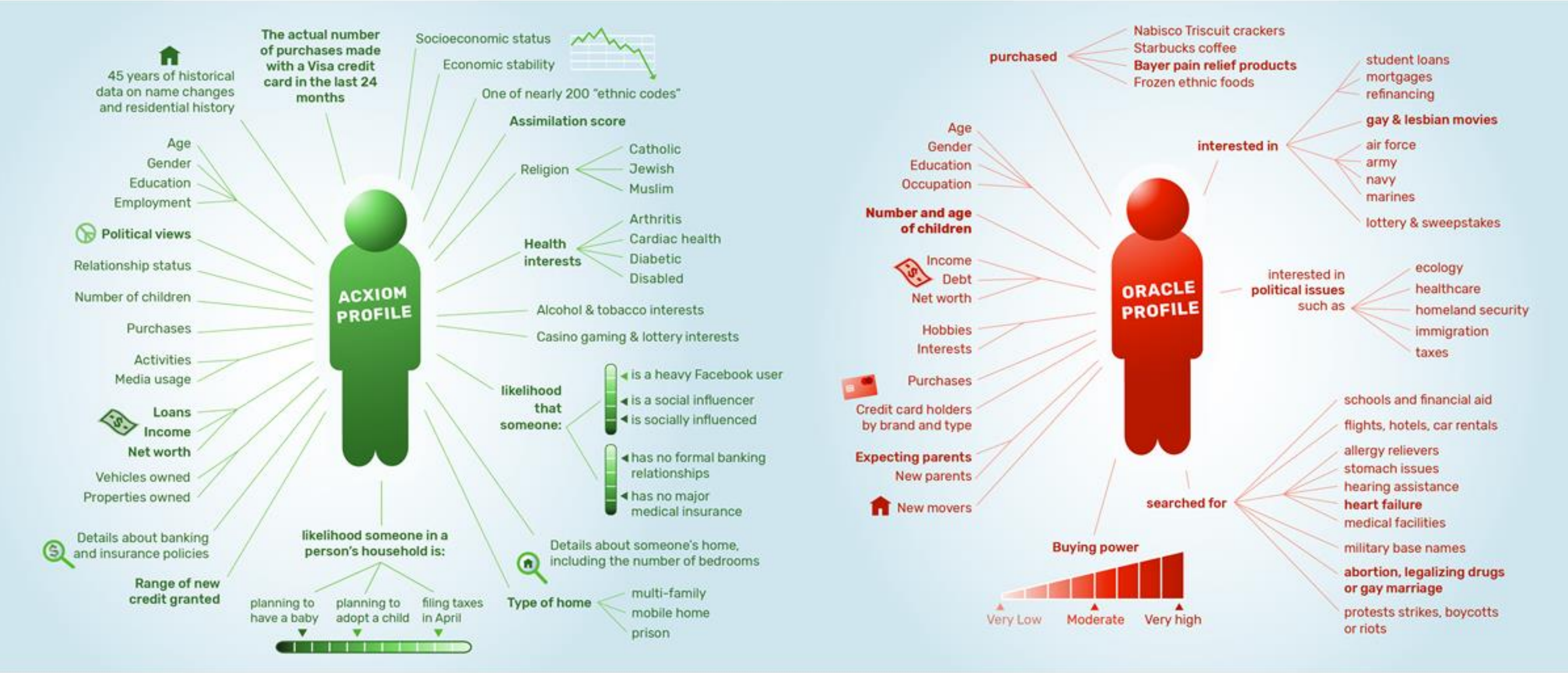
Wolfie Christl

CORPORATE SURVEILLANCE IN EVERYDAY LIFE

How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions



A REPORT BY CRACKED LABS

Vienna, June 2017

Author: Wolfie Christl
Contributors: Katharina Kopp, Patrick Urs Riechert
Illustrations: Pascale Osterwalder

Wolfie Christl

Corporate Surveillance in Everyday Life

How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions
A Report by Cracked Labs, Vienna, June 2017

Author: Wolfie Christl
Contributors: Katharina Kopp, Patrick Urs Riechert
Illustrations: Pascale Osterwalder

Every effort has been made to ensure the accuracy of the texts in this report. The author and the publisher accept no liability in the case of eventual errors.

Cracked Lab – Institute for Critical Digital Culture
Annagasse 8/1/8, 1010 Vienna, Austria
http://crackedlabs.org

Wolfie Christl
is a digital rights activist, technologist, researcher, writer and educator, based in Vienna
http://twitter.com/WolfieChristl
http://wolfie.crackedlabs.org

The production of this report was supported by the Open Society Foundations.

Table of Contents

1. Background and Scope

In recent months, a large ride-hailing platform has been accused of abusing its data power to identify, block and undermine regulators[1], suppliers,[2] and rivals[3], as well as to manipulate both its drivers and riders.[4] As Ryan Calo and Alex Rosenblat suggest in their outstanding paper on "Uber, Information, and Power"[5], such platforms may be "leveraging their access to information about users and their control over the user experience to mislead, coerce, or otherwise disadvantage sharing economy participants". Generally, firms could increasingly "use what they know about consumers" to not only "match them to content they might prefer" but also to "nudge consumers to pay more, to work for less, and to behave in other ways that advantage a firm".

This encapsulates one of the key concerns that scholars, privacy activists and consumer rights advocates have been raising for years. When a rapidly growing number of daily interactions and behaviors undergo unrestricted digital monitoring, analysis, and assessment, corporate actors can systematically abuse their resultant unprecedented data wealth for their economic advantage. Omer Tene and Jules Polonetsky compare the relationship between data companies and individuals to a "game of poker where one of the players has his hand open and the other keeps his cards close"[6] As Calo and Rosenblat write, when a company "can design an environment from scratch, track consumer behavior in that environment, and change the conditions throughout that environment based on what the firm observes, the possibilities to manipulate are legion". For example, they may "reach consumers at their most vulnerable, nudge them into overconsumption, and charge each consumer the most he or she may be willing to pay".[7]

The consequences of pervasive consumer surveillance. However, these practices are not the domain of a one large company. In recent years, a vast landscape of partially interconnected databases has emerged that consists not only of large players such as Facebook and Google but also of thousands of other companies from various industries that collect, analyze, acquire, share, trade, and utilize data on billions of people.[8] Further major concerns about the possible implications of these opaque "networks of corporate surveillance"[9] focus on their potential to

[1] https://www.nytimes.com/2017/03/03/technology/uber-greyball-program-evade-authorities.html

[2] https://mobile.nytimes.com/2017/04/23/technology/travis-kalanick-pushes-uber-and-himself-to-the-precipice.html

[3] https://www.theguardian.com/technology/2017/apr/13/uber-allegedly-used-secret-program-to-cripple-rival-lyft

[4] http://boingboing.net/2017/03/09/weaponized-information-asymmet.html

[5] Calo, Ryan and Rosenblat, Alex (2017): The Taking Economy: Uber, Information, and Power (March 9, 2017). Columbia Law Review, Vol. 117, 2017; University of Washington School of Law Research Paper No. 2017-08. Available at: https://ssrn.com/abstract=2929643

[6] 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).p. 255. Available at: http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

[7] Calo, Ryan and Rosenblat, Alex (2017)

[8] Christl, Wolfie and Sarah Spiekermann: Networks of Control. A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy. Facultas, Vienna 2016, p. 7. Available at: http://crackedlabs.org/en/networksofcontrol

[9] Ibid, p. 10

make data-driven decisions about people that are inaccurate, arbitrary, or biased.[10] This may limit the chances and choices of individuals and lead to the discrimination and social exclusion of whole population groups.[11] Algorithmic decisions based on digital profiles may reinforce existing biases and social inequalities and even become self-fulfilling prophecies.[12] Furthermore, much of corporate data collection and utilization happens invisibly, often with neither knowledge nor consent of the subjects.[13] When people know that they are being constantly monitored, this can produce chilling effects on forms of action or expression. Finally, refusing to participate in today's digital tracking can have negative consequences, too.[14] Consumers can "hardly avoid privacy contracts" because "almost all banks, software and hardware vendors, social networking sites, digital content services, retail loyalty programs, and telecommunications providers employ them"[15]. As Maciej Cegłowski summarizes, "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 means "opting out of much of modern life".[16]

Online advertising: a major driver but minor problem? In crucial areas of life such as finance, insurance, housing, healthcare, welfare, law enforcement, and employment, many would agree that inaccurate, biased or discriminatory automated decisions based on digital profiles about consumers can harm individuals and vulnerable population groups. Similarly, when a single large company possesses fine-grained data on and control over interactions with consumers, the danger of abusive nudging and manipulation appears apparent. In contrast, today's online advertising ecosystem – one of the major drivers of ubiquitous corporate digital tracking and profiling – features a large number of tech companies constantly aggregating, combining, sharing, and trading profiles about people. Many people consider the intrusive data collection for online advertising as merely posing a minor problem. Yet the inner workings of today's advertising technology are barely understood outside of this business, and even fewer people grasp its social, economic and ethical consequences.

The pervasive real-time surveillance machine that has been developed for online advertising is rapidly expanding into other fields, from pricing to political communication to credit scoring to risk management. Many companies have already aggregated enormous, extensive databases of sensitive behavioral data on billions of people, which they freely make use of for any purpose that happens to fit their economic interests. In addition, online advertising itself has already morphed into something other than what has, until recently, been understood under

[10] Ibid, p. 126, 127

[11] Lyon, David (2010): Surveillance, Power and Everyday Life. In: Kalantzis-Cope, Phillip; Gherab-Martín, Karim (eds): Emerging Digital Spaces in Contemporary Society: Properties of Technology, Palgrave 2010, p. 107-120

[12] Christl and Spiekermann (2016), p. 125

[13] Ibid., p. 121-123

[14] Ibid., p. 127

[15] Rhoen, Michiel (2016): Beyond consent: improving data protection through consumer protection law. Internet Policy Review, 5(1), p2. Available at: http://policyreview.info/articles/analysis/beyond-consent-improving-data-protection-throughconsumer-protection-law

[16] Cegłowski, Maciej (2016): The Moral Economy of Tech. Text version of remarks, SASE conference, Berkeley, June 26, 2016. Available at: http://idlewords.com/talks/sase_panel.htm

"advertising"; today's online marketing technologies go far beyond simply displaying ads. Based on data-driven predictive analytics, personalization, measurement, and testing, they aim to influence behavior at scale. In the background, consumers are constantly evaluated, sorted, categorized, and ranked in order to treat them on a case-by-case basis as best fits a company's business interests. Yet as individuals are made ever more transparent, corporate practices remain largely opaque.

This report examines today's corporate networks of digital tracking and profiling and the implications that extensive collection and utilization of personal information have for individuals, groups of individuals, and society at large. It examines how companies record, combine, share, and trade personal data on billions, and aims to better map today's personal data ecosystem. Based on previous research on corporate surveillance and privacy undertaken by the author together with Sarah Spiekermann[17], it further investigates and summarizes the structure and scope of today's personal data ecosystem, as well as of other relevant industries, business models, platforms, services, devices, technologies, and data flows. On a broader level, this report is part of a larger effort that aims to identify key issues relevant to fundamental rights and social justice, as well as to encourage further work by civil society, media, and other critical voices. The key question is how commercial digital profiling and data-driven algorithmic decisions affect equality, freedom, autonomy, democracy, and human dignity on both individual and societal levels. With this in mind, this report focuses on the actual practices and inner workings of today's personal data industry and explores relevant recent developments.

2. Introduction

It is not easy to gain an overview of how information about individuals is collected, analyzed, shared, and utilized in the commercial sphere today – and not only because of the industry's non-transparency.[18] Previous reports and investigations on today's personal data ecosystem and its societal implications have focused on specific practices (such as credit scoring[19] or voter targeting[20]), devices (such as smart TVs[21] or activity trackers[22]), technologies (such as face

[17] Christl, Wolfie and Sarah Spiekermann (2016): Networks of Control. A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy. Facultas, 2016. Available at: http://crackedlabs.org/en/networksofcontrol

[18] Ibid., p. 121

[19] E.g. Ferretti, F. (2009): The credit scoring pandemic and the European vaccine: Making sense of EU data protection legislation. Journal of Information, Law & Technology, 2009. Available at: http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2009_1/ferretti

[20] E.g. Bennett, Colin (2016): Is Your Neighbor and Democrat or a Republican? Lateral Voter Surveillance and the Political Culture of Modern Election Campaigns (April 20, 2016). Available at: https://ssrn.com/abstract=2776308

[21] E.g. Irion, Kristina and Natali Helberger (2016): Smart TV and the online media sector: User privacy in view of changing market realities. Telecommunications Policy, Volume 41, Issue 3, April 2017, Pages 170–184. Available at: http://doi.org/10.1016/j.telpol.2016.12.013

[22] E.g. Norwegian Consumer Council (2016): Consumer protection in fitness wearables. November, 2016. Available at: https://fil.forbrukerradet.no/wp-content/uploads/2016/11/2016-10-26-vedlegg-2-consumer-protection-in-fitness-wearables-forbrukerradet-final-version.pdf

recognition[23]), environments (such as mobile apps[24]), types of data (such as health[25] or location data[26]), harms (such as exclusion[27] or manipulation[28]), groups of people affected (such as minorities[29] or students[30]), industries (such as telecom providers[31] or data brokers[32]), or the dominant companies (such as Facebook[33] and Google[34]).

While such a detailed focus on different components is urgently needed to shed light on the opaque methods of data exploitation and its implications, one runs the risk of losing sight of the bigger picture. Today's personal data ecosystem is diverse, fragmented, quickly developing, and driven by various, often conflicting, interests. Some players, such as Facebook and Google, are very visible for consumers; many others, though, operate behind the scenes and away from the public's attention. The rise of social networks, mobile communication, government mass surveillance, and regular data breaches has precipitated an ongoing debate on privacy. It has even focused the public's attention on previously less known industries, such as consumer reporting agencies and marketing data brokers.[35] And yet, one cannot properly evaluate the data ecosystem without understanding all the players, large and small, visible and hidden. To understand the societal implications of today's personal data ecosystem, it is essential to look at it from both a distance and up close.

[23] E.g. United States Government Accountability Office (2015): Facial Recognition Technology. Commercial Uses, Privacy Issues, and Applicable Federal Law. Report to the Ranking Member, Subcommittee on Privacy, Technology and the Law, Committee on the Judiciary, U.S. Senate, July 2015. Available at: http://www.gao.gov/assets/680/671764.pdf

[24] E.g. 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. Available at: http://techscience.org/a/2015103001

[25] E.g. Tanner, Adam (2017): Our Bodies, Our Data. How companies make billions selling our medical records. Beacon Press, Jan 10, 2017

[26] E.g. Rothmann, Robert; Sterbik-Lamina, Jaro; Peissl, Walter; as, Johann (2012): Aktuelle Fragen der Geodaten-Nutzung auf mobilen Geräten – Endbericht. Bericht-Nr. ITA-PB A63; Institut für Technikfolgen-Abschätzung (ITA): Vienna. Available at: https://www.oeaw.ac.at/itaen/projects/the-use-of-geodata-on-mobile-devices/overview/

[27] E.g. Seeta Peña Gangadharan (2012): Digital inclusion and data profiling. First Monday, Volume 17, Number 5 - 7 May 2012. Available at: http://firstmonday.org/ojs/index.php/fm/article/view/3821/3199

[28] E.g. Calo, Ryan (2013): Digital Market Manipulation. 82 George Washington Law Review 995 (2014); University of Washington School of Law Research Paper No. 2013-27, August 15, 2013. Available at: https://ssrn.com/abstract=2309703

[29] E.g. Angwin, Julia; Jeff Larson; Lauren Kirchne; Surya Mattu (2017): Minority Neighborhoods Pay Higher Car Insurance Premiums Than White Areas With the Same Risk. ProPublica, April 5, 2017. Available at: https://www.propublica.org/article/minority-neighborhoods-higher-car-insurance-premiums-white-areas-same-risk

[30] E.g. Frida Alim, Nate Cardozo, Gennie Gebhart, Karen Gullo, and Amul Kalia (2017): Spying on Students: School-Issued Devices and Student Privacy. Electronic Frontier Foundation, April 13, 2017. Available at: https://www.eff.org/wp/school-issued-devices-and-student-privacy

[31] E.g. Ammari, Nader; Gustaf Björksten; Peter Micek; Deji Olukotun (2015): The Rise of Mobile Tracking Headers: How Telcos Around the World Are Threatening Your Privacy. Access Now, August 2015. Available at: https://www.accessnow.org/cms/assets/uploads/archive/AIBT-Report.pdf

[32] E.g. Rieke, Aaron; Harlan Yu; David Robinson; Joris von Hoboken (2016): Data Brokers in an Open Society. An Upturn Report, prepared for the Open Society Foundations. November 2016. Available at: https://www.opensocietyfoundations.org/sites/default/files/data-brokers-in-an-open-society-20161121.pdf

[33] E.g. Lee, Newton (2014): Facebook Nation. Total Information Awareness. Springer, 2014.

[34] E.g. Tene, Omer (2007): What Google Knows: Privacy and Internet Search Engines (October 1, 2007). Utah Law Review. Available at SSRN: https://ssrn.com/abstract=1021490

[35] E.g. Singer, Natasha (2012): Mapping, and Sharing, the Consumer Genome. New York Times, June 16, 2012. Available at: http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-ofconsumer-database-marketing.htm

Therefore, **chapter 3** takes a big picture view, focusing on the wide range of players within the business of personal data, including key industries such as media, retail, telecom and financial services. Beside the large online platforms and other data companies in the fields of credit reporting, direct marketing, analytics and online advertising, businesses in all industries are increasingly involved in today's digital tracking and profiling landscapes. Without a thriving ecosystem of companies collecting extensive amounts of data on consumers, the data giants would not be able to aggregate, combine, analyze, rent, and sell their extensive amounts of personal data. However, many of these businesses and organizations are even more secretive about their data collecting and sharing than the data and analytics industry itself, which, after all, has to promote and advertise its services to its clients.

Subsequently, **chapter 4** examines companies, services, and technologies within the risk data industry, starting with a summary of how credit and consumer reporting agencies use data to rate people. Even as traditional methods of credit scoring still raise many questions pertaining to accuracy, fairness, equity, transparency, explainability, and accountability, companies are expanding the types of data they use to predict the creditworthiness of individuals. Some have started to incorporate behavioral data derived from e.g. web searches, social media, smartphone usage, and other sources in their scoring mechanisms. Traditional identity verification and fraud prevention vendors now monitor and evaluate how people surf the web and use their mobile devices. Furthermore, they have started to link digital behavioral data with the vast amounts of offline identity information they have been collecting for decades. Google's captcha service is now able to identify whether someone is a legitimate human being or not without any explicit user interaction, such as a quiz or a click on a checkbox, based only on the company's far-reaching behavioral monitoring capabilities.

Chapter 5 turns to the marketing data industry and shows how long-established practices of consumer segmentation and database marketing join forces with the pervasive tracking and profiling embedded in today's online advertising. Companies can now find and target users with specific characteristics and behaviors in real-time, regardless of which service or device is used, which activity is pursued, or where the user is located at a given moment. Within milliseconds, digital profiles about consumers are auctioned and sold to the highest bidder. Large consumer data brokers have started to partner with hundreds of advertising technology firms as well with platforms such as Google and Facebook. They combine data about offline purchases with online behaviors and provide services that allow other companies to recognize, link, and match people across different corporate databases. Businesses in all industries can use the services of data companies to seamlessly collect rich data about consumers, add additional information on them, and utilize the enriched digital profiles across a wide range of technology platforms.

After **chapter 6** presents two case studies that examine the data services, providers, and partners of the consumer data brokers Acxiom and Oracle as illustrative examples of wider practices, **chapter 7** summarizes relevant key developments from recent years:

- Established commercial practices of data collection on consumers have rapidly evolved into pervasive **networks of digital tracking and profiling** that exist in a complex and vast landscape of corporate players who continuously monitor the lives of billions of people. Companies are now able to identify and address consumers on an individual level in a growing number of everyday situations.
- For this purpose, businesses **aggregate and link personal identifiers** such as email addresses, phone numbers, smartphone IDs, and platform user account IDs and use them to recognize people, as well as to link and combine profiles about them across different services, platforms, devices, and life contexts. Since names are not very useful for identifying people in the digital world, data brokers have introduced globally unique corporate IDs for consumers. All of these identifiers can be connected to digital profiles distributed across multiple companies, which are continuously updated based on large-scale data collection and advanced classification and prediction methods.
- Hundreds of large companies, including online platforms, data brokers and advertising technology firms, sort consumers into **tens of thousands of categories**. Digital profiles about individuals include attributes and scores about education, employment, political views, ethnicity, religion, health interests, media usage, purchases, income, economic stability, and personality. Myriads of behavioral data categories refer to website visits, video views, app usage, movements, and web searches, including sensitive "interests" such as for abortion, legalizing drugs, gay marriage, military bases, protests, heart failure, and more.
- Based on the sophisticated ways of linking and combining data across different sources, companies can utilize today's **ubiquitous behavioral data streams** from myriad services to monitor and analyze all activities and behaviors of consumers that might be relevant to their business interests. With the help of data vendors, companies try to **capture and measure every interaction** with a consumer in real-time, including on websites, platforms, and devices that they do not control themselves.
- Data can now be used not only to display ads on websites or within mobile apps, but also to dynamically **personalize** the contents, options, and choices offered to consumers on, for example, a company's website. Companies can define complex sets of business rules that define how to automatically react to certain kinds of consumer activities across the digital world and beyond; for instance, they can personalize prices in online shops by predicting how valuable someone might be as customer in the long term or how much someone is probably willing to pay in a specific moment. Similarly, politicians can target voters with messages personalized to a voter's personality and political views on certain issues.
- Personalization based on rich profile information and pervasive real-time monitoring has become a powerful tool set to influence people's behavior e.g., to make consumers visit a website, click on an ad, register for a service, subscribe to a newsletter, download an app, or purchase a product. To further improve this, companies have started to continuously **experiment on people**. They conduct tests with different variations of functionalities, website designs, user interface elements, headlines, button texts, images, or even different

discounts and prices, and then carefully monitor and measure how different groups of users interact with these variations. In this way, companies systematically optimize their modes of data-driven persuasion.

Mission creep. Not at the least, information about people's behaviors, social relationships, and most private moments is increasingly used in contexts or for purposes completely different from those in which it was recorded – for instance, to make automated decisions in crucial areas such as finance, insurance, healthcare, employment, and law enforcement. Conversely, data that has been collected in the contexts of credit scoring, identity verification, fraud prevention, payment processing, or even healthcare is increasingly used for customer relationship management, online targeting, and other marketing purposes. Furthermore, data and analytics companies that make eligibility decisions and perform risk assessments of individuals have started to provide their clients with data services that integrate and unify data for risk management, customer relationship management, and marketing.

Perhaps most concerning, today's digital fraud detection services use incredibly invasive technologies to evaluate billions of online transactions and to collect vast amounts of information in order to identify devices, individuals, and suspicious behaviors. They combine and link data collected for online fraud with offline identity records, information about someone's financial situation, and many other types of data; this is then used for purposes other than pure fraud detection – for example, to decide which payment options someone gets in an online shop. In addition, there is some evidence that the extensive data collected for digital fraud detection is also used to recognize consumers and link information about them for marketing purposes, and vice-versa.[36] When the two faces of pervasive digital tracking – online advertising that constantly evaluates each individual's economic potential and fraud detection that gauges a given person's risk potential – are tied together, information society moves significantly further toward surveillance society.

[36] See chapter 5 and section 7.8

3. Relevant players within the business of personal data

Today's large online platforms[37], such as Google and Facebook, have extensive information about the everyday lives of billions of people around the globe. They are the most visible, pervasive, and – aside from online advertisers and intelligence contractors – perhaps the most advanced players in the personal data business. However, in contrast to popular belief, they do not directly, for the most part, *sell* and *share* their detailed digital consumer profiles to third parties, at least not in the form of unified dossiers.[38] Instead, the large online platforms mostly let other companies *utilize* their data without fully transferring it, and they let them use their infrastructure to collect more data, to the benefit of both the client companies and the platforms themselves.[39] The data the platforms have about their users is, after all, their most valuable asset, and, as such, one that they do not want to give it away.[40] Their dominance allows them to control how anyone – consumers and third party companies alike – can use the data they have collected. Of course, the large online platforms are not the only ones that collect extensive information about people and let others utilize it to take advantage; however, in contrast to the platform space, which is currently dominated by only a few actors, the ecosystem of players sharing and selling consumer data consists of a greater number of players.

Before the Internet, four basic consumer surveillance streams existed, as Bruce Schneier summarizes in his book "Data and Goliath".[41] The first came from companies that kept records on customers, e.g. through loyalty programs. The second was **direct marketing**, a sector which included the trading of lists about consumers with certain characteristics. The third flowed from **credit bureaus** that collected detailed credit information on people. The fourth consumer surveillance stream originated from publicly available records about citizens such as birth certificates and various licenses, which companies were increasingly able to download or purchase. Today's **data and analytics companies** often combine information from all four of these streams. While some data companies focus more on personal information for marketing, including customer data management and online advertising, others, which are usually referred to as consumer reporting agencies, focus on different aspects of risk management such as credit and insurance scoring, identity verification, and fraud detection. Some, such as Experian, cover almost all major areas of consumer data.

[37] Regarding the concept of "platforms" see e.g.: Evans, Peter C. and Annabelle Gawer (2016): The Rise of the Platform Enterprise. A Global Survey. The Emerging Platform Economy Series No. 1, The Center for Global Enterprise, January 2016. Available at: http://thecge.net/wp-content/uploads/2016/01/PDF-WEB-Platform-Survey_01_12.pdf

[38] The data industry often refers to Facebook and Google's "walled gardens of data", consisting of their "tremendous first party data assets and reach". See e.g. http://www.businessinsider.com/luma-partners-state-of-digital-media-2016-presentation-2016-5

[39] For example, by embedding Facebook and Google's data-gathering software into their websites and ads, including embedded videos and Like buttons, ad and analytics tags, and other many other services.

[40] See e.g. https://adexchanger.com/data-driven-thinking/facebook-and-google-are-bringing-walled-gardens-back

[41] Schneier, Bruce (2015): Data and Goliath. The Hidden Battles to Collect Your Data and Control Your World. W. W. Norton & Company

Besides the large platforms and the consumer data and analytics industry, which later chapters will examine in further detail, the following sections take a big picture view on other key industries in this space, including retail, media, digital publishers, telecom, device providers, and financial services. Businesses in many industries are increasingly involved in today's pervasive digital tracking and profiling landscape. Without a thriving multitude of companies collecting extensive amounts of data on consumers, the data giants would not be able to aggregate, combine, analyze, rent, and sell extensive amounts of personal data that they have amassed. Therefore, the next sections focus on the role of corporate players in different sectors of today's personal data ecosystem. After a quick review of how crucial societal domains such as politics, healthcare, and national security might already be involved in or affected by the commercial tracking and profiling landscape, possible future developments are discussed.

Figure 1 shows a rough map of today's commercial digital tracking and profiling landscape. Which industries are relevant in the context of the large-scale commercial collection and utilization of consumer data? Which key societal domains are affected?[42]

[42] Since many companies operate across industries, and industries are often related, this chart is far from perfect. For example, Google and Facebook could also be seen as the largest players in ad technology. In addition, regarding data protection, the societal and regulatory environments differ widely between different world regions, as well as with regards to society and economy at large. However, the mapping can hopefully support the further exploration of the structure and scope of today's personal data ecosystem. The chart is based on previous own research (see Christl and Spiekermann 2016), previous attempts to map the personal data ecosystem, and an extensive literature review considering academic work, as well as many industry reports, including, but not limited to: FTC "Personal Data Ecosystem", 2012 (https://www.ftc.gov/sites/default/files/documents/public_events/exploring-privacy-roundtable-series/personaldataecosystem.pdf), FTC "Data Brokers", 2014 (https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf), D-CENT Research on Identity Ecosystem, 2015 (https://dcentproject.eu/wp-content/uploads/2015/10/research_on_digital_identity_ecosystems.pdf), European Data Market studies by IDC and Open Evidence, 2016-2017 (http://www.datalandscape.eu/study-reports), works by Kaliya "Identity Woman" Young (see e.g. https://www.slideshare.net/Kaliya/ethical-market-models-in-the-personal-data-ecosystem), LUMA Partners' sector landscapes, 2017 (http://www.lumapartners.com/resource-center/lumascapes-2/)

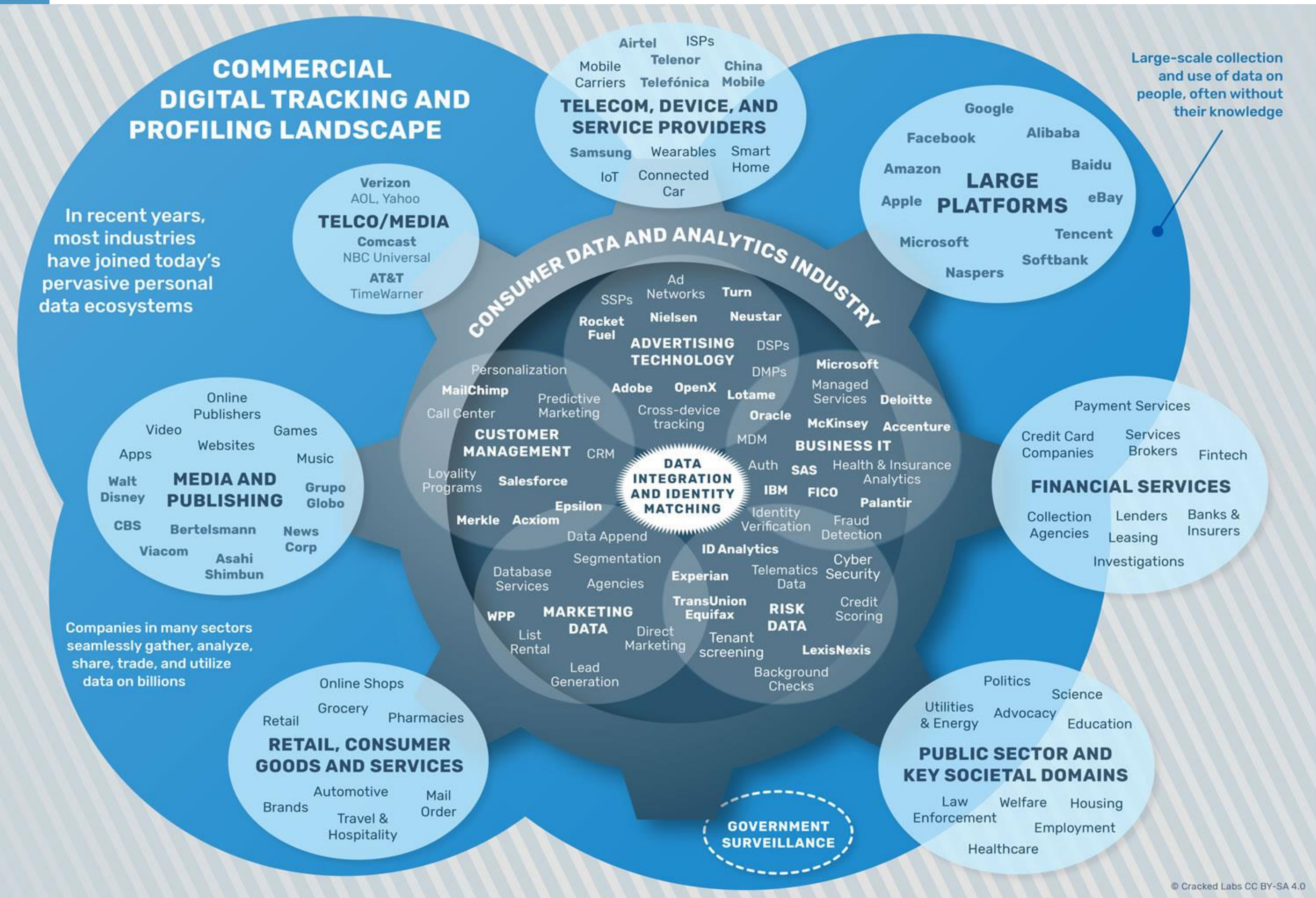


Figure 1: Mapping the commercial digital tracking and profiling landscape

3.1 Businesses in all industries

Businesses in e.g. the retail, travel, consumer goods, media, telecommunication, banking, and insurance sectors that sell consumer products or services have all been collecting, using, and partly also sharing, information about prospects and customers for decades. Database marketing and loyalty programs have intensified not only the amount and detail of collected data, but also companies' ability to make use of it. Companies learned how to use information and data mining techniques to identify, acquire, and retain profitable customers, to calculate a person's future "customer lifetime value"[43], and to "effectively allocate resources" only to the "most

[43] Christl and Spiekermann (2016), p.79

profitable group of customers".[44] They began analyzing the interests, needs, and weaknesses of consumers and then sorting, categorizing, ranking, and treating them accordingly. They learned how to include or exclude groups of customers from certain efforts, how to better influence their behaviors based on data, and how to measure and optimize the outcomes.[45]

After airlines and hotels started **loyalty programs** in the 1980s, retailers followed suit in the early 1990s, and the financial services industry in the late 1990s; today companies in virtually all industries run such programs.[46] In 2015/2016, consumers in the US had 3.3 billion loyalty program memberships total,[47] with an average of seven active memberships per person, and a broad range of sectors, ranging from financial services to travel, pharmacy chains, grocery stores, and other consumer brand retail outlets, offering loyalty programs.[48]

Coalition loyalty programs operated by more than one company or by independent vendors often include sharing customer data among several companies.[49] Companies with related, complementary target groups have been generally sharing certain customer data with each other for many years. In this sense, retailers would share with consumer brands, car dealers with auto manufacturers, and, of course, publishers with advertisers.[50] They also traded lists of names and addresses of people with specific interests, both with each other and with consumer data brokers. Such lists included information on newspaper and magazine subscribers, book and movie club members, catalog and mail order buyers, travel agency bookers, seminar and conference participants, and about consumers filling out warranty card product registrations.[51] [52]

Many **retailers** sell more or less aggregated forms of consumer purchase data to market research companies and consumer data brokers.[53] For example, the US-based company IRI accesses data from more than 85,000 grocery, mass merchandise, drug, club, dollar, convenience, liquor, and pet stores.[54] Oracle claims to have data on billions of purchase transactions from

[44] Ngai, E. W. T., Li Xiu, and D. C. K. Chau (2009): Review: Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications 36, 2 (March 2009), 2592-2602. http://dx.doi.org/10.1016/j.eswa.2008.02.021

[45] Christl and Spiekermann (2016), p. 125; 127-129

[46] Kumar, V. (2008): Managing Customers for Profit: Strategies to Increase Profits and Build Loyalty. FT Press, p. 13

[47] https://www.colloquy.com/latest-news/2015-colloquy-loyalty-census/

[48] http://info.bondbrandloyalty.com/hubfs/Resources/2016_Bond_Loyalty_Report_Executive_Summary_US_Launch_Edition.pdf

[49] Capizzi, Michael T. and Rick Ferguson (2005): Loyalty trends for the twenty-first century. Journal of Consumer Marketing, 22/2, 72-80

[50] https://www.signal.co/blog/data-sharing-second-party-data [28.04.2017]

[51] CIPPIC (2006): On the data trail: How detailed information about you gets into the hands of organizations with whom you have no relationship. A report on the Canadian data brokerage industry, April 2006. Available at: https://idtrail.org/files/DatabrokerReport.pdf

[52] Allenby, Greg M (2010): Perspectives on Promotion and Database Marketing: The Collected Works by Robert C Blattberg. World Scientific, 2010.

[53] http://www.cpgdatainsights.com/get-started-with-nielsen-iri/what-is-syndicated-data

[54] https://www.quantcast.com/audience-grid/data-partner/iri-audience-data/ [25.04.2017]

"1500 leading retailers".[55] Nielsen claims to collect sales information from worldwide 900,000 stores in more than 100 countries.[56] According to the retail analytics expert Emmett Cox, only one notable US retailer, Walmart, has stopped "participating in the data selling" – and only "because they felt that their business was so significantly larger than anyone else's".[57] Of course, large retailers have also created their own data and analytics departments, or they acquire companies in these fields. The large British retailer Tesco, for example, outsourced its Clubcard and data activities into a subsidiary company, Dunnhumby, whose slogan is "transforming customer data into customer delight".[58] In 2014, Dunnhumby acquired the German online advertising technology firm Sociomantic. According to a press release, Dunnhumby will "combine its extensive insights on the shopping preferences of 400 million consumers" with Sociomantic's "real-time data from more than 700 million online consumers" to plan, personalize and evaluate advertising.[59] It is not clear which data about the shopping preferences of 400 million consumers they use, but it is safe to assume that data from Tesco's Clubcard program is included.

Types of data collected. Personal information collected by companies is generally grouped into a variety of types. "Volunteered" (or "declared") data denotes data explicitly shared by individuals; conversely, "observed" data is captured through recording their activities. While "actual" data refers to factual information about individuals, "inferred" (or "modeled") data results from drawing inferences about characteristics or predicted behaviors of consumers based on actual data. Furthermore, data companies often provide "segments" grouping consumers with shared characteristics or predicted behaviors, as well as "scores" indicating the likelihood that an individual exhibits certain characteristics or predicted behaviors. While "first-party data" signifies data collected by companies who have a direct relationship with consumers, "third-party data" is either processed on behalf of other companies or collected, acquired, purchased, or licensed from others.[60]

Online advertising per industry. When it comes to online advertising, which largely depends on tracking and profiling consumers, the share of spending on digital ads might be a good indicator of how connected different industries are to the advertising data ecosystem. In 2016, the retail industry had a share of about 20% of all digital ad spending in the US. Consumer product manufacturers, including consumer electronics, had a share of about 16%; the automotive, travel, telecom, and financial services industries each accounted for about 10%, and media and entertainment each about 5%.[61] This also means that without these industries buy-

[55] Oracle (2015): Oracle Data Cloud Data Directory

[56] http://www.nielsen.com/us/en/solutions/measurement/retail-measurement.html [3.5.2017]

[57] Cox, Emmett (2011): Retail Analytics: The Secret Weapon. Wiley, November 2011, p. 93

[58] https://www.dunnhumby.com/solutions/capabilities [28.04.2017]

[59] https://www.dunnhumby.com/dunnhumby-acquires-sociomantic-revolutionise-digital-advertising [28.04.2017]

[60] Christl and Spiekermann (2016), p. 84-85

[61] http://www.emarketer.com/Chart/US-Digital-Ad-Spending-Share-by-Industry-2016-of-total/190798

ing into the extensive data collected about consumers within the world of advertising technology, today's tracking and profiling networks would not exist.

3.2 Media organizations and digital publishers

Publishers and media organizations have long been selling lists of their subscribers to marketers as well as collecting information about their readers[62], viewers, and listeners through surveys and other forms of audience measurement.[63] Today, measurement takes the form of software embedded into websites and mobile apps that collect extensive demographic and behavioral data. Publishers share this data with advertisers and many other companies, and allow them to utilize, combine, and link it with digital profiles from other sources.[64]

While a lot has been said about the decline of newspapers[65] and their increasing dependence on platforms such as Facebook for traffic[66], not all digital publishers play the same role. From an online advertising perspective, a **publisher** is simply a "distribution system for online advertising", in which "the advertiser pays for messages sent by publishers to customers".[67] Of course this includes traditional major media conglomerates such as Walt Disney, Bertelsmann, Viacom, CBS, Grupo Globo, but also a large variety of other actors such as blogs, online communities, app developers, audio and video platforms, gaming vendors, and many other digital services providers.

Web and mobile tracking. It is widely unknown exactly which kinds of personal data digital publishers share and how third parties use this data. Several academic studies have tried to examine this. Although such investigations are limited to technical information accessible from outside the "black boxes", they can show clear evidence of some aspects of data sharing. A recent study examining one million different websites has, for example, found that there are more than 80,000 third parties that receive data about the visitors of these one million websites. Around 120 of these tracking services were found on more than 1% of websites.[68] Another study on 200,000 different users from Germany visiting roughly 21 million web pages showed that third-party trackers were present on 95% of the pages visited.[69] Similarly, most mobile

[62] Generally, see e.g. Donatello, Michael (1998): Audience research for newspapers. In: Blanchard, Margaret (eds): History of the Mass Media in the United States: An Encyclopedia, Fitzroy Dearborn Publishers, September 1998.

[63] Generally, see e.g.: Napoli, Philip M. (2011): Ratings and Audience Measurement. In: Nightingale, Virginia (eds): The Handbook of Media Audiences, April 2011, Wiley-Blackwell.

[64] See e.g. Forbes in Oracle's data directory (p. 81): http://www.oracle.com/us/solutions/cloud/data-directory-2810741.pdf [28.04.2017]

[65] E.g. https://www.nytimes.com/2008/10/28/business/media/28circ.html

[66] E.g. https://www.theguardian.com/media/2016/jun/15/facebooks-news-publishers-reuters-institute-for-the-study-of-journalism

[67] McConnell, Ted (2015): The Programmatic Primer. A Marketer's Guide to the Online Advertising Ecosystem, p. 17

[68] Narayanan, Arvind; Dillon Reisman (2017): The Princeton Web Transparency and Accountability Project. Pre-published book chapter. Available at: http://randomwalker.info/publications/webtap-chapter.pdf

[69] Yu, Zhonghao; Sam Macbeth, Konark Modi, and Josep M. Pujol (2016): Tracking the Trackers. In Proceedings of the 25th International Conference on World Wide Web (WWW '16). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 121-132. Available at: http://www2016.net/proceedings/proceedings/p121.pdf

apps share information about their users with other companies. A 2015 study of popular apps in Australia, Brazil, Germany, and the US found that between 85% and 95% of free apps, and even 60% of paid ones, connect to third parties that collect personal data.[70]

A prominent example of a digital publisher that sells data about its users is the streaming platform Spotify. Since 2016, it shares "insights on their users' mood, listening and playlist behavior, activity and location" with the data division of the advertising giant WPP, which now "has unique listening preferences and behaviors of Spotify's 100 million users".[71] Digital publishers selling data about their users are not restricted to website and mobile app providers. The marketing intelligence company SimilarWeb, for example, receives data from "hundreds of thousands of direct measurement sources from websites and apps"[72], but also from desktop software and browser extensions.[73]

The large **media conglomerates** are also deeply embedded in today's tracking and profiling ecosystems; moreover, they have often developed or acquired data and tracking capabilities themselves. For example, Time Inc. has acquired Adelphic, a major cross-device tracking and ad technology company,[74] as well as Viant, a company claiming to have "access to over 1.2 billion registered users".[75]

Of course, the **large online platforms**[76] are large publishers as well. According to a report by Zenith, Alphabet (Google) was the world's largest media company in 2016, when measured in terms of revenue that derives from business that support advertising; it was followed by Walt Disney, Comcast, 21st Century Fox, Facebook and Bertelsmann. The Chinese web search platform Baidu ranked as the 9th largest, Yahoo (now owned by Verizon) as the 15th largest, and Microsoft as the 17th largest media company.[77] With Comcast acquiring NBC Universal, Verizon acquiring both AOL and Yahoo, and AT&T most likely acquiring Time Warner, the large telecom and broadband companies in the US are also becoming giant publishers, creating a powerful portfolio of content, audience data, cross-device ad delivery infrastructure, as well as data analytics and targeting capabilities.[78]

[70] 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. Available at: https://www.researchgate.net/publication/282356703_A_measurement_study_of_tracking_in_paid_mobile_applications

[71] http://www.thedataalliance.com/blog/wpps-data-alliance-and-spotify-announce-global-data-partnership/ [08.05.2017]

[72] https://www.similarweb.com/ourdata [26.04.2017]

[73] https://www.similarweb.com/downloads/our-data-methodology.pdf [26.04.2017]

[74] https://adexchanger.com/ad-exchange-news/time-inc-acquire-adelphic-build-people-based-dsp/

[75] http://www.adelphic.com/2017/01/time-inc-s-viant-acquire-adelphic/ [25.04.2017]

[76] Evans, Peter C. and Annabelle Gawer (2016): The Rise of the Platform Enterprise. A Global Survey. The Emerging Platform Economy Series No. 1, The Center for Global Enterprise, January 2016

[77] http://www.zenithoptimedia.cz/en/zenith/news/detail/102-Digital%20gi

[78] http://www.cnbc.com/2016/10/24/how-atts-time-warner-deal-strengthens-its-position-in-advertising.html

3.3 Telecom companies and Internet Service Providers

Telecommunication companies and Internet Service Providers (ISPs) play an essential role in providing access to today's information and communication networks, from phone to mobile to broadband internet. They have deep-level control over the data streams of billions of consumers.[79] For this reason the privacy of telecommunications has historically been subject to special protection.[80] Nevertheless, phone companies have always utilized or shared data about their subscribers to some extent[81]. It has been suggested that even metadata pertaining to numbers called and the time and duration of calls was available for sale in the US.[82]

As telephone and cable television companies evolved to **mobile and broadband ISPs**, they began lobbying the US government for regulation that puts them on an equal playing field with the lightly regulated Silicon Valley big players so that they, too, can commercially exploit customer data such as web browsing and app history data.[83] Rather than being regulated by the Federal Communications Commission, which had reclassified broadband internet service as a utility and proposed strong privacy rules, they argued against the rules and for Federal Trade Commission oversight, on par with their Silicon Valley competitors. And indeed broadband providers, such as AT&T, Comcast, and Verizon, successfully achieved the repeal of the privacy rules that the Federal Communications Commission had introduced only months before.[84] Broadband internet access service providers in the US are now able to sell sensitive customer information to the highest bidder without even requiring customer consent.

In Europe, by contrast, telecommunications and broadband providers have long been subject to relatively strict privacy laws on several levels. But unlike the upcoming EU General Data Protection Regulation, which has already been adopted, the additional new ePrivacy regulation covering telecommunications privacy is at this stage only a proposal[85] and as such still subject to lobbying.[86] In this way, the European telecom lobby echoes its US counterpart, warning that

[79] Ammari, Nader; Gustaf Björksten; Peter Micek; Deji Olukotun (2015): The Rise of Mobile Tracking Headers: How Telcos Around the World Are Threatening Your Privacy. Access Now, August 2015. Available at: https://www.accessnow.org/cms/assets/uploads/archive/AIBT-Report.pdf

[80] See e.g. Rodríguez-Ruiz, Blanca (1997): Privacy in telecommunications: a European and an American approach. The Hague, Boston, Kluwer Law International, 1997.

[81] For example, they have shared phone numbers with third-party phone directory services (https://consumerist.com/2012/04/10/man-changes-number-to-avoid-telemarketers-only-ends-up-with-more-unwanted-calls), or account and payment histories with consumer reporting agencies (http://files.consumerfinance.gov/f/201604_cfpb_list-of-consumer-reporting-companies.pdf)

[82] See e.g. https://consumercal.org/about-cfc/cfc-education-foundation/your-cell-records-are-for-sale/ or http://www.dmnews.com/marketing-strategy/data-brokers-and-telephone-records/article/91887/

[83] See e.g. https://www.privateinternetaccess.com/blog/2017/03/telecom-lobby-fcc-web-browsing-app-usage-history-not-sensitive-information/

[84] http://www.politico.com/story/2017/03/broadband-data-victory-republicans-236760

[85] See e.g. https://euinternetpolicy.wordpress.com/2017/04/01/us-congress-repeal-of-fcc-broadband-privacy-rules-what-do-eu-laws-say-about-broadband-privacy/

[86] https://edri.org/massive-lobby-personal-communications-security-started/

a "restrictive and conservative policy approach would effectively exclude the European telecoms industry from the data economy".[87]

According to a market research company, the global market for **telecom data** was worth $24 billion in 2015, growing to $79 billion in 2020.[88] Already in 2015 telecommunication service providers all over the world aimed to profit from their customer's data. Mobile carriers in at least 10 countries, including AT&T, Bell Canada, Bharti Airtel, Cricket, Telefonica, Verizon and Vodafone, had been caught injecting special tracking mechanisms into their user's web surfing activities. Because they are embedded at the network level, users cannot block these "supercookies". Such tracking mechanisms allow for the creation of rich user data profiles and for their use in advertising and tracking.[89] The German software corporation SAP already "aggregates and analyzes billions of anonymized consumer data points from mobile operator networks" for its mobile data analytics tool "Consumer Insight 365".[90] According to a company representative, SAP provides "interconnection services to about 1,000 mobile carriers around the world, directly or indirectly" and plans to "partner with the carriers and share revenue from its sales of the data".[91] The marketing intelligence company SimilarWeb states that it receives "anonymous clickstream data"[92] from "local internet service providers (ISPs) located in many different countries".

Many large telecom companies and ISPs have acquired ad technology and data companies. For example, Millennial Media, a subsidiary of Verizon's AOL, is a mobile ad platform collecting data from more than 65,000 apps from different developers, and claims to have "access to approximately 1 billion global active unique users".[93] Telenor, a large telecommunications company based in Norway, has acquired Tapad, one of the leading cross-device tracking companies[94]. Tapad, which has data on 2 billion devices around the globe, is able to link desktop computers, laptops, smartphones and tablets belonging to the same person.[95] The Singapore-based telecom corporation Singtel acquired Turn,[96] an advertising and data platform that "pulls together data from all sources" and builds a "composite master ID of every consumer".[97] It gives marketers access to "4.3 billion addressable device and browser IDs"[98] and "90,000 demo-

[87] https://etno.eu/datas/press_corner/press-releases/ETNO%20GSMA%20Joint%20High-Level%20Letter%20on%20e-Privacy%2020.12.2016.pdf

[88] Kingstone, Sheryl; Rich Karpinski; Brian Partridge (2015): Monetizing 'Telecom Data as a Service'. Sep 2015, 451 Research. Executive overview available at: https://451research.com/report-long?icid=3534

[89] Ammari, Nader; Gustaf Björksten; Peter Micek; Deji Olukotun (2015), p. 2

[90] http://news.sap.com/mobile-data-analysis-sap-consumer-insight-365/ [26.04.2017]

[91] http://www.cio.com/article/2385645/big-data/sap-to-crunch-and-sell-carriers--data-on-mobile-use.html

[92] https://www.similarweb.com/downloads/our-data-methodology.pdf [26.04.2017]

[93] http://investors.millennialmedia.com/phoenix.zhtml?c=238412&p=irol-newsArticle&id=2085090

[94] https://www.tapad.com/device-graph/ [26.04.2017]

[95] https://www.tapad.com/press-release/magnetic-and-tapad-forge-strategic-alliance-to-advance-search-retargeting [26.04.2017]

[96] https://adexchanger.com/online-advertising/singtels-amobee-snaps-turn-310m/

[97] https://www.turn.com/platform/products [26.04.2017]

[98] https://www.turn.com/company [26.04.2017]

graphic, behavioral, and psychographic attributes".[99] In addition, large mobile network providers have already started to partner with companies that use their data for credit scoring. For example, Telefonica, Airtel (India) and Globe Telecom (Philippines) have partnerships with the US-based company Cignifi, which predicts the creditworthiness of phone users based on their phone call records. Cignifi sees itself as the "ultimate data monetization platform for mobile network operators".[100]

3.4 Devices and Internet of Things

In terms of devices, smartphones may be contributing the most to today's pervasive tracking and profiling ecosystems. The information recorded by mobile phones provides detailed insights into the user's personality and everyday life. Usually, users must have a Google, Apple, or Microsoft account to make their smartphones work. Additionally, most smartphone apps transmit data to third-party companies.[101]

In recent years, many other kinds of **devices with sensors and network connections** have entered the everyday lives of consumers, adding another dimension to the tracking ecosystem. These include everything from e-readers and wearables to smart TVs, meters, thermostats, smoke alarms, fridges, glasses, toothbrushes, and toys. Like smartphones, these devices give companies unprecedented access to consumer behavior across a wide variety of life contexts. The recorded data is either controlled by a single company, as with Amazon's Kindle e-reader, or the devices are strategically opened up to third-parties, which might provide apps or make other kinds of data contracts with consumers.[102] Additionally, service providers, or so-called "IoT platforms" that provide the infrastructure for analytics and device and data management, often sit between product vendors and consumers.[103]

Many experts and regulators see the **Internet of Things** as a major challenge for consumer privacy.[104] For example, as a recent study on privacy and car telematics shows, many different parties are interested in the data generated and recorded by the "connected car". Car data is attractive not only for automakers and their partners but also for car dealers, insurance companies, financial lenders, telematics service providers, call center operators, third-party app developers, vehicle infotainment content providers, mobile network operators, and mobile device system providers like Google or Apple. Third parties outside the telematics industry itself are also standing in line to access telematics data; as such, actors including local retailers and merchants, online advertising agencies, data brokers, law enforcement agencies, debt collec-

[99] https://www.turn.com/platform/products [26.04.2017]
[100] http://cignifi.com [26.04.2017]
[101] Christl and Spiekermann (2016), p. 46-51
[102] Ibid, p. 69-71
[103] https://www.accenture.com/nz-en/insight-iot-platforms-agile-innovation-scale
[104] Christl and Spiekermann (2016), p. 71-72

tors, fraud investigators, litigants, and many more can be added to the list of potential bidders for the data.[105]

3.5 Financial services and insurance

Banks, insurers, and other financial services companies play an essential role in the lives of consumers. They make important decisions that can have far-ranging consequences for consumers, including whether services are provided, insurance claims are accepted, loans are called in, or payments are authorized. They decide the terms under which financial services companies offer services, including loans and insurance policies. At the same time, they have access to very sensitive information about people.[106] For this reason, many countries have developed relatively strict financial privacy regulations in the past decades, especially regarding payments data.[107] Nevertheless, the financial services companies, their affiliates, as well as related their third parties have always been on the forefront of personal data collection, use, and analytics.

Banks and payment companies have extensive amounts of identity data on their customers – as well as rich transactional data on, among other things, their payments, account balances, or interactions with customer services.[108] **Financial institutions** use data about consumers for risk management, such as credit scoring and fraud detection, as well as for marketing, customer acquisition, and retention.[109] They supplement their own data with external data from consumer reporting agencies, data brokers and marketing data companies. Conversely, banks, lenders, credit card issuers, and other financial institutions share specific financial data about consumers with credit reporting agencies.[110] PayPal, the biggest name in online payments[111], shares personal information with more than 600 third parties including other payment providers, credit reporting agencies, identity verification and fraud detection companies, as well as with the most advanced players within the online tracking ecosystem.[112]

Banks and credit card issuers also run loyalty programs, including co-branded debit and credit cards that they issue in partnership with retailers or other companies. For example, Wells

[105] 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. Available at: https://fipa.bc.ca/wordpress/wp-content/uploads/2015/03/CC_report_lite.pdf

[106] European Banking Authority (2016): Discussion Paper on innovative uses of consumer data by financial institutions, EBA/DP/2016/01, 04 May 2016. Available at: https://www.eba.europa.eu/documents/10180/1455508/EBA-DP-2016-01+DP+on+innovative+uses+of+consumer+data+by+financial+institutions.pdf

[107] Jentzsch, Nicola (2003): The Regulation of Financial Privacy: The United States vs Europe. ECRI Research Reports No. 5, 1 June 2003. Available at: http://aei.pitt.edu/9430/2/9430.pdf

[108] European Banking Authority (2016)

[109] Hormozi, A. M. & Giles, S. (2004): Data Mining: A Competitive Weapon for Banking and Retail Industries.. IS Management, 21, 62-71. Available at: http://cjou.im.tku.edu.tw/dm200707/dm_application.pdf

[110] Rieke, Aaron; Harlan Yu; David Robinson; Joris von Hoboken (2016), p. 29

[111] https://www.nytimes.com/interactive/2016/04/07/business/dealbook/The-Fintech-Power-Grab.html

[112] According to a list, which is part of their German privacy policy: https://www.paypal.com/de/webapps/mpp/ua/third-parties-list?locale.x=en_US [20.04.2017]

Fargo, Citigroup, and Discover reportedly run programs through which they charge retailers a fee to use and access the financial services' payments data; in this way, the retailers can better sell their products, often in partnership with intermediary companies.[113] One of the intermediary companies that help accomplish this is Cardlytics, a firm that runs reward programs with 1,500 financial institutions[114] such as the Bank of America and MasterCard[115]. Cardlytics promises financial institutions that it will "generate new revenue streams using the power of [their] purchase data"[116] and claims to "see debit, credit, ACH and bill pay spend for tens of millions of individual consumers in the US and UK".[117] Cardlytics is also a partner of LiveRamp, the Acxiom subsidiary that links online and offline consumer data.[118]

Credit card networks have also started to make the data about their customers' purchases available to the online tracking and profiling universe. For example, Visa provides data on 14 billion purchase transactions to the data broker Oracle and combines it with demographic, financial, and other data[119] in order to help companies better categorize and target consumers in the digital world.[120] For MasterCard, selling products and services created from data analytics might even become its "core business" given that "information products, including sales of data" already represent a considerable and growing share of its revenue.[121] Google recently stated that it captures "approximately 70% of credit and debit card transactions in the United States" through "third-party partnerships" in order to track purchases, but did not disclose its sources.[122]

Insurance companies were amongst the first to use statistical models to predict the behavior of consumers based on their demographic attributes. Already in the late nineteenth century life insurers began trying to predict people's life spans and their relative risk of death.[123] Ever since, collecting data for individual risk assessments has always been a core element of their activities.[124] Depending on the legislation they operate under, insurers might only be allowed to use certain types of information for underwriting and for evaluating claims, or might only acquire it in certain ways. Being a heavily regulated industry, large insurers, as well as large

[113] http://money.cnn.com/2011/07/06/pf/banks_sell_shopping_data/
[114] http://www.cardlytics.com/about-us/our-story/ [24.04.2017]
[115] http://www.cardlytics.com/financial-institutions/fi-solutions/ [24.04.2017]
[116] http://www.cardlytics.com/ [24.04.2017]
[117] http://www.cardlytics.com/about-us/our-story/ [24.04.2017]
[118] https://liveramp.com/partner/cardlytics/ [24.04.2017]
[119] See p. 170: http://www.oracle.com/us/solutions/cloud/data-directory-2810741.pdf [24.04.2017]
[120] Visa emphasizes to only use "anonymized and aggregated spend data": https://usa.visa.com/run-your-business/commercial-solutions/solutions/advertising-loyalty-card-programs.html [24.04.2017]
[121] http://paymentweek.com/2014-6-16-for-mastercard-processing-and-analytics-go-hand-in-hand-4908/
[122] https://adwords.googleblog.com/2017/05/powering-ads-and-analytics-innovations.html
[123] Bouk, Dan (2015): How Our Days Became Numbered. Risk and the Rise of the Statistical Individual. University of Chicago Press.
[124] However, guaranteed-issue insurers such as most European health insurance programs only use it to calculate their risk internally, and not for underwriting, see e.g. Ridic, Goran; Suzanne Gleason; Ognjen Ridic (2012): Comparisons of Health Care Systems in the United States, Germany and Canada. Mater Sociomed. 2012; 24(2): 112–120. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3633404

banks, may be slower in employing “alternative” and more controversial digital information than native Internet companies such as social networks, online advertisers, and “Fintech” startups. Still, perhaps typical for a traditional insurer, the Boston Consulting Group sees the highest potential for the use of these new data in the areas of risk assessment and pricing, marketing and sales, claims prevention and mitigation, and fraud detection.[125]

Although controversial, insurance companies have used information about an individual’s credit worthiness for underwriting in the US since the 1990s.[126] In recent years, insurers all over the world have started to use digital data about everyday behaviors for risk assessment and pricing. Insurance companies already offer their customers programs involving real-time tracking of behavior such as e.g. driving behavior, health activities, grocery purchases, or visits to the fitness studio.[127] In this way insurers not only obtain the real-time behavioral data that eluded them in the past[128], but also are able to shape consumer behaviors and lower costs through sophisticated reward programs.

3.6 Public sector and key societal domains

On a global level, the relationship between today’s pervasive consumer data ecosystem and crucial societal areas such as healthcare, welfare, education, housing and employment varies a great deal, depending on how societies and regulatory environments are organized in a specific region or country. In the US, for example, the extent of commercial data collection and uses in the areas of **healthcare**[129], **education**[130], **housing**[131] and **employment**[132] is, from a privacy perspective, alarming. While consumer reports are used for eligibility decisions in many societal contexts in the US[133], this is not allowed in many European countries.[134] In addition, healthcare is, for example, widely run as a publicly funded service in Europe[135] which makes eligibility decisions based on behavioral data nonessential. In today’s work environments, vast

[125] Brat, Eric; Stephan Heydorn, Matthew Stover, and Martin Ziegler (2013): Big Data: The Next Big Thing for Insurers? Boston Consulting Group, March 25, 2013. Available at: https://www.bcgperspectives.com/content/articles/insurance_it_performance_big_data_next_big_thing_for_insurers

[126] FTC (2007): Credit-based Insurance Scores: Impacts on Consumers of Automobile Insurance. A Report to Congress by the Federal Trade Commission, July 2007. Available at: https://www.ftc.gov/reports/credit-based-insurance-scores-impacts-consumers-automobile-insurance-report-congress-federal

[127] For a detailed description of how insurance programs involving car telematics and wearables see Christl and Spiekermann (2016), p. 52-68

[128] Brat, Eric; Stephan Heydorn, Matthew Stover, and Martin Ziegler (2013)

[129] See e.g. https://www.theguardian.com/technology/2017/jan/10/medical-data-multibillion-dollar-business-report-warns

[130] See e.g. https://www.equalfuture.us/2015/09/23/clever-student-data-schools-tech-firms/

[131] See e.g. https://www.privacyrights.org/consumer-guides/renters-guide-tenant-privacy-rights

[132] See e.g. http://www.nbcnews.com/business/consumer/exclusive-your-employer-may-share-your-salary-equifax-might-sell-f1B8173066

[133] U.S. Consumer Financial Protection Bureau (2016): List of consumer reporting companies. Retrieved from: http://files.consumerfinance.gov/f/201604_cfpb_list-of-consumer-reporting-companies.pdf

[134] Ferretti, Federico (2015): Credit Bureaus Between Risk-Management, Creditworthiness Assessment and Prudential Supervision. EUI Department of Law Research Paper No. 2015/20, p. 13-16. Available at: https://ssrn.com/abstract=2610142

[135] Ridic, Goran; Suzanne Gleason; Ognjen Ridic (2012): Comparisons of Health Care Systems in the United States, Germany and Canada. Mater Sociomed. 2012; 24(2): 112–120. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3633404

amounts of information about **employees** are collected and analyzed all over the world.[136] As modern surveillance arguably has roots in Taylorist "scientific management"[137], the numerous interconnections between the tracking of customers and the tracking of employees are hardly surprising.[138] The use of data in **political election campaigns** has raised serious concerns, especially in the last year.[139] Notably, the **nonprofit and advocacy** sectors have also been collecting and acquiring marketing data for campaigning purposes for decades. Charity and donation records are prominently represented within the data inventory of consumer data brokers.[140] In the **national security and law enforcement** contexts, both excessive mass surveillance and targeted investigations depend on the extensive data collections by corporate players.[141]

A detailed assessment of the degree to which these different societal fields are already connected to today's networks of digital tracking and profiling lies beyond the scope of this report. However, some examples are further examined in the following chapters on the risk and marketing data industries.

3.7 Future developments?

It is difficult to predict the future of the business of personal data and rapidly changing roles of its main players. Nobody knows whether Facebook will turn its patent for assessing creditworthiness based on someone's friends[142] into reality. Few people could have predicted how Uber and Airbnb took over taxi services and the booking of travel accommodations a few years ago.[143] It is equally unclear how Google will change the future of driving or – with its 2016 acquisitions of eight insurance technology startups – insurance.[144] Google has also entered several other areas such as education. In the US, a significant percentage of primary- and secondary-school students already use Google's laptops and education apps.[145] All of the tech giants are actively shaping public policy and engage in lobbying lawmakers and regulators to advance their position.[146] The US telecom industry's recent successes with having customer privacy protections eliminated demonstrate the effectivity of these efforts. At the same time, different

[136] Kidwell, Roland E. and Sprague, Robert (2009): Electronic Surveillance in the Global Workplace: Laws, Ethics, Research and Practice. New Technology, Work and Employment, Vol. 24, Issue 2, pp. 194-208, July 2009. Available at SSRN: https://ssrn.com/abstract=1487801

[137] Sprague, Robert D. (2007): From Taylorism to the Omnipticon: Expanding Employee Surveillance Beyond the Workplace, 25 J. Marshall J. Computer & Info. L. 1, 2007. Available at: http://repository.jmls.edu/jitpl/vol25/iss1/1/

[138] For example, systems that collect data for both customer and workforce management in retail or call centers, see e.g. Christl and Spiekermann (2016), p. 32; 75

[139] See e.g. https://www.privacyinternational.org/node/1440

[140] See e.g. the lists offered in Infocore's data catalog: http://www.infocore.com/insights/data-catalogs.htm

[141] https://theintercept.com

[142] http://www.theatlantic.com/technology/archive/2015/09/facebooks-new-patent-anddigital-redlining/407287

[143] http://www.independent.co.uk/news/business/comment/hamish-mcrae/facebook-airbnb-uber-and-the-unstoppable-rise-of-the-content-non-generators-10227207.html

[144] https://www.equities.com/news/insuring-the-future

[145] https://www.nytimes.com/2017/05/13/technology/google-education-chromebooks-schools.html

[146] http://www.politico.com/story/2015/01/tech-lobby-apple-amazon-facebook-google-114468

industry sectors fight each other with legal means, such as when the data broker Oracle prodded regulatory examinations of Google's user profiling practices in both Europe and the US.[147]

Regardless, the trend is clear. Business consultants advise companies in all industries that they must join the personal data economy, that they should collect as much data as possible about their customers and prospects, and, of course, that they also should find ways to reap the benefits of these riches. In an article entitled "How to Monetize Your Customer Data", the consulting company Gartner suggests that companies "need to start treating information as a corporate asset that generates tangible future benefits".[148] Consumers, in contrast, have fewer and fewer options to resist the power of this data ecosystem. Often, they have no choice but to accept the services on offer.[149] Similarly, many companies that are not central players in this data driven world, especially smaller ones, have no choice other than to accommodate themselves with the new rules, structures, and standards set by the large corporate players. Companies in many industries are strategically acquiring other companies with access to data, customer relationships, and technology. While some commentators see finance "being taken over by tech"[150], banks are investing in "fintech" startups.[151] Many companies openly try to amass hundreds of millions of data contracts with consumers, from loyalty programs to "free" online services. Yet others aim to secretly track consumer behavior. Many website providers and mobile app developers actively allow third-party companies to collect and share information about consumers in utterly irresponsible ways.

Google and Facebook, the de facto "duopoly" in the digital world, dominate today's online advertising markets[152] through their extensive data assets about billions of consumers – and clearly have much further-reaching ambitions. Yet in spite of their dominance, some of the traditional industry players are in an excellent position to join the game on a large scale – or have done so. What started with newspaper subscriptions and loyalty programs in travel, retail, and finance has become part of the business model of many of today's companies: enticing large numbers of consumers into data contracts that include frequent interactions and access to extensive behavioral data.

The European Union's **General Data Protection Regulation (GDPR)**, which will become effective in 2018, will certainly have a considerable impact on personal data practices not only in Europe but also on a global level.[153] The US legal and regulatory framework, on the other hand, has enabled the growth of this data driven world without any effective consumer safeguards – and there is little in sight that will bring about a fundamental change for consumers. Some

[147] http://fortune.com/2017/01/25/google-super-profiles/
[148] http://www.gartner.com/smarterwithgartner/how-to-monetize-your-customer-data/
[149] Christl and Spiekermann (2016), p. 123
[150] https://www.ft.com/content/2f6f5ba4-dc97-11e6-86ac-f253db7791c6
[151] http://www.pymnts.com/news/b2b-payments/2016/kpmg-pulse-of-fintech-report-cb-insights-bank-corporate-venture-capital-vc-investment/
[152] http://adage.com/article/digital/verizon-chases-digital-duopoly-facebook-google/305258/
[153] See e.g. https://iapp.org/resources/article/top-10-operational-impacts-of-the-gdpr/

commentators warn that the political process and democracy in general have lost control over the tech industry.[154]

The next two chapters further investigate today's personal data ecosystem alongside the practices of data and analytics companies and their clients, both in the fields of marketing and risk, under particular consideration of the broader environment described above as well as from a consumer perspective.

[154] See e.g. https://medium.com/humane-tech/how-do-we-reform-tech-581d58ee11fd

4. The Risk Data Industry

4.1 Rating people in finance, insurance and employment

As Richard Cordray, the director of the US Consumer Financial Protection Bureau, has recently stated, consumer reporting agencies exert a "tremendous influence over the ways and means of people's financial lives" through the data they manage.[155]

A typical **credit report** contains information about a consumer's payment and debt history as provided by banks, lenders, collection agencies, and other institutions; this includes, for instance, the number and type of accounts, the dates they were opened, and information about bankruptcies, liens, and judgments.[156] Consumer reporting agencies in turn provide these reports to creditors and potential creditors, including credit card issuers, car loan lenders, mortgage lenders, but also to retailers, utility companies, mobile phone service providers, collection agencies, and any entities with a court order. Experian, the world's largest credit reporting agency,[157] and, next to Equifax and TransUnion, one of the three major agencies in the United States, has credit data on 918 million people.[158]

Several US consumer reporting agencies also provide reports for employers, landlords, insurance companies, government agencies, and other entities. For example, **tenant screening** reports contain information about someone's address and rent payment history, eviction, and criminal information. Reports from **employment screening** companies additionally contain information pertaining to someone's employment, salary, education and driving history, professional licenses, health screening, drug and alcohol testing information, participation in volunteer activities, and sometimes even their fingerprints. Reports for **personal property insurers** can contain previous coverage and losses, vehicle ownership, and driving history. Some companies even provide reports containing medical conditions, prescription drug purchase histories, as well as "hazardous avocations", to **life, long-term care, disability income, and health insurers**. Specialized companies collect information about telecom and utility payments, as well as product returns in the retail business.[159] Many of these agencies are organized as associations, exchanges, or "pools" wherein the participating companies can retrieve records about consumers in exchange for information they contribute themselves.[160]

[155] https://www.consumerfinance.gov/about-us/newsroom/prepared-remarks-cfpb-director-richard-cordray-consumer-advisory-board-meeting-march-2017/

[156] U.S. Consumer Financial Protection Bureau (2016): List of consumer reporting companies. Available at: http://files.consumerfinance.gov/f/201604_cfpb_list-of-consumer-reporting-companies.pdf

[157] https://www.wsj.com/articles/experian-fined-over-alleged-deception-in-credit-score-marketing-1490297908

[158] https://www.experianplc.com/media/2744/discover-experian-fy17.pdf [03.04.2017]

[159] US Consumer Financial Protection Bureau (2016)

[160] Hunt, Robert (2005): A century of consumer credit reporting in America, No 05-13, Working Papers, Federal Reserve Bank of Philadelphia, p. 17

In Europe, the situation is quite different. Many EU countries have public credit bureaus; Belgium and France operate only public credit registers. The types of data that credit reporting agencies may collect, and the purposes the records may be used for, also differ. While in the UK even non-credit data may be used, in France only "negative" information may be included as opposed to "positive" information about payments made on time. There are only a few EU countries – such as the UK, Germany, Sweden and Finland – in which credit reports may be used for purposes other than credit assessment.[161]

Credit reports themselves are rarely used to make decisions about consumers. Since the introduction of **credit scoring** decades ago, the raw information is typically fed into mathematical algorithms, transformed into a number or score, and then used to make predictions about consumers' possible future financial behavior. The use of simple numbers instead of long reports improved the efficiency of automated decision-making for the assessment of credit applications and determinations of creditworthiness; in this way, credit ratings and the thereby evaluated individuals became comparable. In fact, rather than merely being based off of information pertaining to an individual's characteristics and past behaviors, a credit score also judges a person in relation to other people.[162] It is a "way of recognising different groups in a population according to certain features expressed by a combination of personal financial and other non-personal data".[163]

Resident, driver and insurance scores. Data brokers also offer scores for purposes other than credit assessment. For example, TransUnion's "ResidentScore" promises to "predict negative residence outcomes" for the rental industry.[164] Its "Vehicle History Score" for insurance policy risk assessment does "not use consumer credit information", but is instead based on vehicle identification numbers (VIN)[165] and data from CARFAX, a company that aggregates vehicle history information from motor vehicle agencies, auto auctions, fire and police departments, collision repair facilities, fleet management, and rental agencies.[166] TransUnion's "DriverRisk" product combines personal driving records with vehicle records.[167] LexisNexis Risk Solutions provides insurance scores that are "designed for the relative rank ordering of applications and policyholders", including for auto, homeowners, and renters insurance.[168] The company's background checks and scoring solutions for both "pre-employment and post-hire information

[161] Ferretti, Federico (2015): Credit Bureaus Between Risk-Management, Creditworthiness Assessment and Prudential Supervision. EUI Department of Law Research Paper No. 2015/20, p. 13-16. Available at: https://ssrn.com/abstract=2610142

[162] Ferretti, F. (2009): The Credit Scoring Pandemic and the European Vaccine: Making Sense of EU Data Protection Legislation, 2009(1) Journal of Information, Law & Technology (JILT). Available at: http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2009_1/ferretti

[163] Ferretti, Federico (2014): EU Competition Law, the Consumer Interest and Data Protection. The Exchange of Consumer Information in the Retail Financial Sector. SpringerBriefs in Law, 2014, p. 17

[164] http://rentalscreening.transunion.com/solutions/residentscore [03.04.2017]

[165] See: https://en.wikipedia.org/wiki/Vehicle_identification_number

[166] https://www.carfax.com/company/about [03.04.2017]

[167] https://www.transunion.com/product/driverrisk [03.04.2017]

[168] http://www.lexisnexis.com/risk/products/insurance/attract.as [03.04.2017]

needs" include on-going criminal record reviews and "medical compliance monitoring".[169] LexisNexis' resident data services provide "continuous resident monitoring" and should help renters protect their property from "problem renters".[170]

Similar services are also available in some European countries. For example, the German company Arvato – a subsidiary of the Bertelsmann media conglomerate and large service provider in marketing, financial services and customer relationship management that has 70.000 employees in 40 countries – claims to perform 100 million credit checks per year. The company runs a pool containing information on consumers with negative payment behavior in the telecom area and also offers tenant screening.[171] In addition, Arvato runs a system in which all German insurers exchange information about consumers.[172]

Regulation and accuracy. In the US, the collection and use of data for credit reporting is subject to several domain-specific laws protecting some individual rights. These restrict the types of information consumer reporting agencies are allowed to use and provide "some guarantees of access and accuracy".[173] In the EU, by contrast, there are currently no harmonized domain-specific legal frameworks and the protection of the rights of individuals in the context of credit reporting largely relies on data protection legislation and diverse national laws.[174] Despite assessments of data brokerage in the credit context being "well regulated"[175], surveys both in the US and in Europe found evidence that credit reports and credit scores contain incorrect data. In 2012, the FTC conducted a study on credit report accuracy and found that 26% of survey participants had at least one error in one of three credit reports.[176] A German study on credit scoring, published by the Federal Ministry of Justice and Consumer Protection and the Federal Ministry of the Interior, found that scores are often based on estimations and their validity on an individual level should be questioned.[177]

The problem with credit scores is more complex. Although they may provide access to financial services for many, in spite of existing regulation, credit scores remain opaque and may produce arbitrary outcomes. Moreover, the pervasiveness of scores in our **scored society** can make them "self-fulfilling prophecies, creating the financial distress they claim merely to indicate"

[169] https://www.lexisnexis.com/government/solutions/literature/screening.pdf [03.04.2017]
[170] Ibid.
[171] Christl and Spiekermann (2016), p. 104
[172] https://www.arvato.com/finance/de/industries/versicherungen.html [03.04.2017]
[173] Rieke, Aaron; Harlan Yu; David Robinson; Joris von Hoboken (2016), p. 33
[174] Ferretti, Federico (2014): The Legal Framework of Consumer Credit Bureaus and Credit Scoring in the European Union: Pitfalls and Challenges—Overindebtedness, Responsible Lending, Market Integration, and Fundamental Rights. Suffolk University Law Review, Vol. XLVI:791, p. 807. Available at: http://suffolklawreview.org/wp-content/uploads/2014/01/Ferretti_Lead.pdf
[175] Rieke, Aaron; Harlan Yu; David Robinson; Joris von Hoboken (2016), p. 33
[176] Christl and Spiekermann (2016), p. 85
[177] Ibid.

and might systematize discriminatory practices.[178] The European consumer-finance scholar Federico Ferretti points to the fact that while credit scoring was originally intended to "minimise the percentage of consumers who default", lenders now use them to "identify the customers who are most profitable and to maximise profits through risk based pricing", while "blurring this all with direct marketing activities".[179] He argues that credit scoring would emphasize the minimization of business risk and increased profitability – which is certainly a legitimate business interest, but should not disproportionally limit the rights of individuals.[180] In addition, he fundamentally questions the "capability of credit data to prevent future over-indebtedness", because the use of credit data could not foresee many major causes of over-indebtedness, such as illnesses, divorce, job losses, and poor market conditions.[181]

4.2 Credit scoring based on digital behavioral data

Credit reporting agencies and financial technology companies are rapidly expanding the types of data they use to predict individual creditworthiness far beyond traditional data elements. This use of non-credit related behavioral data for credit scoring further undermines the validity of these scores. TransUnion, for instance, includes "alternative data" such as address stability and "club and subscription activity".[182] Smaller companies already use data from mobile phone, social media, and web usage to predict people's credit risk, and have started to partner with larger companies in consumer reporting, telecommunications, and e-commerce.

Cignifi, a US based company with clients around the world, analyzes patterns from phone call records such as call durations and times, who a call or text message was initiated by, and the numbers frequently called, to calculate credit risk scores for mobile phone users.[183] The company partners with large mobile network providers such as Telefonica[184] and has announced a "multi-year" partnership with the large credit reporting agency Equifax. Together with Cignifi, Equifax will provide credit scores to banks, retailers, and insurers in Latin America who can now underwrite individuals who had no prior formal payment history.[185]

Singapore-based **Lenddo** for instance calculates credit scores based on online behavior, including mobile data, browser data, application data, transactional data from telecom companies, as well as data from web publishers and social networks, including information from someone's

[178] 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. Available at: http://ssrn.com/abstract=2376209

[179] Ferretti, Federico (2015): Credit Bureaus Between Risk-Management, Creditworthiness Assessment and Prudential Supervision. EUI Department of Law Research Paper No. 2015/20. Available at: https://ssrn.com/abstract=2610142

[180] Ferretti, Federico (2009)

[181] Ferretti, Federico (2015)

[182] https://www.transunion.com/resources/transunion/doc/products/resources/product-creditvision-link-as.pdf [03.04.2017]

[183] Christl and Spiekermann (2016), p. 30

[184] Ibid.

[185] https://venturebeat.com/2016/03/30/cignifi-and-equifax-partner-to-bring-next-generation-credit-scores-to-unbanked-population-in-latin-america/

online friends. The company even includes computer mouse click data and data about how people fill out web forms[186]. In this vein, “always running out of battery” might impact one’s credit score; conversely, an extremely well-maintained smartphone might raise a red flag in the system, too.[187] Lenddo operates in 20 countries such as India, South Korea, Mexico and Philippines. According to their chairman, they help dozens of banks analyze data from millions of smartphones globally.[188]

The US company **ZestFinance** predicts consumer creditworthiness based on data from “thousands of sources”, including customer support data, purchase transactions, and how a customer fills out a form or navigates on a lender’s website.[189] As part of partnerships with China’s leading e-commerce and web search companies such as jd.com and Baidu, ZestFinance provides credit scoring based on data about online shopping habits and web searches.[190]

Most of these companies aggressively advertise their products in the name of financial inclusion and in support of the “underbanked”, i.e. people with little or no credit history – both in the world’s richest societies and in the global south. In Kenya, the Commercial Bank of Africa and the mobile network operator Safaricom offer the mobile banking product **M-Shwari**, which is connected to M-PESA, a mobile money service used by two-thirds of Kenyan adults. Their credit scoring algorithm, which is used to assess new applicants and to assign individual credit limits, is based on telecommunications data from Safaricom, including airtime, M-PESA data, and the length of the customer relationship history.[191] Another company active in both Kenya and Tanzania uses GPS data, call logs, and social network data for credit scoring, including the “grammar or punctuation” of text messages.[192]

Financial inclusion is certainly a crucial challenge for billions of people in the world. However, the use of personal information about activities unrelated to financial behavior, such as social relationships and people’s most private, everyday moments, constitutes a massive invasion of privacy. Moreover, the use of opaque, automated, predictive, and classifying algorithms for this purpose leaves open many questions about accuracy, fairness, equity, transparency, explainability and accountability, as well as about the informational power imbalance between people and the businesses that purport to serve them.

[186] Christl and Spiekermann (2016), p. 29

[187] http://money.cnn.com/2016/08/24/technology/lenddo-smartphone-battery-loan/

[188] Christl and Spiekermann (2016), p. 29

[189] http://www.businesswire.com/news/home/20170214005357/en/ZestFinance-Introduces-Machine-Learning-Platform-Underwrite-Millennials

[190] Christl and Spiekermann (2016), p. 28

[191] Cook, Tamara and Claudia McKay (2015): How M-Shwari Works: The Story So Far. Access to Finance FORUM Reports by CGAP and Its Partners. No. 10, April 2015. Available at: https://www.cgap.org/sites/default/files/Forum-How-M-Shwari-Works-Apr-2015.pdf

[192] https://www.devex.com/news/how-alternative-credit-scoring-is-transforming-lending-in-the-developing-world-88487

4.3 Identity verification and fraud prevention

With the rise of technology-mediated services, verifying consumer identities and preventing fraud have both become increasingly important and challenging issues, especially in the light of cybercrime and automated fraud.[193] At the same time, today's data-driven identity and fraud analysis systems have aggregated giant databases with sensitive information about entire populations, ranging from names, addresses, and relationships between people to extensive behavioral and biometric data.[194] Many of these systems are operated by private companies and cover a wide range of use cases, including proof of identity for financial services, assessing insurance and benefits claims, analyzing payment transactions, and evaluating a large variety of online transactions. They decide whether an application or transaction is accepted or rejected, or which payment and shipping options are available for someone during an online transaction. Commercial services for identity verification and fraud analytics are also used in areas such as law enforcement and national security. The line between commercial applications of identity and fraud analytics and those used by government intelligence services is increasingly blurring.[195]

When people are singled out by such opaque systems, they might get flagged as suspicious and warranting special treatment or investigation – or they may be rejected without explanation. They might get an email, a phone call, a notification, an error message – or, the system may simply withhold an option, without the user ever knowing of its existence for others. Inaccurate assessments may spread from one system to another. It is often difficult, if not impossible, to object to such negative assessments that exclude or deny,[196] especially because of how hard it is to object to mechanisms or decisions that someone does not know about at all. Of course, nobody wants fraud. However, the more digital technologies determine lives, the more important it becomes that those systems also become more transparent and accountable, and that people can object to arbitrary decisions. Most importantly, companies should not be allowed to use the vast amounts of sensitive data that they collect for fraud detection and security purposes for other purposes such as marketing.

Most large consumer and credit reporting agencies also offer identity verification, fraud detection, and other risk analytics services. For example, Experian's "Precise ID Platform" provides consumer verification and identity screening, based on data about auto registrations, property ownership, cellphones, credit records[197], including social security numbers, name

[193] EU Fraud Prevention Expert Group (2007): Report on Identity Theft/Fraud. Brussels, 22 October 2007. Available at: http://ec.europa.eu/internal_market/fpeg/docs/id-theft-report_en.pdf

[194] Christl and Spiekermann (2016), p. 106

[195] Ibid., p. 38-40

[196] E.g. https://www.theguardian.com/money/2016/mar/18/banned-by-amazon-returning-faulty-goods-blocked-credit-balance

[197] http://www.experian.com/decision-analytics/identity-and-fraud/identity-authentication.html [23.04.2017]

variants, and current and previous addresses[198]. In addition, the company uses fraud records[199] from its cross-industry "National Fraud Database", which contains addresses, phone numbers, social security numbers, and driving licenses contributed to by companies in banking, telecommunications, and retail.[200] Equifax, which claims to have data on 820 million consumers[201], also offers several identity verification[202] and fraud detection[203] products using credit, employment, utilities, and telecommunication account data, as well as information about household and familial relationships.[204] Acxiom, which is mainly in the marketing data business, provides risk mitigation products, too. Its identity verification services are based on data records about individual demographics, personal assets, education, licenses, filings, business, bankruptcies, judgments, liens, and other "risk flags".[205] The telecom giant Verizon also offers identity verification services. For example, Verizon has been one of several identity providers within "GOV.UK Verify", a system that allows citizens to access digital services run by the UK government.[206]

LexisNexis Risk Solutions, a data company which works for all of the 50 US largest banks and for most local government authorities and federal agencies in the US[207], claims to have data on 500 million consumer identities.[208] Through its "TrueID" system, the company offers the ability to verify identities against a database of 34 billion records from 10,000 sources. Identity documents can be checked against a database of 4,100 ID types from 200 countries. Identities can be linked to payment cards, checks, loyalty cards, and other customer data. The system is also suitable for age verification. In addition, biometric fingerprints can be integrated.[209] Moreover, Lexis Nexis provides biometric services for voice recognition using "the sound, pattern and rhythm of an individual's voice".[210] With its "LexID" the company uses its own unique identifier to store and link records about consumers. As these identity profiles are continuously updated, it claims that they paint an "extremely accurate picture that accounts for identity changes over time".[211] In addition, LexisNexis' government division provides fraud prevention systems for many application areas of government services, including food stamps,

[198] https://www.experian.com/credit_report_basics/pdf/samplecreditreport.pdf [23.04.2017]
[199] http://www.experian.com/decision-analytics/identity-and-fraud/identity-authentication.html [23.04.2017]
[200] http://www.experian.com/decision-analytics/national-fraud-database.html [23.04.2017]
[201] https://www.equifax.com/about-equifax/company-profile [23.04.2017]
[202] http://www.equifax.com/assets/IFS/efx-952-adv-acctVerify.pdf [23.04.2017]
[203] http://www.equifax.com/assets/IFS/Fraud/efx-usa-2038_suspicious_id_ps.pdf [23.04.2017]
[204] http://www.equifax.com/assets/IFS/efx_identity_proofing.pdf [23.04.2017]
[205] https://www.acxiom.com/what-we-do/identity-verification-authentication/ [23.04.2017]
[206] 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, p. 63. Available at: http://dcentproject.eu/wpcontent/uploads/2015/10/research_on_digital_identity_ecosystems.pdf
[207] http://www.lexisnexis.com/risk/about/default.aspx [23.04.2017]
[208] http://www.lexisnexis.com/risk/identity/verification.aspx [23.04.2017]
[209] http://www.lexisnexis.com/risk/downloads/literature/trueid.pdf [23.04.2017]
[210] http://www.lexisnexis.com/risk/products/voice-biometrics.aspx [23.04.2017]
[211] http://www.lexisnexis.com/risk/about/lexid.aspx [23.04.2017]

healthcare, retirement, public housing, student loans, tax refunds, and unemployment insurance.[212]

Biometrics, or the use of "any automatically measurable, robust and distinctive physical characteristic or personal trait" of a person to "identify an individual or verify the claimed identity of an individual"[213], is now used in many governmental and commercial fields. Credit card networks such as MasterCard, started to use fingerprints or facial recognition to verify the cardholder's identities for their payment solutions.[214] In the UK, a private company has deployed a facial recognition system that is connected to the CCTVs of more than 10,000 participating retailers, and manages a blacklist of suspicious faces.[215] Facebook states that its facial recognition technology is able to identify people with 83% accuracy, even if their faces are partly covered on a photo.[216] Generally, biometrics can be based on many physical and behavioral traits or characteristics, such as fingerprints, shapes of hands or faces, iris and DNA composition, as well as how someone talks, walks, or types on a keyboard. Biometrics is used for physical access control, border control, and law enforcement, as well as in healthcare, finance, and marketing.[217] In 2016, a report[218] found that a facial recognition database used by the FBI contained photos of half of all adults in the US, without their knowledge or consent. Notably, the detection algorithm was wrong nearly 15% of time and more likely to misidentify African Americans.[219]

4.4 Online identity and fraud scoring in real-time

In addition to traditional identity verification and fraud prevention technologies, a large industry focusing on the digital world has emerged in recent years, evaluating billions of online transactions every day, and sometimes linking their vast amounts of digital information with offline identity data.

Trustev, an online identity verification and fraud prevention company based in Ireland, which was acquired by TransUnion in 2015, evaluates digital transactions for clients in financial services, government, healthcare, and insurance in real-time[220] by analyzing digital behaviors, identities, and devices such as phones, tablets, laptops, game consoles, TVs, and even refrigera-

[212] http://blogs.lexisnexis.com/identitygov/identity-challenges/ [23.04.2017]

[213] Woodward, J. D. , N. M. Orlans, P.T. Higgins (2003): Biometrics: Identity Assurance in the Information Age, McGraw, Hill Osborne Media, 2003

[214] http://newsroom.mastercard.com/eu/press-releases/mastercard-makes-fingerprint-and-selfie-payment-technology-a-reality/ [23.04.2017]

[215] http://boingboing.net/2015/12/17/uks-unaccountable-crowdsourc.html

[216] https://www.newscientist.com/article/dn27761-facebook-can-recognise-you-in-photos-even-if-youre-not-looking

[217] See e.g. http://findbiometrics.com/applications/

[218] Garvie, Clare; Alvaro Bedoya; Jonathan Frankle (2016): The Perpetual Line-Up. Unregulated Police Face Recognition in America. Georgetown Law Center on Privacy & Technology, October 18, 2016. Available at: https://www.perpetuallineup.org/

[219] https://www.theguardian.com/technology/2017/mar/27/us-facial-recognition-database-fbi-drivers-licenses-passports

[220] http://newsroom.transunion.com/transunion-expands-fraud-and-identity-management-solutions-with-acquisition-of-trustev [23.04.2017]

tors.[221] The company offers corporate clients the ability to analyze how visitors click and interact with websites and apps, and uses a wide range of data to assess users, including phone numbers, email addresses, postal addresses, browser and device fingerprints, credit checks, transaction histories across merchants, IP addresses, mobile carrier details, and cell location.[222] To help "approve future transactions" every device is assigned a "unique device fingerprint".[223] During an online transaction, all data gets combined and compiled into a score of 1-100, which is then used to allow, deny, or flag it. In addition, Trustev offers a "social fingerprinting" technology, which includes "friend list analysis" and "pattern identification" analyzing social media content. The latter, at least, is only used with the "full permission" of individuals through a "voluntary social network login".[224]

The credit reporting agency TransUnion, which states that it obtains data from 90,000 sources on one billion consumers globally,[225] has integrated Trustev technology into its own identity and fraud solutions.[226] Its "IDVision" product combines information about personal identity (e.g. demographics, licenses, accounts) and reputation (e.g. shared fraud and watch lists) with data about digital identity (e.g. location, mobile, devices) and transactions (e.g. behaviors, scoring history, inquiries).[227] TransUnion also runs a "fraud prevention exchange" which utilizes real-time transaction data contributed by industry clients.[228]

Other credit reporting agencies have developed similar capabilities by connecting online fraud detection with their extensive consumer identity databases. Equifax, for instance, states that it has data on "nearly 1 billion devices" and can validate "where a device really is and whether it is associated with other devices used in known fraud". By combining this data with "billions of identity and credit events to find suspicious activity" across industries, and with information about employment and about relationships between households, families and associates, Equifax claims to be able to "identify devices as well as individuals".[229] In Germany, Arvato's profile tracking product similarly promises to have the "ability to recognise individual internet access devices on the basis of their fingerprint" for fraud detection.[230] ID Analytics, a US-based credit and fraud risk data company recently acquired by Symantec, runs an "ID Network" with "100 million identity elements coming in each day from leading cross-industry

[221] Trustev Sales Pack. REAL TIME ONLINE IDENTITY VERIFICATION. PDF Brochure. Personal copy of file with author Wolfie Christl

[222] http://www.trustev.com/how-it-works [23.04.2017]

[223] Trustev Sales Pack.

[224] Trustev Sales Pack.

[225] http://www.annualreports.com/HostedData/AnnualReports/PDF/NYSE_TRU_2016.pdf [23.04.2017]

[226] http://newsroom.transunion.com/transunion-expands-fraud-and-identity-management-solutions-with-acquisition-of-trustev [23.04.2017]

[227] https://www.transunion.com/idvision [23.04.2017]

[228] http://transunioninsights.com/fraud-exchange/ [23.04.2017]

[229] http://www.equifax.com/assets/IFS/efx_identity_proofing.pdf [23.04.2017]

[230] https://www.arvato.com/content/dam/arvato/documents/financial-solutions/Onepager_Profile_Tracking_ohne_RSS_EN.pdf [23.04.2017]

organizations"[231], containing data about 300 million consumers.[232] They aggregate data about consumer behaviors, ranging from "credit card and wireless phone applications, to changes in checking account information, to sub-prime loans and eCommerce transactions".[233] 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.[234] The company's "ID Score" evaluates information about digital devices, as well as personal data such as names, social security numbers, postal and email addresses.[235] ID Analytics' partners include CoreLogic, TransUnion, Visa, and ThreatMetrix.

The cyber security company **ThreatMetrix** runs a "Digital Identity Network" which "captures millions of daily consumer transactions including logins, payments and new account originations" and maps the "ever-changing associations between people and their devices, locations, account credentials, and behavior" in order to prevent online fraud, verify and authenticate digital identities.[236] The company emphasizes that it incorporates only "anonymized, non-regulated personal information such as user name [and] email address"[237]. Similarly, data passed from online businesses is "anonymised and encrypted".[238] However, its software allows companies to "understand the intricate connections between individuals, devices, behaviors, locations and any present threats" and to "discover hidden relationships between activities".[239] According to an industry report, ThreatMetrix processes more than two billion transactions per month involving 1.4 billion "unique user accounts" across "thousands of global websites".[240] The company's partners include Equifax, Lexis Nexis, TransUnion, and other financial services companies[241]. It helps Visa "bridge the gap between static identity assessment data and online identities".[242] Other clients include Netflix[243] and companies in fields as diverse as gaming, government services, and healthcare.

Device intelligence, credit reporting and marketing. Similarly, the large credit reporting agency Experian provides "device intelligence solutions" that allow online services to "invisibly protect customers applying online" "without stopping their online activity".[244] One product "establishes a reliable ID for the device and collects rich device data," "identifies every device

[231] http://www.idanalytics.com/data-and-technology/idnetwork/ [23.04.2017]
[232] http://www.idanalytics.com/media/VA-Resolve360-Datasheet.pdf [23.04.2017]
[233] http://www.idanalytics.com/data-and-technology/ [23.04.2017]
[234] http://www.idanalytics.com/media/Exploring-the-Impact-of-SSN-Randomization.pdf [23.04.2017]
[235] http://www.idanalytics.com/media/Fraud-ID-Score-9.0-Datasheet.pdf [23.04.2017]
[236] https://www.threatmetrix.com/cyber-security-software/ [24.04.2017]
[237] https://www.threatmetrix.com/cyber-crime/solution-brief/global-shared-intelligence/ [24.04.2017]
[238] https://www.threatmetrix.com/digital-identity-blog/travel/fighting-fraud-data-science-innovations-digital-identity/ [24.04.2017]
[239] https://www.threatmetrix.com/cyber-security-software/decision-management/ [24.04.2017]
[240] The Paypers (2016): Web Fraud Prevention & Online Authentication Market Guide 2016/2017.
[241] https://www.threatmetrix.com/partners/ [24.04.2017]
[242] https://www.threatmetrix.com/transaction-fraud/case-study/threatmetrix-supporting-visa-consumer-authentication-service/ [24.04.2017]
[243] https://www.threatmetrix.com/transaction-fraud/case-study/netflix-prevents-new-account-fraud/ [24.04.2017]
[244] http://www.experian.co.uk/identity-and-fraud/device-intelligence.html [24.04.2017]

on every visit in milliseconds” and “gives unparalleled visibility into the person behind the payment” for online fraud detection purposes.[245] According to Experian, its product protects “7+ billion transactions”[246] a year and is used by “120 of the world’s most recognized online merchants”, including 5 of the 6 “tier one global banks” and 4 of the top 5 global airlines.[247] Experian states that their fraud management solution for account opening “combines behavioral and device data, combined with user information”, including “blacklisted data, inconsistent locale indicators, risky locales and physical addresses, links to known fraud, and device intelligence attributes”. It “brings together information from multiple data sources” and consumers can also be screened “against bureau data and past application history”.[248]

Experian’s device identification technology comes from 41st parameter, a web fraud detection provider that the company acquired in 2013.[249] 41st parameter has also offered a second product marketed under the brand “AdTruth”, which provides "universal device recognition for effective digital advertising".[250] AdTruth promises to “identify users” across “both desktop and mobile web and apps”.[251] According to a product sheet, the company created a “unique user ID” for each device, by collecting information such as IP addresses, device models and device settings.[252] The service is now marketed as “AdTruth by Experian”. It provides “universal device recognition” and promises to “recognise devices whether in mobile, web or apps while upholding consumer privacy and choice”.[253] In 2015, Experian announced “AdTruth Resolve”, which is able to “reconcile and 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”.[254]

I am not a robot. Although most people may not be aware of it, Google’s **reCaptcha** product provides similar functionality. It is embedded into millions of websites and helps website providers decide whether a visitor is a legitimate human being or not. Until recently, users had to solve several kinds of quick challenges such as deciphering letters on a picture, choosing objects in a grid of pictures, or simply clicking on a “I’m not a robot” checkbox.[255] In 2017, Google introduced the “invisible” version of reCaptcha, explaining that from now on “human users will be let through” without any user interaction, while “suspicious ones and bots” would still

[245] http://www.experian.co.uk/assets/identity-and-fraud/device-insight-for-payments-one-pager.pdf [24.04.2017]
[246] Ibid.
[247] http://www.experian.com/decision-analytics/41st-parameter.html [24.04.2017]
[248] http://www.experian.com/decision-analytics/identity-and-fraud/account-opening.html [24.04.2017]
[249] https://adexchanger.com/data-exchanges/experian-buys-device-id-firm-41st-parameter-for-324m-gets-adtruth-in-the-bargain/ [24.04.2017]
[250] https://web-beta.archive.org/web/20140208125208/http://adtruth.com/what-we-do/differentiators
[251] https://web-beta.archive.org/web/20140208070551/http://adtruth.com/what-we-do/what-is-adtruth
[252] https://struqmarket.files.wordpress.com/2013/11/ad-truth-fact-sheet-uk.pdf [24.04.2017]
[253] http://www.experian.co.uk/marketing-services/products/adtruth-device-recognition.html [24.04.2017]
[254] https://www.experianplc.com/media/news/2015/adtruth-resolve/ [24.04.2017]
[255] https://arstechnica.com/gadgets/2017/03/googles-recaptcha-announces-invisible-background-captchas/

have to solve a challenge or click on a checkbox.[256] Google states that it uses a "combination of machine learning and advanced risk analysis" but does not disclose which kinds of user data and behaviors they use to "identify humans".[257] Critical investigations by journalists, and ironically also by Experian's AdTruth, suggest that Google does not only use IP addresses, browser fingerprints, the way users type, move their mouse, or use their touchscreen "before, during, and after" a reCaptcha interaction, but also several cookies set by Google's services.[258] It is not clear whether people without Google user accounts face a disadvantage, whether Google is able to identify specific individuals rather than only "humans", or whether Google also uses the data recorded within reCaptcha for other purposes than for bot detection. Either way, Google decides which individuals are considered valid and which are classified as suspicious or fraudulent on millions of websites.

4.5 Investigating consumers based on digital records

Other fraud prevention systems are not primarily optimized for real-time automatic detection of suspicious behavior, but rather help investigators of banks, insurers, law enforcement agencies, or other organizations analyze large amounts of digital records.

For example, as a part of its i2 platform, **IBM** provides software which helps clients analyze a wide range of data including "telephone call records, financial transactions, computer IP logs and mobile forensics data" in order to "identify, predict, prevent and disrupt criminal, terrorist and fraudulent activities". Their software is used by intelligence agencies, police departments, prisons, as well as by insurance companies to "Identify key people, events, connections and patterns" and to "highlight key individuals and relationships and their connections to key events".[259] Another IBM analytics product promises governmental agencies that it helps prevent "social program waste and abuse" by better combining eligibility and identity information, including insights into "familial relationships".[260]

Sentinel Visualizer, a software tool to "discover hidden relationships, connections, and patterns among people, places, and events" in all kinds of data[261], including "massive number of phone calls"[262], is used by both intelligence and law enforcement agencies. At the same time, large banks, insurance companies, and healthcare organizations, including Capital One and CIGNA, use the product for customer relationship mining and fraud detection.[263] Similarly, the data mining services provided by the highly controversial[264] analytics company **Palantir** are

[256] https://www.google.com/recaptcha/intro/invisible.html [24.04.2017]
[257] Ibid.
[258] http://www.businessinsider.com/google-no-captcha-adtruth-privacy-research-2015-2
[259] http://www.ibm.com/software/industry/i2software [24.04.2017]
[260] https://www.ibm.com/analytics/us/en/industry/government/social-programs/ [24.04.2017]
[261] http://www.fmsasg.com/ [24.04.2017]
[262] http://www.fmsasg.com/LinkAnalysis/telephone_logs/call_data_records.htm [24.04.2017]
[263] http://www.fmsasg.com/LinkAnalysis/Commercial/Solutions.asp [24.04.2017]
[264] Christl and Spiekermann (2016), p. 108

used by intelligence agencies, as well as by banks, insurers and healthcare organizations. Palantir's insurance analytics product, for example, enables clients to "integrate and analyze health data to identify fraudulent schemes, improve patient outcomes, and identify meaningful trends".[265]

Together with **Social Intelligence**, a company that provides technology for social media searches and insurance claim investigations, LexisNexis Risk Solutions offers a product to "discover an individual's online presence in real-time". This so-called "Social Media Locator" provides "up-to-the-minute intelligence on individuals" based on "searches of hundreds of social networks and millions of websites" that leave "virtually no stone unturned when it comes to identifying fraudulent claims". Social Intelligence promotes its capabilities to "gather, aggregate, analyze and distribute social data to improve fraud detection". The company aims to make "social data available and actionable for decision-making processes in every business environment" and also provides "social media risk scoring to improve risk assessment".[266] Many **business intelligence and analytics** tools including data mining, predictive analytics, data management and integration, provide similar functionalities as IBM, Sentinel, and Palantir's software. These tools often integrate many different services from risk mitigation to marketing. The market leaders in business intelligence and analytics solutions are Oracle, SAP, IBM, Microsoft, SAS, Teradata, Salesforce, Adobe, Tableau Software, and Informatica.[267]

[265] https://www.palantir.com/solutions/ [24.04.2017]

[266] http://insurancenewsnet.com/press-releases/social-intelligence-and-lexisnexis-risk-solutions-integrate-social-data-with-fraud-prevention-platform

[267] IDC (2016): Worldwide Business Analytics Software Market Shares, 2015: Healthy Demand Despite Currency Exchange Rate Headwinds. July 2016, IDC.

5. The Marketing Data Industry

5.1 Sorting and ranking consumers for marketing

The marketing data industry is arguably the main driving force behind ubiquitous consumer surveillance. It consists of a wide range of different types of companies, including marketing and advertising agencies, list brokers, database management providers, online and mobile advertisers, and firms engaged in direct mail, telephone sales services, and data-driven commerce, as well as companies offering loyalty programs.[268] Marketing data companies, which are often called "data brokers", mostly offer a smaller or bigger selection of these services.

Client firms pay for the use of the data brokers' data, marketing, analytics, and technology services to find, attract, and target valuable new customers, but also to retain existing customers and prospects and to maximize the profitability or "lifetime value" that they derive from them.[269] They do so by utilizing data and analytics to sort and rank both customers and prospects for prioritization purposes and personalized treatment, such as customized telemarketing scripts, promotional materials, online content, ads, offers, discounts and pricing.[270] Data broker clients can also buy additional consumer data and append it to their own data.[271] Based on this data and analytics services, businesses in all industries try to sell their customers complementary or more expensive products, to prevent customer attrition, and, generally, to persuade consumers to act in certain ways.[272]

A data broker can be defined as a "company or business unit that earns its primary revenue by supplying data or inferences about people gathered mainly from sources other than the data subjects themselves".[273] According to the US Federal Trade Commission (FTC), data brokers collect "massive amounts of data, from online and offline sources, and combine them into profiles about each of us".[274] While "multiple layers of data brokers" are "providing data to each other", consumers are "largely unaware" of their activities.[275] Some of them store data indefi-

[268] Deighton, John; Peter A. Johnson (2013): The Value of Data: Consequences for Insight, Innovation & Efficiency in the U.S. Economy. October 8, 2013. Available at: https://www.ipc.be/~/media/documents/public/markets/the-value-of-data-consequences-for-insight-innovation-and-efficiency-in-the-us-economy.pdf

[269] In marketing, customer lifetime value (CLV or often CLTV), lifetime customer value (LCV), or life-time value (LTV) is a prediction of the net profit attributed to the entire future relationship with a customer. See e.g. Abdolvand, Neda; Amir Albadvi; Hamidreza Koosha (2014): Customer Lifetime Value: Literature Scoping Map, and an Agenda for Future Research. International Journal of Management Perspective, Vol. 1, No.3, pp. 41-59. Available at: https://www.researchgate.net/publication/277843944_Customer_Lifetime_Value_Literature_Scoping_Map_and_an_Agenda_for_Future_Research

[270] Christl and Spiekermann (2016), p. 41-44

[271] Ibid., p. 82

[272] Ibid., p. 127-128

[273] Upturn's report about data brokers conducted an evaluation of different definitions of this term, see: Rieke, Aaron; Harlan Yu; David Robinson; Joris von Hoboken (2016), p. 4

[274] FTC, US Federal Trade Commission (2014): Data Brokers. A Call for Transparency and Accountability, p. C3. Available at: http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

[275] Ibid., p. iv

nitely and make inferences about consumers, including “potentially sensitive” ones.[276] The FTC’s report is based on the examination of nine companies, including companies engaged in marketing data, identity verification, and fraud detection, but doesn’t cover credit reporting agencies. While some large data brokers like Acxiom originated from direct marketing[277], others, like Experian, come from credit and consumer reporting but also offer marketing data services.[278] The information that consumer data brokers collect from commercial, governmental and public sources includes, but is not limited to, data about demographic attributes, occupation, education, purchases, property ownership, income, interests, and ethnicity, as well as religious and political affiliation.[279]

Lists of people with names, addresses, or other contact information that group consumers by specific characteristics have historically been an important product sold by marketing data companies. Today these lists include people with low credit scores and with specific conditions such as cancer or depression.[280] They originate from third-party companies which sell or rent information about their customers to data brokers. Lists have been used to sell products and services and to send direct mail to consumers, but also as a basis for other applications. In 2017, for instance, Amnesty International was offered a list of 1.8 million US Muslims; during an investigation on data companies, the organization also discovered offers for lists of “Americans with Bosnian Muslim Surnames” or “Unassimilated Hispanic Americans”.[281] The website dmdatabases.com offers email and mailing lists of wheelchair and insulin users, of people addicted to alcohol, drugs, and gambling, as well as of people suffering from breast cancer, HIV, clinical depression, impotence, and vaginal infections.[282] Nextmark offers consumer lists titled “Pay Day Loan Central – Hispanic”[283], “Help Needed – I am 90 Days Behind With Bills”, “One Hour Cash”[284], “High Ranking Decision Makers in Europe”,[285] or “Identity Theft Protection Responders”.[286]

Unified consumer databases with unified identifiers. Data brokers have different kinds of contracts with their data providers. They might claim “ownership” of the acquired data or only the right to use it or resell it for a specific time period.[287] Similarly, their clients may be re-

[276] Ibid.

[277] http://www.fundinguniverse.com/company-histories/acxiom-corporation-history [20.04.2017]

[278] https://www.experianplc.com/media/1323/8151-exp-experian-history-book_abridged_final.pdf [20.04.2017]

[279] FTC (2014): Data Brokers

[280] 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, p. 5. Available at: 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

[281] https://medium.com/amnesty-insights/data-brokers-data-analytics-muslim-registries-human-rights-73cd5232ed19

[282] http://dmdatabases.com/databases/consumer-mailing-lists/ailments-lists [20.04.2017]

[283] https://lists.nextmark.com/market?page=order/online/datacard&id=311835 [20.04.2017]

[284] https://lists.nextmark.com/market?page=order/online/datacard&id=261706 [20.04.2017]

[285] https://lists.nextmark.com/market?page=order/online/datacard&id=340254 [20.04.2017]

[286] https://lists.nextmark.com/market?page=order/online/datacard&id=290341 [20.04.2017]

[287] FTC (2014): Data Brokers, p. 16

stricted in how the acquired, rented, or licensed data can be used.[288] In addition to the data itself, data brokers provide software to manage their client's customer databases, enhance data, merge lists, remove duplicate records, and to sort them into groups with specific characteristics. In 1999, large companies such as Acxiom and Experian introduced their master customer database products "AbiliTec" and "TruVue" as single sources for information on individuals and households in which every person gets a unique code. This allowed them to link together records from different sources by combining key identifiers such as names, addresses, and zip codes. Providing other companies with "identity resolution"[289] and linking capabilities, based on highly accurate centralized databases which are constantly updated with information about marriages, divorces, births, deaths, name and address changes, became an important service in itself.[290] [291] The respective national postal services were always an important source for this kind of information and they play a vital role in the consumer data business themselves. For example, while national postal services in both Europe[292] and North America[293] offer consumers change-of-address services to forward mail to their new addresses, they also sell this information to other companies. Data brokers also receive data about address changes from companies when consumers update their magazine subscriptions, phone connections, credit card details, or other services.[294]

Combining data about neighborhoods, buildings, households and individuals. Data companies have never only collected individual-level data on consumers. When they started to segment and cluster consumers, they generally relied on census data[295] to attribute specific characteristics and behaviors to postal addresses in certain neighborhoods, and thus to households. Today, all manner of larger-scale information are still in use, including census data, market research data from consumer panels,[296] and surveys, as well as aggregated data on households that was originally collected from individuals. Especially in most European countries, data brokers rely more on aggregated data to profile individuals due to stronger data protection regulations that limit the direct collection, use, and sharing of individual-level data.

For example, the German consumer data broker Arvato AZ Direct, a subsidiary of the media giant Bertelsmann, offers different kinds of data at the individual, household, building, and neighborhood levels – the latter of which consist of either 5, 20, 70, or 500 households. The

[288] Ibid, p. 40

[289] Identity resolution might be defined as a "data management process through which an identity is searched and analyzed between disparate data sets and databases to find a match and/or resolve identities. Identity resolution enables an organization to analyze a particular individual's or entity's identity based on its available data records and attributes": https://www.techopedia.com/definition/29011/identity-resolution

[290] http://blogs.gartner.com/martin-kihn/a-brief-incredible-history-of-marketing-data-matching/

[291] FTC (2014): Data Brokers

[292] http://www.sueddeutsche.de/wirtschaft/verbraucherschutz-kritik-am-daten-handel-der-post-1.196084

[293] https://www.forbes.com/sites/adamtanner/2013/07/08/how-the-post-office-sells-your-new-address-with-anyone-who-pays-and-the-little-known-loophole-to-opt-out

[294] Ibid.

[295] Senate Committee on Commerce, Science, and Transportation (2013), p. 23

[296] E.g. http://www.nielsen.com/apac/en/solutions/measurement/consumer-panels.html [20.04.2017]

company also offers data at the zip code level or at the level of geographic grid cells, sized from 125x125m to 10x10km.[297] Arvato provides attributes such as gender and age at the individual level, but their "geoscore", predicting creditworthiness, is provided only for groups of approximately 20 households. However, when using this data for targeting prospects or to sort customer databases, this group-level credit rating can be assigned to a person at the individual again, as it is originally based on individual-level data. In their marketing data catalog, people with the best credit rating are labeled with an "A" and "VIP clients", while those with worst receive a "G" and "postpone processing".[298] According to Arvato, detailed information about buildings which individuals inhabit is the basis for many of the other attributes they provide.[299] Aggregated data about registered vehicles are used to predict individuals' social standing, purchasing power, and attitudes.[300] Overall, the company provides 600 attributes on 70 million consumers in Germany, assigning every person, household, and building a unique ID.[301]

At global level, an industry report by Forrester lists eleven major consumer data brokers and analytics companies, which it refers to as "customer insights services providers": Acxiom, Ansira, Epsilon, Experian Marketing Services, Harte Hanks, Merkle, Precision Dialogue, Rapp, Targetbase, Wunderman, and Yes Lifecycle Marketing.[302] Client companies use their services to, for example, manage their customer and prospect databases as well as to analyze, segment and sort customers regarding their characteristics, behaviors, profitability and lifetime value. In addition, clients can buy consumer data and append it to their own data.[303] Acxiom, for example, manages customer databases for 7,000 clients, including 47 or the Fortune 100 companies.[304] Similarly, Experian manages 7,500 customer databases of large companies.[305] Merkle states that it manages more than 3.7 billion customer records for clients, including for Dell, Nespresso, Microsoft, Marriott, Chase, American Express, and Universal.[306] The Bertelsmann subsidiary Arvato "maintains relationships with over 600 million consumers and business cli-

[297] Arvato AZ Direct (2015): AZ DIAS PROFILDATEN. Merkmalskatalog. Personal copy of file with author Wolfie Christl

[298] Ibid., in German: "Der Informa-Geoscore prognostiziert die Zahlungsausfallwahrscheinlichkeit auf Mikrozellenebene (mit im Schnitt 20 Haushalten je Mikrozelle). Er basiert im Gegensatz zu vielen anderen externen Daten auf validen adress- bzw. personenbezogenen Informationen, welche auf Mikrozellenebene aggregiert werden", „1 Nahezu kein Risiko (VIP-Kunden, A)", „7 Höchstes Risiko (Bearbeitung zurückstellen, G)"

[299] Ibid., p. 9

[300] Ibid., p. 52

[301] Hüffner, W. (2015): Datenschutzkonformes Smart Data und Data Pooling. arvato Digital Marketing, Mar. 05, 2015. Available at: 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 [20.04.2017]

[302] Forrester (2015): The Forrester Wave™: Customer Insights Services Providers, Q4 2015. November 10, 2015.

[303] Ibid., p. 5

[304] Acxiom (2015): Annual Report 2014. Available at: http://files.shareholder.com/downloads/ACXM/0x0x763250/A1DBFBD8-E136-4701-B0F2-3DC695E5ED08/acxiom2014_Annual_Report_FINAL_RRD_PDF_.pdf [08.04.2017]

[305] Forrester (2015): The Forrester Wave™: Customer Insights Services Providers

[306] https://www.merkleinc.com/sites/default/files/Sum14_Agency-Overview_5.23.14_v2.pdf [20.04.2017]

ents" on "behalf of its clients".[307] At the same time, these "customer insights services providers" also have their own databases with extensive information on consumers across the world. According to Forrester, Yes Lifecycle Marketing, for example, has "business and consumer data in 225 countries".[308]

5.2 The rise of programmatic advertising technology

Over the last ten years the collection, integration, analysis, and utilization of digital information about consumers has become embedded into many areas of life, a development accompanied by unprecedented levels of data volume, velocity, and variety. Some aspects of this are obvious: the ascent of the Internet, the pervasive nature of social media, the spread of smartphones, the ubiquity of online advertising, and the variety and number of applications and Internet connected devices visibly represent the rapid evolution of data-driven services. Nevertheless, many of the technical and commercial developments that enable today's digital data economy occurred in the background, and remain opaque and barely understood by not only most consumers, but also journalists and policymakers.

The online advertising sector has become one of the most advanced industries aiming at the extraction of value from and monetization of data. While the industry consists of a plethora of different kinds of companies, technologies, and practices interacting with one another, only a tip of this iceberg – both in terms of complexity as well as sheer volume – is visible to consumers. Every time a user visits a website or uses an app, data is not only transmitted to the publishers, but often additionally to circa 30 (or more) third-party services.[309] This occurs because publishers and other service providers incorporate tiny pieces of software that transmit data to tracking services into their websites and apps. Some of these tracking services – such as Facebook Like buttons, embedded YouTube videos, or other kinds of widgets – provide user functionality or are otherwise visible to users. Many, however, are invisible to users visiting a website or using a mobile app.[310] Those tiny pieces of embedded software are usually referred to as web bugs, beacons or tags.[311]

Marketing and ad technology. Ensighten, a "tag management platform", provides publishers with an easy way to embed tracking tags from 775 different third-party vendors into their websites and mobile apps in order to transfer user data to third parties.[312] Many of those third-party services facilitate certain aspects of online advertising. Others provide analytics, testing, personalization, or fraud detection services. Some allow publishers to transmit user data into

[307] https://www.arvato.com/en/about/press/2017/arvato-systems-and-arvato-crm-solutions-again-receive-german-top-employer-award.html [20.04.2017]
[308] Forrester (2015): The Forrester Wave™: Customer Insights Services Providers
[309] Christl and Spiekermann (2016), p. 45
[310] Narayanan, Arvind; Dillon Reisman (2017): The Princeton Web Transparency and Accountability Project. Pre-published book chapter. Available at: http://randomwalker.info/publications/webtap-chapter.pdf
[311] Christl and Spiekermann (2016), p. 45, 49
[312] https://www.ensighten.com/products/enterprise-tag-management/integrations/ [20.04.2017]

their cloud-based customer databases or email newsletter and marketing services.[313] Most of these services collect, analyze, and share extensive user data across many websites, mobile apps, and platforms. Sometimes the only reason a website and app publisher embeds such tags is in order to sell the data about their visitors or users.[314] This ecosystem of third-party vendors is often referred to with the buzzwords "advertising technology" or "ad tech", as well as "marketing technology". Although all three are sometimes used interchangeably, "ad tech" often refers more specifically to technologies that address groups of consumers, while "marketing technology" relates to technologies that pertain to single individuals.[315]

Programmatic bidding on profiles. Most of today's advertising occurs through highly automated real-time auctions between publishers and advertisers, a process often referred to as "programmatic advertising".[316] When a person visits a website, it sends metadata about the contents of the page, user profile data, and, usually, some kind of user identifier to multiple advertisers. Then, advertisers who are interested in delivering an ad to this particular user at this particular time and in the context of this particular page make a bid. The highest–bidding advertiser wins and gets to place the ad. This so-called "real-time bidding" happens within the milliseconds of a website loading.[317] Similarly, advertisers can bid for user profiles and ad placements within mobile apps. For the most part, however, this process doesn't happen directly between publishers and advertisers: rather, it involves a range of advertising technology companies so wide that sometimes even experts in online marketing have difficulty fully comprehending it.[318] Some of the common types of services are:

- **Ad servers** and **ad networks**, which help publishers manage advertising requests from many advertisers, and vice versa. **Ad exchanges** connect several ad networks.
- **Sell-side platforms (SSP)**, which allow publishers to sell user profiles and ad placements to many ad networks, exchanges, and demand-side platforms (DSP).
- **Demand-side platforms (DSP)**, which allow advertisers to bid on ad placements and user profiles with specific characteristics from many publishers, ad servers, exchanges, and sell-side platforms (SSP).

The terms being used to describe[319] these main types of ad technology providers are not always consistent. The lines between the different kinds of vendors, such as ad exchanges, SSPs, or DSPs, are often overlapping and blurred. The consulting firm LUMA Partners provides popular maps containing extensive lists of relevant companies in different fields.[320] An additional dis-

[313] See e.g. the list of integrations of the tag manager Segment: https://segment.com/catalog [20.04.2017]

[314] http://www.gartner.com/smarterwithgartner/how-to-monetize-your-customer-data/

[315] See e.g. slide 7: https://www.slideshare.net/tkawaja/luma-digital-brief-010-power-to-the-people

[316] McConnell, Ted (2015): The Programmatic Primer. A Marketer's Guide to the Online Advertising Ecosystem, p. 8

[317] Ibid., p. 7

[318] http://blogs.gartner.com/martin-kihn/ad-tech-worse-wall-street/

[319] Gallagher, Kevin (2017): Ad Tech explainer. How innovation is changing the digital advertising business and creating new opportunities for disruption. Business Insider, BI Intelligence, January 2017.

[320] http://www.lumapartners.com/resource-center/lumascapes-2/ [20.04.2017]

tinction is usually made between companies that focus on tracking and advertising alongside search results[321], general ads on the web[322], mobile ads[323], video ads[324], social network ads[325], and ads within games[326]. In any case, Google dominates the online advertising landscape. The company offers several types of services for both advertisers and publishers, and in 2016 accounted for 32.8% of worldwide digital ad revenues, followed by Facebook with 14.1%.[327]

Profiling users requires tracking online behaviors over time. The delivery of ads based on a user's web searches, browsing, or app usage behavior is referred to as behavioral advertising.[328] In contrast, when serving contextual ads, a publisher does not need to know much about the site visitor or app user, but simply serves ads that relate to the content on the site – such as, for instance, serving an automotive ad on a car-racing site. To profile web or mobile app users, all parties involved in online advertising have developed sophisticated methods to accumulate, compile, and link information from different companies to create extensive long-term profiles about users' behaviors. Companies use cookie syncing[329] to "anonymously" follow individuals across websites, or link different user accounts, devices and even offline data, such as store purchases. To achieve this, they make use of pseudonymous identifiers that refer to individuals, mainly based on email addresses, phone numbers, and cookie and device IDs.[330]

Only a few companies have all the collected profile information in one place.[331] Often profiles about individuals are put together only for a single interaction by combining information from multiple companies in the moment. In this way advertisers can find and target users with specific characteristics and behaviors in milliseconds, regardless of which service or device is used, which activity is pursued, or where the user is located. Many of the participating services might not know the name of the person to whom the ad was served. Nevertheless, an individual's profile or a specific attribute that was assigned to that profile may follow her everywhere and can have a significant impact on that person across ad networks and other data companies.

[321] http://www.lumapartners.com/lumascapes/search-lumascape/ [20.04.2017]
[322] http://www.lumapartners.com/lumascapes/display-ad-tech-lumascape/ [20.04.2017]
[323] http://www.lumapartners.com/lumascapes/mobile-lumascape/ [20.04.2017]
[324] http://www.lumapartners.com/lumascapes/video-lumascape/ [20.04.2017]
[325] http://www.lumapartners.com/lumascapes/social-lumascape/ [20.04.2017]
[326] http://www.lumapartners.com/lumascapes/gaming-lumascape/ [20.04.2017]
[327] https://www.emarketer.com/Chart/Net-Digital-Ad-Revenue-Share-Worldwide-by-Company-2016-2019-of-total-billions/205364
[328] Yan, Jun; Ning Liu, Gang Wang, Wen Zhang, Yun Jiang, and Zheng Chen (2009): How much can behavioral targeting help online advertising?. In Proceedings of the 18th international conference on World wide web (WWW '09). ACM, New York, NY, USA, 261-270. Available at: http://dl.acm.org/citation.cfm?id=1526745
[329] Narayanan, Arvind; Dillon Reisman (2017), p. 13
[330] Christl and Spiekermann (2016), p. 90
[331] Large platforms such as Google and Facebook, or large data companies such as Acxiom

5.3 Connecting offline and online data

Until recently, marketers who used Facebook, Google or other online ad networks, could only target individual profiles based solely on online behavior. In 2012, Facebook started to allow companies to upload their own lists of email addresses and phone numbers from their customer databases to its platform.[332] Even though these email addresses and phone numbers are converted into pseudonymous codes,[333] Facebook has since then been able to directly link third party customers with their Facebook accounts, assuming they are present on the platform. In this way Facebook enables companies to find and target exactly those persons they have email addresses or phone numbers on. As a result, companies can also use Facebook's powerful filter criteria to target – or systematically exclude – only a subset of these persons.

Facebook's ability to exactly target people based on email addresses and phone numbers is perhaps more powerful than it appears at first glance. It lets companies systematically connect their own customer data with Facebook's data. Moreover, it allows also other ad tech vendors to synchronize with the platform's databases and tap into its capacities, essentially providing a kind of real-time remote control for Facebook's data universe. Today companies can capture information about very particular activities – such as specific webpage activities, swipes in a smartphone app, or types of purchases – in real-time and tell Facebook to immediately find and target the persons who performed these activities. Google[334] and Twitter[335] launched similar features in 2015. Today, most online advertising platforms allow companies to pass different kinds of pseudonymous identifiers (mostly generated from email addresses, phone numbers, or device IDs) to each other.

In addition, Facebook started partnering with the four large consumer data brokers Acxiom, Epsilon, Datalogix, and BlueKai in 2013,[336] the latter two of which Oracle subsequently acquired. As of 2017, six data brokers provide "audience data" to Facebook to help the platform better sort and categorize its users. Currently, these six data brokers are Acxiom, Epsilon, Experian, Oracle, CCC Marketing (Japan), and Quantium (Australia).[337]

Data management platforms. Beside the large online platforms and the different types of advertising technology providers, another type of service has, in recent years, become essential to online advertising, and digital tracking and profiling in general. Data management platforms

[332] Facebook introduced its "custom audiences": http://techcrunch.com/2012/10/11/facebook-custom-audience-ads/

[333] Email addresses and phone numbers are "hashed": https://www.facebook.com/ads/manage/customaudiences/tos.php [20.04.2017], for details see also section 7.3

[334] Google introduced "custom match": https://adexchanger.com/mobile/google-allows-targeted-ads-based-on-first-party-data/

[335] Twitter introduced its "partner audiences": http://venturebeat.com/2015/03/05/twitters-new-partneraudiences-will-help-more-advertisers-track-you-outside-twitter/

[336] https://techcrunch.com/2013/02/27/facebook-ad-data-providers/

[337] https://facebookmarketingpartners.com/marketing-partners/#ad=&co=&in=&la=&qss=&sp%5B%5D=Audience+Data+Providers [20.04.2017]

(DMPs) have assumed the role of "central hubs" that aggregate, integrate, manage and deploy different sources of data about consumers.[338] They provide the following functionalities:

- **First**, DMPs allow companies to upload their own data about customers and prospects in real-time, including demographic data and real-time information about purchase activities, website visits, app usage, and email responses. This data can then be combined and linked with data from myriads of third-party vendors.
- **Second**, DMPs help companies analyze, sort, and categorize consumers, add or remove them from lists of individuals with certain characteristics ("audiences" and "segments"), and select them for specific treatment. Thus, a company can address certain people in certain ways with certain messages on certain channels or devices.[339] It could, for example, target a group of existing customers who have visited a certain page on its website, and are predicted as valuable and likely to respond, with personalized content or a discount – either on Facebook, in a mobile app, or on the company's own website.
- **Third**, DMPs provide ways to find and target people with similar characteristics and behaviors as the people provided ("lookalikes")[340] as well as to exclude certain kinds of people from targeting ("audience suppression").[341]

In recent years, larger companies such as software vendors, data brokers, direct marketing agencies, and customer relationship management (CRM) companies have acquired some of these data management platforms. Oracle (see chapter 6), Adobe, Salesforce (Krux), and Wunderman (KBM Group/Zipline) run the most relevant DMPs, but other vendors, such as Neustar, Lotame, and Cxense also provide such platforms.[342]

Lotame, for example, offers its clients the ability to provide real-time data from their customer databases, email systems, websites, apps, and ad campaigns to its data management platform.[343] The uploaded data can then be combined with "billions of consumer profiles, each with thousands of behavioral attributes" originating from both data collected by Lotame itself and by third-party sources[344]. Client companies can then profile, sort, and rank their users, customers, and prospects with regard to their demographics, interests, and behaviors[345] in order to target them with ads through software integrations with many other advertising technology providers.[346] They can share this data with partner companies or feed it into personali-

[338] Winterberry Group (2012): The Data Management Platform: Foundation for Right-Time Customer Engagement. A Winterberry Group Whitepaper. Available at: http://www.iab.net/media/file/Winterberry_Group_White_Paper-Data_Management_Platforms-November_2012.pdf

[339] http://blogs.gartner.com/martin-kihn/data-management-platform

[340] For example: http://www.krux.com/data-management-platform-solutions/lookalikes/ [28.04.2017]

[341] https://blogs.adobe.com/digitalmarketing/advertising/increase-ad-roi-audience-suppression/ [28.04.2017]

[342] Forrester (2015): The Forrester Wave™: Data Management Platforms, Q4 2015. Available at: http://www.krux.com/upload/image/company/The_Forrester_Wave_Data_Management_Platforms_Q4_2015.pdf

[343] https://www.lotame.com/data-management-platform/ [28.04.2017]

[344] http://resources.lotame.com/data-stream-content-access [28.04.2017]

[345] https://www.lotame.com/data-management-platform/ [28.04.2017]

[346] https://www.lotame.com/data-management-platform/ [28.04.2017]

zation and product recommendation engines, CRM, and business intelligence systems or even use it for fraud prevention.[347] **Neustar**, in contrast, runs not only an ad tech DMP, but also provides fraud detection and identity verification services. The company promises to "link traditional off-line identity verification (name, address, phone, email, etc.) with 21st century digital identity data (IP address, geo-location, cookie, device ID, synthetic ID, and more) to give a holistic view of the consumer and the device".[348]

The emergence of data management platforms marks a key moment in the development of pervasive commercial behavioral tracking. With their help, companies in all industries across the globe can seamlessly combine and link the customer and prospect data that they have been collecting for years with billions of digital profiles. Traditional consumer data brokers play a key role in this evolution. They have morphed from old-school direct mail marketing agencies selling consumer lists into high-tech data analytics companies.

5.4 Recording and managing behaviors in real-time

The example of the customer data platform mParticle shows how today's companies can seamlessly collect rich data about their customers and others in real-time, add more information about them from third parties, and utilize the enriched profiles within the marketing and ad technology ecosystem.

Real-time data collection. With the mParticle platform companies can collect customer data including interactions such as views, clicks, searches, social media posts or shares, menu or tab navigation[349], subscribes, app installs, purchases, and refunds[350] – from apps, websites, and several third-party tools.[351]. They can collect information about whether somebody removed an item from a shopping cart and many other custom-defined "events".[352] If available, a user's location can be logged for every event.[353] Screen tracking allows clients to monitor how users navigate through an app.[354] Additionally, they can import data feeds from third-party advertising, marketing, and customer relationship management tools such as Salesforce.[355] The company also supports collecting data from Roku video channels.[356]

Identifying and profiling. Each user profile on mParticle may contain, in addition to device and carrier information,[357] a customer ID, email address, and user IDs from Google, Apple, Mi-

[347] http://resources.lotame.com/data-stream-content-access [28.04.2017]
[348] https://www.neustar.biz/risk/fraud-detection [28.04.2017]
[349] http://docs.mparticle.com/#event-type [22.04.2017]
[350] http://docs.mparticle.com/#selecting-data-criteria [22.04.2017]
[351] https://www.mparticle.com/ [22.04.2017]
[352] http://docs.mparticle.com/#selecting-data-criteria [22.04.2017]
[353] http://docs.mparticle.com/#location-tracking [22.04.2017]
[354] http://docs.mparticle.com/#screen-tracking [22.04.2017]
[355] http://docs.mparticle.com/#feed-configurations [22.04.2017]
[356] http://docs.mparticle.com/#roku [22.04.2017]
[357] http://docs.mparticle.com/#selecting-data-criteria [22.04.2017]

crosoft, Facebook, Twitter, Yahoo as well as "any other identifier that can contribute to user identification".[358] Other available user attributes include name, address, gender, and phone number[359] as well as the "customer lifetime value".[360] As soon as data is captured, mParticle enriches it with data from providers such as Oracle's Datalogix[361], adding information about occupation, income, brand preferences, interests, hobbies, and credit cards used. Altogether, it makes "thousands of 3rd party data segments" available to its clients.[362]

Managing groups of people. Companies can create "audiences"[363], i.e. lists of user profiles with certain kinds of identifiers, and then "ship" those audiences to a wide selection of platforms and marketing technology companies, including Adobe, AppNexus, Facebook, Google, Inmobi, Lotame, Millennial Media (Verizon), Pinterest, Snapchat, Tapad, and Twitter[364]. A company could, for example, create an "audience" consisting of high-income users who installed an app in the last 72 hours, have used this app less than three times and additionally have visited a certain location. This "audience" consisting of a dynamic list of users might then be forwarded to Facebook, which can identify these people by matching the email addresses or mobile identifiers to its own userbase[365] in order to address them with specific messages.[366]

Real-time behavioral feeds. This is very similar to how companies have used mailing lists with names and addresses for ages – with the crucial difference that mParticle's "audiences" are not simply static lists of people. They are dynamic feeds about groups of people with certain characteristics and behaviors, continuously updated by real-time interactions captured in different contexts, and are used to make automated decisions on those people across devices and platforms in order to influence their behavior according to complex sets of rules and instructions. The recorded user interactions and events can also be directly transmitted to other platforms such as Abakus (SAP), Bluekai (Oracle), or Krux (SalesForce).

5.5 Collecting identities and identity resolution

A report on the "strategic role of identity resolution" in marketing does a good job describing how companies should proceed with collecting data on their customers and others as well as how to best utilize this data.[367]

[358] http://docs.mparticle.com/#user-identity [22.04.2017]
[359] http://docs.mparticle.com/#user-tags-and-attributes [22.04.2017]
[360] http://docs.mparticle.com/#lifetime-value [22.04.2017]
[361] http://docs.mparticle.com/#user-insights [22.04.2017]
[362] http://blog.mparticle.com/mparticle-feature-spotlight-third-party-segment-insight/ [22.04.2017]
[363] http://docs.mparticle.com/#audiences [22.04.2017]
[364] http://docs.mparticle.com/#audience-configurations [22.04.2017]
[365] To be precise, pseudonymous identifiers based on hashed email addresses or mobile identifiers
[366] http://docs.mparticle.com/#audiences [22.04.2017]
[367] Stanhope, Joe; Mary Pilecki; Fatemeh Khatibloo; Tina Moffett; Arleen Chien; Laura Glazer (2016): The Strategic Role Of Identity Resolution. Identity Is Context In The Age Of The Customer. Forrester, October 17, 2016.

Aggregating identifiers. The report suggests that the "ability to accurately identify customers" constitutes the "most basic prerequisite for marketing analytics, orchestration, and execution" today. Marketing has always been about "knowing the customer[s]", which means "being able to identify them". Identification, though, is "more than name recognition". Thus, complete identity resolution should marry "multiple sources of identifier and interaction information" in order to "build robust customer profiles based on multiple data sources and interactions". To enable identity resolution as a "strategic capability that facilitates the links between systems and interactions", the report suggests a four-step process:

- **First**, companies should "collect identity keys" such as names, addresses, social handles, device IDs, cookie IDs, IP addresses, MAC addresses, and other types of identity keys that refer to consumers, devices, or other entities. Every interaction in a "customer life cycle" would generate or leverage such keys. Keys might also be sourced from other companies.
- **Second**, companies should "link the keys together", for example by "associating a cookie ID with an email address, or a phone number with a loyalty card number", possibly supported by "second- or third-party data sets". Generally, "more keys and linkages" will create "increasingly robust identities".
- **Third**, companies should "connect identities to descriptive data" in order to create profiles, for example by "aggregating device and channel data, behavior and interactions, self-submitted information, purchase and customer data, and third-party data sources". This will also require the "ability to continuously update profiles as identity keys and personal data changes over time".
- **Fourth**, companies should "apply identities to marketing use cases". For example, they might use them for individual targeting and recognition across channels, devices, and "touchpoints", for measurement, analytics, personalization, fraud detection, and transaction authorization. Companies might also use it to "proactively omit consumers from campaigns" and make "online data available for offline applications". In addition, companies should consider incorporating "emerging channels" such as Smart TVs or IoT devices. Eventually, they should also consider monetizing their identity data by selling it to other companies.

The report adds that identity resolution capabilities require a balance between "complete and accurate data" and "return on investment". In this way, it implies that it is acceptable for some data to be inaccurate, as long as it adds enough to the bottom line. As the above summary suggests, the report largely praises extensive corporate data collection, although, at least, the authors repeatedly warn that collecting and using consumer data carries the "risk of damaging brands and incurring legal consequences".

Accordingly, the consumer data broker **Acxiom** has rebranded itself as an "identity resolution" company.[368] Their website states that their identity resolution services would help clients "connect data to consumers in a privacy-compliant way", so they can "create a single view of the customer. Clients could "match data at scale across a broad set of identifiers including email addresses, postal addresses, phone numbers, and digital IDs", in order to "recognize consumers accurately across offline and digital channels".[369]

The data broker **Experian** explains in a brochure that a "consumer" in the context of customer databases would be "any uniquely identifiable individual who has at all interacted with a brand". The "persistent consumer ID for each channel or device" would "broadly speaking" be "the equivalent of a postal address for that channel".[370] Experian's "Truvue" system provides "persistent ID linkage" between individuals, addresses, and households to "accurately identify individual customers", based on Experian's "vast name and address history" which is continuously updated with "reliable and verifiable data" from "thousands of contributors".[371] Similarly, Experian's "persistent linkage platform" ConsumerView helps companies "recognise customers across channels" and "link fragmented and incomplete data including email, social and personally identifiable information (PII), across channels and devices" based on data on "8 billion name and address combinations".[372]

5.6 Managing consumers with CRM, CIAM and MDM

Businesses in all industries use different kinds of software and services from third party vendors to collect, manage, utilize, and share customer data on the enterprise level. **Customer relationship management (CRM)** software accounts for a key piece in this puzzle. The research firm Gartner considers CRM as enabling a "business strategy that optimizes revenue and profitability while promoting customer satisfaction and loyalty".[373] CRM technologies "enable strategy, and identify and manage customer relationships, in person or virtually" by providing functionalities for sales, marketing, customer service, and digital commerce. The CRM software market is dominated by Salesforce, followed by SAP, Oracle, Microsoft, and Adobe.[374] CRM tools and services are also provided by other large IT corporations like IBM and SAS, data brokers and analytics companies like Acxiom, Experian, Epsilon, and FICO, as well as by large business consulting companies like Accenture, Capgemini, Deloitte, and McKinsey.[375]

[368] In the HTML title tag of their website they use the phrase "Identity Resolution – Acxiom": https://www.acxiom.com [30.04.2017]

[369] https://www.acxiom.com/what-we-do/identity-resolution/ [30.04.2017]

[370] http://www.experian.co.uk/assets/marketing-services/white-papers/wp-cross-device-identification.pdf [30.04.2017]

[371] http://www.experian.com/marketing-services/customer-data-management-integration-persistent-id.html [30.04.2017]

[372] http://www.experian.com.au/marketing-services/consumerview.html [30.04.2017]

[373] Gartner (2016): The Gartner CRM Vendor Guide. Gartner, 9 May 2016

[374] https://www.forbes.com/sites/louiscolumbus/2016/05/28/2015-gartner-crm-market-share-analysis-shows-salesforce-in-the-lead-growing-faster-than-market

[375] Gartner (2016)

Customer value analytics, personalization and price optimization. Gartner's CRM vendor guide lists hundreds of companies in different segments such as digital commerce, loyalty management, lead management, customer segmentation, email marketing, contact center routing, customer engagement, customer journey analytics, and customer experience management, among many others. For example, solutions for "customer value analytics" provide predictions and scores about the current and future profitability of customers. So-called "personalization engines" are used to create "relevant, individualized interaction" based on customer data or data from "similar individuals". Software for "real-time decisioning" combines data with business strategy to "identify the optimal customer treatment that applies broadly across the enterprise"; this is mostly deployed in contact centers, retail stores, bank branches, and on websites. It also includes applications for "prioritizing customer support opportunities, fraud detection and service personnel alignment". Tools for "price management and optimization" provide data-driven analytics, ranging from "pricing guidance" to "digitally enabled and algorithm-based pricing".[376]

Consumer identity and access management (CIAM). In recent years, another genre of software related to consumer data has arisen from a completely different field – namely, that of security and authentication within enterprises. Some of the traditional software solutions for "identity and access management" (IAM), which are designed to "provision, authenticate, authorize, and store information about employee users", have evolved into platforms for "Consumer Identity and Access Management" (CIAM). Instead of user account management and authentication for a company's employees and business partners, these platforms provide the same features for customer and prospect data management in the consumer sphere. In addition, CIAM platforms support collecting and analyzing information about consumers from many sources, in both the realms of fraud prevention and marketing.[377] According to a report from a consulting firm, they are "merging and absorbing functionality from CRM and marketing automation" but also provide features with regards to "Know Your Customer" (KYC), an especially important service for banks and other actors in the financial sector. CIAM software vendors include ForgeRock, Salesforce, Microsoft, IBM, PingIdentity, Okta, Janrain, LoginRadius, iWelcome, SecureAuth, and SAP.[378]

Master data management (MDM). Generally, software for master data management, which supports the "global identification, linking and synchronization of master data domains across heterogeneous data sources" within large corporations, including customer data, is provided by companies such as IBM, Informatica, Oracle, SAP, and SAS.[379]

[376] Ibid.
[377] Tolbert, John (2016): CIAM Platforms. Leaders in innovation, product features, and market reach for Consumer Identity and Access Management Platforms. Your compass for finding the right path in the market. KuppingerCole Report, November 2016.
[378] Ibid.
[379] Ibid.

6. Examples of Consumer Data Broker Ecosystems

This chapter summarizes how the data brokers Acxiom and Oracle collect data on consumers, who they partner with, and how their clients use their data services. These two large companies were selected as illustrative examples of wider practices based on previous research.[380] The findings are based on publicly available information from e.g. corporate websites, marketing materials, brochures, data catalogs, case studies, developer guides, and API docs, as well as from reports, news, and trade magazine articles.

6.1 Acxiom, its services, data providers, and partners

Founded in 1969,[381] Acxiom is one of the most cited examples of a large US data broker that collects, analyzes and trades vast amounts of consumer information.[382] The company mainly provides data and marketing services, but also risk mitigation, identity verification, and fraud detection solutions.[383] Acxiom helps its clients analyze, enhance, sort, and utilize consumer data, as well as link and combine their client's customer data to data from other sources[384]. The company's clients include large organizations in the fields of financial services, insurance, telecommunications, retail, healthcare, and government.[385] It manages 15,000 customer databases and 2.5 billion customer relationships for 7,000 clients, including for 47 of the Fortune 100 companies.[386] With regards to consumer data, Acxiom claims to provide access to up to 5,000 data elements on 700 million people worldwide from "thousands of sources"[387] in many countries[388], including data about consumers in the US, the UK,[389] and Germany.[390]

Acxiom attaches every person in its database to a globally unique identifier.[391] These personal ID numbers within the company's "Abilitec Link" system refer to persons with certain names, postal addresses, phone numbers, email addresses, and other identifiers,[392] and are

[380] Christl and Spiekermann (2016), p. 94-100

[381] Grant, Tina (2001): International Directory of Company Histories, Vol. 35, St. James Press, 2001. Retrieved from: http://www.fundinguniverse.com/company-histories/acxiom-corporation-history/

[382] http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-ofconsumer-database-marketing.html

[383] https://www.acxiom.com/what-we-do/identity-verification-authentication/ [14.05.2017]

[384] Christl and Spiekermann (2016), p. 94-97

[385] Acxiom (2015): Annual Report 2014. Available at: http://files.shareholder.com/downloads/ACXM/0x0x763250/A1DBFBD8-E136-4701-B0F2-3DC695E5ED08/acxiom2014_Annual_Report_FINAL_RRD_PDF_.pdf [08.04.2017]

[386] Ibid.

[387] Acxiom (2017): Annual Report 2016. Available at: https://s21.q4cdn.com/580938034/files/doc_financials/annual_reports/ACXM_Annual_Report_FINAL_RRD_Printers_Proof_6-17-16_.pdf

[388] http://ec.europa.eu/justice/news/consulting_public/0003/contributions/organisations/acxiom_en.pdf

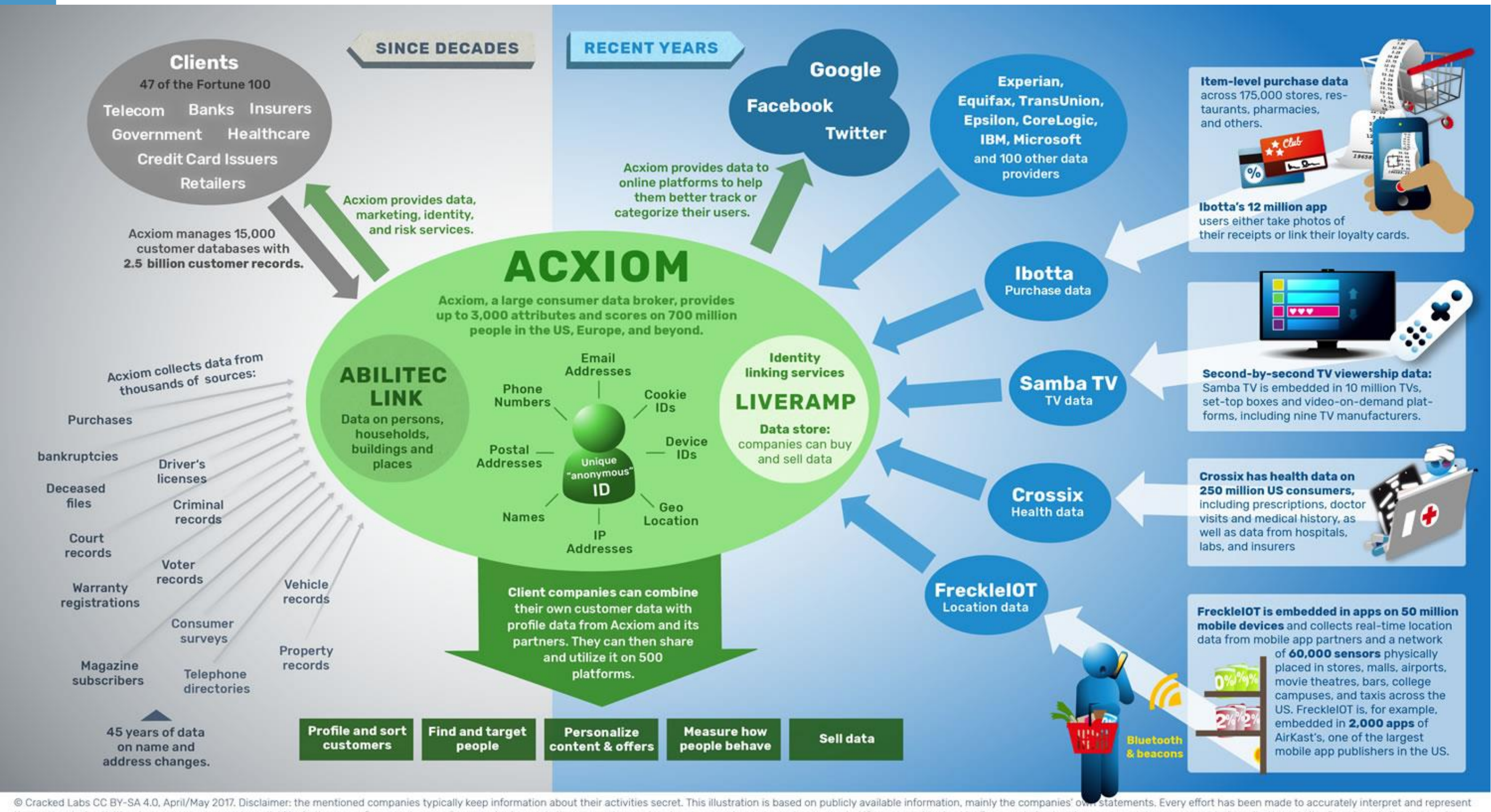
[389] http://dq2qu0j6xxb34.cloudfront.net/wp-content/uploads/2014/01/Facebook-Case-Study-Final-no-bleed.pdf [22.01.2016]

[390] Video from the panel discussion "Die Strategien der großen Daten-Anbieter" at the d3con 2014 conference. In German: "Heute haben wir in Deutschland Offline-Profildaten verfügbar für nahezu jeden Haushalt": https://www.youtube.com/watch?v=W41HcRo-3P8 [22.01.2016]

[391] https://developer.myacxiom.com/code/api/data-bundles/bundle/abilitec [11.05.2017]

[392] Ibid.

based on name change data and residential histories that go back decades.[393] Acxiom has long been collecting data from public records as well as from other companies, including magazine subscriber lists, consumer surveys and questionnaires, product and warranty registrations, and purchases.[394] When asked about a specific person, Acxiom provides, for example, one of 13 religious affiliations including “Catholic”, “Jewish” and “Muslim” and one of nearly 200 “ethnic codes”.[395] Overall, the company provides hundreds of data attributes and scores about consumers; these include information about education, occupation, children, political views, interests, activities, purchases, media usage, properties and vehicles owned, banking and insurance policies, income, loans, health interests, and economic stability as well as scores, such as the “likelihood” that someone in a person’s household “will change careers or jobs”.[396]


SINCE DECADES
RECENT YEARS
Clients
47 of the Fortune 100
Telecom
Banks
Insurers
Government
Healthcare
Credit Card Issuers
Retailers
Acxiom manages 15,000 customer databases with 2.5 billion customer records.
Acxiom provides data, marketing, identity, and risk services.
Acxiom provides data to online platforms to help them better track or categorize their users.
Google
Facebook
Twitter
Experian, Equifax, TransUnion, Epsilon, CoreLogic, IBM, Microsoft and 100 other data providers
Item-level purchase data across 175,000 stores, restaurants, pharmacies, and others.
Club
Ibotta's 12 million app users either take photos of their receipts or link their loyalty cards.
ACXIOM
Acxiom, a large consumer data broker, provides up to 3,000 attributes and scores on 700 million people in the US, Europe, and beyond.
Ibotta
Purchase data
Acxiom collects data from thousands of sources:
Purchases
bankruptcies
Driver's licenses
Deceased files
Criminal records
Court records
Voter records
Warranty registrations
Vehicle records
Consumer surveys
Magazine subscribers
Telephone directories
Property records
45 years of data on name and address changes.
ABILITEC LINK
Data on persons, households, buildings and places
Email Addresses
Phone Numbers
Cookie IDs
Postal Addresses
Device IDs
Unique “anonymous” ID
Names
Geo Location
IP Addresses
Identity linking services
LIVERAMP
Data store: companies can buy and sell data
Samba TV
TV data
Second-by-second TV viewership data: Samba TV is embedded in 10 million TVs, set-top boxes and video-on-demand platforms, including nine TV manufacturers.
Crossix
Health data
Crossix has health data on 250 million US consumers, including prescriptions, doctor visits and medical history, as well as data from hospitals, labs, and insurers
FreckleIOT
Location data
FreckleIOT is embedded in apps on 50 million mobile devices and collects real-time location data from mobile app partners and a network of 60,000 sensors physically placed in stores, malls, airports, movie theatres, bars, college campuses, and taxis across the US. FreckleIOT is, for example, embedded in 2,000 apps of AirKast's, one of the largest mobile app publishers in the US.
Bluetooth & beacons
Client companies can combine their own customer data with profile data from Acxiom and its partners. They can then share and utilize it on 500 platforms.
Profile and sort customers
Find and target people
Personalize content & offers
Measure how people behave
Sell data

© Cracked Labs CC BY-SA 4.0, April/May 2017. Disclaimer: the mentioned companies typically keep information about their activities secret. This illustration is based on publicly available information, mainly the companies' own statements. Every effort has been made to accurately interpret and represent the companies' activities, but we cannot accept any liability in the case of eventual errors. Sources: Acxiom website, press releases, brochures, annual reports, and response to US congress inquiry. LiveRamp website, brochures, press releases, presentations. Ibotta website. Crossix website, press releases. FreckleIOT press releases. For details about the sources see the report "Corporate Surveillance in Everyday Life".

Figure 2: Acxiom and some of its data providers, partners and services.

[393] https://lp.liveramp.com/rs/982-LRE-196/images/Identity%20Graph%20One%20Pager.pdf [14.05.2017]

[394] See https://www.acxiom.com/wp-content/uploads/2013/09/Acxiom-Marketing-Products.pdf, and Acxiom (2012): Response letter to U.S. congress inquiry, August 15, 2012. Available at: http://web.archive.org/web/20130425093308/http://markey.house.gov/sites/markey.house.gov/files/documents/Acxiom.pdf

[395] https://developer.myacxiom.com/code/api/data-bundles/bundle/eTechDemographics [20.04.2017]

[396] See note 494

Connecting to the digital data ecosystem. Since the acquisition of the online data company LiveRamp in 2014,[397] Acxiom has made major efforts to connect its decade-spanning database to today's pervasive digital tracking and profiling universe. With technologies such as LiveRamp's "Identity Graph" and "IdentityLink", the company is now able to "connect offline data and online data back to a single identifier".[398] In other words, clients can "resolve all" of their "offline and online identifiers back to the individual consumer".[399] To achieve this, the company matches the consumer profiles in its "Abilitec Link" system, which contains names[400], postal addresses, email addresses, phone numbers[401], geo locations[402] and IP addresses[403], to online and mobile identifiers, such as cookie and device IDs.[404] As a result, LiveRamp is now able to link digital profiles across hundreds of data and advertising companies.[405]

Clients can upload their own consumer data to LiveRamp, combine it with data from more than 100 third-party data providers such as Equifax, Experian, TransUnion and Epsilon[406], and then utilize it on more than 500 marketing technology platforms.[407] They can use this data to find and target people with specific characteristics[408], to recognize and track consumers across devices and platforms[409], to profile and categorize them[410], to personalize content for them, and to measure how they behave.[411] For example, clients could "recognize a website visitor" and "provide a customized offer"[412] based on extensive profile data, without requiring said user to log in to the website.[413] Furthermore, LiveRamp's "IdentityLink" technology enables other companies to "buy and sell valuable customer data" in its "Data Store".[414]

Data providers and partners. Examples of companies which Acxiom's LiveRamp has recently promoted as data providers include Ibotta, Freckle IOT, Samba TV, and Crossix:[415]

[397] https://www.acxiom.com/news/acxiom-completes-liveramp-acquisition/ [14.05.2017]
[398] https://liveramp.com/discover-identitylink/identitylink-features/build-an-omnichannel-view/ [14.05.2017]
[399] https://lp.liveramp.com/rs/982-LRE-196/images/Identity%20Graph%20One%20Pager.pdf [14.05.2017]
[400] LiveRamp (2016): IdentityLink Data Store. Providing access to Identity-based data and buyers across the marketing ecosystem. PDF brochure, p. 2. Personal copy of file with author Wolfie Christl
[401] https://liveramp.com/discover-identitylink/identitylink-features/identity-graph/ [14.05.2017]
[402] https://developer.myacxiom.com/code/api/endpoints/match [14.05.2017]
[403] LiveRamp (2016): IdentityLink Data Store, p. 2
[404] https://liveramp.com/discover-identitylink/identitylink-features/identity-graph/ [14.05.2017]
[405] https://liveramp.com/discover-identitylink/identitylink-features/ [14.05.2017]
[406] Several companies are listed as "data providers" on https://liveramp.com/partners, including Equifax, Experian, TransUnion, Corelogic, Epsilon, IBM and Microsoft [09.05.2017]
[407] https://liveramp.com/discover-identitylink/identitylink-features/ [14.05.2017]
[408] https://liveramp.com/discover-identitylink/ [14.05.2017]
[409] https://liveramp.com/discover-identitylink/identitylink-features/ [14.05.2017]
[410] https://liveramp.com/discover-identitylink/identitylink-features/build-an-omnichannel-view/ [14.05.2017]
[411] https://liveramp.com/discover-identitylink/ [14.05.2017]
[412] Ibid.
[413] https://liveramp.com/discover-identitylink/identity-resolution/people-based-marketing/personalization/ [14.05.2017]
[414] https://liveramp.com/discover-identitylink/identitylink-features/ [14.05.2017]
[415] Some of the "unique data sets available" on LiveRamp include Ibotta, Samba TV and Freckle IOT (https://www.acxiom.com/news/liveramp-launches-identitylink-for-data-owners/). Ibotta and Samba TV are promoted as "new data spotlights" for "data buyers" in LiveRamp's "IdentityLink Data Store" brochure from 2016. Crossix is listed as "data provider" in LiveRamp's partner directory (http://liveramp.com/partners) [16.04.2017]

- **Ibotta** provides “item-level purchase data from millions of mobile users”, according to a LiveRamp press release[416]; said data is recorded by the company’s smartphone app. Ibotta’s 12 million monthly active users[417] can earn rewards by either taking photos of receipts after shopping, or by directly linking store loyalty cards with their app.[418] According to a LiveRamp brochure, Ibotta receives purchase data from 1.3 million users who have linked their loyalty cards and additionally scans 1.25 million new receipts per week, overall collecting purchase information from more than 175,000[419] stores, including retailers, restaurants, bars, movie theatres, and pharmacies.[420]
- **Freckle IOT** collects real-time location data,[421] including mobile identifiers,[422] from mobile app partners and a network of beacon sensors placed in stores, restaurants, bars, malls, airports, movie theatres, college campuses, and even in street furniture.[423] The company partners, for example, with Blue Bite, a firm which operates 60,000 beacons across North America, and uses NFC, Bluetooth, WiFi, and other technologies to track the location of mobile phone users on a “hyper-local” level.[424] [425] Freckle’s services are embedded into mobile apps of partner companies, altogether installed on 50 million devices.[426] They are, for example, integrated in more than 2,000 apps of Airkast’s, a large mobile publisher in the US that provides apps for broadcasters and media companies including Cumulus Media, ESPN, CBS News Radio, and Fox News Radio.[427] Freckle provides profile data “built from mobile and beacon data”[428] to LiveRamp[429], as well as data about people who have visited specific locations, via “one to one matching of real-time user data”.[430]
- **Samba TV** records “second-by-second” TV viewing behaviors from more than 10 million “opted-in” TVs, set-top boxes, and video-on-demand platforms, including data from 9 major TV manufacturers.[431] The company provides content recommendation software,

[416] https://www.acxiom.com/news/liveramp-launches-identitylink-for-data-owners [16.04.2017]

[417] http://www.businessinsider.com/groupon-and-jetcom-sign-partnership-with-mobile-shopping-app-ibotta-2016-9

[418] https://ibotta.com/how [16.04.2017]

[419] LiveRamp (2016): IdentityLink Data Store.

[420] https://ibotta.com/where/in-store [16.04.2017]

[421] http://freckleiot.com/httpfreckleiot-comfreckles-industry-first-attribution-tag-supported-by-five-global-dsps/ [16.04.2017]

[422] http://freckleiot.com/privacy-policy/ [16.04.2017]

[423] http://www.prnewswire.com/news-releases/freckle-iot-and-blue-bite-partner-to-create-north-americas-largest-proximity-network-300063315.html [16.04.2017]

[424] Ibid.

[425] See also: http://adage.com/article/datadriven-marketing/mobile-data-marketplace-creates-direct-path-app-dollars/305947/

[426] http://freckleiot.com/freckle-iot-partners-with-liveramp-to-bring-in-store-1st-party-data-segments-to-the-marketing-ecosystem [16.04.2017]

[427] http://freckleiot.com/freckle-partners-with-airkast-to-further-cement-freckle-iot-as-largest-beacon-attribution-company-in-the-world [16.04.2017]

[428] https://www.acxiom.com/news/liveramp-launches-identitylink-for-data-owners [16.04.2017]

[429] http://www.prnewswire.com/news-releases/freckle-iot-partners-with-liveramp-to-bring-in-store-1st-party-data-segments-to-the-marketing-ecosystem-300418304.html [16.04.2017]

[430] https://liveramp.com/partner/freckle-iot/ [16.04.2017]

[431] LiveRamp (2016): IdentityLink Data Store.

which is behind many apps on Roku and also used on smart TV platforms from Samsung, LG, and Sony.[432] According to the company's privacy policy, the recorded information includes titles and amounts of time the content is watched, as well as "actions taken while viewing the content" and other data.[433] LiveRamp lists Samba TV as one of "the unique data sets available"[434] and as a "new data spotlight" for "data buyers".[435] Samba TV's own data management platform allows "1:1 TV to digital match"[436] and is powered by LiveRamp's IdentityLink platform.[437]

- **Crossix**, a company which is listed as a "data provider" by LiveRamp[438], claims to have health data on more than 250 million US consumers, including doctor visits, treatment history, medical claims, prescriptions, and patient health information such as blood pressure, cholesterol level, and diagnoses.[439] Recently, Crossix announced the ability to access additional data from hospitals, labs, insurers, and medical devices.[440] According to a press release, Acxiom and Crossix are "connecting consumer data to predictive health data" for marketing purposes and to provide people with "relevant messaging, support, and behavioral modification content".[441] The mobile advertising company 4Info, for example, uses Crossix' data to target mobile ads at consumers who show a "propensity" to have a certain disease, such as diabetes.[442] For this purpose, Acxiom "matches its household-level data with 4Info's household-level data associated with mobile device IDs, and in turn connects it with a unique Crossix ID", according to Advertising Age.[443] To find out a mobile users' home address, 4Info reportedly "tracks locations of mobile devices when people open certain apps. After a device has been spotted several times during certain hours in a particular residential location, it is determined to be a home address". In addition, the company "verifies home location data through mapping apps that give directions to or from a home address provided directly by the user".[444] All involved parties emphasize many times that only "de-identified" data is used.

Partnerships with the large online platforms. Acxiom also partners with platforms such as Facebook,[445] Google[446] and Twitter[447] in several ways. For example, the company helps them

[432] http://rethinkresearch.biz/articles/samba-tv-takes-connected-tvs-by-storm/ [16.04.2017]
[433] https://samba.tv/privacy-policy/ [16.04.2017]
[434] https://www.acxiom.com/news/liveramp-launches-identitylink-for-data-owners [16.04.2017]
[435] LiveRamp (2016): IdentityLink Data Store.
[436] https://samba.tv/advertising/tv-dmp/ [16.04.2017]
[437] https://samba.tv/samba-tv-launches-tv-dmp-bring-speed-accuracy-digital-television/ [16.04.2017]
[438] https://liveramp.com/partner/crossix/ [16.04.2017]
[439] http://www.crossix.com/platform.aspx [16.04.2017]
[440] http://www.businesswire.com/news/home/20161117005239/en/Crossix-Completes-Largest-Ever-Expansion-Connected-Health-Data
[441] http://www.crossix.com/Press-Releases/Crossix-Solutions-and-Acxiom-Partnership.aspx
[442] http://adage.com/article/dataworks/data-partners-tie-mobile-ads-drug-refills-doc-visits/302937/
[443] Ibid.
[444] Ibid.
[445] https://www.acxiom.com/news/acxiom-becomes-audience-data-provider-facebook-marketing-partner-program/ [14.05.2017]

track and categorize their users even better, based on data collected from beyond these platforms.[448]

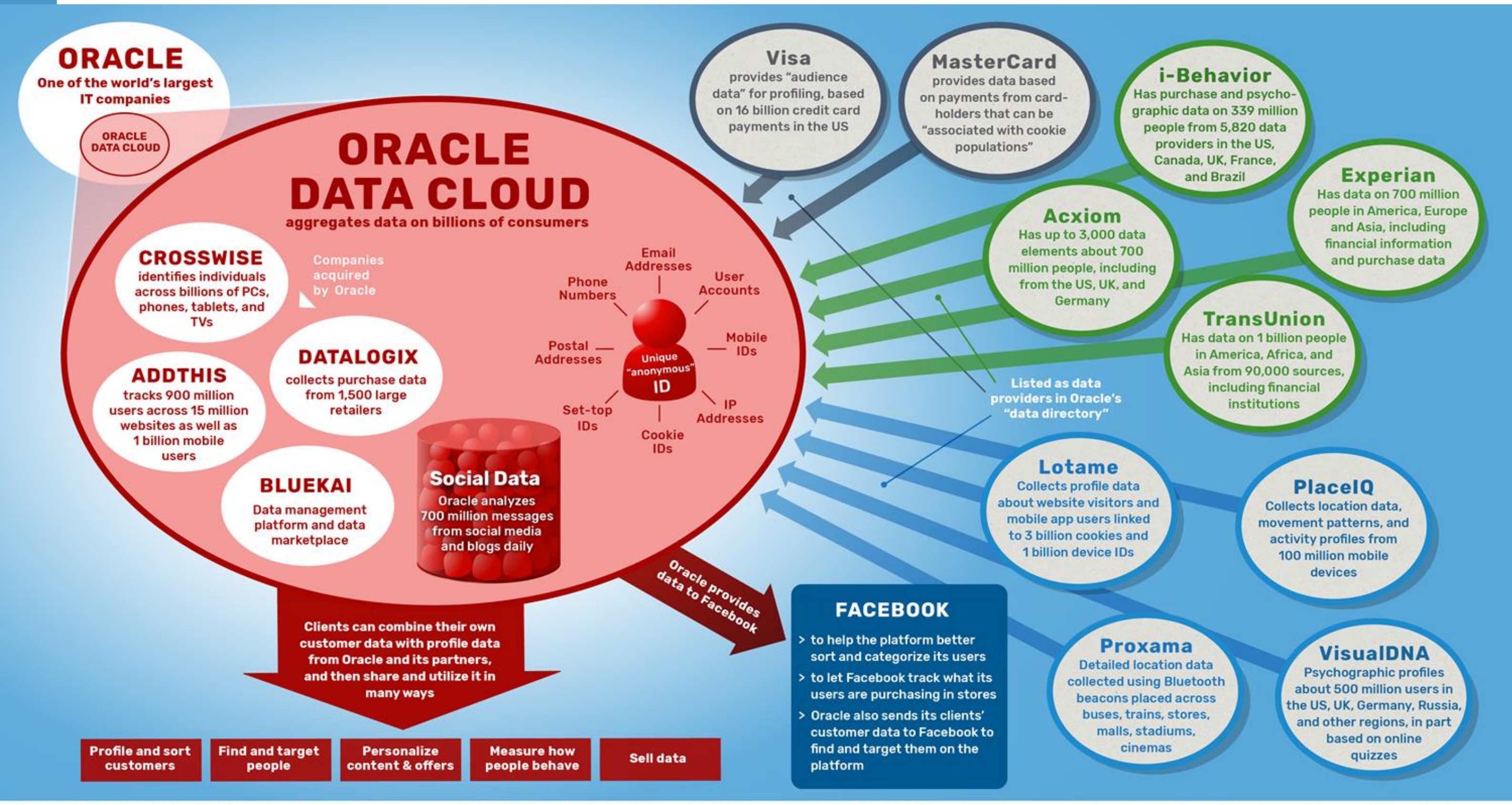
6.2 Oracle as a consumer data platform

Oracle, one of the world's largest business software and database vendors[449], has recently become one of the largest consumer data brokers as well. In recent years, the company acquired several data companies, including Datalogix, AddThis, Crosswise, and BlueKai, a data management platform and data marketplace[450]. While **Datalogix** aggregates data on billions of purchase transactions across 50 grocery chains[451] and 1,500 large retailers[452], the social bookmarking service **AddThis** tracks 900 million users across 15 million websites as well as 1 billion mobile users[453]. **Crosswise** collects activity data across billions of devices and identifies which PCs, phones, tablets, and TVs are being used by an individual consumer[454]. In addition, Oracle aggregates and analyzes "700 million social messages daily"[455] from social media networks, message boards, blogs, consumer review sites, and video platforms.[456] Overall, Oracle's data marketplace provides "more than 30,000 data attributes on **two billion consumer profiles**".[457]

Oracle sorts consumers into thousands of categories. These include characteristics such as age, gender, family composition, parenthood, education, occupation, political views, media usage, income, loans, net worth and purchases, but also data about online searches. For example, Oracle tracks who has searched for issues such as abortion, legalizing drugs, gay marriage, military bases, protests, heart failure, or medical facilities.[458] Together with its data partners the company provides over 50,000 different categories that may be assigned to consumers.[459] Within its **data cloud**, Oracle claims to record "what consumers do", "what consumers say", and "what consumers buy" in order to allow companies to find and target people across devices and platforms, personalize interactions, and measure consumer behavior.[460] Oracle's clients

[446] https://www.acxiom.com/news/liveramp-extends-data-connectivity-partnership-with-google/ [14.05.2017]
[447] https://www.acxiom.com/news/acxiom-expands-its-global-data-partnership-with-twitter/ [14.05.2017]
[448] See the press releases above.
[449] http://www.computerworld.com/article/2489278/itmanagement/oracle-overtakes-ibm-as-second-largest-software-vendor--gartner-says.html
[450] http://www.oracle.com/us/assets/general-presentation-2150582.pdf [12.05.2017]
[451] http://www.oracle.com/us/corporate/acquisitions/datalogix/general-presentation-2395307.pdf [12.05.2017]
[452] http://www.oracle.com/us/products/applications/audience-guide-3034880.pdf [12.05.2017]
[453] http://www.oracle.com/us/products/applications/audience-guide-3034880.pdf [12.05.2017]
[454] https://www.oracle.com/corporate/acquisitions/crosswise/index.html [12.05.2017]
[455] http://www.oracle.com/us/solutions/cloud/daas-for-business-2245611.pdf [12.05.2017]
[456] http://www.oracle.com/us/products/social-relationship-mgmt-brief-1915605.pdf [12.05.2017]
[457] https://www.oracle.com/corporate/pressrelease/eyeota-011917.html [12.05.2017]
[458] Supra note 495
[459] Oracle (2017): Taxonomy Changes Report, 31.01.2017. For a list of all data categories offered by Oracle, see the Inventory worksheet in its "taxonomy changes report": https://docs.oracle.com/cloud/latest/marketingcs_gs/OMCDA/Resources/xls/TaxonomyChangesReport2017-01-31.xlsx via https://docs.oracle.com/cloud/latest/marketingcs_gs/OMCDA/Help/AudienceDataMarketplace/Taxonomy_Updates/taxonomy_updates_january_2017.html [12.05.2017]
[460] http://www.oracle.com/us/corporate/acquisitions/datalogix/general-presentation-2395307.pdf [12.05.2017]

can import their own data about customers, website visitors, and app users into Oracle's data cloud,[461] combine it with data from Oracle and its partners,[462] and then transfer and utilize it on hundreds of other marketing and advertising services.[463] Furthermore, companies can also sell their data about consumers on Oracle's data marketplace.[464]


ORACLE
One of the world's largest IT companies
ORACLE DATA CLOUD
ORACLE DATA CLOUD
aggregates data on billions of consumers
Companies acquired by Oracle
CROSSWISE
identifies individuals across billions of PCs, phones, tablets, and TVs
ADDTHIS
tracks 900 million users across 15 million websites as well as 1 billion mobile users
DATALOGIX
collects purchase data from 1,500 large retailers
BLUEKAI
Data management platform and data marketplace
Social Data
Oracle analyzes 700 million messages from social media and blogs daily
Email Addresses
Phone Numbers
User Accounts
Postal Addresses
Mobile IDs
Unique "anonymous" ID
Set-top IDs
IP Addresses
Cookie IDs
Clients can combine their own customer data with profile data from Oracle and its partners, and then share and utilize it in many ways
Profile and sort customers
Find and target people
Personalize content & offers
Measure how people behave
Sell data
Oracle provides data to Facebook
FACEBOOK
> to help the platform better sort and categorize its users
> to let Facebook track what its users are purchasing in stores
> Oracle also sends its clients' customer data to Facebook to find and target them on the platform
Visa
provides "audience data" for profiling, based on 16 billion credit card payments in the US
MasterCard
provides data based on payments from card-holders that can be "associated with cookie populations"
i-Behavior
Has purchase and psycho-graphic data on 339 million people from 5,820 data providers in the US, Canada, UK, France, and Brazil
Experian
Has data on 700 million people in America, Europe and Asia, including financial information and purchase data
Acxiom
Has up to 3,000 data elements about 700 million people, including from the US, UK, and Germany
TransUnion
Has data on 1 billion people in America, Africa, and Asia from 90,000 sources, including financial institutions
Listed as data providers in Oracle's "data directory"
Lotame
Collects profile data about website visitors and mobile app users linked to 3 billion cookies and 1 billion device IDs
PlaceIQ
Collects location data, movement patterns, and activity profiles from 100 million mobile devices
Proxama
Detailed location data collected using Bluetooth beacons placed across buses, trains, stores, malls, stadiums, cinemas
VisualDNA
Psychographic profiles about 500 million users in the US, UK, Germany, Russia, and other regions, in part based on online quizzes

© Cracked Labs CC BY-SA 4.0, April/May 2017. Disclaimer: the mentioned companies typically keep information about their activities secret. This illustration is based on publicly available information, mainly the companies' own statements. Every effort has been made to accurately interpret and represent the companies' activities, but we cannot accept any liability in the case of eventual errors. Sources: Oracle website, press releases, data directory, brochures, presentations. MasterCard website. Acxiom annual report. TransUnion annual report. Lotame website. VisualDNA brochure. Facebook website. ProPublica article. For details about the sources see the report "Corporate Surveillance in Everyday Life".

Figure 3: Oracle and some of its data providers, partners and services.

To identify, link, and match user profiles across different companies, platforms, devices, and contexts in real-time, Oracle uses its "Identity Graph" or "ID Graph". This technology promises to "unify addressable identities across all devices, screens and channels" and to "identify customers and prospects everywhere". It "unites all interactions across various channels to create one addressable consumer profile", by linking several types of identifiers referring to individuals, including postal addresses, email addresses, user accounts for online services, mobile IDs, set-top IDs, and cookie IDs.[465] When data providers or clients want to connect their online, of-

[461] p. 4-1 https://docs.oracle.com/en/cloud/saas/data-cloud/dsmkt/using-oracle-data-cloud.pdf [12.05.2017]
[462] p. 523 http://docs.oracle.com/cloud/latest/marketingcs_gs/OMCDA/OMCDA.pdf [12.05.2017]
[463] https://www.oracle.com/marketingcloud/products/data-management-platform/ecosystem.html [12.05.2017]
[464] p. 3-107 https://docs.oracle.com/en/cloud/saas/data-cloud/dsmkt/using-oracle-data-cloud.pdf [12.05.2017]
[465] http://www.oracle.com/us/corporate/acquisitions/datalogix/general-presentation-2395307.pdf [12.05.2017]

fline, or mobile data to Oracle's data cloud, they can send "match keys" that "identify" their users "in both the online and offline spaces".[466] Oracle will then "synchronize" these match keys to its "network of user and statistical IDs that are linked together in the Oracle ID Graph", which is "used to manage IDs and user attributes".[467] To create these match keys, Oracle's data providers and clients have to convert email addresses, phone numbers, postal addresses, IP addresses[468] and mobile device IDs[469] into pseudonymous codes[470]. These codes are still unique references to certain email addresses, phone numbers, and other identifiers; they are pseudonyms. When a company sends these pseudonyms to Oracle, they are "are mapped to the network of BlueKai anonymous user profiles, anonymous user IDs, and statistical IDs in the Oracle ID graph".[471] Although these ostensibly "anonymous" user IDs still refer to certain individuals and can be used to recognize and single out individuals[472], Oracle no longer considers the processed data as "personally identifiable information".[473]

Oracle also partners with large online platforms. For example, it provides data to Facebook in order to help the platform better sort and categorize its users with data collected from beyond Facebook,[474] as well as to track its user's purchases in stores.[475] In addition, Oracle provides data about its clients' customers to Facebook in order to find and target these customers on Facebook.[476]

Conversely, many companies provide consumer data to Oracle. Nearly 100 companies are listed as **third-party data providers** in Oracle's "data directory". Examples of companies that provide different types of consumer data to Oracle's data cloud[477] include:

- Beside other consumer data brokers such as **Acxiom**, which provides behavioral, demographic, financial, political, retail and other data[478], all three large credit reporting

[466] p.4-7 https://docs.oracle.com/en/cloud/saas/data-cloud/dsmkt/using-oracle-data-cloud.pdf [15.05.2017]
[467] Ibid, p. 4-8
[468] Ibid. p.4-7
[469] Ibid. p. 4-20
[470] Ibid. p. 4-7; For example, clients can send "Oracle Hashed IDs (oHashes)", which are "MD5 and SHA-256 hashed email addresses or phone numbers that have been automatically generated from raw, personally identifiable information", "Encrypted hashed UUIDs", which are "Encrypted hashed email addresses, phone numbers, physical addresses, and client account numbers", or "IP Addresses" ("The user's IP address, which is collected from the IP header and encrypted")
[471] Ibid. p. 4-48
[472] Ibid. p. 4-27; the "Oracle ID graph seamlessly pulls together the many IDs across marketing channels and devices that comprise a given person, enabling marketers to tie their interactions to an actionable customer profile. The Oracle ID graph helps marketers connect identities across disparate marketing channels and devices to one customer"
[473] E.g. "No personally identifiable information (PII) may be sent to BlueKai or stored in the BlueKai platform. All IDs derived from PII must be hashed in the browser or on your servers before being sent to BlueKai" (https://docs.oracle.com/en/cloud/saas/data-cloud/dsmkt/using-oracle-data-cloud.pdf)
[474] https://www.propublica.org/article/facebook-doesnt-tell-users-everything-it-really-knows-about-them
[475] https://facebookmarketingpartners.com/marketing-partners/datalogix/ [16.04.2017]
[476] Ibid.
[477] Oracle Data Cloud (2017): Data Directory. Available at: http://www.oracle.com/us/solutions/cloud/data-directory-2810741.pdf [13.05.2017]
[478] Ibid.

agencies[479] are listed in Oracle's data directory. **Experian**, which has marketing data on 700 million people in America, Europe and Asia,[480] provides data on finances, purchases, mortgages, property, psychographics, and more. Similarly, **Transunion**, which has data on 1 billion people in America, Africa and Asia from 90,000 sources[481], provides financial and other data.[482] **IXI**, a subsidiary of the credit reporting agency **Equifax**, provides credit, financial, mortgage, retail and other data.[483] KBM Group's subsidiary **i-Behavior**, which belongs to the advertising and marketing giants Wunderman and WPP[484] collects purchase data on 339 million people from 5,820 data contributors in the US, Canada, UK, France and Brazil and provides financial, political, retail, and other data.[485]

- In addition to traditional consumer data brokers, many of the data companies that have emerged in recent years also appear as data providers in Oracle's data directory.[486] **VisualDNA**, for example, has psychographic profiles on 500 million users in the US, UK, Germany, Russia, and other regions, in parts collected through online quizzes.[487] **Lotame**, one of the largest independent data management platforms, collects extensive profile data about website visitors and mobile app users which is linked to 3 billion cookies and 1 billion device IDs.[488] **PlaceIQ**, which collects location data, movement patterns, and activity profiles from 100 million mobile devices[489], provides "real-world visitation" data to Oracle. **Proxama**'s "transport-based mobile proximity services" provide location data collected using Bluetooth beacons placed across buses, trains, stores, malls, stadiums, and cinemas.[490]

Both Visa and MasterCard are listed as data providers, too. **Visa**, which provides "audience data" based on 16 billion credit-card payments in the US, combines its transactional data with demographic, purchase and other data from Oracle. The company emphasizes that it "aggregates and de-identifies all transactional data output".[491] **MasterCard**, in contrast, seems to go one step further, explaining that its data gets "associated with cookie populations" through a

[479] Experian, Equifax and TransUnion. See e.g. http://www.myfico.com/credit-education/questions/why-are-my-credit-scores-different-for-3-credit-bureaus/

[480] Experian Brochure "Discover Experian 2016" https://www.experianplc.com/media/2744/discover-experian-fy17.pdf [13.05.2017]

[481] TransUnion Annual Report 2015 http://s21.q4cdn.com/588148537/files/doc_financials/2015/YE/TRU-2015-Annual-Report-FINAL.PDF

[482] Oracle Data Cloud (2017): Data Directory

[483] Ibid.

[484] https://www.i-behavior.com/about-us/our-history/ [13.05.2017]

[485] Oracle Data Cloud (2017): Data Directory

[486] Ibid.

[487] https://d3i35f0m1a0vwn.cloudfront.net/wp-content/uploads/2016/02/VisualDNA-Data-Jan-161.pdf [16.04.2017]

[488] https://www.lotame.com/resource/cross-device/ [16.04.2017]

[489] Oracle Data Cloud (2017): Data Directory

[490] Ibid.

[491] Ibid.

"double blind matching process"[492], which suggests that purchases might get directly linked with an individual's online activities. Overall, MasterCard's professional services division offers "actionable intelligence based on 95 billion anonymized, real transactions from 2 billion cardholders in 210 countries worldwide" in order to "forecast consumer behavior" and help companies "make better decisions with real-time intelligence".[493]

6.3 Examples of data collected by Acxiom and Oracle

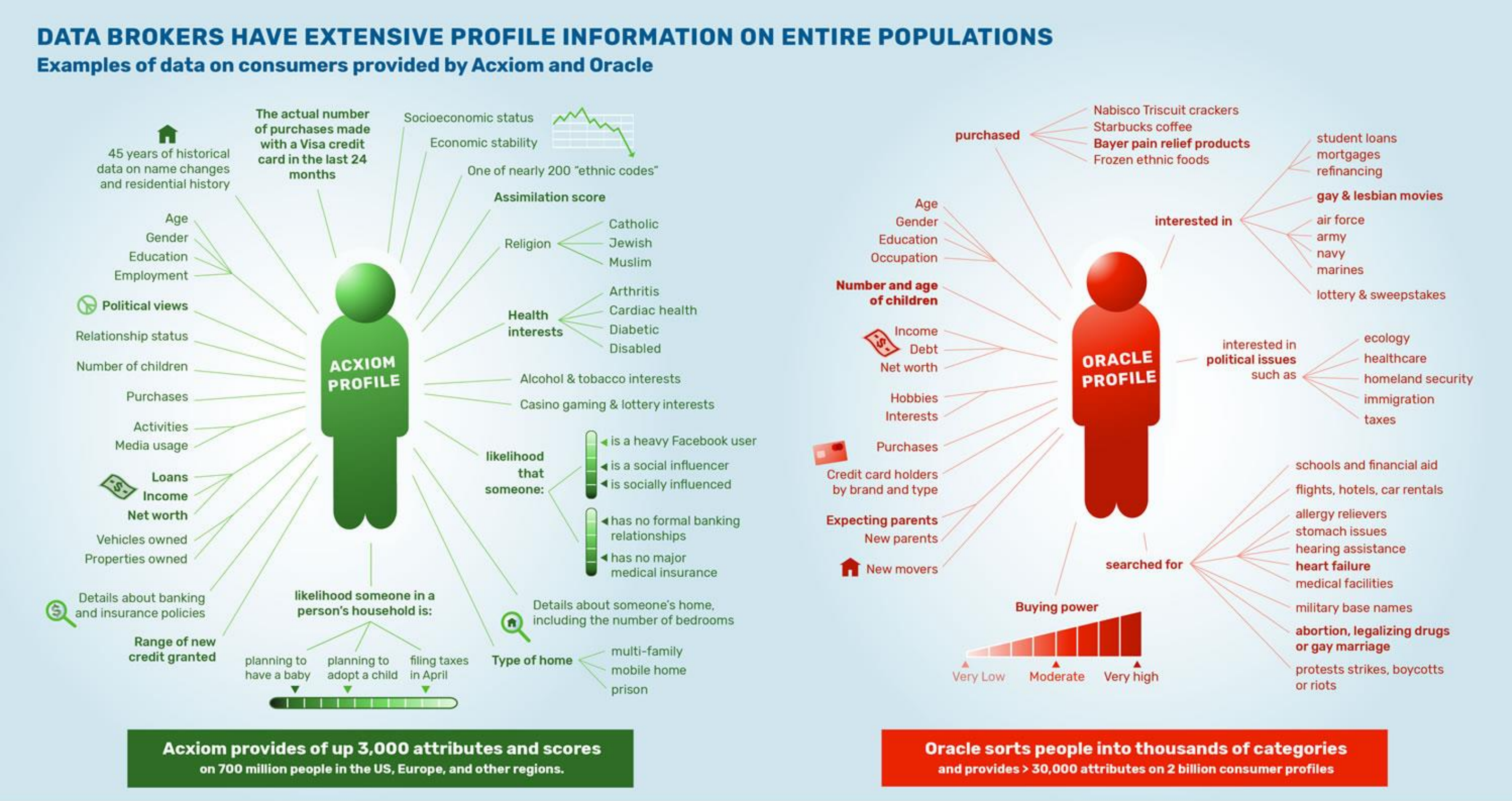
Figure 4 shows examples of data on consumers provided by Acxiom and Oracle. [494] [495]

[492] Ibid.

[493] http://www.mastercardadvisors.com/information-services.html [16.04.2017]

[494] Acxiom provides thousands of attributes and scores about consumers, including: name, postal address, zip, city (https://developer.myacxiom.com/code/api/data-bundles/bundle/postalContact), email (https://developer.myacxiom.com/code/api/data-bundles/bundle/emailContact), phone number (https://developer.myacxiom.com/code/api/data-bundles/bundle/phoneContact), age, gender, education level, employment status, occupation, political party, marital status, number of children (https://developer.myacxiom.com/code/api/data-bundles/bundle/basicDemographics), ethnicity, assimilation index, religion (https://developer.myacxiom.com/code/api/data-bundles/bundle/eTechDemographics), activities, interests, media usage, purchases (see also menu on the left side: https://developer.myacxiom.com/code/api/data-bundles/level1bundle/spending), loans (https://developer.myacxiom.com/code/api/data-bundles/bundle/mortgagesAndLoans), income, net worth, economic stability (https://developer.myacxiom.com/code/api/data-bundles/bundle/investmentsAndAssets), socioeconomic status (https://developer.myacxiom.com/code/api/data-bundles/bundle/areaDemographicsGS), details about banking & insurance policies (https://developer.myacxiom.com/code/api/data-bundles/level1bundle/financial), details about vehicles owned (https://developer.myacxiom.com/code/api/data-bundles/level1bundle/vehicles), details about properties owned (https://developer.myacxiom.com/code/api/data-bundles/level1bundle/property), alcohol interests (https://developer.myacxiom.com/code/api/data-bundles/bundle/alcohol), tobacco interests (https://developer.myacxiom.com/code/api/data-bundles/bundle/tobacco), casino gaming & lottery interests (https://developer.myacxiom.com/code/api/data-bundles/bundle/gambling), health interests like arthritis/mobility, cardiac health, diabetic, disabled (https://developer.myacxiom.com/code/api/data-bundles/bundle/healthAndMedical), details about someone's home, including the "number of bedrooms"; the type of the home e.g. "multi-family residential", "mobile home" or "prison" (https://developer.myacxiom.com/code/api/data-bundles/bundle/propertyDescription), information about whether someone "has been reported as deceased" (https://developer.myacxiom.com/code/api/data-bundles/bundle/basicDemographics), the "actual number of purchases made with a Visa credit card in the last 24 months", the "max range of new credit granted for an individual" (https://developer.myacxiom.com/code/api/data-bundles/bundle/creditAndBankCards), the "likelihood someone has no formal banking relationships" (https://developer.myacxiom.com/code/api/data-bundles/bundle/bankingAndServices), the "likelihood to have no major medical insurance" (https://developer.myacxiom.com/code/api/data-bundles/bundle/healthAndMedical), information about the "presence of an expectant parent in the household", a "divorce in the household in the last 12 months" or a "recent college graduate in the household"; the "likelihood that someone in the household" is "planning to have a baby", is "planning to adopt a child", "will become a grandparent" or "will change careers or jobs" (https://developer.myacxiom.com/code/api/data-bundles/bundle/basicDemographics), the "likelihood of individual" to "be a heavy user of Facebook", to "respond to posts on social media", to "have influence in the social media universe" or to "be influenced by social media" (https://developer.myacxiom.com/code/api/data-bundles/bundle/socialMedia) [11.05.2017]

[495] Oracle sorts consumers into thousands of categories, including about age, gender, number and age of children, education, occupation, number of bedrooms, income, net worth, hobbies, interests, purchases, loans, new parents, new movers, credit card holders by brand and type; people who are interested in gay & lesbian films; people who are interested in political issues like ecology, healthcare, immigration, taxes, homeland security, jobs, welfare; people who are interested in air force, army, marines, navy, military bases; people who purchased Bayer pain relief products; people who searched for flights, hotels, car rentals; people who searched for schools or financial aid (http://www.oracle.com/us/products/applications/audience-guide-3034880.pdf), expecting parents; social influencers; people who searched for allergy relievers, stomach issues, hearing assistance, sleep issues, medical facilities; people who searched for cardiology and heart failure; people who searched for military base names; people who searched for abortion, legalizing drugs and gay marriage; people who searched for protests, strikes, boycotts and riots (https://docs.oracle.com/cloud/latest/marketingcs_gs/OMCDA/Resources/xls/TaxonomyChangesReport2017-01-31.xlsx) [12.05.2017]



DATA BROKERS HAVE EXTENSIVE PROFILE INFORMATION ON ENTIRE POPULATIONS
Examples of data on consumers provided by Acxiom and Oracle
45 years of historical data on name changes and residential history
The actual number of purchases made with a Visa credit card in the last 24 months
Socioeconomic status
Economic stability
One of nearly 200 "ethnic codes"
Assimilation score
Age
Gender
Education
Employment
Political views
Relationship status
Number of children
Purchases
Activities
Media usage
Loans
Income
Net worth
Vehicles owned
Properties owned
Details about banking and insurance policies
Range of new credit granted
ACXIOM PROFILE
Religion
Catholic
Jewish
Muslim
Health interests
Arthritis
Cardiac health
Diabetic
Disabled
Alcohol & tobacco interests
Casino gaming & lottery interests
likelihood that someone:
is a heavy Facebook user
is a social influencer
is socially influenced
has no formal banking relationships
has no major medical insurance
likelihood someone in a person's household is:
planning to have a baby
planning to adopt a child
filing taxes in April
Details about someone's home, including the number of bedrooms
Type of home
multi-family
mobile home
prison
Acxiom provides of up 3,000 attributes and scores
on 700 million people in the US, Europe, and other regions.
purchased
Nabisco Triscuit crackers
Starbucks coffee
Bayer pain relief products
Frozen ethnic foods
Age
Gender
Education
Occupation
Number and age of children
Income
Debt
Net worth
Hobbies
Interests
Purchases
Credit card holders by brand and type
Expecting parents
New parents
New movers
ORACLE PROFILE
interested in
student loans
mortgages
refinancing
gay & lesbian movies
air force
army
navy
marines
lottery & sweepstakes
interested in political issues such as
ecology
healthcare
homeland security
immigration
taxes
searched for
schools and financial aid
flights, hotels, car rentals
allergy relievers
stomach issues
hearing assistance
heart failure
medical facilities
military base names
abortion, legalizing drugs or gay marriage
protests strikes, boycotts or riots
Buying power
Very Low
Moderate
Very high
Oracle sorts people into thousands of categories
and provides > 30,000 attributes on 2 billion consumer profiles

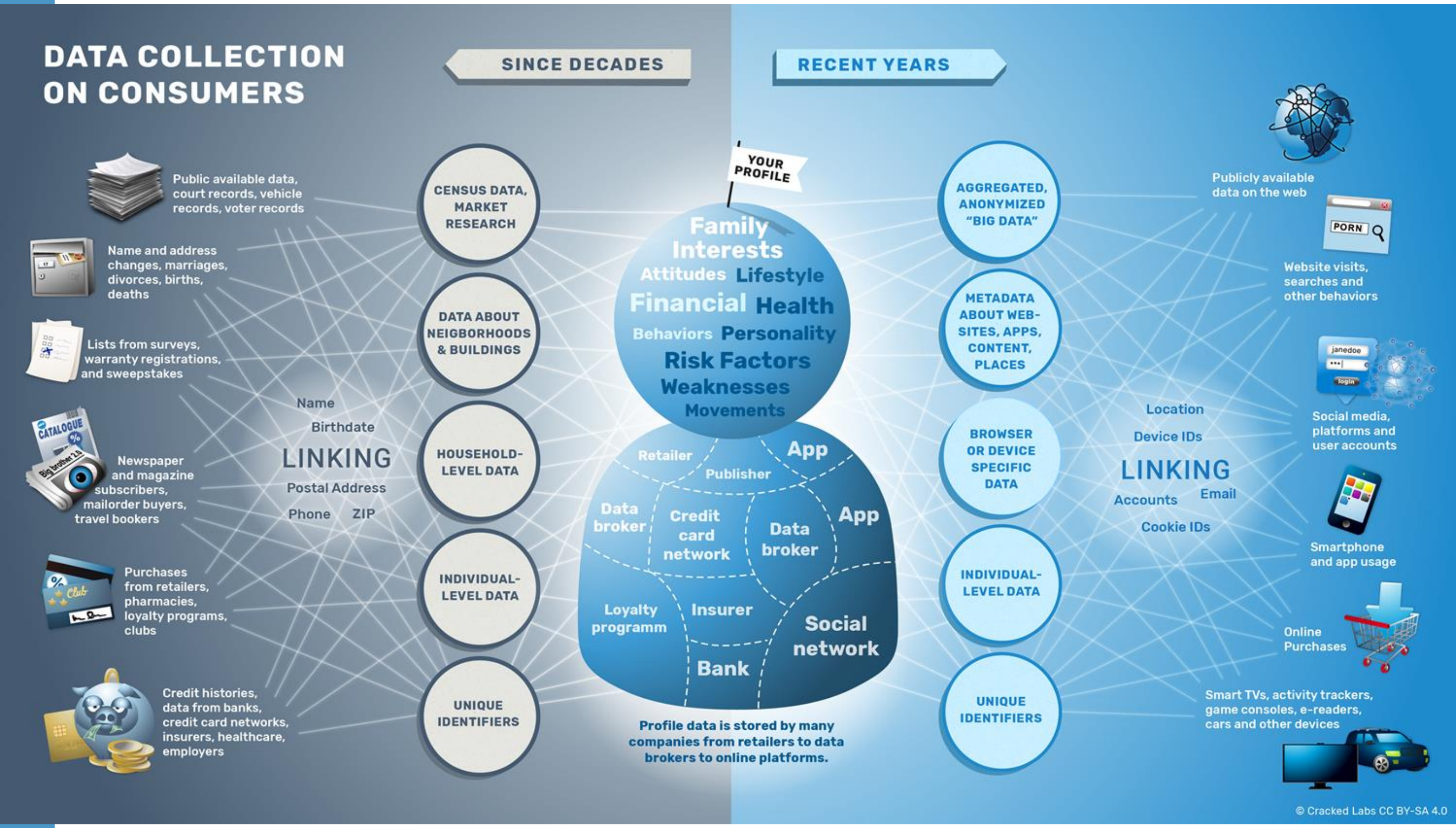
© Cracked Labs CC BY-SA 4.0, April/May 2017. Disclaimer: the mentioned companies typically keep information about their activities secret. This illustration is based on publicly available information by Acxiom and Oracle. Every effort has been made to accurately interpret and represent the companies' activities, but we cannot accept any liability in the case of eventual errors. Sources: Acxiom annual reports, developer website (API docs), Oracle press release, help center website, audience playbook, taxonomy updates for January, 2017 (Excel document). For details about the sources see the report "Corporate Surveillance in Everyday Life".

Figure 4: Examples of data on consumers provided by Acxiom and Oracle.

7. Key Developments in Recent Years

7.1 Networks of digital tracking and profiling

In recent years, pre-existing practices of commercial consumer data collection have rapidly evolved into pervasive networks of digital tracking and profiling. Today, a vast and complex landscape of corporate players continuously monitors the lives of billions. A first big step into systematic consumer surveillance occurred in the 1990s through database marketing, loyalty programs, and advanced consumer credit reporting. After the rise of the Internet and online advertising in the early 2000s and the advent of social networks, smartphones, and programmatic advertising in the late 2000s, we are now witnessing a new phase defined by the traditional consumer data industry joining forces with the new data industry.



DATA COLLECTION ON CONSUMERS
SINCE DECADES
RECENT YEARS
Public available data, court records, vehicle records, voter records
Name and address changes, marriages, divorces, births, deaths
Lists from surveys, warranty registrations, and sweepstakes
CATALOGUE
Big brother
Newspaper and magazine subscribers, mailorder buyers, travel bookers
Purchases from retailers, pharmacies, loyalty programs, clubs
Club
Credit histories, data from banks, credit card networks, insurers, healthcare, employers
Name
Birthdate
LINKING
Postal Address
Phone
ZIP
CENSUS DATA, MARKET RESEARCH
DATA ABOUT NEIGBORHOODS & BUILDINGS
HOUSEHOLD-LEVEL DATA
INDIVIDUAL-LEVEL DATA
UNIQUE IDENTIFIERS
YOUR PROFILE
Family
Interests
Attitudes
Lifestyle
Financial
Health
Behaviors
Personality
Risk Factors
Weaknesses
Movements
Retailer
App
Publisher
Data broker
Credit card network
Data broker
App
Loyalty programm
Insurer
Social network
Bank
Profile data is stored by many companies from retailers to data brokers to online platforms.
AGGREGATED, ANONYMIZED "BIG DATA"
METADATA ABOUT WEBSITES, APPS, CONTENT, PLACES
BROWSER OR DEVICE SPECIFIC DATA
INDIVIDUAL-LEVEL DATA
UNIQUE IDENTIFIERS
Location
Device IDs
LINKING
Accounts
Email
Cookie IDs
Publicly available data on the web
PORN
Website visits, searches and other behaviors
janedoe
login
Social media, platforms and user accounts
Smartphone and app usage
Online Purchases
Smart TVs, activity trackers, game consoles, e-readers, cars and other devices
© Cracked Labs CC BY-SA 4.0

Figure 5: Different levels, realms and sources of consumer data collection

Companies have never only used data collected on a strictly personal level to label, profile, sort and rank individuals. Information about neighborhoods, buildings, and registered vehicles has been used to characterize consumers at a household address level for decades. Today, postal addresses continue playing their long-standing role as a key attribute that enables combining and linking data about consumers from different sources. In addition to information about the types of buildings people live in, companies now profile people with information about the types of websites they surf, metadata about the videos they watch, the apps they use, and the geographic locations they visit. The scale and depth of behavioral data streams generated by all manner of everyday life activities, such as web browsing, social media use, and device usage by smartphones, tablets, PCs, and many other "smart" devices has changed rapidly. The online advertising industry is the pioneering force in developing sophisticated technologies that combine and link digital profiles across different companies such as data brokers and data aggregators; it has made created the infrastructure that enables real-time profiling of individuals across different types of technologies, platforms, and customer databases.

One single interaction of a consumer, such as a website visit, might trigger a wide range of data flows and a chain of hidden events across many different parties. Profile data distributed across several services gets dynamically interlinked and combined in order to make many automated decisions, both trivial and consequential, about people every day. While some actors – such as the large platforms, data brokers, and other companies with hundreds of millions of customers – have a unique position in terms of the scale and depth of their consumer profiles, the data used to make decisions on people is often not held in one place, but assembled from several places in the moment and as needed. While a wide range of data, analytics, and marketing services provide the means to deploy these powerful technologies, businesses in all industries and their consumer databases play an equally important role in making this happen.

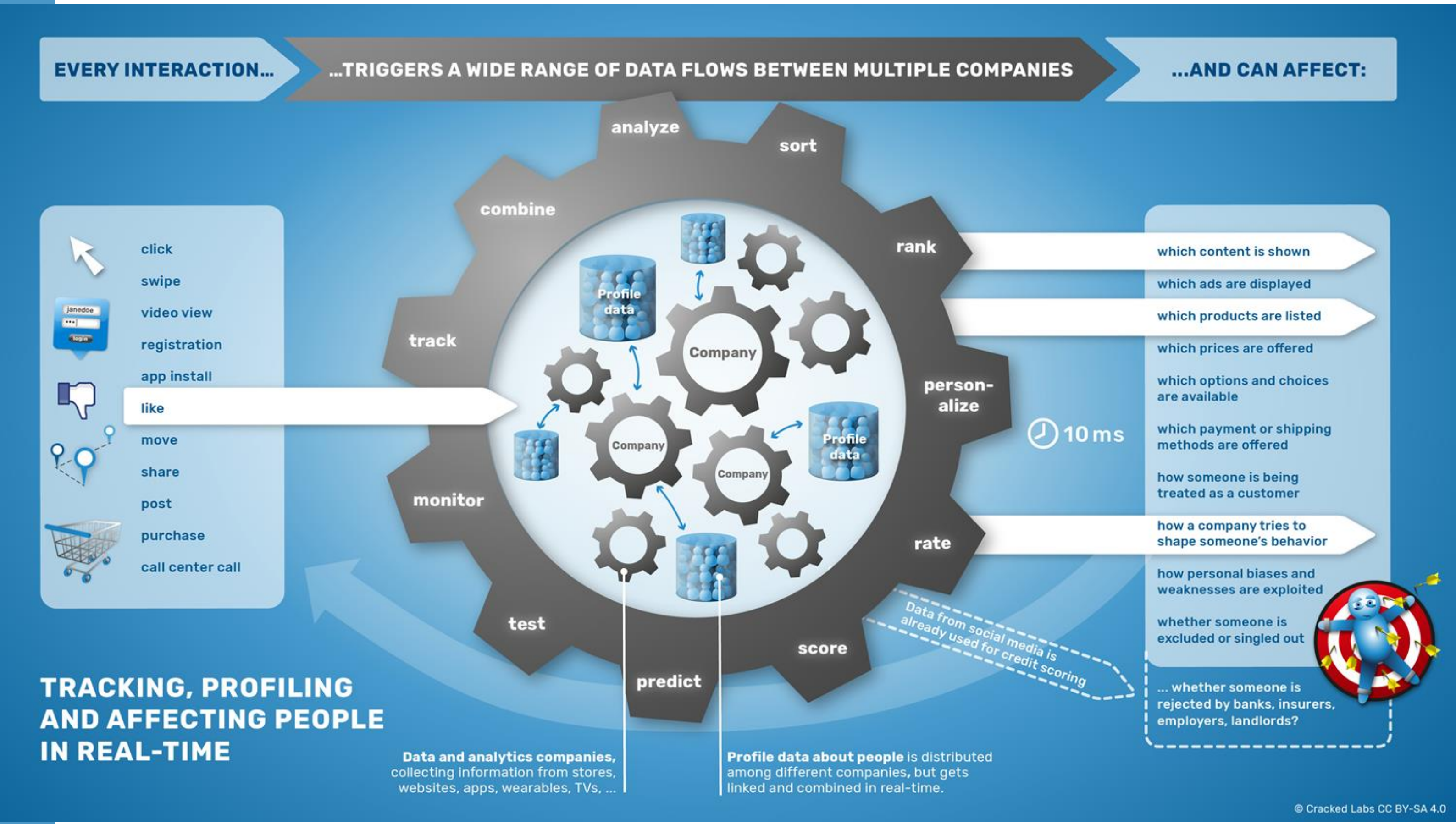

EVERY INTERACTION...
...TRIGGERS A WIDE RANGE OF DATA FLOWS BETWEEN MULTIPLE COMPANIES
...AND CAN AFFECT:
click
swipe
video view
registration
app install
like
move
share
post
purchase
call center call
janedoe
login
analyze
sort
combine
rank
track
person-alize
monitor
rate
test
predict
score
Profile data
Company
Company
Company
Profile data
10 ms
which content is shown
which ads are displayed
which products are listed
which prices are offered
which options and choices are available
which payment or shipping methods are offered
how someone is being treated as a customer
how a company tries to shape someone's behavior
how personal biases and weaknesses are exploited
whether someone is excluded or singled out
... whether someone is rejected by banks, insurers, employers, landlords?
Data from social media is already used for credit scoring
TRACKING, PROFILING AND AFFECTING PEOPLE IN REAL-TIME
Data and analytics companies, collecting information from stores, websites, apps, wearables, TVs, ...
Profile data about people is distributed among different companies, but gets linked and combined in real-time.
© Cracked Labs CC BY-SA 4.0

Figure 6: Tracking, profiling and affecting people in real-time.

While this goes on behind the scenes, consumers are left in the dark, with little understanding of these technological processes, and with even less of a view on the short- and long-term impacts that these mechanisms have on their lives. One reason for this certainly lies in the high levels of complexity and abstraction in play. Perhaps more importantly, though, companies make no effort to improve transparency or understanding; on the contrary, they inform consumers incompletely, inaccurately, or not at all, often employing ambiguous, misleading, and obfuscating language. Whether in user interfaces or in contracts, the disclosures that do exist – such as privacy policies and terms of service – are difficult to understand, obscure, and use hypothetical language. Moreover, companies often systematically trick consumers into data contracts. As soon as privacy advocates, consumer rights organizations, regulators, scholars, and journalists express their concerns and ask for more information, companies decline to answer, arguing that their data practices constitute trade secrets and must therefore be protected and kept proprietary.[496]

7.2 Large-scale aggregation and linking of identifiers

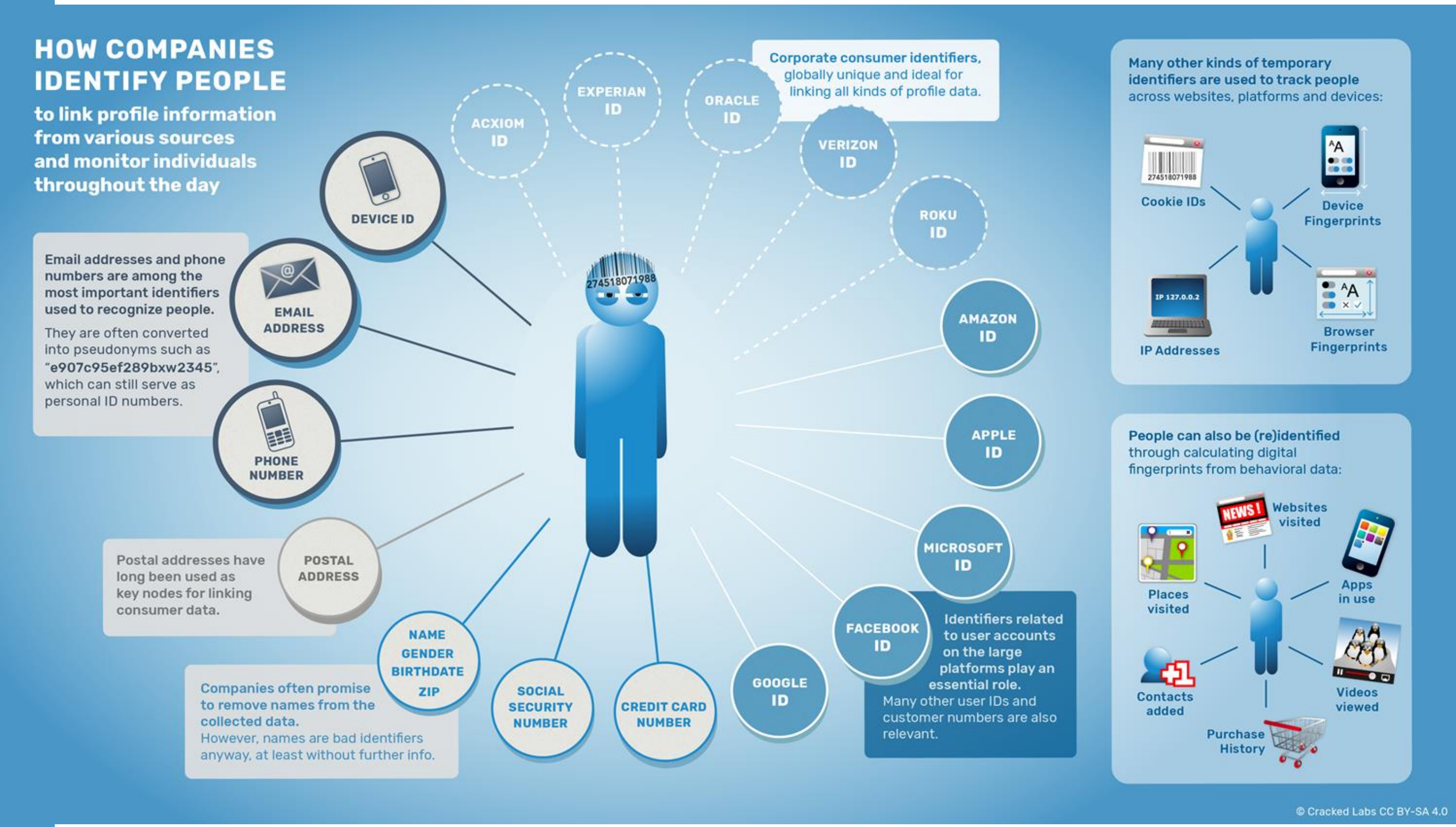
One of the major developments in recent years is that companies can now address, identify, and recognize consumers on an individual level across a growing number of disparate situations in their lives. Therefore, they increasingly aggregate data suitable for combining, linking, and cross-referencing profile data from different sources, such as email addresses and phone numbers. After Facebook started allowing other firms to upload their customer data to its platform in 2012, and began partnering with large consumer data brokers in 2013 in order to combine online and offline data, many other companies followed suite. Today, many platforms and firms provide ways to integrate, enhance, and utilize digital profiles across different technologies and contexts.

Basic individual attributes such as name, gender, birthdate, postal address, and ZIP code still play a crucial role. Any combination of two or three of these attributes can be used to uniquely identify individuals with relatively high certainty.[497] For decades, both consumer reporting agencies and direct marketing companies amassed and maintained databases containing these basic identifiers on entire populations, including name variants and previous addresses, in order to help other businesses organize and manage their consumer data.

Today, large numbers of companies are busy aggregating a wide range of digital identifiers pointing to billions of consumers. The most important identifiers used to link profiles across different services, platforms, devices, and life contexts are **email addresses**, **phone numbers**, and those that refer to **devices** such as smartphones. Semi-temporary identifiers such as cookie IDs, which are attached to users surfing the web get synchronized between different services and interlinked with other identifiers. Also, the **user account IDs** of the large platforms such as

[496] Christl and Spiekermann (2016), p. 121-123
[497] Ibid., p. 21-23

Google, Facebook, Apple, and Microsoft play an important role. Google, Apple, Microsoft and Roku also provide so-called “advertising IDs” for users, which are now widely used to match and link digital profiles on an individual level instead of relying on hardware device identifiers.[498] [499]



HOW COMPANIES IDENTIFY PEOPLE

to link profile information from various sources and monitor individuals throughout the day

Email addresses and phone numbers are among the most important identifiers used to recognize people.

They are often converted into pseudonyms such as “e907c95ef289bxw2345”, which can still serve as personal ID numbers.

Postal addresses have long been used as key nodes for linking consumer data.

Companies often promise to remove names from the collected data. However, names are bad identifiers anyway, at least without further info.

Corporate consumer identifiers, globally unique and ideal for linking all kinds of profile data.

Identifiers related to user accounts on the large platforms play an essential role. Many other user IDs and customer numbers are also relevant.

Many other kinds of temporary identifiers are used to track people across websites, platforms and devices:

People can also be (re)identified through calculating digital fingerprints from behavioral data:

© Cracked Labs CC BY-SA 4.0

Figure 7: How companies identify consumers and link profile information about them.

Similarly, some large data companies such as Acxiom,[500] Experian,[501] and Oracle[502] have introduced their own **proprietary identifiers** for people, which are used to link their extensive consumer profile information with data managed by other companies, and then to link it with the advertising data ecosystems around the globe. Often, these companies use two different identifiers, one for data that they see as “personally-identifiable information”, such as names, and one that is used for other digital profile data. However, both identifiers are linkable. This should not be underestimated; the ability to link the private “population registers” that data

[498] Christl and Spiekermann (2016), p. 92
[499] https://sdkdocs.roku.com/display/sdkdoc/Roku+Advertising+Framework []
[500] See section 6.1
[501] See section 5.5
[502] See section 6.2

companies have maintained for decades with both corporate customer databases and the digital world brings pervasive consumer surveillance to a completely new level.

Cross-device matching. Other companies provide services to link digital profiles across platforms, devices and contexts by running data cooperatives or by applying machine learning technologies to large amounts of data in order to ascertain which devices and user fingerprints belong to the same person. For example, Adobe runs a service through which participating member companies contribute knowledge about login IDs and other user data. Subsequently, the company is able to link groups of devices to persons and households, which allows other companies to "measure, segment, target and advertise directly to individuals across all of their devices".[503] In contrast, Tapad, which has been acquired by the Norway-based telecom giant Telenor, analyzes "billions of non-PII data points" and uses "behavioral and relationship-based patterns" to statistically discover the probability of certain computers, tablets, phones, and other devices belonging to one person.[504] Companies have even started to use hidden ultrasonic audio signals, which cannot be heard, but are played by the speakers on one device (e.g. a computer, phone or TV) and then recorded and recognized by another device (e.g. a smartphone app) in order to track exactly which TV commercials and programs someone has seen across devices.[505]

7.3 "Anonymous" recognition

When data brokers and online companies write about their services on their websites, in brochures, and in their privacy policies, they often claim to only process "anonymized" or "de-identified" data on individuals. Indeed, they tend to remove names and convert email addresses and phone numbers into unique alphanumeric string codes by using, for example, a cryptographic hash function such as MD5.[506] In theory, hashing is a one-way operation and cannot be reversed. However, in most cases it is misleading to consider this data as "de-identified" or even "anonymized" for several reasons. First, these hashed identifiers still persistently refer to unique individuals and therefore cannot be considered as anonymized, but rather should be understood as **pseudonyms**.[507] Second, most companies use identical and deterministic methods to create – or calculate – these unique codes; therefore, they can match and link profiles as soon as one of these pseudonymous identifiers appears within the digital data ecosystem.

[503] https://marketing.adobe.com/resources/help/en_US/mcdc/mcdc-overview.html [02.05.2017]

[504] https://www.tapad.com/device-graph [02.05.2017]

[505] See e.g. Calabrese, C., McInnis, K. L., Hans, G. S., Norcie, G. (2015): Comments for November 2015. Workshop on Cross-Device Tracking. Center For Democracy & Technology. Available at: https://cdt.org/files/2015/10/10.16.15-CDT-Cross-Device-Comments.pdf

[506] Christl and Spiekermann (2016), p. 90

[507] 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". When additional information, for example how names relate to pseudonyms, is known, pseudonymity can be easily reverted. For more details see Christl and Spiekermann (2016), p. 21-23

Third, many companies have billions of records that map email addresses and phone numbers to these unique codes which undermines any claims of "anonymization" at all.

The third aspect aside, some of the companies collecting data might indeed not know anything other than that a person connected to such an identifier has searched for e.g. “breast cancer”. In an extreme case, neither the website collecting and sharing this information with other services, nor the company targeting individuals who searched for “breast cancer”, knows much more about a person – yet is still able to effectively target her. Nevertheless, the person’s identifier is irrevocably connected with the “breast cancer” search term across several databases. Whenever this person interacts with a digital service, this identifier remains inextricably linked to her previous search for “cancer”. In this way, even though companies use “hashed” or “encrypted” email addresses and phone numbers, they can recognize consumers again upon their using another service linked with the same email addresses or phone numbers. Marketing scholar Joseph Turow states that when 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’”.[508]

In his paper “Singling out people without knowing their names” privacy scholar Frederik Borgesius concludes that a name is “merely one of the identifiers that can be tied to data about a person, and it is not even the most practical identifier” in today’s online tracking economy.[509] Several scholars, privacy regulators, and even some industry representatives share the same conclusion. If a unique identifier consistently links digital profiles to the same person across the data ecosystem, these digital profiles should be considered personal data regardless of whether the identifiers are stored in clear text or hashed.[510]

7.4 Analyzing, categorizing, rating and ranking people

A list of identifiers might already provide certain insights, especially when said list indicates that the individuals it refers to share certain characteristics. For example, a data collector might have a list of names and addresses from a magazine subscriber database, or a list of cookie IDs referring to people that have visited a certain website. Similarly, a data collector might know which people are living at a certain address and assign them to a household ID. Much of today’s corporate data collection and analysis happens on such a seemingly trivial, but in fact very powerful level. Companies, however, try to extract many additional insights from the collected data by using mathematical and statistical methods, data mining, and predictive

[508] Joseph Turow (2011): The Daily You. Cited from: Senate Committee on Commerce, Science, and Transportation (2013, p. 32)
[509] 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). Available at: http://ssrn.com/abstract=2733115
[510] For a more detail explanation see: Christl and Spiekermann (2016), p. 90-91

analytics.[511] However, most companies keep the details of how they analyze consumer data secret.

Scientific studies have shown that many kinds of personal characteristics can be inferred from transactional and behavioral data such as web searches, browsing histories, product purchases, Facebook likes, and viewing or listening behaviors. For example, personal attributes such as ethnicity, religious and political views, relationship status, sexual orientation, and alcohol, cigarette, or drug use as well as personality traits such as emotional stability, life satisfaction, impulsivity, depression, and "sensationalist interest" can be inferred reasonably accurately from Facebook likes. Personality traits can also be inferred from information about the websites someone has visited, as well as from phone call records and mobile app usage. Information about someone's occupation and educational level can be inferred from the browsing history. Canadian researchers have even successfully predicted emotional states such as confidence, nervousness, sadness, and tiredness by analyzing the rhythm of typing patterns on a computer keyboard.[512]

These predictions are based on **statistical correlations** with certain probability levels. Although they predict these attributes and personality traits significantly above chance, they are not accurate in every case. Nevertheless, companies use similar methods to categorize, sort, rate, and rank people. For example, without actually using data related to financial transactions, such as whether somebody pays off the bills on time, data based on the timing and frequency of phone call records, social media data, or web searches are already used to predict an individual's creditworthiness.[513] Companies even claim to use facial recognition technologies to predict smoking habits and mortality risk[514], as well as whether someone might be a terrorist[515]. LexisNexis Risk Solutions offers a health scoring product that predicts an individual's health risk based on vast amounts of consumer data, including purchase activities.[516]

These are not extreme outliers. Platforms such as **Facebook** use similar methods to categorize their users by, for example, ethnicity[517] or political views.[518] An investigation of ProPublica found 52,000 unique attributes Facebook uses to categorize its users, about 350 of which are provided by Oracle.[519] Overall, the data broker Oracle claims to provide 40,000 data attributes about hundreds of millions of consumers to other companies.[520] This data originates from many different sources. Most of today's data providers create extensive **taxonomies** by which

[511] Christl and Spiekermann (2016) , p. 11-12

[512] For a more detailed description of these academic studies see: Christl and Spiekermann (2016), p. 13-21

[513] See section 4.2

[514] https://lapetussolutions.com/assets/atto/file/chronos_brochure_v01_6.pdf

[515] http://www.faception.com/ [02.05.2017]

[516] http://www.lexisnexis.com/risk/downloads/literature/health-care/socioeconomic-data-coverages-br.pdf [02.05.2017]

[517] https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race

[518] https://www.theverge.com/circuitbreaker/2016/8/24/12621784/facebook-political-preferences-ads

[519] https://www.propublica.org/article/facebook-doesnt-tell-users-everything-it-really-knows-about-them

[520] https://www.oracle.com/us/assets/lad-2015-ses16178-toledo-2604862.pdf [02.05.2017]

they categorize and group consumers with regard to certain characteristics or (predicted) behaviors, such as “hispanics”, “low income”, “new parents”, “interested in military” or “purchased pain relief products”.[521]

Generally, data and advertising companies either provide each other with **segments**, which are basically lists of recognizable or addressable individuals assigned to these categories[522] or they directly provide profile attributes for specific individuals. The attributes provided might represent actual data collected about consumers, such as their age, educational level, or the “actual number of purchases made with a Visa credit card in the last 24 months”. They might also represent predictive scores that only tell about certain probabilities, such as “is likely against immigration” or “has very likely no formal banking relationships”. **Dynamic real-time scoring** has emerged from “static segments” in recent years, according to Acxiom’s LiveRamp.[523] Once, the process of large-scale corporate data collection from different sources and its utilization was complicated and slow. A few big companies provided “static segments” about consumers, which only contained “on/off” information about whether a person did or did not have certain assigned traits. These segments were compiled and curated by humans, which took several weeks and only represented a one-way data flow from corporate data collectors to data users. Today, thousands of data companies provide real-time data “signals”, which are transformed into scores that predict the probability of whether a person does or does not have certain traits. By applying machine learning and other technologies, calculating scores about consumers happens within milliseconds and includes two-way data flows.[524] Consequently, the lines between corporate data collectors and data users are increasingly blurring.

Accuracy? Except, perhaps, for the companies themselves, nobody really knows how accurately the digital profiles based on large-scale data collection and predictive analytics really represent an individual’s characteristics and behaviors. Some evidence suggests that data used for online[525] and mobile targeting[526] may often be largely inaccurate. One reason for this might be that marketers are “typically seeking only marginal gains in customer sales and acquisition” and therefore often accept that wildly inaccurate data as long as it helps them meet their business goals.[527] Another reason may lie in the fact that most of the technologies used are still at an early stage. Either way, ultimately, profile data is attributed to individuals and used to sort

[521] See chapter 6

[522] More precisely, not lists of individuals, but lists of identifiers. Marketers describe a “segment” e.g. as an " identifiable group of individuals who share similar characteristics or needs and generally respond in a predictable matter to a marketing stimulation": http://www.visualiq.com/resources/marketing-attribution-newsletter-articles/30-attribution-terms-every-business-should-know [02.05.2017]

[523] LiveRamp (2016): IdentityLink Data Store. Providing access to Identity-based data and buyers across the marketing ecosystem. PDF brochure. Personal copy of file with author Wolfie Christl

[524] Ibid.

[525] E.g. https://www.nytimes.com/2016/11/08/us/politics/facebook-ads-campaign.html

[526] E.g. https://adexchanger.com/data-exchanges/half-age-data-mobile-exchanges-inaccurate/

[527] Hoofnagle, Chris Jay (2016): Federal Trade Commission Privacy Law and Policy - Chapter 6 Online Privacy (February 1, 2016). Chris Jay Hoofnagle, Federal Trade Commission Privacy Law and Policy (Cambridge University Press 2016); UC Berkeley Public Law Research Paper No. 2800276, p. 174 Available at: https://ssrn.com/abstract=2800276

and rank them according to its effect on the company's bottom line. As Mark Andrejevic states, to "someone who has been denied health care, employment, or credit, the difference between a probabilistic prediction and a certainty is, for all practical purposes, immaterial".[528] Regardless of whether data is accurate or inaccurate, today's pervasive tracking and profiling ecosystem leads to a massive invasion of privacy and threatens the fundamental rights of individuals.

7.5 Real-time monitoring of behavioral data streams

The technologies currently in place allow for the monitoring and analysis of an individual's behavior in nearly any situation. Companies have always been interested in understanding how their advertising and marketing efforts lead to additional customers, subscribers, registrations, store visits and purchases and, ultimately, affect their bottom lines. Generally speaking, advertisers and marketers want to influence people's behavior and measure the outcomes, ideally at the individual level. To achieve that, companies have made efforts in the past to link a particular piece of advertising or marketing to an eventual purchase. Mail order catalogue companies, for example, printed unique codes on their direct mail pieces and asked their customers to provide these codes when purchasing products. In the past these efforts were sporadic, disjointed, spread across various companies and databases, and often remained incomplete. A TV advertisement, for example, could not be directly measured for its effectiveness in increasing sales.

Today, the **ubiquitous streams of behavioral data** collected at the individual level by myriads of services across many fields of life are linked and utilized to monitor and analyze every interaction of a consumer that might be relevant to a company's customer management and acquisition efforts. A crucial milestone occured in 2012, when Facebook started to link its profile data with information about offline purchases in stores. Provided by Oracle's Datalogix, companies could, for the first time, measure how Facebook ads affect store visits and purchases.[529] Today, companies try to capture as many "touchpoints" across the whole "customer journey", from online and mobile to in-store purchases, direct mail, TV ads, and call center calls, as possible.

Tracking the consumer journey. The online data management platform Krux, which is owned by the leading customer data management company Salesforce, describes on its website how it might track and associate very different behaviors across a person's day, including interactions that involve desktop computers, mobile phones, tablets, TVs, and store visits. A person may search for a product on the web, "like" a product and share it on social media, view a TV commercial, find a coupon on the web, and use it at a grocery store for a purchase. Krux explains that by monitoring the whole of such a "consumer journey" and getting a "granular un-

[528] Andrejevic, Mark (2014): The Big Data Divide. International Journal of Communication 8 (2014), p. 1677. Available at: http://ijoc.org/index.php/ijoc/article/download/2161/1163

[529] https://www.forbes.com/sites/kashmirhill/2012/09/26/facebook-is-tracking-what-users-buy-in-stores-to-see-whether-its-ads-work

derstanding of individual interests and behaviors", companies can "support [their] best customers at the most critical points in their journey" and "influence high-potential prospects exactly when they're ready to engage".[530]

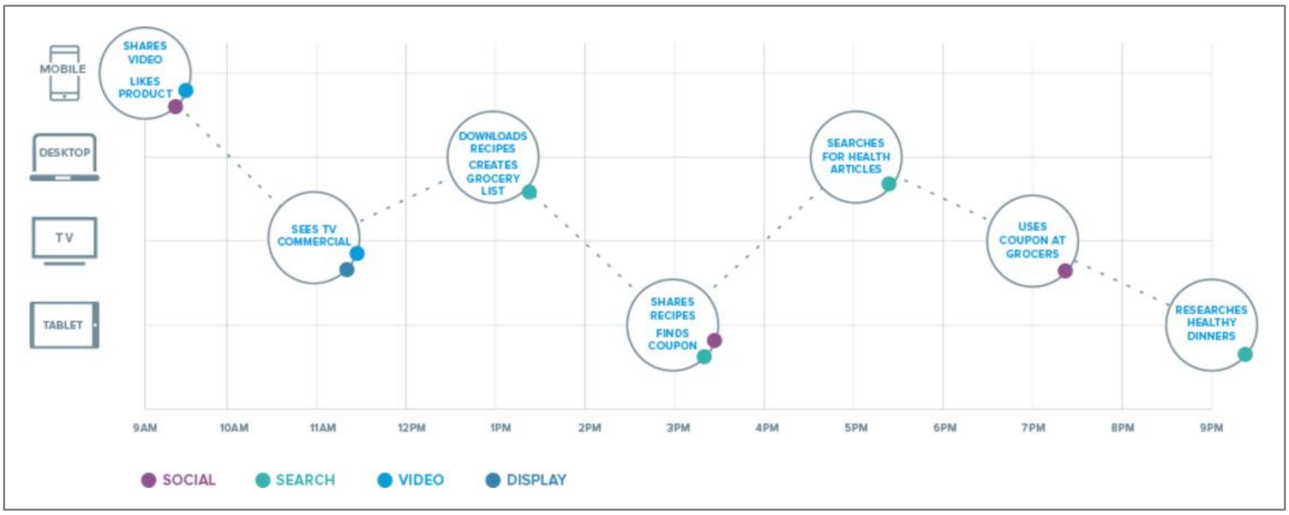


Figure 8: How the data management platform Krux explains to track the "consumer journey" of an individual across channels and devices. Source: Krux Website (http://www.krux.com/data-management-platform-solutions/customer-journeys)

The connected event stream. Similarly, the customer data management company Merkle explains how their "connected event stream" may trace all sorts of behaviors of a person looking for auto insurance. They describe how an "anonymous consumer" sees an ad on a car shopping site, searches for "insurer x" with Google, clicks on an "organic search link", and then browses the website of "insurance x" on the work PC. The next day, the consumer researches insurance offers on a different website, navigates back to the website of "insurance x" and then starts to fill in the application on the home PC. Two days later, the person calls at the insurer's call center and leaves personal information, including an email address. The company sends an email including a link to its website with an embedded account number, which the person opens on a smartphone. Later the person returns to the insurer's website using the work PC. Merkle explains that their "connected recognition linking" makes it possible to "connect nearly all of these interactions across multiple devices back to a single master identifier" in order to "maintain a traceable 1:1 relationship across classic CRM data and online engagement".[531]

These examples show how companies have moved from creating static consumer profiles to continuously updating dynamic profiles. They try to capture every interaction, whether relevant or not, including those on websites, platforms, and devices that they do not control them-

[530] http://www.krux.com/data-management-platform-solutions/customer-journeys [14.05.2017]

[531] https://www.merkleinc.com/what-we-do/marketing-technology/merkle-data-management-cloud/connected-recognition [02.05.2017]

selves. By putting these interactions in a temporal relation, profiles become dynamic. As section 5.4 shows in detail, today's customer management platforms allow the definition of complex sets of rules about how to automatically react to certain kinds of consumer activities based on profile data from different sources and real-time tracking of behaviors across the digital world and beyond.

In today's digital world consumers never know whether their behavior may have triggered an action from any of those interconnected, opaque systems, how this may lead to them getting taken advantage of, and, ultimately, how this causal chain may affect their emotions and behavior. On an aggregate level, it is also entirely unclear how these processes affect groups of people that, without their knowing, are treated in similar ways over time, and how societies as a whole are affected by those consequences.

7.6 Mass personalization

Large online platforms were the first companies with the capability to provide personalized services at a large scale, based on a large volume and variety of data about their users – for example, Google with its personalized search, Facebook with its news feed, and Amazon, YouTube and Netflix's recommendation systems. A report by GroupM, a subsidiary of the marketing giant WPP, estimates that Facebook makes **200 trillion decisions a day** that decide which content may be relevant to each of its users. Relevance, though, is a two way street; the report suggests that what is relevant for both Facebook and Google "is about economic outcomes for the company as much as it is about quality of user experience".[532]

The same applies to targeted ads delivered by others. Calculations aimed at maximizing efficiency and profit determine which ad someone sees, including whether said individual might be 'worth' the ad exposure in the first place at a certain moment. The profile data used for personalization is based on a broad variety of data – seemingly innocuous to most people – including websites visited, terms searched for, apps used, products purchased, friends added on a social network, or places visited. Together with sophisticated yet myopically and single-mindedly consumption-focused modeling and classification schemes, technology companies, data brokers, and many other businesses are now able to learn what people are seemingly interested in, what they did that day, what they will likely do the next, and how much they might be worth as a customer. Data can now be used not only to display ads on websites or within mobile apps, but also on a company's own website, to dynamically personalize the contents, options, and choices offered to a seemingly "anonymous" visitor. For example, online stores can, on an individual basis, personalize how they address someone, which products they display prominently, and even the prices of products or services.[533]

[532] https://groupmp6160223111045.azureedge.net/cmscontent/admin.groupm.com/api/file/2618 [03.05.2017]
[533] Christl and Spiekermann (2016), p. 41

Instant personalization. Direct marketing has long been working on personalizing direct mail, call center and email communication, and customer treatment in general. Now, companies can do personalization in real-time, across devices and communication channels, based on complex rules about how their systems ought to react to different kinds of interactions by consumers. Three types of software services play an important role for this kind of instant personalization. In the first two steps, software mentioned earlier plays a role; first, companies use advanced Customer Relationship Management (CRM) systems to manage their customer data. Second, they use Data Management Platforms (DMPs) and other data platforms to connect their own data to the digital advertising ecosystem, and gain additional profile information about their customers. Finally, as a third step, they use "predictive marketing platforms", which help them to compile the "right message to the right person at the right time".[534]

Predictive marketing. For example, a company might employ RocketFuel, which claims to have "2.7 billion unique profiles" in its data store[535] and offers clients to "bring together trillions of digital and real-world signals to create individual profiles and deliver personalized, always-on, always-relevant experiences to the consumer".[536] RocketFuel says that it "scores every impression for its propensity to influence the consumer".[537] Another platform, Optimizely, can help personalize a website for "first time visitors", based on extensive digital profiles about those visitors provided by Oracle.[538] The predictive marketing platform TellApart, which is owned by Twitter, promises to create a "customer value score" for each shopper and product combination, a "compilation of likelihood to purchase, predicted order size, and lifetime value", based on "100s of online and in-store signals about a particular anonymous customer". Subsequently, TellApart helps automatically assemble pieces such as "product imagery, logos, offers and any metadata" into personalized content for ads, emails, and websites.[539]

Personalized pricing and election campaigns. Similar methods can be used to personalize prices in online shops by, for example, making predictions as to how valuable someone might be as customer in the long-term, or how much someone may be willing to pay at that moment. Strong evidence suggests that online shops already show differently priced products to different consumers, or even different prices for the same products, based on individual characteristics and past behaviors.[540] However, since companies such as Amazon already vary their prices up to 2.5 million times on an average day[541], it will be difficult to prove whether and, if so, to what extent they might incorporate user behavior into their dynamic pricing one day. Another similar field is the use of personalization during election campaigns. Targeting voters with

[534] Standard phrase often used in online marketing:
https://www.google.at/search?q=right+message+to+the+right+person+at+the+right+time

[535] http://rocketfuel.com/wp-content/uploads/DSP_2015.pdf [03.05.2017]

[536] https://rocketfuel.com/predictive-marketing [03.05.2017]

[537] http://rocketfuel.com/wp-content/uploads/DSP_2015.pdf [03.05.2017]

[538] https://www.optimizely.com/de/partners/technology/bluekai/ [03.05.2017]

[539] https://www.tellapart.com/platform/ [03.05.2017]

[540] Christl and Spiekermann (2016), p. 41

[541] https://www.digitalcommerce360.com/2014/12/30/right-price-not-lowest-price/

personalized messages adapted to their personality and political views on certain issues has already raised massive debates about its potential for political manipulation.[542]

Generally speaking, as Ryan Calo has summarized, the "digitization of commerce dramatically alters the capacity of firms to influence consumers at a personal level".[543] The more companies know about individuals, such as their "personal biases and weaknesses"[544], the better they can "change people's actual behavior at scale".[545] Shoshana Zuboff points to the fact that we are not only witnessing the rise of "markets for personal data" but also of "markets for behavioral control", which are "composed of those who sell opportunities to influence behavior for profit and those who purchase such opportunities".[546]

7.7 Testing and experimenting on people

During the US congressional elections 2010, Facebook showed 61 million randomly chosen users a box that suggested going out to vote; users were additionally given a way to tell their friends when they did. After a comparison with a control group and with other data, Facebook estimated that this nudging increased voter turnout by at least 340,000.[547] In another experiment on 1.9 million of users, Facebook secretly manipulated what was shown in their users' news feeds, e.g. decreasing the amount of personal posts such as baby photos in favor of more "hard" news; this, too, led to increased voter turnout during the 2012 elections .[548] In the same year, Facebook conducted its notorious "**mood experiment**" on nearly 700,000 users that involved manipulating the amount of emotionally positive and negative posts in the users' news feeds, which influenced how many emotionally positive and negative messages the users posted themselves.[549]

After massive criticism, not only, but also because all of these experiments were conducted without the participants' knowledge, the online dating platform OkCupid followed up with a provocative blog post defending such practices, stating "we experiment on human beings" and "so does everybody else". They reported an experiment in which they had manipulated the "match" percentage shown to pairs of users. When they showed a 90% match to pairs that actually had a much lower score, these users exchanged significantly more messages with each

[542] See e.g. https://medium.com/@privacyint/cambridge-analytica-explained-data-and-elections-6d4e06549491

[543] Calo, Ryan (2013)

[544] Helberger, Natali (2016): Profiling and Targeting Consumers in the Internet of Things – A New Challenge for Consumer Law. Feb 6, 2016, p. 15. Available at: http://ssrn.com/abstract=2728717

[545] Zuboff, Shoshana (2016): The Secrets of Surveillance Capitalism. Frankfurter Allgemeine Zeitung, 05.03.2016. Available at: http://www.faz.net/aktuell/feuilleton/debatten/the-digital-debate/shoshana-zuboff-secrets-ofsurveillance-capitalism-14103616.html?printPagedArticle=true

[546] 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, p. 85. Available at: http://ssrn.com/abstract=2594754

[547] Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012): A 61-million-person experiment in social influence and political mobilization. Nature, 489(7415), 10.1038/nature11421. http://doi.org/10.1038/nature11421

[548] http://www.motherjones.com/politics/2014/10/can-voting-facebook-button-improve-voter-turnout

[549] Kramer, Adam D. I.; Jamie E. Guillory; Jeffrey T. Hancock (2014): Experimental evidence of massive-scale emotional contagion through social networks. PNAS vol. 111 no. 24, 8788–8790. Available at: http://www.pnas.org/content/111/24/8788.full

other. OkCupid stated that when they would “tell people” they were a “good match”, those people would “act as if they are”.[550] All these ethically highly questionable experiments demonstrate that data-driven personalization clearly has the potential to influence behavior on a massive scale. This form of experimentation on unaware users has become the new normal in the digital world. Facebook stated in 2014 to run “over a thousand experiments each day” in order to “optimize specific outcomes” or to “inform long-term design decisions”.[551] Testing is at the core of today’s digital tracking economy.

Today, online advertisers routinely show different ads to different groups of users in order to test which one makes more people click and interact. Similarly, more and more news organizations, including large outlets such as the Washington Post[552], use different versions of article headlines to figure out which variation performs better.[553] Generally, when companies use two different variations of functionalities, website designs, user interface elements, headlines, button texts, images, or even different discounts and prices, and then carefully track and measure how different groups of users are interacting, this is known as “**A/B testing**” or “split testing”. Tests with more than two variations are known as “multivariate”. Optimizely, one of the major technology providers for such tests, offers its clients the ability to “experiment broadly across the entire customer experience, on any channel, any device, and any application”.[554]

Usually, marketers focus on optimizing “**conversion rates**”, which describe the percentage of people, who were persuaded to act exactly in the way marketers wanted them to act. For example, companies might want users to visit a website, click on an ad, register for a service, subscribe to a newsletter, download an app, or purchase a product. To achieve this, they conduct tests with different variations of some element and then carefully monitor and analyze which version “converts” more users. The most basic requirement for these kinds of tests is data on how users interact, ideally by tracking and measuring every single interaction on an individual level.

Maurits Kaptein et al point to the concept of “persuasion profiles” as “sets of estimates on the effectiveness of particular influence-strategies on individuals, based on their past responses to these strategies”.[555] Pervasive behavioral tracking, in combination with real-time personalization and testing, has become a powerful tool set to systematically influence people’s behavior.

[550] https://theblog.okcupid.com/we-experiment-on-human-beings-5dd9fe280cd5 [03.05.2017]

[551] https://www.facebook.com/notes/facebook-data-science/big-experiments-big-datas-friend-for-making-decisions/10152160441298859/ [03.05.2017]

[552] https://www.wsj.com/articles/washington-posts-bandit-tool-optimizes-content-for-clicks-1454960088

[553] https://blog.upworthy.com/why-the-title-matters-more-than-the-talk-867d08b75c3b

[554] https://www.optimizely.com [03.05.2017]

[555] Kaptein, Maurits; Dean Eckles, and Janet Davis (2011): Envisioning Persuasion Profiles: Challenges for Public Policy and Ethical Practice. 9/10 Interactions 66-69, 66; Available at: https://www.semanticscholar.org/paper/Envisioning-persuasion-profiles-challenges-for-KapteinEckles/fe5f2029df491bdea2cf46697b2e4145c1e226f2/pdf

7.8 Mission creep – everyday life, risk assessment and marketing

Today, the very normal everyday lives of consumers are monitored and tracked by a wide range of companies that provide them with services, both directly and indirectly, via data transfer to third parties. Opting out is difficult or impossible without also opting out of much of modern life. Extensive data is recorded when people use the web, social media, or smartphones, when they purchase something in a store, or when they listen to music. Furthermore, many other kinds of devices with sensors and network connections, including wearables, e-readers, smart TVs, thermostats, and cars comprehensively record and report information. Companies often collect and store a range of consumer data far wider than that needed to provide the services they offer. Much of this excessive data collection has been introduced as a byproduct of online advertising and approaches as well as popular "more data is always better"[556] imperatives.

At the same time, information about people's behaviors, social relationships, and most private moments is increasingly applied in contexts or for purposes completely different from those for which it was recorded.[557] Most notably, data about everyday life behaviors is increasingly used to make automated decisions about individuals in crucial areas of life such as finance, insurance, employment, and law enforcement.

Risk assessment based on everyday life data. Companies have started to predict consumers' creditworthiness based on factors such as the timing and frequency of phone call records, GPS location, customer support data, online purchases, web searches, and data from social networks, including information about someone's social network connections. Also, behavioral data about how someone fills out an online form or navigates on a website is already being used to calculate credit scores, as is the grammar and punctuation of text messages and even an individual's phone battery status. Most of these services are provided by startup companies rather than large banks. However, many large companies from telecom network providers to credit card networks to consumer reporting agencies have started partnering with such startups. The larger-scale application of such services is particularly on the rise in countries of the global south, as well as for vulnerable population groups in other regions.[558]

In addition, large insurers in the US and in Europe have introduced programs that allow consumers to get significant discounts on their insurance premiums if they agree to provide real-time data about car driving behavior and activities such as their steps, grocery purchases, or fitness studio visits.[559] In 2016, Facebook blocked a program by the large UK insurer Admiral, who wanted to price auto insurance based on personality assessments calculated of their cus-

[556] http://www.nytimes.com/2012/03/25/business/factuals-gil-elbaz-wants-to-gather-the-datauniverse.html

[557] As Helen Nissenbaum has argued, contextual integrity of data is a cornerstone for the protection of peoples' privacy. See Nissenbaum, H. (2004): Privacy as Contextual Integrity. Washington Law Review (79:1), pp 101-139.

[558] See section 4.2

[559] Christl and Spiekermann (2016), p. 52-68

tomers' Facebook profile information.[560] However, Facebook itself has registered a patent for automated loan decisions based on credit ratings of someone's friends on a social network,[561] casting this seemingly beneficent decision in a new light.

In healthcare, data companies and insurers are working on programs that use everyday life data about consumers to predict someone's health risks. For example, the large insurer Aviva, together with the consulting firm Deloitte, predicted individual health risks for e.g. diabetes, cancer, high blood pressure, and depression for 60,000 insurance applicants based on marketing consumer data that they had purchased from a data broker.[562] Similarly, the consulting firm McKinsey helped predict the hospital costs of patients based on consumer data for a "large US payor" in healthcare.[563] Using information about demographics, family structure, purchases and car ownership and other data, they stated that such "insights could help identify key patient subgroups before high-cost episodes occur".[564] GNS healthcare, a US health analytics company, also calculates individual health risks for patients from a wide range of data, including from genomics, medical records, lab data, mobile health devices, – and consumer behavior. The company partners with insurers such as Aetna, provides a score that identifies people likely to participate in interventions, and offers to predict the progression of illnesses and intervention outcomes.[565] LexisNexis Risk Solutions offers a health scoring product that predicts individual health risks of people based on vast amounts of consumer data, including purchase activities.[566]

Risk data for marketing and customer relationship management. Key players in data, analytics, and technology that provide risk assessments of individuals in important fields of life such as credit and insurance assessment mostly also provide marketing solutions. For example, the large credit reporting agency TransUnion offers marketing and audience segmentation products with a focus on consumer financial behaviors[567], including "offline-to-online matching" and "insights into consumers who are actively engaged in searches for insurance, mortgages, auto loans and more".[568] TransUnion's "risk triggers" provide "consumer and account monitoring" to alert companies about "changes for collections, account management and marketing", for example, about when "consumers' credit activities indicate receptiveness to marketing messages".[569] In addition, TransUnion partners with the ad technology firm Rocket

[560] https://www.theguardian.com/money/2016/nov/02/facebook-admiral-car-insurance-privacy-data

[561] http://www.theatlantic.com/technology/archive/2015/09/facebooks-new-patent-anddigital-redlining/407287

[562] Christl and Spiekermann (2016), p. 35

[563] Ibid., p. 36

[564] http://healthcare.mckinsey.com/sites/default/files/791750_Changing_Patient_Behavior_the_Next_Frontier_in_Healthcare_Value.pdf [03.05.2017]

[565] Christl and Spiekermann (2016), p. 36

[566] http://www.lexisnexis.com/risk/downloads/literature/health-care/socioeconomic-data-coverages-br.pdf [02.05.2017]

[567] https://www.transunion.com/solution/marketing-audience-segmentation [03.05.2017]

[568] https://www.transunion.com/product/adsurety [03.05.2017]

[569] https://www.transunion.com/product/triggers [03.05.2017]

Fuel to help banks and insurers track consumers across the digital world[570]. The company also provides "aggregated consumer credit data segments" to Oracle's online data marketplace[571] and is listed as a "data provider" for LiveRamp, the data matching platform of Acxiom[572].

Ultimately, TransUnion's "DecisionEdge" product for financial services and retail "unifies" marketing, risk management, and a company's economic goals. It allows companies to implement "marketing and risk strategies tailored for customer, channel and business goals"[573] based on credit data. It also promises "unique insights into consumer behavior, preferences, and risk to drive profitable growth".[574] It includes consumer "prescreening" and "prequalification" during customer acquisition by, for example, letting consumers "choose from a suite of offerings that are tailored to their needs, preferences and risk profile" or by letting companies "evaluate a customer for multiple products across channels, and then only present the offer(s) that are most relevant to them, and profitable" for the company.[575]

Similarly, Experian provides solutions for "credit marketing" to "identify profitable populations" such as consumer prescreening and prequalification.[576] Amongst their offers for customer acquisition they list marketing solutions which allow companies to "identify prospective customers matching [their] business objectives" using Experian's "specialized research, market surveys, psychographic profiles and analytics tools to segment audiences". The company also provides solutions that allow clients to "determine which businesses and consumers make profitable new customers and present the least amount of risk".[577] Another product combines "consumer credit and marketing information that is compliantly available from Experian".[578] The company's "Identity Element Network", a fraud detection solution, provides not only a "predictive identity fraud risk score" describing the likelihood that a customer's identity has been compromised, but also allows the client to "rank-order" the "best and riskiest accounts" and to prioritize "less risky" customers in costly operational processes.[579]

Most marketing data brokers also trade many kinds of sensitive information about consumers, including about their financial situation. Acxiom, for example, provides data and scores about someone's income, net worth, economic stability[580], socioeconomic status[581], loans[582], banking

[570] http://rocketfuel.com/uk/transunion-rocket-fuel-combine-technologies-to-enable-ai-powered-marketing-initiatives/ [03.05.2017]
[571] http://www.oracle.com/us/solutions/cloud/data-directory-2810741.pdf [03.05.2017]
[572] https://liveramp.com/partner/transunion/ [03.05.2017]
[573] https://www.transunion.com/product/decisionedge [03.05.2017]
[574] https://www.transunion.com/resources/transunion/doc/products/resources/product-decisionedge-overview [03.05.2017]
[575] Ibid.
[576] http://www.experian.com/consumer-information/lending-marketing.html [03.05.2017]
[577] http://www.experian.com/business-services/customer-acquisition.html [03.05.2017]
[578] http://www.experian.co.uk/integrated-marketing/delphi-for-mailing.html [03.05.2017]
[579] http://www.experian.com/decision-analytics/identity-and-fraud/current-account-fraud.html [24.04.2017]
[580] https://developer.myacxiom.com/code/api/data-bundles/bundle/investmentsAndAssets [09.05.2017]
[581] https://developer.myacxiom.com/code/api/data-bundles/bundle/areaDemographicsGS [09.05.2017]
[582] https://developer.myacxiom.com/code/api/data-bundles/bundle/mortgagesAndLoans [09.05.2017]

and insurance policies,[583] and the “max range of new credit granted for an individual”[584] for marketing purposes. Similarly, the German marketing data broker Arvato AZ Direct, a subsidiary of the media giant Bertelsmann, provides information about creditworthiness. In its marketing data catalog, those with the best credit rating are labeled with an “A” and “VIP clients”, the worst with G and “postpone processing”.[585] Information about people’s creditworthiness has already even entered online advertising. Thanks to data provided by Oracle’s Datalogix, ads on Twitter, for example, can now be targeted by users’ predicted creditworthiness.[586]

Online fraud detection and the marketing data dragnet. Conversely, the ubiquitous streams of behavioral data in the digital world are being fed into fraud detection systems. The marketing data platform Segment, for example, offers clients easy ways to send data about their customers, website, and mobile app users to many different marketing technology services, but also to fraud detection companies. One of them is Castle, which uses “customer behavioral data to predict which users are likely a security or fraud risk”[587]. Another one, Smyte, helps “prevent scams, spam, harassment and credit card fraud”.[588] Generally, today’s online fraud detection services use highly invasive technologies to evaluate billions of digital transactions and collect vast amounts of information about devices, individuals, and suspicious behaviors.

Large data and analytics companies often provide both marketing and risk services. Experian, for example, offers a service that provides “universal device recognition” across mobile, web, and apps for digital marketing[589]. The company promises to “reconcile and associate” their client’s “existing digital identifiers”, including “cookies, device IDs, IP addresses, and more”, providing marketers with a “ubiquitous, consistent and persistent link across all channels”.[590] Experian’s device identification technology comes from 41st parameter, an online fraud detection company that Experian acquired in 2013.[591] Based on 41st parameter’s technology, Experian also offers a “device intelligence” solution for fraud detection during online payments, which “establishes a reliable ID for the device and collects rich device data”, “identifies every device on every visit in milliseconds”, and “gives unparalleled visibility into the person behind

[583] https://developer.myacxiom.com/code/api/data-bundles/level1bundle/financial [09.05.2017]

[584] https://developer.myacxiom.com/code/api/data-bundles/bundle/creditAndBankCards [09.05.2017]

[585] In German: “Der Informa-Geoscore prognostiziert die Zahlungsausfallwahrscheinlichkeit auf Mikrozellenebene (mit im Schnitt 20 Haushalten je Mikrozelle). Er basiert im Gegensatz zu vielen anderen externen Daten auf validen adress- bzw. personenbezogenen Informationen, welche auf Mikrozellenebene aggregiert werden“, „1 Nahezu kein Risiko (VIP-Kunden, A)“, „7 Höchstes Risiko (Bearbeitung zurückstellen, G)“. Source: Arvato AZ Direct (2015): AZ DIAS PROFILDATEN. Merkmalskatalog. Personal copy of file with author Wolfie Christl

[586] https://twitter.com/WolfieChristl/status/850467843430912000

[587] https://segment.com/integrations/castle [18.05.2017]

[588] https://segment.com/integrations/smyte [18.05.2017]

[589] http://www.experian.co.uk/marketing-services/products/adtruth-device-recognition.html [03.05.2017]

[590] https://www.experianplc.com/media/news/2015/adtruth-resolve/ [03.05.2017]

[591] https://adexchanger.com/data-exchanges/experian-buys-device-id-firm-41st-parameter-for-324m-gets-adtruth-in-the-bargain/

the payment".[592] It is not clear whether Experian uses the same data for its device identification services in fraud detection and marketing.

Fraud prevention, network and information security have a special status in the forthcoming European General Data Protection Regulation, which makes it easier for companies to process data for these purposes without user consent.[593] Therefore, any data collected for fraud prevention and information security may not be used for any other purpose. The same holds true for other online fraud detection services examined in this report, including ThreatMetrix and Google's reCaptcha.[594]

This is especially important since, generally, behavioral data recorded by all kinds of services and devices is increasingly used for risk management. Conversely, data that has been collected in the context of fraud prevention, identity verification, credit scoring, or payment processing is increasingly used for customer relationship management, online targeting, and other marketing purposes. Many companies in data, customer relationship management, and business analytics provide services for marketing as well as for risk management, including law enforcement and intelligence.

[592] http://www.experian.co.uk/assets/identity-and-fraud/device-insight-for-payments-one-pager.pdf [03.05.2017]
[593] European Commission (2016), Recitals 47,49
[594] See also section 4.4

8. Conclusion

This investigation aimed to examine how commercial actors collect, analyze, share, and utilize information about individuals and to better map the structure and scope of today's personal data ecosystem, placing a strong focus on elucidating its implications for individuals, groups of people, and society at large. The findings suggest that networks of online platforms, advertising technology providers, data brokers, and businesses in many industries can now monitor, recognize, and analyze individuals in many aspects of their lives. Information about personal characteristics and behaviors of individuals is linked, combined, and utilized across companies, databases, platforms, devices, and services in real-time. Based on data and guided by their business interests and economic goals, companies have constructed an environment in which individuals are constantly surveyed and evaluated, investigated and examined, categorized and grouped, rated and ranked, numbered and quantified, included or excluded, and, as a result, treated differently.

Pervasive digital tracking and profiling, in combination with personalization and testing, are not only used to monitor, but also to systematically influence people's behavior. Companies are also increasingly using behavioral data about everyday life situations to make both trivial and consequential automated decisions about people, which may lead to cumulative disadvantages for and discrimination against groups of individuals, and may reinforce or even worsen existing inequalities. These developments affect everyone, whether as individuals or as members of groups – memberships of which one often remains unaware – and, ultimately, the dynamics of society at large.

While companies have been collecting and utilizing data about individuals for decades, several key developments in recent years have rapidly introduced completely unprecedented new qualities of ubiquitous corporate surveillance. These include the rise of social media and networked devices, the real-time tracking and linking of behavioral data streams, the merging of online and offline data, and the dissolution of the distinction between marketing and risk management data.

Although the degree to which different industries are connected to the invasive tracking ecosystems developed within online advertising varies, businesses across a diverse set of sectors are joining unconditionally. Both consumer facing companies with direct customer relationships and hidden third-party tracking services for the most part do not consider either any kinds of ethical and societal implications or possible legal consequences. At the same time, only the tip of the iceberg of data and profiling activities is visible to individuals. Much of it occurs in the background and remains opaque and barely understood by most consumers, as well as by civil society, journalists, and policymakers. Companies inform people incompletely, inaccurately, or not at all about their data practices. They use ambiguous, misleading, and obfuscating language in both user interfaces and contracts such as privacy policies and terms of service. Moreover, companies systematically trick consumers into data contracts.

At the same time, people have ever fewer options to resist the power of this data ecosystem; opting out of pervasive tracking and profiling has essentially become synonymous with opting out of much of modern life. Even though corporate leaders argue that "privacy is dead"[595] (while also caring a great deal about their own privacy),[596] people do indeed perceive the power asymmetries in today's digital world and react to them. The research of Mark Andrejevic suggests that people feel "frustration over a sense of powerlessness in the face of increasingly sophisticated and comprehensive forms of data collection and mining".[597] He argues in a compelling way that users "operate within structured power relations that they dislike but feel powerless to contest".[598]

This report focuses on the actual practices and inner workings of today's personal data industries. It explores relevant recent developments, as well as technologies, business models, platforms, services, devices, and data flows. While the picture is becoming clearer, large parts of the systems in place still remain in the dark. Enforcing transparency about corporate data practices remains a key prerequisite to resolving the massive information asymmetries and power imbalances between data companies and the individuals that they process data on. Beside corporate transparency, there is also a strong need for a much better general understanding of today's pervasive tracking and profiling technologies, as well as their impact on and consequences for individuals and society on a broad level. Existing metaphors and narratives that describe violations of personal privacy, such as being watched naked by a stranger, do not help much in comprehending the privacy implications of interlinked databases that analyze ubiquitous data streams on everyone.

Admittedly, this investigation falls short of fully addressing the relationship between the examined corporate practices on the one hand and the existing and upcoming regulatory frameworks on the other. Hopefully, however, this report's findings will encourage and contribute to further work by researchers, scholars, journalists, and stakeholders in the fields of civil rights, data protection, consumer protection, and hopefully also of policymakers and the companies themselves. In 1999, Lawrence Lessig famously predicted that "left to itself, cyberspace will become a perfect tool of control", shaped by commerce and the market's "invisible hand". He suggested that we could "build, or architect, or code cyberspace to protect values that we believe are fundamental, or we can build, or architect, or code cyberspace to allow those values to disappear".[599] Today, the latter has nearly been made reality – by billions of dollars in venture capital poured into funding business models based on the unscrupulous mass exploitation of data. The shortfall of privacy regulation in the US and the absence of its enforcement in Eu-

[595] https://www.theguardian.com/technology/2010/jan/11/facebook-privacy

[596] http://www.msn.com/en-us/money/companies/facebooks-zuckerberg-sues-hundreds-of-hawaiians-to-force-property-sales-to-him/ar-AAm1TvX

[597] Andrejevic, Mark (2014): The Big Data Divide. International Journal of Communication 8 (2014), p. 1685.. Available at:: http://ijoc.org/index.php/ijoc/article/download/2161/1163

[598] Ibid., p. 1678

[599] Lessig, Lawrence (1999): Code and Other Laws of Cyberspace

rope has actively impeded the emergence of other kinds of digital innovation, that is, of practices, technologies, and business models that preserve freedom, democracy, social justice, and human dignity.

The upcoming new European privacy legislation, which supersedes the 1995 Data Protection Directive and includes both the already adopted EU General Data Protection Regulation (GDPR), as well the still disputed ePrivacy Regulation, will become effective in May 2018. Although far from perfect and characterized by many years of negotiations, lobbying, and compromises, it might come just in time to prevent the worst. Still, because of its complexity, the implications on a broad scale are difficult to predict. Interpretations such as that of Doc Searls, who expects[600] that "surveillance capitalism" will become "illegal a year from now in the EU", may be too optimistic. Nevertheless, the GDPR might ban or at least slow down the most irresponsible and invasive practices of third-party tracking and open up ways to make corporate data practices more transparent and accountable. Furthermore, its impact will most likely reach beyond the European Union. The US legal and regulatory framework, in contrast, has enabled the growth of today's data-driven world without any effective consumer safeguards and there is little in sight that will bring about a fundamental change. The recent backlash regarding broadband privacy shows that rather the opposite is the case.

On a broader level, data protection legislation alone may not mitigate the consequences a data-driven world has on individuals and society, whether in the US or Europe. In particular, preventing the dominant data platforms and conglomerates from abusing the unprecedented data power that they have consolidated based on extensive behavioral information on billions of people presents a major challenge. While consent and choice are crucial principles to resolve some of the most urgent problems of intrusive data collection, they can also produce an "illusion of voluntariness".[601] For example, when insurers offer programs without all-day monitoring of a consumer's physical activity only for significantly higher prices, a consumer's choice would end up being between participation and punishment and thereby shift the default toward the former. As such, while important in many respects, a policy discourse that focuses *only* on individual empowerment and control over one's data may not lead to effective long-term solutions. Besides additional regulatory instruments such as anti-discrimination, consumer protection, and competition law, it will generally require a major collective effort to make a positive vision of a future information society reality. Otherwise, we might soon end up in a society of pervasive digital social control, where privacy becomes – if it remains at all – a luxury commodity for the rich. The building blocks are already in place.

[600] https://blogs.harvard.edu/doc/2017/03/23/brands-need-to-fire-adtech/

[601] Davis, Simon G. (1997): Re-Engineering the Right to Privacy: How Privacy Has Been Transformed from a Right to a Commodity. In: Technology and Privacy: The New Landscape, eds. Agre, Philip E. and Marc Rotenberg. Cambridge, MA: The MIT Press. Pp. 143-162.

Figures

References

Allenby, Greg M (2010): Perspectives on Promotion and Database Marketing: The Collected Works by Robert C Blattberg. World Scientific, 2010.

Ammari, Nader; Gustaf Björksten; Peter Micek; Deji Olukotun (2015): The Rise of Mobile Tracking Headers: How Telcos Around the World Are Threatening Your Privacy. Access Now, August 2015. Available at: https://www.accessnow.org/cms/assets/uploads/archive/AIBT-Report.pdf

Andrejevic, Mark (2014): The Big Data Divide. International Journal of Communication 8 (2014). Available at: http://ijoc.org/index.php/ijoc/article/download/2161/1163

Angwin, Julia; Jeff Larson; Lauren Kirchne; Surya Mattu (2017): Minority Neighborhoods Pay Higher Car Insurance Premiums Than White Areas With the Same Risk. ProPublica, April 5, 2017. Available at: https://www.propublica.org/article/minority-neighborhoods-higher-car-insurance-premiums-white-areas-same-risk

Arvato AZ Direct (2015): AZ DIAS PROFILDATEN. Merkmalskatalog. Personal copy of file with author Wolfie Christl

Bennett, Colin (2016): Is Your Neighbor and Democrat or a Republican? Lateral Voter Surveillance and the Political Culture of Modern Election Campaigns (April 20, 2016). Available at: https://ssrn.com/abstract=2776308

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012): A 61-million-person experiment in social influence and political mobilization. Nature, 489(7415), 10.1038/nature11421. http://doi.org/10.1038/nature11421

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). Available at: http://ssrn.com/abstract=2733115

Bouk, Dan (2015): How Our Days Became Numbered. Risk and the Rise of the Statistical Individual. University of Chicago Press.

Brat, Eric; Stephan Heydorn, Matthew Stover, and Martin Ziegler (2013): Big Data: The Next Big Thing for Insurers? Boston Consulting Group, March 25, 2013. Available at: 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. Available at: http://dcentproject.eu/wpcontent/uploads/2015/10/research_on_digital_identity_ecosystems.pdf

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

Calo, Ryan (2013): Digital Market Manipulation. 82 George Washington Law Review 995 (2014); University of Washington School of Law Research Paper No. 2013-27, August 15, 2013. Available at: https://ssrn.com/abstract=2309703

Calo, Ryan and Rosenblat, Alex (2017): The Taking Economy: Uber, Information, and Power (March 9, 2017). Columbia Law Review, Vol. 117, 2017; University of Washington School of Law Research Paper No. 2017-08. Available at: https://ssrn.com/abstract=2929643

Capizzi, Michael T. and Rick Ferguson (2005): Loyalty trends for the twenty-first century. Journal of Consumer Marketing, 22/2, 72-80

Cegłowski, Maciej (2016): The Moral Economy of Tech. Text version of remarks, SASE conference, Berkeley, June 26, 2016. Available at: http://idlewords.com/talks/sase_panel.htm

Christl, Wolfie and Sarah Spiekermann: Networks of Control. A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy. Facultas, Vienna 2016. Available at: http://crackedlabs.org/en/networksofcontrol

CIPPIC (2006): On the data trail: How detailed information about you gets into the hands of organizations with whom you have no relationship. A report on the Canadian data brokerage industry, April 2006. Available at: https://idtrail.org/files/DatabrokerReport.pdf

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. Available at: http://ssrn.com/abstract=2376209

Cook, Tamara and Claudia McKay (2015): How M-Shwari Works: The Story So Far. Access to Finance FORUM Reports by CGAP and Its Partners. No. 10, April 2015. Available at: https://www.cgap.org/sites/default/files/Forum-How-M-Shwari-Works-Apr-2015.pdf

Cox, Emmett (2011): Retail Analytics: The Secret Weapon. Wiley, November 2011

Davis, Simon G. (1997): Re-Engineering the Right to Privacy: How Privacy Has Been Transformed from a Right to a Commodity. In: Technology and Privacy: The New Landscape, eds. Agre, Philip E. and Marc Rotenberg. Cambridge, MA: The MIT Press.Pp. 143-162.

Deighton, John; Peter A. Johnson (2013): The Value of Data: Consequences for Insight, Innovation & Efficiency in the U.S. Economy. October 8, 2013. Available at: https://www.ipc.be/~/media/documents/public/markets/the-value-of-data-consequences-for-insight-innovation-and-efficiency-in-the-us-economy.pdf

Donatello, Michael (1998): Audience research for newspapers. In: Blanchard, Margaret (eds): History of the Mass Media in the United States: An Encyclopedia, Fitzroy Dearborn Publishers, September 1998.

EU Fraud Prevention Expert Group (2007): Report on Identity Theft/Fraud. Brussels, 22 October 2007. Available at: http://ec.europa.eu/internal_market/fpeg/docs/id-theft-report_en.pdf

European Banking Authority (2016): Discussion Paper on innovative uses of consumer data by financial institutions, EBA/DP/2016/01, 04 May 2016. Available at: https://www.eba.europa.eu/documents/10180/1455508/EBA-DP-2016-01+DP+on+innovative+uses+of+consumer+data+by+financial+institutions.pdf

Evans, Peter C. and Annabelle Gawer (2016): The Rise of the Platform Enterprise. A Global Survey. The Emerging Platform Economy Series No. 1, The Center for Global Enterprise, January 2016. Available at: http://thecge.net/wp-content/uploads/2016/01/PDF-WEB-Platform-Survey_01_12.pdf

Ferretti, F. (2009): The Credit Scoring Pandemic and the European Vaccine: Making Sense of EU Data Protection Legislation, 2009(1) Journal of Information, Law & Technology (JILT). Available at: http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2009_1/ferretti

Ferretti, Federico (2014): EU Competition Law, the Consumer Interest and Data Protection. The Exchange of Consumer Information in the Retail Financial Sector. SpringerBriefs in Law, 2014

Ferretti, Federico (2014): The Legal Framework of Consumer Credit Bureaus and Credit Scoring in the European Union: Pitfalls and Challenges—Overindebtedness, Responsible Lending, Market Integration, and Fundamental Rights. Suffolk University Law Review, Vol. XLVI:791, p. 807. Available at: http://suffolklawreview.org/wp-content/uploads/2014/01/Ferretti_Lead.pdf

Ferretti, Federico (2015): Credit Bureaus Between Risk-Management, Creditworthiness Assessment and Prudential Supervision. EUI Department of Law Research Paper No. 2015/20. Available at: https://ssrn.com/abstract=2610142

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. Available at: https://fipa.bc.ca/wordpress/wp-content/uploads/2015/03/CC_report_lite.pdf

Forrester (2015): The Forrester Wave™: Customer Insights Services Providers, Q4 2015. November 10, 2015.

Forrester (2015): The Forrester Wave™: Data Management Platforms, Q4 2015. Available at: http://www.krux.com/upload/image/company/The_Forrester_Wave_Data_Management_Platforms_Q4_2015.pdf

Frida Alim, Nate Cardozo, Gennie Gebhart, Karen Gullo, and Amul Kalia (2017): Spying on Students: School-Issued Devices and Student Privacy. Electronic Frontier Foundation, April 13, 2017. Available at: https://www.eff.org/wp/school-issued-devices-and-student-privacy

FTC (2007): Credit-based Insurance Scores: Impacts on Consumers of Automobile Insurance. A Report to Congress by the Federal Trade Commission, July 2007. Available at: https://www.ftc.gov/reports/credit-based-insurance-scores-impacts-consumers-automobile-insurance-report-congress-federal

FTC (2014): Data Brokers. A Call for Transparency and Accountability, p. C3. Available at: http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

Gallagher, Kevin (2017): Ad Tech explainer. How innovation is changing the digital advertising business and creating new opportunities for disruption. Business Insider, BI Intelligence, January 2017.

Gangadharan, Seeta Peña (2012): Digital inclusion and data profiling. First Monday, Volume 17, Number 5 - 7 May 2012. Available at: http://firstmonday.org/ojs/index.php/fm/article/view/3821/3199

Garvie, Clare; Alvaro Bedoya; Jonathan Frankle (2016): The Perpetual Line-Up. Unregulated Police Face Recognition in America. Georgetown Law Center on Privacy & Technology, October 18, 2016. Available at: https://www.perpetuallineup.org/

Helberger, Natali (2016): Profiling and Targeting Consumers in the Internet of Things – A New Challenge for Consumer Law. Feb 6, 2016, Available at: http://ssrn.com/abstract=2728717

Hoofnagle, Chris Jay (2016): Federal Trade Commission Privacy Law and Policy - Chapter 6 Online Privacy (February 1, 2016). Chris Jay Hoofnagle, Federal Trade Commission Privacy Law and Policy (Cambridge University Press 2016); UC Berkeley Public Law Research Paper No. 2800276. Available at: https://ssrn.com/abstract=2800276

Hormozi, A. M. & Giles, S. (2004): Data Mining: A Competitive Weapon for Banking and Retail Industries.. IS Management, 21, 62-71. Available at: http://cjou.im.tku.edu.tw/dm200707/dm_application.pdf

Hunt, Robert (2005): A century of consumer credit reporting in America, No 05-13, Working Papers, Federal Reserve Bank of Philadelphia

Irion, Kristina and Natali Helberger (2016): Smart TV and the online media sector: User privacy in view of changing market realities. Telecommunications Policy, Volume 41, Issue 3, April 2017, Pages 170–184. Available at: http://doi.org/10.1016/j.telpol.2016.12.013

Jentzsch, Nicola (2003): The Regulation of Financial Privacy: The United States vs Europe. ECRI Research Reports No. 5, 1 June 2003. Available at: http://aei.pitt.edu/9430/2/9430.pdf

Kaptein, Maurits; Dean Eckles, and Janet Davis (2011): Envisioning Persuasion Profiles: Challenges for Public Policy and Ethical Practice. 9/10 Interactions 66-69. Available at: https://www.semanticscholar.org/paper/Envisioning-persuasion-profiles-challenges-for-KapteinEckles/fe5f2029df491bdea2cf46697b2e4145c1e226f2/pdf

Kidwell, Roland E. and Sprague, Robert (2009): Electronic Surveillance in the Global Workplace: Laws, Ethics, Research and Practice. New Technology, Work and Employment, Vol. 24, Issue 2, pp. 194-208, July 2009. Available at: https://ssrn.com/abstract=1487801

Kihn, Martin (2016): A Brief, Incredible History of Marketing Data Matching. Blog article at Gartner for Marketers, May 16, 2016. Available at: http://blogs.gartner.com/martin-kihn/a-brief-incredible-history-of-marketing-data-matching/

Kingstone, Sheryl; Rich Karpinski; Brian Partridge (2015): Monetizing 'Telecom Data as a Service'. Sep 2015, 451 Research. Executive overview available at: https://451research.com/report-long?icid=3534

Kramer, Adam D. I.; Jamie E. Guillory; Jeffrey T. Hancock (2014): Experimental evidence of massive-scale emotional contagion through social networks. PNAS vol. 111 no. 24, 8788–8790. Available at: http://www.pnas.org/content/111/24/8788.full

Kumar, V. (2008): Managing Customers for Profit: Strategies to Increase Profits and Build Loyalty. FT Press

Lee, Newton (2014): Facebook Nation. Total Information Awareness. Springer, 2014.

Lessig, Lawrence (1999): Code and Other Laws of Cyberspace

LiveRamp (2016): IdentityLink Data Store. Providing access to Identity-based data and buyers across the marketing ecosystem. PDF brochure. Personal copy of file with author Wolfie Christl

Lyon, David (2010): Surveillance, Power and Everyday Life. In: Kalantzis-Cope, Phillip; Gherab-Martín, Karim (eds): Emerging Digital Spaces in Contemporary Society: Properties of Technology, Palgrave 2010

McConnell, Ted (2015): The Programmatic Primer. A Marketer's Guide to the Online Advertising Ecosystem

Napoli, Philip M. (2011): Ratings and Audience Measurement. In: Nightingale, Virginia (eds): The Handbook of Media Audiences, April 2011, Wiley-Blackwell.

Narayanan, Arvind; Dillon Reisman (2017): The Princeton Web Transparency and Accountability Project. Pre-published book chapter. Available at: http://randomwalker.info/publications/webtap-chapter.pdf

Ngai, E. W. T., Li Xiu, and D. C. K. Chau (2009): Review: Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications 36, 2 (March 2009), 2592-2602. Available at: http://dx.doi.org/10.1016/j.eswa.2008.02.021

Nissenbaum, H. (2004): Privacy as Contextual Integrity. Washington Law Review (79:1)

Norwegian Consumer Council (2016): Consumer protection in fitness wearables. November, 2016. Available at: https://fil.forbrukerradet.no/wp-content/uploads/2016/11/2016-10-26-vedlegg-2-consumer-protection-in-fitness-wearables-forbrukerradet-final-version.pdf

Oracle (2016): The Audience Playbook. Available at: http://www.oracle.com/us/products/applications/audience-guide-3034880.pdf

Oracle (2017): Oracle BlueKai User Guide. Available at: http://docs.oracle.com/cloud/latest/marketingcs_gs/OMCDA/OMCDA.pdf

Oracle (2017): Taxonomy Changes Report, 31.01.2017. Available at: https://docs.oracle.com/cloud/latest/marketingcs_gs/OMCDA/Resources/xls/TaxonomyChangesReport2017-01-31.xlsx (via https://docs.oracle.com/cloud/latest/marketingcs_gs/OMCDA/Help/AudienceDataMarketplace/Taxonomy_Updates/taxonomy_updates_january_2017.html)

Oracle (2017): Using Oracle Data Cloud. Available at: https://docs.oracle.com/en/cloud/saas/data-cloud/dsmkt/using-oracle-data-cloud.pdf

Oracle (2017): Data Directory. Available at: http://www.oracle.com/us/solutions/cloud/data-directory-2810741.pdf

Rhoen, Michiel (2016): Beyond consent: improving data protection through consumer protection law. Internet Policy Review, 5(1). Available at: http://policyreview.info/articles/analysis/beyond-consent-improving-data-protection-throughconsumer-protection-law

Ridic, Goran; Suzanne Gleason; Ognjen Ridic (2012): Comparisons of Health Care Systems in the United States, Germany and Canada. Mater Sociomed. 2012; 24(2): 112–120. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3633404

Rieke, Aaron; Harlan Yu; David Robinson; Joris von Hoboken (2016): Data Brokers in an Open Society. An Upturn Report, prepared for the Open Society Foundations. November 2016. Available at: https://www.opensocietyfoundations.org/sites/default/files/data-brokers-in-an-open-society-20161121.pdf

Rodríguez-Ruiz, Blanca (1997): Privacy in telecommunications: a European and an American approach. The Hague, Boston, Kluwer Law International, 1997.

Rothmann, Robert; Sterbik-Lamina, Jaro; Peissl, Walter; as, Johann (2012): Aktuelle Fragen der Geodaten-Nutzung auf mobilen Geräten – Endbericht. Bericht-Nr. ITA-PB A63; Institut für Technikfolgen-Abschätzung (ITA): Vienna. Available at: https://www.oeaw.ac.at/itaen/projects/the-use-of-geodata-on-mobile-devices/overview/

Schneier, Bruce (2015): Data and Goliath. The Hidden Battles to Collect Your Data and Control Your World. W. W. Norton & Company

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. Available at: 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, 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. Available at: https://www.researchgate.net/publication/282356703_A_measurement_study_of_tracking_in_paid_mobile_applications

Singer, Natasha (2012): Mapping, and Sharing, the Consumer Genome. New York Times, June 16, 2012. Available at: http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-ofconsumer-database-marketing.htm

Sprague, Robert D. (2007): From Taylorism to the Omnipticon: Expanding Employee Surveillance Beyond the Workplace, 25 J. Marshall J. Computer & Info. L. 1, 2007. Available at: http://repository.jmls.edu/jitpl/vol25/iss1/1/

Tanner, Adam (2017): Our Bodies, Our Data. How companies make billions selling our medical records. Beacon Press, Jan 10, 2017

Tene, Omer (2007): What Google Knows: Privacy and Internet Search Engines (October 1, 2007). Utah Law Review. Available at: https://ssrn.com/abstract=1021490

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). Available at: http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

The Paypers (2016): Web Fraud Prevention & Online Authentication Market Guide 2016/2017.

Tolbert, John (2016): CIAM Platforms. Leaders in innovation, product features, and market reach for Consumer Identity and Access Management Platforms. Your compass for finding the right path in the market. KuppingerCole Report, November 2016.

Trustev Sales Pack. REAL TIME ONLINE IDENTITY VERIFICATION. PDF Brochure. Personal copy of file with author Wolfie Christl

Turow, Joseph (2011): The Daily You.

U.S. Consumer Financial Protection Bureau (2016): List of consumer reporting companies. Available at: http://files.consumerfinance.gov/f/201604_cfpb_list-of-consumer-reporting-companies.pdf

United States Government Accountability Office (2015): Facial Recognition Technology. Commercial Uses, Privacy Issues, and Applicable Federal Law. Report to the Ranking Member, Subcommittee on Privacy, Technology and the Law, Committee on the Judiciary, U.S. Senate, July 2015. Available at: http://www.gao.gov/assets/680/671764.pdf

Winterberry Group (2012): The Data Management Platform: Foundation for Right-Time Customer Engagement. A Winterberry Group Whitepaper. Available at: http://www.iab.net/media/file/Winterberry_Group_White_Paper-Data_Management_Platforms-November_2012.pdf

Woodward, J. D. , N. M. Orlans, P.T. Higgins (2003): Biometrics: Identity Assurance in the Information Age, McGraw, Hill Osborne Media, 2003

Yan, Jun; Ning Liu, Gang Wang, Wen Zhang, Yun Jiang, and Zheng Chen (2009): How much can behavioral targeting help online advertising?. In Proceedings of the 18th international conference on World wide web (WWW '09). ACM, New York, NY, USA, 261-270. Available at: http://dl.acm.org/citation.cfm?id=1526745

Yu, Zhonghao; Sam Macbeth, Konark Modi, and Josep M. Pujol (2016): Tracking the Trackers. In Proceedings of the 25th International Conference on World Wide Web (WWW '16). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 121-132. Available at: http://www2016.net/proceedings/proceedings/p121.pdf

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. Available at: 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. Available at: http://ssrn.com/abstract=2594754

Zuboff, Shoshana (2016): The Secrets of Surveillance Capitalism. Frankfurter Allgemeine Zeitung, 05.03.2016. Available at: http://www.faz.net/aktuell/feuilleton/debatten/the-digital-debate/shoshana-zuboff-secrets-ofsurveillance-capitalism-14103616.html?printPagedArticle=true