040.12.18.13-senate-commerce-committee-report-on-data-broker-industry.pdf

Seneviratne, S. (2012) OW. Augmenting the web with accountability. In: 21st International Conference Companion on World Wide Web, Lyon, France. 2188006. ACM; 2012. p. 185e9.

Seneviratne, Suranga; Harini Kolamunna, and Aruna Seneviratne (2015): A measurement study of tracking in paid mobile applications. In Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks (WiSec '15). ACM, New York, NY, USA, Article 7, 6 pages. Online: https://www.researchgate.net/publication/282356703_A_measurement_study_of_tracking_in_paid_mobile_applications

Simonite, Tom (2013): Ads Could Soon Know If You're an Introvert (on Twitter). MIT Technology Review, 08.11.2013. Online: http://www.technologyreview.com/news/520671/ads-could-soon-know-if-youre-an-introvert-on-twitter/

Singer, Natasha (2012): You for Sale. Mapping, and Sharing, the Consumer Genome. New York Times, 16.06.2012. Online: http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html

Solon, Olivia (2015): Wearable Technology Creeps Into The Workplace. Bloomberg, August 7, 2015. Online: http://www.bloomberg.com/news/articles/2015-08-07/wearable-technology-creeps-into-the-workplace

Solove, Daniel J. and Schwartz, Paul M., An Overview of Privacy Law (2015): Chapter 2 of PRIVACY LAW FUNDAMENTALS (published by IAPP, 2015); GWU Law School Public Law Research Paper No. 2015-45; GWU Legal Studies Research Paper No. 2015-45. Online: http://ssrn.com/abstract=2669879

Solove, Daniel J. (2004). The Digital Person: Technology and Privacy in the Information Age. New York University Press, p. 97 et seq.

Solove, Daniel J. (2006). A Taxonomy of Privacy. University of Pennsylvania Law Review, Vol. 154, No. 3, January 2006, pp. 477-560

Sommerville, I. (2011): Software Engineering, (Nine ed.) International, Pearson.

Spiekermann S, Dickinson I, Günther O, Reynolds D. (2003): User agents in e-commerce environments: Industry vs. Consumer perspectives on data exchange. In: Eder J, Missikoff M, editors. Lncs. Berlin: Springer; 2003. p. 696e710.

Spiekermann, S. and Pallas, F. (2005): Technology Paternalism - Wider Implications of Ubiquitous Computing. Poiesis & Praxis: International Journal of Ethics of Science and Technology Assessment, Vol. 4, 2005. Online: http://edoc.hu-berlin.de/oa/articles/reFDoa9WpdlyU/PDF/27aIqGTe3Neo.pdf

Spiekermann, S., and Novotny, A. (2015): "A vision for global privacy bridges: Technical and legal measures for international data markets," Computer Law and Security Review (31:2), pp 181-200.

Spiekermann, S., Aquisti, A., Böhme, R., and Hui, K.-L. (2015) (eds.): Person Data Markets and Privacy. Elsivier, 2015.

Spiekermann, S., and Korunovska, J. (2016): "Towards a Value Theory for Personal Data," Journal of Information Technology (JIT) (forthcoming).

Steinberg, Gregory B.; Bruce W. Church, Carol J. McCall, Adam B. Scott, Brian P. Kalis (2014): Novel Predictive Models for Metabolic Syndrome Risk: A "Big Data" Analytic Approach. The American Journal of managed care, Vol. 20, No. 6, June 2014. Online: http://www.ajmc.com/journals/issue/2014/2014-vol20-n6/novel-predictive-models-for-metabolic-syndrome-risk-a-big-data-analytic-approach

Su, Xing; Hanghang Tong, and Ping Ji (2014): Activity Recognition with Smartphone Sensors. In: Tsinghua Science and Technology, ISSN ll1007-0214ll02/11llpp235-249, Volume 19, Number 3, June 2014. Online: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6838194

Suler, J. (2004): "The Online Disinhibition Effect," Cyper Psychology & Behavior (7:3), pp 321-326.

Sweeney, L. (2002): k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; pp. 557-570. Online: http://epic.org/privacy/reidentification/Sweeney_Article.pdf

Sweeney, L. (2014): Health Data Flows. Presentation at "Consumer Generated and Controlled Health Data", May 7, 2014. Online: https://www.ftc.gov/system/files/documents/public_events/195411/consumer-health-data-webcast-slides.pdf

Talbot, David (2012): A Phone that Knows Where You're Going. MIT Technology Review, 09.07.2012. Online: http://www.technologyreview.com/news/428441/a-phone-that-knows-where-youre-going/

Tene, Omer and Jules Polonetsky (2013): Big Data for All: Privacy and User Control in the Age of Analytics. 11 Nw. J. Tech. & Intell. Prop. 239 (2013). Online: http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

Tessman, L. (2005): Burdened Virtues - Virtue Ethics for Liberatory Struggles Oxford, UK, Oxford University Press.

Thurm, S.; Kane, Y. (2010): Your Apps Are Watching You, Wall Street Journal. Online: http://online.wsj.com/article/SB10001424052748704694004576020083703574602.html, In: What They Know. Online: http://blogs.wsj.com/wtk-mobile/

Tisserand, J.-C. (2014): "Ultimatum game: A meta-analysis of the past three decades of experimental research," Proceedings of International Academic Conferences, International Institute of Social and Economic Sciences.

Traung, P. (2012): "The Proposed New EU Data Protection Regulation," CRi - A Journal for Information Law and Technology: 2).

Tufekci, Zeynep (2014): Engineering the public: Big data, surveillance and computational politics. First Monday, Volume 19, Number 7, 7 July 2014. Online: http://firstmonday.org/article/view/4901/4097

ULD (2014), Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, GP Forschungsgruppe: Scoring nach der Datenschutz-Novelle 2009 und neue Entwicklungen. Abschlussbericht der Studie im Auftrag der Bundesanstalt für Landwirtschaft und Ernährung. Online: http://www.bmi.bund.de/SharedDocs/Downloads/DE/Nachrichten/Kurzmeldungen/studie-scoring.pdf?__blob=publicationFile

Urban, Jennifer M.; Hoofnagle, Chris Jay; Li, Su: Mobile Phones and Privacy. Berkeley Consumer Privacy Survey, BCLT Research Paper, 11.07.2012. Online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2103405

Valentino-Devries, Jennifer; Singer-Vine, Jeremy; Soltani, Ashkan (2012): Websites Vary Prices, Deals Based on Users' Information. Wall Street Journal. Online: http://online.wsj.com/news/articles/SB10001424127887323777204578189391813881534

Varian, H.R., and Shapiro, C. 1999 Information Rules - A Strategic Guide to the Network Economiy Boston, Massachusetts, Harvard Business Books Press.

Venkatesan, R. and Kumar, V. (2004): A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. Journal of Marketing, Vol. 68, No. 4, pp. 106-125

Wall Street Journal (2010): What They Know. Online: http://blogs.wsj.com/wtk

Weichert, Thilo (2013): Big Data und Datenschutz. Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein. Online: https://www.datenschutzzentrum.de/bigdata/20130318-bigdata-und-datenschutz.pdf

Weiser, Mark (1991): The Computer for the 21st Century. Scientific American, September 1991. Draft version Online: http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html

Weiser, M., Gold, R., and Brown, J.S. (1999): "The origins of ubiquitous computing research at PARC in the late 1980s," IBM Systems Journal (38:4), pp 693-696.

Whitson, Jennifer R. (2013): Gaming the Quantified Self. Surveillance & Society. Special Issue on Surveillance Futures 11(1/2). Online: http://library.queensu.ca/ojs/index.php/surveillance-and-society/article/view/gaming

Wolf, Gary (2010): The Data-Driven Life. New York Times, 02.05.2010. Online: http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html?pagewanted=all

WEF, World Economic Forum (2011): Personal Data: The Emergence of a New Asset Class. Online: http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf

WEF, World Economic Forum (2012): Rethinking Personal Data. Strengthening Trust. Online: http://www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf

WEF, World Economic Forum (2014): Rethinking Personal Data: Trust and Context in User-Centred Data Ecosystems. (in co-operation with Microsoft), Davos.

Yee, N. (2014): The Protheus Paradox New Haven, Yale University Press.

Youyou W., Kosinski M., Stillwell D. (2015): Computer-based personality judgments are more accurate than those made by humans. PNAS vol. 112 no. 4 1036–1040, doi: 10.1073/pnas.1418680112. Online: http://www.pnas.org/content/112/4/1036

Zang, Jinyan, Krysta Dummit, James Graves, Paul Lisker, and Latanya Sweeney (2015): Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps. Technology Science, 2015103001, 30.10.2015. Online: http://techscience.org/a/2015103001

Zuboff, Shoshana (2015): Big Other: Surveillance Capitalism and the Prospects of an Information Civilization (April 4, 2015). Journal of Information Technology (2015) 30, 75–89. doi:10.1057/jit.2015.5. Online: http://ssrn.com/abstract=2594754

Zuboff, Shoshana (2016): The Secrets of Surveillance Capitalism. Frankfurter Allgemeine Zeitung, 05.03.2016. Online: http://www.faz.net/aktuell/feuilleton/debatten/the-digital-debate/shoshana-zuboff-secrets-of-surveillance-capitalism-14103616.html?printPagedArticle=true